# SESSION

# CRYPTOGRAPHIC TECHNOLOGIES

## Chair(s)

### Dr. Levent Ertaul

# A Behavior Based Covert Channel within Anti-Virus Updates

**D. Anthony, D. Johnson, P. Lutz and B. Yuan**
Department of Networking, Security and Systems Administration, Rochester Institute of Technology,
Rochester, New York, USA

**Abstract -** *This paper presents a new behavior based covert channel utilizing the database update mechanism of anti-virus software. It is highly covert due to unattended, frequent, automatic signature database update operations performed by the software. The design of the covert channel is described; its properties are discussed and demonstrated by a reference implementation. This paper uses these points to strengthen the inclusion of behavior-based covert channels within standard covert channel taxonomy.*

**Keywords:** Covert Channels, Security, Data Hiding

## 1 Introduction

B.W. Lampson [2] originally defined covert channels as communication channels not intended for communication at all. Covert channels are typically categorized into one of two traditionally accepted definitions. Storage based covert channels are described as any direct (or otherwise) writing of information value(s) into a legitimate overt channel by a sender, and the direct (or otherwise) reading of those information value(s) by a receiver [3]. Timing based covert channels are channels in which the message sender relays information via the modulation of resources such as CPU usage, or modulating the arrival of information such as network packets over time which allows a receiver who also knows the modulation method used to decode the message [3]. These two definitions adequately describe the vast majority of covert channels; however this paper aims to strengthen the argument for the inclusion of a third classification standard for covert channels.

## 2 Previous Covert Channels

There have been multiple covert channels designed and implemented by researchers in the past which fit the standard Storage or Timing based definitions. Rowland describes how small messages could be sent during the initial 3-way handshake of TCP connections[9]. Specifically, the initial sequence number could be modified to relay data covertly. The authors in [10] examine 22 potential network storage covert channels within IPv6 headers. Okamura and Oyama discovered a scenario in which two distinct and isolated virtual machines running on the Xen Hypervisor may be able to transmit covert messages between each other through the timing modulation of CPU loads [8].

## 3 Behavior Based Covert Channels

Behavior based covert channels are a relatively recent development in the covert channels taxonomy. Johnson, Lutz and Yuan [4] explain how the purposeful alteration of the internal states or behavior of an application can allow the leakage of information between two parties [4]. A game could be used as a covert channel if two parties agree upon a handshake initialization as well as protocol for encoding a message through moves made within the game [4]. The authors use a game called Magneton to demonstrate the channel. This behavior based model needs to be recognized as a completely separate method of classifying a covert channel, and can be used to uniquely identify the Anti-Virus Covert Channel (AVCC, a new covert channel presented in this paper). AVCC contains qualities that could be neither described as storage based or timing based alone. Instead, it is the *behavior* of the anti-virus scanning engine after a database update has been applied which will allow a covert message to be transmitted.

## 4 Signature Based Anti-Virus System

To understand how an Anti-Virus Covert Channel would function, Anti-Virus signatures must be discussed. Anti-Virus signatures typically will take the form of a set of unique cryptographic hashes or pieces of code which corresponds to a known malicious file[6]. These signatures will typically reside within a database that an Anti-Virus application will use to identify known malware that resides on a system. If a match is discovered, the Anti-Virus application will alert the user that a malicious file has been identified and quarantine the file. These Anti-Virus databases must be kept up-to-date to ensure virus scans are accurate and effective. There are many Anti-Virus products on the market and each will have their own method of updating the signature database. ClamAV, a popular free and open source Anti-Virus application, makes use of a tool called Freshclam for this task. Freshclam is a highly customizable tool which may be run interactively from a command line shell or alternatively run as a daemon process on the host machine [7]. Freshclam can also be configured to automatically check for new updates to the signature database as many as 50 times per day [5]. ClamAV provides additional security metrics such as digital signatures on all updates to be distributed. Updates to the any signature database would be  considered normal network traffic and not arouse suspicion from security administrators. New malware is discovered daily and generates constant pressure for AV Companies to identify the latest threats. Database updates occur on a regular basis and may contain a large amount of information depending on the specific Anti-Virus application. For these reasons, the updating of an Anti-Virus database may be seen as an ideal method for transmitting a covert message.

# 5 A New Covert Channel

The Anti-Virus Covert Channel (AVCC) can be classified as an exemplary layer 7 or application layer behavior-based covert channel which can be used to subvert the existing security policy of an organization. ClamAV was chosen as the example software to demonstrate this new covert channel due to its open source, cross-platform nature combined with its user friendly signature database. It should be noted however,  the design of this covert channel is not specific to a certain brand of Anti-Virus, and the ideas presented could theoretically be applied to any AV implementation.

The Prisoner's Problem mentioned in [1] is used to describe the method in which this covert channel may be applied. One node on the internet will be sending data to a second node while attempting to prevent this communication from detection by a third party. The sender will have access to either an official anti-virus signature distribution point or own a location that can be used for updates. The sender will then encode specially crafted signatures into updates which will later be used to relay a message to the receiver. Each signature added into the update will be representative of a binary “1” within the covert message. At designated intervals the receiver will update his/her anti-virus signatures which will force the  retrieval of the database update crafted by the original sender node. The receiver will then scan a unique directory of files with his/her updated anti-virus application. The results of the directory scan will be used to decode the covert message. It is this *behavior* of the anti-virus program during the scan, not solely the signatures within the update which relay the covert message. This flow of data transfer is pictured in Figure 1. A proof-of- concept implementation will be discussed later in this paper.

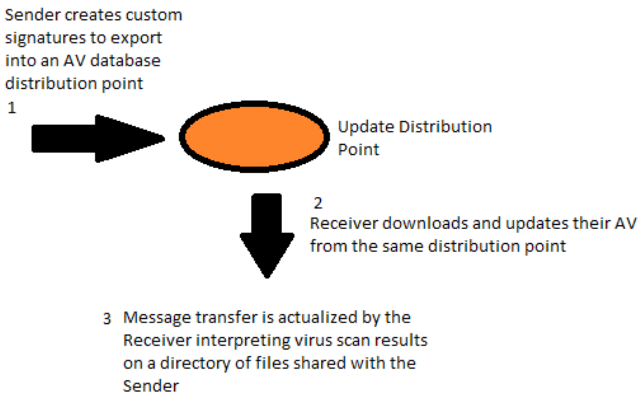

Sender creates custom signatures to export into an AV database distribution point
1
Update Distribution Point
2
Receiver downloads and updates their AV from the same distribution point
3 Message transfer is actualized by the Receiver interpreting virus scan results on a directory of files shared with the Sender

Figure 1. Data Transfer

## 5.1 Covertness

The transferred covert message is never fully contained within the updates which precludes classification as a strictly storage based covert channel. Only binary “1's” from the message are transferred through the update mechanism- not “0's”. Even if a third party was able to capture the network traffic and knew how the channel operated, the covert message would be unable to be revealed from physical inspection. Additionally, this covert channel is not based upon the modulation of timed events which implies that it is not a timing based covert channel. If a third party is able to observe the timing of all network events, no irregularity can be detected as AV signature updates are a normal regular network event and in some cases may be required by an organization's security policy.

## 5.2 Data Rate

Anti-virus database updates vary in size depending upon the specific application being discussed. These updates are often based upon the discovery of new pieces of malware which varies and can have an impact on the amount of data contained within an update. Some days may have more information stored within an update than others. There is no theoretical limit to the size of a signature database. In practice, each specific anti-virus product will have an average update size. The data rate of the AVCC will be closely tied to the specific anti-virus application which is chosen for covert channel use. Covert messages transferred using the AVCC will also have different data rates depending upon the type of data being sent. According to channel design, instances of a binary “1” within a message will require a signature being added to a database update. To remain covert, the modified database updates must remain similar in size to normal updates. It may arouse suspicion if certain updates are exceedingly large. Since each  binary “1” within a covert message correlates to a new signature being created, the data rate of the AVCC may be measured by the number of signatures that can be added while limiting the size characteristic to remain within the confines of a normal database update. Once this limit has been determined, a node is only limited to sending messages containing binary “1's” which fall below that threshold.

## 5.3 Robustness

If packets do not arrive in a specific order or network timing is disrupted, the integrity of a timing based covert channel can suffer; many storage based covert channels do not survive packet regeneration performed by proxies or routers. The AVCC design does not succumb to these types of obstacles. Update mechanisms of application layer software will typically make use of the Transmission Control Protocol (TCP) to insure reliable data delivery. Two ways in which a covert message could be modified or tampered with is the complete prevention of data transfer or removal of a sender's specially crafted signatures before the transmission. If the data sent in an update maintains integrity and reaches the receiver, the message will be relayed. This gives the AVCC a high score in the robustness category.

## 5.4 Limitations

The primary concern of the AVCC design is the issue of a sender controlling a database update location. It may not be ideal or feasible in certain situations for a user to be an operator or have access to locations in which database updates can be distributed. In cases regarding proprietary software this access becomes even more difficult. Another limitation that should be noted is that the type of data being transferred will make a significant impact on the data rate. As each binary “1” in a message corresponds to a virus signature that must be added, data made up of many “1s” will increase the size of updates to a greater degree. Staying inconspicuous requires staying beneath an update size threshold, thus will have an impact on the data throughput of the covert channel. A minor area of concern could be the simplex style of communication signature updates provide, however this same characteristic also makes AVCC a good choice for the infiltration of messages or data into an organization.

# 6 Assumptions

Before a Proof of Concept was created and tested, the following assumptions are applied. The model assumed in this paper precludes a third party from physical tampering with the machines involved in the covert message sending/receiving, including modification of information during transit. This covert channel does not currently offer the security characteristic of authentication or data integrity- simply transfer of covert messages. It is also assumed that persons wishing to use the AVCC will be able to access the distribution site for ClamAV or another anti-virus and be able to direct their host AV software to receive updates from this same location. This will also attempt to accurately represent the use of this covert channel in the real world where network defense/defenders would not initially be aware of AVCC's existence.

# 7 Proof of Concept

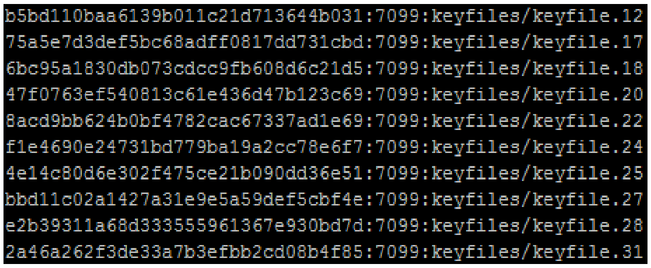
## 7.1 Software Design

The AVCC has been created and implemented in Python as a proof of concept. The implementation script (avcc.py) is a “wrapper” to the tools and binaries that ClamAV provides with a default installation. The creation of signatures is managed with *sigtool* which is the custom signature and database management tool that accompanies ClamAV. Sigtool will allow senders to easily create and add signatures which avcc.py uses to automatically encode/decode the covert message. Alice and Bob have a shared secret which takes the form of a directory of files which will henceforth be referred to as keyfiles. They each share the exact same number, order, and md5sum property for the directory of keyfiles. These keyfiles are what will be scanned with ClamAV when the update has been retrieved. This proof of concept makes use of ASCII encoded messages for all covert communication. The avcc.py script contains methods for the encoding of a message, decoding of a message, and the creation of unique keyfiles based upon a single unique executable.

## 7.2 Example Signatures

The avcc.py implementation of AVCC works by creating signatures for specific files in the keyfiles directory. ClamAV updates can make use of many different types of databases, however this proof of concept will specifically make use of the .hdb database which defines MD5 based signatures. All signatures in this database refer to an executable file.

This database contains the most basic form of an Anti-Virus signature that ClamAV implements and is perfect for the proof-of-concept due to simplicity and ease of creation. Figure 2 displays example signatures that have been added to the database by the avcc.py script.



```
b5bd110baa6139b011c21d713644b031:7099:keyfiles/keyfile.12
75a5e7d3def5bc68adff0817dd731cbd:7099:keyfiles/keyfile.17
6bc95a1830db073cdcc9fb608d6c21d5:7099:keyfiles/keyfile.18
47f0763ef540813c61e436d47b123c69:7099:keyfiles/keyfile.20
8acd9bb624b0bf4782cac67337ad1e69:7099:keyfiles/keyfile.22
f1e4690e24731bd779ba19a2cc78e6f7:7099:keyfiles/keyfile.24
4e14c80d6e302f475ce21b090dd36e51:7099:keyfiles/keyfile.25
bbd11c02a1427a31e9e5a59def5cbf4e:7099:keyfiles/keyfile.27
e2b39311a68d333555961367e930bd7d:7099:keyfiles/keyfile.28
2a46a262f3de33a7b3efbb2cd08b4f85:7099:keyfiles/keyfile.31
```

Figure 2. Example Signatures

These virus signatures are made up of three colon delimited fields. The first field is the MD5 sum of the contents of a file. The second and third sections are the size of the file and its name (given by sigtool) respectively. As it may not be practical to have an identical directory of unique executable files on both ends of transmission, the proof-of-concept is able to generate a keyfiles directory. Keyfile creation with avcc.py is based upon the modification of a single shared executable file. This allows the sender and receiver to produce the same set of keyfiles from one original executable file. This also prevents the need to send a shared directory of keyfiles beforehand and potentially arouse suspicion. In real world scenarios, sender and receiver could arbitrarily choose an executable to generate the keyfiles from. They may even decide to use a commonly shared file on their machines and prevent the need of transferring a file at all. During keyfile generation avcc.py modifies the contents of the base executable slightly to ensure uniqueness between keyfiles. The example keyfile directory shown in Figure 2 contains files named in the fashion *keyfile.number*.

## 7.3 Encoding

Encoding of the message can be demonstrated by the following pseudo code example:

```
- Populate a keyfile array with keyfile names
- Accept User Input (string of chars)
- Translate input into binary array
- Iterate through binary array
- If ( current value == 0 )
   - Do nothing
- If ( current value == 1)
   - Create signature for the keyfile in the corresponding element of keyfile array
- Distribute created signatures
```

Avcc.py has been created to accept user input in the form of an ASCII string. The string will then be converted into a binary array. The keyfiles directory will be examined and an array will be created. Each entry within the keyfile array contains the name of a keyfile. The binary array will be examined and encoding takes place based upon this array. If the current entry in the binary array is a 0, that means that no signature will be created for the file in the keyfile array. If the current entry is a 1, then a signature is required to be added to the database update. This encoding method will create 1 signature to be added to the .hdb database file contained within a ClamAV update for every binary "1" of the data being transferred.

## 7.4 Decoding

Decoding is the reverse of the encoding process. When the receiver has updated his database of signatures from the sender, a virus scan is run upon receiver's directory of keyfiles. The decoding can be explained via the following pseudo code example:

```
- Download database update
- Create empty array
- Create keyfile array from names of keyfiles
- Iterate through keyfile array
- Scan current element
- If ( Virus Detected )
   - Push "1" onto empty array
- If ( No Virus Detected )
   - Push "0" onto empty array
- Convert binary array into real binary data representation
- Print binary data representation as string
- Message is revealed
```

Avcc.py has been created to automatically reveal the hidden message within a signature database update. The receiver's directory of keyfiles will be scanned in order with ClamAV. Every keyfile that triggers a virus-alert will be represented as a binary "1" and keyfiles with no detection are represented as a "0". This sequence of 1s and 0s is representative of the binary data being transferred. Avcc.py will keep track ClamAV behavior and this order-specific sequence of bits and reveal the hidden message to the receiver.

## 7.5 Performance

A quantitative analysis of avcc.py performance was done with the assistance of Nathaniel Morefield using the Minitab statistics software. Avcc.py uses ASCII encoded strings for each covert communication. Representations of ASCII characters contain an average of 3.5 binary "1s". Over the course of 5 weeks, 37 ClamAV database updates were collected from which conclusions about update size were drawn. The mean size of the .hdb file contained within a ClamAV update was 30,229 bytes and contained a mean of 503 unique signatures. In order to keep the additions to the update inconspicuous, total update size *including the message* should be no more than two standard deviations from the mean. The standard deviation of this set of data was 5,950 bytes. This allows for an additional 11,900 bytes of data added to an average update using the assumptions above. A regression test was performed to determine the average size of a single virus signature. The test was restricted so that the best fit line passed through the origin, since an update with no virus signatures would have a size of zero bytes.

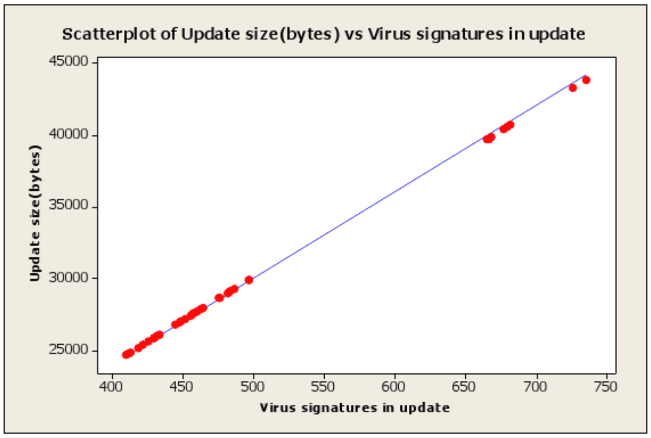


Figure 3. Regression Analysis

Figure 3 shows the scatter plot and regression line associated with the sample data. From this test it has been determined that the average virus signature requires 60.1 bytes of data. Each ASCII character is represented with an average of 3.5 binary “1s”; it will require 210.35 bytes to encode one ASCII character. In order to stay within the two standard deviation size limit of 11,900 bytes, approximately 57 ASCII characters can be effectively hidden within the covert channel.

### 7.6 Environment

The environment that avcc.py has been tested on is Ubuntu 10.04 x86 Server Edition. This platform provided an environment that was stable and easy to work with.

### 7.7 Limitations of Implementation

The implementation avcc.py is a proof of concept of how an anti-virus database update could be used as a covert channel. A fully-functional implementation would require a server that both digitally signs and packs many different types of anti-virus signatures into one file for distribution. This proof of concept strictly details how one may encode covert data within a .hdb. The .hdb file is just one of many specific types of databases that are packed together and digitally signed for a ClamAV signature update. The implementation could be extended to any of the other types of databases, however this proof of concept is not that ambitious. If the methods described in this paper were applied to a fully functional anti-virus signature database distribution point, this covert channel should operate flawlessly.

### 7.8 Detection and Prevention

The implementation of avcc.py is used with unofficial signatures for testing purposes. The files that are used for the creation of signatures are recognized as unofficial signatures by the ClamAV scanning product. The finding of many unofficial signatures could be one way in which anything suspicious could be noted outside of a complete audit of every signature contained within update. Auditing each file in a regular update would be theoretically possible although unfeasible in practice, so this covert channel would be very difficult to detect. A method of prevention could be to audit the locations that a user is updating their signature database from. Restricting update location to approved locations would be one way in which this covert channel could be prevented altogether. This prevention method would be futile if a sender was able to gain access to an official update location. Another method that a sender could use to thwart detection would be to submit new legitimate virus signatures to the official website, which would inevitably be added to the next official update. These “real” signatures could also then be used to transfer a covert message.

## 8 Conclusions

In the future, we would like to explore new possibilities of sending covert communication relating to software updates and further the support for behavior based covert channels being included in standard taxonomy. Anti-Virus software tends to fly under the radar of security programs that monitor for malicious activity. Anti-Virus products are some of the most popular types of applications in use today, and if this technology can be harnessed for use as a covert channel it may prove to be a rich area of further research.

## 9 References

[1] Simmons, G. J. “The Prisoner's Problem and the Subliminal Channel”, Advances in Cryptology: *Proceedings of* CRYPTO '83, Plenum Press, 1984, pp. 51-67.

[2] B. W. Lampson. “A note on the confinement problem”.*Communications of the ACM*, 16(10) pp. 613-615, 1973.

[3] Zander, S., Armitage, G. and Branch, P, “A survey of covert channels and countermeasures in computer network protocols”, *IEEE Communications Surveys & Tutorials,* 9(3) pp. 44-57, 2007.

[4] Johnson, D. and Lutz, P. and Yuan, B. “Behavior-based covert channel in cyberspace”, in: Vanhoof, et al (eds), *Intelligent Decision Making Systems*, World Scientific, 2009, pp. 311-318.

[5] Kojm, Tomasz. “Freshclam(1) – Linux man page”. Internet:http://linux.die.net/man/1/freshclam , Jan. 23, 2012.

[6] Landesman, Mary. “What is a Virus Signature?”. Internet:http://antivirus.about.com/od/whatisavirus/a/virussignature.htm, November 29, 2011

[7] Kojm, Tomasz. “Clam AntiVirus 0.97.4 User Manual”. Internet:http://www.clamav.net/doc/latest/clamdoc.pdf , 2007-2011. Jan 17, 2012

[8] Okamura, Keisuke and Yoshihiro Oyama. "Load-based Covert Channels between Xen Virtual Machines." In *Proceedings of the 2010 ACM Symposium on Applied Computing*. Pp 173-180. 2010.

[9] C. H. Rowland, “Covert channels in the TCP/IP protocol suite.” *First Monday,* vol. 2,no. 5, 1997

[10] Lucena, N., Lewandowski, G., and Chapin, S. “Covert Channels in IPv6”, *Lecture Notes in Computer Science (2006)*. Vol. 3856, Springer, Pages 147-166

# An Evolutionary Approach for the Playfair Cipher Cryptanalysis

**G. Negara**
Faculty of Computer Science, “Al. I. Cuza” University, Iasi, Romania

**Abstract** - *In this paper we describe the Playfair substitution cipher and we propose an evolutionary algorithm for Playfair’s cryptanalysis. The structural properties of the cipher and its enciphering rules determine the suitability of an evolutionary, genetic-like approach for the cipher’s cryptanalysis. Classical cryptographic ciphers like Playfair are no longer used in practice, but evolutionary approaches as the one we propose can be used for modern cryptographic ciphers/algorithms analysis in order to study cryptographic operations, components or modules and for avoiding time consuming approaches like exhaustive enumerations of search spaces (e.g. keys spaces).*

**Keywords:** cryptanalysis, Playfair cipher, genetic algorithm, classical cryptography, substitution cipher

## 1 Introduction

Cryptography means securely communicating in the presence of an adversary. The data transmitted between the communication parties is secured by different techniques. One of the main techniques in classical cryptography is enciphering - concealing a message, where constitutive parts of the message are transposed or substituted by some elements (letters, symbols etc).

The study of classical ciphers represents an adequate technique for understanding the basis of cryptography; classical ciphers can be seen as building block of modern cryptography – principles used in classical ciphers are still used in the design, implementation and analysis of modern cryptographic algorithms/ciphers/systems.

In general, classical ciphers operate on an alphabet of letters and are implemented by hand or with simple mechanical devices. The development of new techniques and technologies made them unreliable. Classical schemes are often susceptible to cipher text-only attacks, sometimes even without knowledge of the system itself, using tools such as frequency analysis [1].

### 1.1 Cryptanalysis of classical ciphers

Classical ciphers are nowadays commonly quite easy to break. Many of the classical ciphers can be broken even if the attacker only knows sufficient enciphered text – the method is called “*ciphertext-only attack*”. Some classical ciphers (e.g. the *Caesar* cipher) have a small key space. These ciphers can be broken using a brute force attack, by simply enumerating all the possible keys. Substitution ciphers can have a large key space, but are often susceptible to frequency analysis because, for example, frequent letters in the plaintext language correspond to frequent letters in the enciphered texts. Polyalphabetic ciphers prevent simple frequency analysis by using multiple substitutions; more advanced techniques can still be used to break these ciphers.

Classical ciphers are not designed to satisfy cryptographically strong criteria; hence they are not currently used in real cryptographic systems. Their design was suitable for manual, writing encryption, tens of years ago. Principles of classical ciphers design were extended and adapted, tough, to the security and efficiency requirements of modern ciphers.

### 1.2 The Playfair Cipher

The *Playfair* cipher is a substitution cipher invented in *1854* by *Charles Wheatstone (1802-1875).* The name of the cipher as it is known in the cryptology literature comes from the name of the lord *Playfair* who strongly promoted the cipher.

The enciphering process is based on a table where one letter of the English alphabet is omitted, and the remaining *25* letters are arranged in a *5x5* grid. Typically, the letter “*J*” is removed from the alphabet and an “*I*” takes its place in the text that is to be enciphered. The grid, with no key, look like the following [2]:

| | | | | |
|---|---|---|---|---|
| *A* | *B* | *C* | *D* | *E* |
| *F* | *G* | *H* | *I* | *K* |
| *L* | *M* | *N* | *O* | *P* |
| *Q* | *R* | *S* | *T* | *U* |
| *V* | *W* | *X* | *Y* | *Z* |

Using the word “*PLAY*” as the key, the enciphering grid becomes:

| | | | | |
|---|---|---|---|---|
| *P* | *L* | *A* | *Y* | *B* |
| *C* | *D* | *E* | *F* | *G* |
| *H* | *I* | *K* | *M* | *N* |
| *O* | *Q* | *R* | *S* | *T* |
| *U* | *V* | *W* | *X* | *Z* |

To encipher, the plain text message is split into two-letter groups (bigrams). Repeated letters in the same group are usually separated by an "*X*" letter.

For example, the message:

*"Like most premodern era ciphers the playfair cipher can be easily cracked if there is enough text"*

would become:

*"LI KE MO ST PR EM OD ER NE RA CI PH ER ST HE PL AY FA IR CI PH ER CA NB EX EA SI LY CR AC KE DI FT HE RE IS EN OU GH TE XT".*

In case of an odd number of letters in the message, another "*X*" letter is used for padding. The next step is to look up for the letters positions in the grid.

The following *4* rules are applied, in order, to each pair of letters in the plaintext:

1. If both letters are the same (or only one letter is left), add an "*X*" after the first letter. Encrypt the new pair and continue. Some variants of *Playfair* use "*Q*" instead of "*X*", but any infrequent letter could be used.
2. If the letters appear on the same row of the enciphering grid, replace them with the letters to their immediate right respectively (circular right shift).
3. If the letters appear on the same column of the enciphering grid, replace them with the letters immediately below respectively (circular down shift).
4. If the letters are not on the same row or column, replace them with the letters placed on the same row respectively but at the other pair of corners of the rectangle defined by the original pair. The order is important – the first letter of the encrypted pair is the one that lies on the same row as the first letter of the plaintext pair.

To decrypt, the inverse operations of the last *3* rules are applied, and also taking into consideration the 1st rule (dropping any extra "*X*"s or "*Q*"s) that don't make sense in the final message when finished).

Following the rules, the resulting enciphered message is:

*"DQ RK HS TO AO FK QC KW KG WE DH CO KW TO KC LA YB EY KQ DH CO KW EP TG FW KE QM AB EO PE RK IQ GS KC WK MQ GK UP CN RG ZS"*

All non-letters are ignored and not enciphered. Numbers, spaces, and punctuation are also skipped. Some other customizations are possible, depending on the cipher variant used.

# 2 A Genetic Algorithm for the Playfair Cipher Cryptanalysis

Evolutionary computation involves iterative processes and models such as growth or development in a population of individuals. Under a certain model, the iterative processes are guided by random techniques. An evolutionary mechanism involves also a search space – usually parallel searching strategies are used to construct partial solutions and finally leading to a solution considered optimal. The solutions are modeled based on the nature of the problem, as elements (individuals, arrays of values etc.) of the search space. Evolution mechanisms and models can produce optimized processes, graphs, networks etc, that is evolutionary algorithms have various applications in computer science, including cryptography and cryptanalysis.

Genetic algorithms, one of the central paradigms of evolutionary computation, became widely known through the work of Holland [3] and Koza [4].

Based on the cipher's definition we propose a genetic algorithm (*GAPFC*) as an evolutionary approach for the *Playfair* cipher cryptanalysis.

## 2.1 GAPFC Algorithm Description

The algorithm starts with random sets of distinct letters as the candidate keys population. The keys are "evolving" by crossover operations (like switching parts of the parent keys) and mutations operations (like switching letters or inversing key parts). The evolution of the population after each round is quantified by evaluating the fitness function (based on the comparison between letter/bigrams frequency in the usual English language and the frequency of letters/bigrams in the deciphered texts corresponding to candidate keys). The best keys are maintained in the population while the worse are eliminated. New candidates are introduced in the population by random candidate keys generation. After a number of iterations the best candidate key found represents the algorithm's solution.

## 2.2 The Fitness Function

The fitness function (1) is based on the comparison of the simple distribution frequency and bigrams frequency obtained from the deciphered texts corresponding to the candidate keys and the "standard" simple letters and bigrams frequencies in the English language.

$$F(key) = \alpha \sum_{i=1}^{n} |f_i^{standard} - f_i^{decipher(key)}|^2 + \beta \sum_{i,j=1}^{n} |f_{ij}^{standard} - f_{ij}^{decipher(key)}|^2 \quad (1)$$

For each letter indexes *i*, *j* (from *1* to *25* letters), $f_i^{standard}$ and $f_{ij}^{standard}$ represents the standard frequency of the *i*$^{th}$ letter of the alphabet and, respectively the standard frequency of the bigram composed by the $i^{th}$ and $j^{th}$ letters, while $f_i^{decipher(key)}$ and $f_{ij}^{decipher(key)}$ represents the corresponding simple and bigrams frequencies observed in the deciphered text obtained from the key under evaluation. α and β are values used for balancing the weights of the simple frequency and of the bigrams frequency in the fitness function. n represents the number of letters from the alphabet - *25* (*"J"* replaced by *"I"*).

For the letters and bigrams frequencies, we are using a *25* length table, respectively a *25 × 25* frequency table, where each table value corresponds to a certain bigram. The tables' values where obtained using a frequency analysis application. As input text, we used an electronic English book[1]. More information related to the frequencies tables is presented in the Appendix.

Frequency tables can be found in complex cryptanalysis projects, such as [5]. In this work, we are using our own analysis of the letters and bigrams frequency of the English alphabet.

## 2.3 GAPFC Algorithm Pseudocode

The pseudocode of the *GAPFC* algorithm is synthesized in the following:

INPUT

1. *keyLen* – the length of the keys (up to *25*, usually up to *10* – the first *2* rows of the enciphering square)
2. *genNum* – the number of generations
3. *keyNum* – the number of candidate keys in the population
4. *cipheredText* – the ciphered text that needs to be deciphered
5. *bestNum* – the percentage of keys selected for crossover and mutation
6. *lettersFreq* – the "standard" letter frequency
7. *bigramsFreq* – the "standard" bigrams frequency

ALGORITHM:

```
1. generate the initial population (keyNum random keys
   of length keyLen – the key squares are filled first
   with the letters from the key and then with the
   remaining letters, in alphabetical order)
2. evaluate each key from the population, based on the
   fitness function
3. sort in ascending order the keys based on the fitness
   values
4. for genNum generations
   a. eliminate the 2 * bestNum worse keys from the
      population
   b. select bestNum best keys
   c. for each best key pair
        i. apply one crossover operation at random –
           obtain 2 new candidate keys
       ii. apply one mutation operations at random for
           the 2 keys
      iii. add the 2 resulting keys in the population
   d. add bestNum new random keys in the population
   e. evaluate each key from the population, based on
      the fitness function (step 2)
   f. sort in ascending order the keys based on the
      fitness values (step 3)
5. select the best key as the final solution and
   determine the corresponding deciphered text
```

OUTPUT

1. *bestKey* – the best key based on the fitness function value
2. *decipheredText* – the deciphered text corresponding to *bestKey*

Crossover operations:

```
1. split each of the 2 keys at random index and switch
   the final parts
   a. Split:
      $key_1 = key_{1,1} | key_{1,2}$
      $key_2 = key_{2,1} | key_{2,2}$
   b. Recombine
      $key_1 = key_{1,1} | key_{2,2}$
      $key_2 = key_{2,1} | key_{1,2}$

2. split each of the 2 keys at random index and switch
   the initial parts
   a. Split:
      $key_1 = key_{1,1} | key_{1,2}$
      $key_2 = key_{2,1} | key_{2,2}$
   b. Recombine
      $key_1 = key_{2,1} | key_{1,2}$
      $key_2 = key_{1,1} | key_{2,2}$
```

Mutation operations:

```
1. split key at random index and switch parts
   a. Split:
      $key = key_1 | key_2$
      $key = key_2 | key_1$
   b. Recombine
      $key = key_2 | key_1$

2. select 2 random letter indexes and switch letters
   a. select indexes $i, j$ at random
      $key = key_1 \mid c_i \mid key_2 \mid c_i \mid key_3$
   b. switch the letters $c_i$ and $c_j$
      $key = key_1 \mid c_j \mid key_2 \mid c_i \mid key_3$
```

[1] Viktor Suvorov, *Spetsnaz. The Inside Story of the Soviet Special Forces* , Hamish Hamilton Ltd, ISBN 0-241-11961-8, 1987.

# 3 Results

## 3.1 Test scenario 1

The same plaintext is successively enciphered using the keys from *3* test sets (*3*, *4* and *5-letter* keys sets, respectively).

| Parameters setting | Key \| # generations for key recovery | | | | | |
|---|---|---|---|---|---|---|
| | keyLen=3 | | keyLen=4 | | keyLen=5 | |
| keyNum<br>1000<br>bestNum<br>4%<br>alpha<br>0.3<br>beta<br>0.7 | cba | 6 | dcba | 9 | woman | 22 |
| | cat | 1 | cats | 8 | bulet | 28 |
| | bus | 17 | rats | 15 | usato | 33 |
| | man | 4 | bats | 16 | frequ | 83 |
| | pen | 9 | myth | 12 | setup | 50 |
| | gun | 1 | hand | 19 | their | 23 |
| | pin | 4 | best | 8 | later | 31 |
| | yes | 2 | made | 38 | house | 16 |
| | cod | 4 | gold | 29 | fathe | 28 |
| | mis | 5 | girl | 29 | encip | 48 |

*Table 1.*

The parameters setting and the number of algorithm's generations for the recovery of *3*, *4*, and *5-letter* keys.

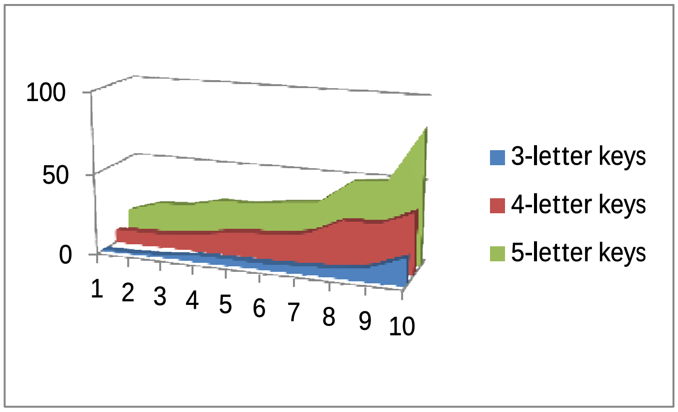


*Figure 1.*

A comparative analysis of the number of generations ( on the vertical axis) for the recovery of the keys from the *3* test sets (*3*, *4* and *5-letter* keys).

## 3.2 Test scenario 2

The same plaintext is successively enciphered using the *20* related keys of length *6*, obtained by successively adding all the possible letters to the word "*relat*" (the letters of the resulting key word must be distinct).

| Parameters setting | Initial key | Result key | Fitness value |
|---|---|---|---|
| keyLen<br>6<br>genNum<br>1000<br>keyNum<br>2000<br>bestNum<br>5%<br>alpha<br>0.3<br>beta<br>0.7 | relatb | retacm | 73.36 |
| | relatc | qcfnru | 78.08 |
| | relatd | relatd | 47.49 |
| | relatf | mrcatf | 73.76 |
| | relatg | mrcatf | 73.76 |
| | relath | relash | 60.58 |
| | relati | ncfrot | 72.09 |
| | relatk | rhuoap | 70.06 |
| | relatm | relatm | 47.49 |
| | relatn | picnrt | 68.87 |
| | relato | relato | 47.49 |
| | relatp | relatp | 47.49 |
| | relatq | frntaq | 72.82 |
| | relats | risnat | 69.11 |
| | relatu | relatu | 47.49 |
| | relatv | voagrq | 80.63 |
| | relatw | grtoaw | 68.58 |
| | relatx | cnyair | 66.97 |
| | relaty | cinytr | 69.18 |
| | relatz | rtoiay | 71.57 |

*Table 2.*

The parameters setting, the keys and the fitness values obtained by successively running the algorithm for the *20* enciphered texts associated to the related keys. The bold results represent the cases where the initial key was fully recovered (in less than *1000* generations).

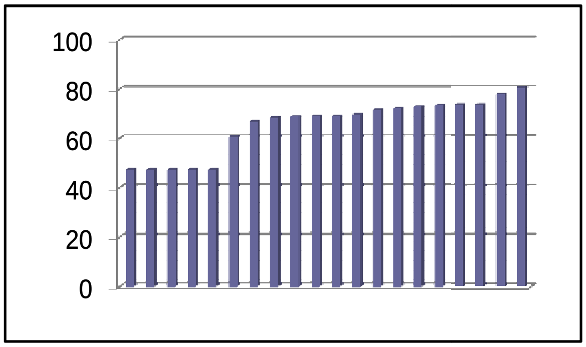


*Figure 2.*

The fitness values (the vertical axis) of the *20* keys obtained by running the algorithm for the *20* related input keys. One can observe the relative proximity of the fitness values of the partially recovered keys. Also, there is a "*gap*" between the fitness value of the fully recovered keys and the other keys, suggesting the appropriate selection of the fitness function.

## 3.3 Test scenario 3

This testing scenario differs from the previous one only by the parameters setting. The purpose of this scenario is to observe the results improvement when using an increased number of generations.

| Parameters setting | Initial key | Result key | Fitness value |
|---|---|---|---|
| keyLen<br>6<br>genNum<br>2000<br>keyNum<br>2000<br>bestNum<br>5%<br>alpha<br>0.3<br>beta<br>0.7 | relatb | relscb | 65.72 |
| | relatc | qfdsup | 71.99 |
| | relatd | relatd | 47.49 |
| | relatf | rmfatb | 69.59 |
| | relatg | relatg | 47.49 |
| | relath | relash | 60.58 |
| | relati | relati | 47.49 |
| | relatk | prunah | 67.63 |
| | relatm | relatm | 47.49 |
| | relatn | relatn | 47.49 |
| | relato | relato | 47.49 |
| | relatp | relatp | 47.49 |
| | relatq | npaeiq | 70.02 |
| | relats | risnat | 69.11 |
| | relatu | relatu | 47.49 |
| | relatv | rncaiv | 71.29 |
| | relatw | relatw | 47.49 |
| | relatx | cnyair | 66.97 |
| | relaty | relatu | 49.07 |
| | relatz | rtoiay | 71.57 |

*Table 3.*

Test scenario 3 results, as in Table 2.

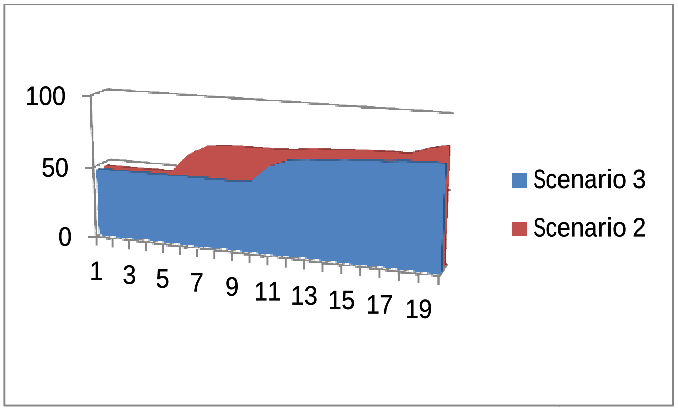


*Figure 3.*

A comparative analysis of the fitness values obtained in test scenarios *2* and *3*. Increasing the numbers of algorithm generations led to better results.

## 3.4 Test scenario 4

A different and longer plaintext is enciphered using the *7-letter* key "*GABRIEL*". The algorithm is run several times for the same enciphered text. The number of generations needed for the key recovery is recorded for each test. The purpose of this test scenario is to demonstrate the stability and the increased efficiency of the algorithm (when looking for longer keys recovery) when a longer enciphered text is used as input.

| Parameters setting | # test | # generations for key recovery |
|---|---|---|
| keyLen<br>7<br>keyNum<br>2000<br>bestNum<br>5%<br>alpha<br>0.3<br>beta<br>0.7 | 1 | 174 |
| | 2 | 116 |
| | 3 | 164 |
| | 4 | 90 |
| | 5 | 78 |
| | 6 | 84 |
| | 7 | 64 |
| | 8 | 100 |
| | 9 | 73 |
| | 10 | 68 |

*Table 4.*

The number of generations needed for the recovery of the *7-letter* key. The same enciphered text is used as algorithm's input for each of the 10 tests.

## 3.5 Test scenario 5

The same plaintext as the one in test scenario *4* is enciphered using the *8-letter* key, "*PLAYFIRC*". Similarly, the algorithm is run several times for the same enciphered text and the number of generations needed for the key recovery is recorded for each test. The parameters' values are the same as in test scenario *4*, except the key length (*8-letter* key).

| Parameters setting | # test | # generations for key recovery |
|---|---|---|
| keyLen<br>8<br>keyNum<br>2000<br>bestNum<br>5%<br>alpha<br>0.3<br>beta<br>0.7 | 1 | 179 |
| | 2 | 196 |
| | 3 | 114 |
| | 4 | 115 |
| | 5 | 100 |
| | 6 | 279 |
| | 7 | 157 |
| | 8 | 100 |
| | 9 | 251 |
| | 10 | 109 |

*Table 5.*

The number of generations needed for the recovery of the *8-letter* key.

The most consuming runs of the algorithm from the above test scenarios took less than *10* seconds, on a processor Intel *Pentium B950, 2.1 GHz, 2 CPUs*. Implementation's efficiency was not the subject of this work. Future work could be focused on improving the algorithm's computational complexity and implementation's efficiency by parallelization and other related techniques [6].

# 4 CONCLUSIONS

The algorithm we proposed demonstrates the suitability of evolutionary approaches for classical substitution ciphers like Playfair.

The results and the algorithm efficiency are influenced by the genetic operators, the parameters settings, the fitness function and the enciphered text length. The fitness values are likely to decrease more when the algorithm is running on longer input enciphered texts (the deciphered texts analyzed are "*closer*", from the letters and bigrams frequencies point of view, to real English texts). The algorithm could also be used in the hypothesis of knowing one or a few words from the plaintext, in which case the fitness function is changed appropriately or using stopping conditions based on these words.

Similar results are obtained if using a different book for computing the frequency tables. The values of frequency tables are close enough when using as input different English text books of relatively large size.

Generic evolutionary schemes such the one used for the *GAPFC* algorithm are potentially useful tools in analyzing and solving both cryptanalysis and cryptographic problems.

# 5 APPENDIX

## 5.1. Letters frequency table

The "*standard*" letters frequency table used in the fitness function (percentage per each letter). In the algorithm's implementation, the frequencies percentages of "*I*" and "*J*" letters are summed ("*J*" is substituted by "*I*"):

*8.161484772*
*1.404417556*
*2.882406354*
*3.402328024*
*12.43162524*
*2.377800630*
*1.909023280*
*4.968110012*
*7.193571679*
*0.104750131*
*0.655577192*
*3.538257045*
*2.557762997*
*7.362594084*
*7.590965780*
*2.384091108*
*0.109126116*
*6.582848328*
*6.972857956*
*9.695540872*
*2.608086820*
*1.214335726*
*1.681745580*
*0.178047873*
*1.697335025*
*0.335309820*

## 5.2. Bigrams frequency table (partial)

The first *20* most frequent bigrams and their frequencies, in descending order:

*th (3.022164362%)*
*he (2.610274812%)*
*er (1.857058463%)*
*in (1.837366532%)*
*an (1.675455102%)*
*es (1.396486084%)*
*re (1.370777175%)*
*en (1.25864257%)*
*nt (1.254266585%)*
*et (1.237309645%)*
*on (1.219258708%)*
*st (1.162370908%)*
*or (1.139943987%)*
*ar (1.118064064%)*
*at (1.097825136%)*
*nd (1.088526168%)*
*to (1.036014353%)*
*ti (1.023980396%)*
*ea (1.008117451%)*
*of (0.975297567%)*

## 5.3. The plaintext used in the test scenarios 1,2 and 3

*"The study of classical ciphers represents an adequate technique for understanding the basis of cryptography classical ciphers can be seen as building block of modern cryptography principles used in classical ciphers are still used in the design implementation and analysis of modern cryptographic algorithms ciphers systems in general classical ciphers operate on an alphabet of letters and are implemented by hand or with simple mechanical devices The development of new techniques and technologies made them unreliable"*

## 5.4. Test example (test scenario 1)

*5 length key "SETUP" found after 50 generations*

*# generations: 50*

| | | | | |
|---|---|---|---|---|
| *S* | *E* | *T* | *U* | *P* |
| *A* | *B* | *C* | *D* | *F* |
| *G* | *H* | *I* | *K* | *L* |
| *M* | *N* | *O* | *Q* | *R* |
| *V* | *W* | *X* | *Y* | *Z* |

*Fitness function value:*
*47.49*

Deciphered text:

*TH ES TU DY OF CL AS XS IC AL CI PH ER SR EP RE SE NT SA NA DE QU AT ET EC HN IQ UE FO RU ND ER ST AN DI NG TH EB AS IS OF CR YP TO GR AP HY CL AS XS IC AL CI PH ER SC AN BE SE XE NA SB UI LD IN GB LO CK OF MO DE RN CR YP TO GR AP HY PR IN CI PL ES US ED IN CL AS XS IC AL CI PH ER SA RE ST IL XL US ED IN TH ED ES IG NI MP LE ME NT AT IO NA ND AN AL YS IS OF MO DE RN CR YP TO GR AP HI CA LG OR IT HM SC IP HE RS XS YS TE MS IN GE NE RA LC LA SX SI CA LC IP HE RS OP ER AT EO NA NA LP HA BE TO FL ET XT ER SA ND AR EI MP LE ME NT ED BY HA ND OR WI TH SI MP LE ME CH AN IC AL DE VI CE ST HE DE VE LO PM EN TO FN EW TE CH NI QU ES AN DT EC HN OL OG IE SM AD ET HE MU NR EL IA BL EX*

## 5.5. The plaintext used in the test scenarios 4 and 5.

*"CNN The legendary Hanging Gardens of Babylon are the inspiration behind an ambitious plan to grow a rooftop forest high above Beiruts crowded streets The cityscape is currently overshadowed by concrete highrises with parkland making up just three percent of Lebanons capital according to a study by the American University of Beirut The lack of greenery has contributed to poor air quality and trapped heat among a host of other environmental issues claims architect Wassim Melki This led him to the the unconventional idea of greening the citys rooftops We want to cover the top of nearly every building in the city with trees said Melki Challenging though it sounds Melkis proposal does not involve complicated drainage systems or engineering We are just talking about planting small to medium sized trees in relatively large pots and securing them to the roofs he added"*

# 6 REFERENCES

[1] Simon Singh, The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography, ISBN 0-385-49532-3 (2000).

[2] Alex Biryukov, Cryptanalysis of the Classical Ciphers, http://www.wisdom.weizmann.ac.il/~albi/cryptanalysis/lect3.htm.

[3] Holland, John H, Adaptation in Natural and Artificial Systems, University of Michigan Press (1975).

[4] Koza, John R., Genetic Programming, MIT Press (1992).

[5] Basic Cryptanalysis, FM 34-40-2, FIELD MANUAL, DEPARTMENT OF THE ARMY http://www.umich.edu/~umich/fm-34-40-2/.

[6] Erick Cantu-Paz, A Survey of Parallel Genetic Algorithms, University of Illinois at Urbana-Champaign, Department of Computer Science and Illinois Genetic Algorithms Laboratory.

# Enabling Searchable Dynamic Data Management for Group Collaboration in Cloud Storages

**Fu-Kuo Tseng, and Rong-Jaye Chen**
Department of Computer Science,
National Chiao Tung University, Hsinchu, Taiwan

**Abstract**— *Cloud storage provides users with high availability, immense storage capacity and abundant processing capability. However, the data in transit or stored in cloud storage could be tempered by unauthorized individuals or even the cloud storage provider. Individual users can utilize data/searchable encryption to securely manage their data. However, enterprise users are required to share their (encrypted) data and search capability within a working group. In this paper, we focus on an enterprise scenario, where search capability and data decryption of stored data can be shared among privileged group users, namely, each data can only be retrieved and decrypted according to a specified policy. We utilize public-key encryption with conjunctive keyword search (PECK) and attribute-based encryption (ABE) to construct our protocol. In addition, we consider dynamic data and group management issues, which are essential for the enterprise application. Finally, we present extensive security analysis and performance evaluation to demonstrate that our design is practical for the enterprise to take advantage of cloud storages.*

**Keywords:** cloud storage, public-key encryption with keyword search, attribute-based encryption, group collaboration

## 1. Introduction

Cloud computing envisions highly-available, on-demand network access to a shared pool of configurable computing resources [1], [2], [3]. Users can enjoy flexible storage capacity and computation capability without paying attention to the construction and maintenance of these infrastructures. While cloud computing brings in promising opportunities, it also brings along new security and privacy issues, which hinder the public to adopt the cloud technologies. The data in transit or stored in cloud storage could be tempered by unauthorized individuals or even the cloud storage provider [4], [5], [6]. A number of encryption techniques are available to protect the security of cloud data and services [7], [8], [9]. However, as these encryption techniques bring along extra complexities, new techniques must be devised to manage encrypted data securely and efficiently.

For an individual cloud storage user, he/she stores his/her data and retrieves part of the stored data later. However, for enterprise users, the stored data should be shared among group members. One type of encryption scheme called *attribute-based encryption (ABE)* could be used to apply fine-grained access control over the shared data [10], [11], [12], [13], [14], [15]. In addition, the dynamics of group members and corresponding stored data should be taken into account to construct a feasible fine-grained access control for the enterprise [16], [17], [18].

Furthermore, given the total amount of data generated and stored in the cloud, accessing data through navigation is time-consuming and bothersome. Accessing cloud data through (keyword) search is thought to be practical and indispensable. However, as the cloud data are secured through cryptographic techniques, which bring in high costs when retrieving through searching. Searchable encryption was introduced to enable users to hide the searchable keywords (of a file) by encryption [19], [20], [21], [22], [23], [24]. Later, users could generate appropriate tokens/trapdoors for specific keywords to retrieve the encrypted data containing these keywords. For enterprise users, searching capability is also shared under fine-grain policies [25], [26], [27], [28]. One user can generate searchable indexes for a file and specify a subset of users who can utilize these searchable indexes. Users outside the specified group cannot search out this file. Moreover, dynamics of group members and searchable indexes should be considered to yield a practical and robust searchable encryption [29], [30].

In this paper, we propose one novel cloud storage construction enabling the management of dynamic searchable data for group collaboration. We make use of *attribute-based encryption scheme (ABE)* and *public-key encryption with conjunctive keyword search (PECK)* to design our protocol. We demonstrate that our scheme uniquely integrates essential functionality for enterprise users, namely, the fine-grained access control for the searchable index and the content of the data. Furthermore, we provide security analysis and conduct extensive performance evaluation to show the feasibility of our design for enterprise users.

The rest of the paper is structured as follows. Related background is described in Section 2, while targeted system models and two cryptographic building blocks are presented in Section 3. Our novel construction is detailed in Section 4. Then the security and performance analysis are shown in Section 5. Finally, our contributions are reiterated and future direction is mentioned to conclude this paper.

## 2. Related Works

In this section, we provide cryptographic background of elliptic curves and bilinear pairings, followed by an introduction of searchable encryptions and attribute-based encryptions, which is relevant to our proposed scheme.

### 2.1 *Elliptic Curve and Bilinear Pairing*

An elliptic curve [31] is the set of all points $(x, y)$ satisfying the equation of the form $y^2 = x^3 + ax + b$ together with an extra point called point at infinity. In cryptography, we usually consider elliptic curves over one finite field $\mathbb{F}_q$ where each coordinate of the point $(x, y)$ is a member of $\mathbb{F}_q$. The set of points on the elliptic curve forms an Abelian group under a certain addition rule where the point at infinity serves as the identity of this group.

Bilinear pairings [32], [33] are critical building blocks that enable identity-based cryptography and its variants. Let $G_1$ be a cyclic additive group generated by $P$, whose order is a prime $p$, and $G_2$ be a cyclic multiplicative group of the same order $p$. A bilinear pairing is a map $\hat{e} : G_1 \times G_1 \rightarrow G_2$ with the following properties: (1) *Bilinear:* $\forall P, Q \in G_1$, $\forall a, b \in \mathbb{F}_q^*$, $\hat{e}(aP, bQ) = \hat{e}(aP, bQ)^{ab}$, (2) *Non-degenerate:* $\forall P, Q \neq 0_{G_1}$, $\hat{e}(P, Q) \neq 1_{G_2}$, and (3) *Computable:* $\hat{e}(P, Q)$ is efficiently computable.

### 2.2 *Attribute-based Encryption*

*Attribute-based Encryption (ABE)* provides a fine-grained access control of shared data. ABE was originated from the work by Sahai and Waters [10]. Later, two tracks of ABE have been developed: *ciphertext-policy ABE (CP-ABE)* [13], [15] and *key-policy ABE (KP-ABE)* [12], [14]. In the CP-ABE scheme, the user is granted attribute keys (associated with attributes), and the access policy could be enforced on the ciphertext. If the user owns the attribute keys satisfying the specified access policy, the user could decrypt the message. A reverse setting is called KP-ABE, which specifies decryption policy on the attribute keys and the ciphertext is tagged with a set of attributes.

However, to deploy in practical applications, managing dynamic access policy is required to support ever-changing access group. The attribute keys should be re-issued and ciphertext be re-encrypted to comply with the current access control policy. In addition, user revocation should be carried out in an efficient way to control the damages. Some ABE suggested that the expiration time is appended with the attribute when generating associated attribute keys [13], [17]. However, the trade-off between the granularity of "window of vulnerability" and the burden to update the attribute keys should be considered. Boldyreva *et al.* [16] proposed an efficient revocation scheme for IBE and KP-ABE, while Yu *et al.* [18] proposed an ABE scheme with attribute revocation. They integrated the *proxy re-encryption (PRE)* with ABE, and enabled the authority to delegate most of the work for key update of the user to proxy servers. Because *role-based access control (RBAC)* [34] is commonly used to restricting system access to authorized users. CP-ABE, which is closely related to RBAC, is chosen as a building block of our scheme for enterprise application scenario.

### 2.3 *Searchable Encryption*

Searchable encryption enables users to hide the searchable keywords (of a file) by encryption. Later, users could generate appropriate tokens/trapdoors for specific keywords to retrieve the encrypted data containing these keywords. Song *et al.* [19] first introduced the concept of searching on encrypted data and provided practical solutions. Goh [20] then formalized the notion of security for this problem and constructed a more efficient scheme using Bloom filter. Following that, some research [22], [23] has been conducted to either improve the efficiency or provide stronger security of searchable encryption. One commonality of these works is that they all supported only single keyword search in the symmetric key setting.

The concept of conjunctive keyword search in symmetric key setting was first introduced by Golle *et al.* [25]. They provided a security notion for conjunctive keyword search over encrypted data and constructed a more efficient scheme compared with the one trivially extended from single keyword search scheme. Later, Ballard *et al.* [26] improved by shortening the trapdoor size and reducing computation/storage overhead. However, due to the symmetric key setting, these schemes only enable one user to store and retrieve his/her own private data. Sharing of index building and searching capability cannot be achieved easily.

Boneh *et al.* [21] first addressed one kind of practical applications called email routing system. The searchable index of a mail can be generated by using the recipient's public key. The recipient can retrieve particular emails from the routing server by delegating related trapdoors. The corresponding emails can be collected. In addition, Boneh *et al.* [35] proposed another application called searching over audit log, where the company can delegate specific trapdoor to the auditor to inspect only audit-related records. However, these schemes supported only single keyword search. There are other applications requiring more expressive search over possible keywords.

To enrich search expressions, Park *et al.* [27] proposed *public-key encryption with conjunctive keyword search (PECK)*. Boneh *et al.* [36] further provided a scheme supporting the conjunction of subset and range queries on encrypted data. Their construction used the bilinear group of composite order, which yields less efficient construction. In addition, they considered only single-user setting, where sharing of searchable index is hard to achieve. Hwang *et al.* [28] provided one efficient PECK and considered a possible extension to multi-user settings [29], [30]. In this paper, we will further consider the sharing of searchable index should be provided to enable group collaboration.

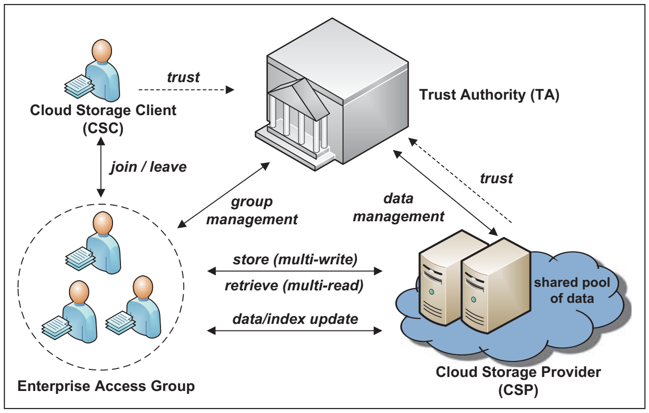


Fig. 1: Enterprise Cloud Storage Access Model

# 3. System Model and Building Blocks

We start by listing the notations used throughout the paper in Table 1. We then describe our cloud storage system model we are using in this paper. Finally, we detail two building blocks used in our proposed scheme.

## 3.1 System Model

We consider a general cloud data storage architecture containing three system entities for enterprise application scenario. (See Fig. 1)

1) *Cloud Storage Client (CSC)* stores a large number of data (files) in the cloud storage server. A CSC could be an individual user or organization. The CSC can play the role as data producer, data consumer, or both.
2) *Cloud Storage Provider (CSP)* provides search-based secure store/retrieval services for cloud storage clients. A CSP has abundant storage space and computation power to manage the data of CSCs.
3) *Trusted Authority (TA)* is trusted by all the other system entities, who seek to resolve their disputes.

In an enterprise application scenario, the enterprise moves its data to the cloud to enjoy on-demand storage and computation resources. The enterprise defines its own access control policies over the data, including who can retrieve the data in cloud storages and who can read the content of the data. These policies reflect the organization hierarchy and information classification. The TA generates corresponding credentials to enable enterprise members (acting as the CSCs) to generate data, encrypt data and construct searchable indexes based on the assigned privileges. The data are encrypted and indexed through fine-grained cryptographic primitives. To retrieve the data in the cloud storages, the CSC picks some keywords for the data so that this data can be searched and retrieved later by specifying any of the keywords. To learn the content of the retrieved data, the CSC needs to have appropriate credentials from TA.

Table 1: Notations

| Notation | Descriptions |
|---|---|
| $PK/MK$ | Public/Private key for attribute-based encryption |
| $AS$ | Access structure for attribute-based encryption |
| $CT$ | Ciphertext for attribute-based encryption |
| $AK$ | User attribute keys in attribute-based encryption |
| $param$ | Public key for attribute-based encryption |
| $W$ | Keyword set for searchable encryption |
| $S$ | Searchable index for searchable encryption |
| $Q$ | The specified keyword set |
| $m$ | Maximum number of users for a group |
| $m'$ | Number of authorized users for a file |
| $n$ | Maximum number of attributes supported |
| $n'$ | Number of attributes of specified trapdoor |
| $W$ | Keyword set for searchable encryption |
| $l$ | Number of attributes in W |
| $T_{j,Q}$ | The trapdoor generated by user $j$ containing $Q$ |
| $\mathcal{QF}$ | The qualified file set returned by CSP |

## 3.2 Design Goals

The cloud storage for the enterprises in group collaboration should meet the following design goals.

1) *Confidentiality*: The data of a group should not be learned by unauthorized users. The content is protected under group access control. Authorized users are granted credentials to find the content.
2) *Retrievability*: The retrieval of group data is managed under group access control. Authorized users are granted credentials to generate valid trapdoors to the cloud to retrieve group data.
3) *Data/Group Dynamics*: The ever-changing data and access group should be carried out in an effective and efficient way. The credentials of eligible users should be re-issued and ciphertexts be re-encrypted to comply with the current access control policy.

## 3.3 Building Blocks - ABEra

*Attribute-based encryption with attribute revocation* denoted as *ABEra* [18] consists of seven algorithms: *Setup, Enc, KeyGen, Decrypt, ReKeyGen, ReKey*, and *ReEnc*. In this scheme, attributes are represented by their index values and the attribute universe is $U = \{1, 2, ..., n\}$, where $n$ is size of $U$. Each attribute can appear in the access policy in one of the following three states: *positive*, *negative*, and *don't care*. The access structure ($AS$) is represented as $\bigwedge_{\tilde{i}\in I} \tilde{i}$ , where $I$ denotes the set of attributes of interest (in $AS$). $\tilde{i}$ represents either the positive state (denoted as $+i$), or the negative state (denoted as $-i$) of attribute $i$. If the attribute does not appear in $AS$, its state is *don't care*.

*Setup*$(1^\lambda)$ Set $PK = (G_1, G_2, \hat{e}, Y, T_1, \ldots, T_{3n}, P)$, where $Y = \hat{e}(P, P)^y$ and $T_i = t_iP$, for $1 \leq i \leq 3n$. Set $ver = 1$ and publish $(ver, PK)$, $(ver, MK)$, where $MK = (y, t_1, \ldots, t_{3n})$.

*Enc*$(M, AS, PK)$ Denote $AS$ as $\bigwedge_{\tilde{i} \in I} \tilde{i}$. Assume $M \in G_2$ and pick a random values $s \in Z_p^*$. Compute $CT = (ver, AS, \tilde{C}, \hat{C}, \{C_i\}_{i \in U})$, where $\tilde{C} = MY^s, \hat{C} = sP$. For each $i \in I, C_i = sT_i$, if $\tilde{i} = +i$, $C_i = sT_{n+1}$, if $\tilde{i} = -i$, and otherwise $C_i = sT_{2n+i}$.

*KeyGen*$(MK, S)$ Pick a random $r_i \in Z_p^*$ for each $i \in U$. Set $r = \sum_{i=1}^n r_i$. Compute user attribute key as $AK = (ver, S, D, \bar{D} = \{D_i, F_i\}_{i \in U})$, where $ver$ is the current version number, $D = (y - r)P$, and $U$ is attribute universe. For each $i \in U$, $F_i = \frac{r_i}{t_{2n+i}} P$; if $i \in S, D_i = \frac{r_i}{t_i} P$, and otherwise $D_i = \frac{r_i}{t_{n+i}} P$.

*Decrypt*$(CT, PK, SK)$ If any two of $CT$, $PK$, and $SK$ have different $ver$, return $\perp$. Otherwise, proceed as follows: Suppose $CT = \{ver, AS, \tilde{C}, \hat{C}, \{C_i\}_{i \in U}\}$, $SK = \{ver, S, D, \bar{D} = \{D_i, F_i\}_{i \in U}\}$, and parse $AS$ as $AS = \bigwedge_{\tilde{i} \in I} \tilde{i}$. For each $i \in I$, (1) if $\tilde{i} = +i$ and $i \notin S$, compute $\widehat{e}(C_i, D_i) = \widehat{e}(st_i P, \frac{r_i}{t_i Q})$. (2) if $\tilde{i} = -i$ and $i \notin S$, $\widehat{e}(C_i, D_i) = \widehat{e}(st_{n+i} P, \frac{r_i}{t_{n+i}} P)$. If $\tilde{i} \notin I$, $\widehat{e}(C_i, D_i) = \widehat{e}(st_{2n+i} P, \frac{r_i}{t_{2n+i}} P)$. Thus, $M = \bar{C} / (\widehat{e}(\hat{C}, \hat{D}) \prod_{i=1}^n \widehat{e}(P, P)^{r_i s})$.

*ReKeyGen*$(\gamma, MK)$ Denote the attribute set for update as $\gamma$. Each $i \in \gamma$ is defined within the range of $[1, 2n]$, the corresponding proxy re-key is generated as follows. Randomly select $t'_i \in Z_p^*$ and compute $rk_i = \frac{t'_i}{t_i}$. For each $i \in \{1, \ldots, 2n\} \backslash \gamma$, $rk = 1$. Output proxy re-key as $rk = (ver, \{rk_i\}_{1 \leq i \leq 2n})$, where $ver$ is current version which is the $ver$ in $MK$ plus one.

*ReKey*$(\bar{D}, rk, \theta)$ Denote the attribute set as $\theta$, where the corresponding rekey is not 1. Each attribute $\in \theta$ is defined within the range of $[1, 2n]$. If $\bar{D}$ and $rk$ have different $ver$, return $\perp$. Otherwise, update $\bar{D}$ as follow. For each $i \in \theta$, $D'_i = rk_i^{-1} D_i$, if $1 \leq i \leq n$, or $D'_{i-n} = rk_i^{-1} D_{i-n}$ if $n < i \leq 2n$. For each $i \in U$, $D'_i = D_i$ if $i \notin \theta$ and $i + n \notin \theta$. $\bar{D}' = \{D'_i, F_i\}_{i \in U}$. $ver$ is incremented by one in corresponding secret key $SK$.

*ReEnc*$(CT, rk, \beta)$ Denote the access structure of $CT$ as $AS = \bigwedge_{\tilde{i} \in I} \tilde{i}$. Denote the attribute set in $AS$ as $\beta$. Each $i \in \beta$ is defined within the range of $[1, 2n]$. If $CT$ and $rk$ contain different version number, use input $CT$ as output. Otherwise, re-encrypt $CT$ as follows. For each $i \in \beta$, $C'_i = rk_i C_i$, if $1 \leq i \leq n$, or $C'_{i-n} = rk_i C_{i-n}$ if $n < i \leq 2n$. For each $i \in U$, $C'_i = C_i$ if $i \notin \beta$ and $i{+}n \notin \beta$, or $i \notin I$. Output $CT' = (ver + 1, AS, \tilde{C}, \hat{C}, \{C'_i\}_{i \in U})$.

### 3.4 Building Blocks - muPECS

*Public-key conjunctive keyword search with multi-user* setting denoted as *muPECKS* [28] consists of four algorithms: *KeyGen, mPECK, Trapdoor*, and *Test*. In this scheme, each group can support up to $m$ users, which is decided at *Setup* phase. Each group member is given a public/private key pair: the public key is used to produce (conjunctive) searchable indexes, while the private key is applied to generate valid trapdoors. The storage performs *Test* using the trapdoor provided and the searchable indexes stored. If the equality holds, the corresponding file is qualified.

*Setup*$(1^\lambda)$ Set $param = (G_1, G_2, \hat{e}, H_1(\cdot), H_2(\cdot), P)$, where $H_1,\ H_2$: $\{0, 1\}^* \longrightarrow G_1$ are different collision-resistance hash functions. Set the maximum size of a group as $m$. Generate $x_1, \ldots, x_m \in Z_p^*$ and compute $y_i = x_i P$. Send the public/private key pairs $(x_i, y_i)$ to user $i$ securely.

*mPECK*$(pk_1, \ldots, pk_m, W)$ Pick keyword set $W = \{w_1, \ldots, w_l\}$ for the file $M$. Pick two random values $s$, $r \in Z_p^*$ and compute $A = rP, B_j = sy_j$, for $1 \leq j \leq m$. $C_i = rU_i + sV_i$, for $1 \leq i \leq l$, where $U_i = H_1(w_i)$ and $V_i = H_2(w_i)$. Output $S = \langle A, B_1, \ldots, B_m, C_1, \ldots, C_l \rangle$.

*Trapdoor*$(sk_j, Q)$ Set $T_{j,Q_1} = tP, T_{j,Q_2} = t \sum_{i=1}^{m'} (U_{I_i})$, where $t \in Z_p^*$, $T_{j,Q_3} = \frac{t}{x_j} \sum_{i=1}^{m'} (V_{I_i})$, where $Q = \{I_1, \ldots I_m, w_{I_1}, \ldots, w_{I_m}\}$. The trapdoor $T_{j,Q} = (T_{j,Q_1}, T_{j,Q_2}, T_{j,Q_3}, I_1, \ldots I_m)$.

*Test*$(pk_j, S, T_Q)$ Include the file to the qualified file set $\mathcal{QF}$ if $\hat{e}(T_{j,Q_1}, \sum_{i=1}^{m'} (C_{I_i}) = \hat{e}(A, T_{j,Q_2}) \cdot \hat{e}(B_j, T_{j,Q_3})$.

## 4. Proposed Scheme

In this section, we propose one novel cloud storage construction enabling the management of searchable dynamic data for dynamic group collaboration. We make use of the *ABE* and *PECK* detailed in Section 3.3 and Section 3.4 respectively. We also provide one practical example to demonstrate how our construction enables dynamic group collaboration by providing search-based retrieval and fine-grained access control of dynamic data for enterprises.

### 4.1 Detailed Construction

Our scheme consists of ten operations: Setup, GrpCreate, GrpStore, Retrieve, Search, GrpDecryt, UpdateAK, UpdateCT, UpdateAU, and UpdateSI. The subscript in each operation denotes its initiator: $s$ is for CSP, $c$ is for CSC, and $t$ is for TA. (See Fig. 2)

Setup$_t(1^\lambda)$: $\langle (PK, MK) \rangle \leftarrow ABEar.Setup\ (i^\lambda)$, $param \leftarrow muPECK.Setup(1^\lambda)$.

GrpCreate$_t(MK, S)$: $AK \leftarrow ABEar.KeyGen(MK,\ S)$. CSP sends $AK$ to qualified CSC.

GrpStore$_c(M, AS, pk_1, \ldots, pk_n, W)$: $CT \leftarrow ABEar. Enc(M, AS, PK), S \leftarrow muPECK.mPECK(pk_1, \ldots, pk_n, W)$. CSC sends $CT$ to CSP.

Retrieve$_c(sk_j, Q)$: $T_{j,Q} \leftarrow muPECK.Trapdoor(sk_j,\ Q)$. CSC sends $T_{j,Q}$ to CSP.

Search$_s(pk_j, S, T_Q)$: $\mathcal{QF} \leftarrow mPECK.Test(pk_j, S, T_Q)$. CSP sends $\mathcal{QF}$ back to CSC.

GrpDecryt$_c(\mathcal{QF}, SK)$: For each encrypted file $CT$ in $\mathcal{QF}$, $M = ABEar.Decrypt(CT, PK, SK)$.

| | |
|---|---|
| Setup$_t(1^\lambda)$: | $\langle(PK, MK)\rangle \leftarrow ABEar.Setup\ (1^\lambda)$, $param \leftarrow muPECK.Setup(1^\lambda)$. |
| GrpCreate$_t(MK, S)$: | $AK \leftarrow ABEar.KeyGen(MK,\ S)$. CSP sends $AK$ to qualified CSC. |
| GrpStore$_c(M, AS,$<br>$pk_1, \ldots, pk_n, W)$: | $CT \leftarrow ABEar.\ Enc(M, AS, PK), S \leftarrow muPECK.mPECK(pk_1, \ldots, pk_n, W)$. |
| Retrieve$_c(sk_j, Q)$: | $T_{j,Q} \leftarrow muPECK.Trapdoor(sk_j,\ Q)$. CSC sends $T_{j,Q}$ to CSP. |
| Search$_s(pk_j, S, T_Q)$: | $\mathcal{QF} \leftarrow muPECK.Test(pk_j, S, T_Q)$. CSP sends $\mathcal{QF}$ back to CSC. |
| GrpDecryt$_c(\mathcal{QF}, SK)$: | For each encrypted file $CT$ in $\mathcal{QF}$, $M = ABEar.Decrypt(CT, PK, SK)$. |
| UpdateAK$_t(SK, rk, \gamma, \theta)$: | $rk \leftarrow ABEar.ReKeyGen\ (\gamma\ ,MK)$, $SK' \leftarrow ABEar.ReKey(SK, rk, \theta)$. CSP sends $SK'$ to CSC. |
| UpdateCT$_c(CT, rk, \beta)$: | $CT' \leftarrow ABEar.ReEnc\ (CT, rk,\ \beta)$. CSC sends $CT'$ to CSP to replace $CT$. |
| UpdateAU$_c(S, i)$: | CSC send $B_i$ to CSP. CSP removes $B_i$ from $S$. |
| UpdateSI$_c(S, i, op)$: | If $op$ = $add$, compute searchable index $C_i$ and add into $S$.<br>Otherwise, remove searchable index $C_i$ from $S$. CSC sends $S$ to CSP. |

Fig. 2: Our Proposed Scheme

UpdateAK$_s(SK, rk, \gamma, \theta)$: $rk \leftarrow ABEar.ReKeyGen(\gamma, MK)$, $SK' \leftarrow ABEar.ReKey(SK, rk, \theta)$. CSP sends $SK'$ to CSC.

UpdateCT$_c(CT, rk, \beta)$: $CT' \leftarrow ABEar.ReEnc\ (CT, rk, \beta)$. CSC sends $CT'$ to CSP to replace $CT$.

UpdateAU$_t(S, i)$: CSC sends $B_i$ to CSP. CSP removes $B_i$ from $S$.

UpdateSI$_c(S, i, op)$: If $op$ = $add$, compute searchable index $C_i$ and add into $S$. Otherwise, remove searchable index $C_i$ from $S$. CSC sends $S$ to CSP.

### 4.2 Application Example

Consider a technology company, where projects are executed by a group of employees including managers, engineers and sales people. Each project has its descriptive attributes. One employee may belong to several project groups, which are formed by the conjunctive combination of attributes. Within the same project group, their data are encrypted and indexed according to the policy of the company. Only the members of that group can search the data and further decrypt the data. For example, the data could be encrypted under the policy like "Member of Project No.45" and "Engineer". The employee who is a member of "Project No.45" and whose title is "Engineer" can decrypt the data. The company can prepare system parameters and create access groups through the operations Setup and GrpCreate. Similarly, the searchable index is also shared among selected employees. We note that the data encryption policy and index processing policy may be different. Usually, we allow users to retrieve the encrypted data first through Retrieve and to obtain the decryption keys later through UpdateAK. Thus, one employee can generate conjunctive searchable indexes like "Member of Project No.45", "Engineer", "Year 2012", and "Important". This data is encrypted as $CT$ and indexed as $S$ through GrpStore. Any employee possessing the valid $sk_j$ can retrieve the data. The employees who hold appropriate $AK$ can decrypt using GrpDecrypt.

One group member may leave the project when he/she finishes his/her work in the Project 45. This member may be excluded from the group because the revocation event happens. CSP has to re-encrypting the data by UpdateCT operation, while TA is required to re-issue attribute keys to all the eligible users by UpdateAK. TA is in charge of generating proxy re-keys and updating user attribute keys. TA also securely transmits re-keys to CSPs to update the ciphertexts using these re-keys. Next TA asks the data owner to adjust authorized user set of his/her data through UpdateAU and to update searchable indexes by UpdateSI. TA can also update authorized user set of a data if the revocation event of some users happens. Therefore, the revoked group member cannot search the files which he/she is not entitled to; neither can he/she learn the content of those files. When a new member joins, he/she is granted by TA the attribute keys for attribute-based decryption and the private key for trapdoor generation.

## 5. Security and Performance

In this section, we demonstrate the security and performance of our protocol. In this demonstration, only authorized group members can 1) search/retrieve the group data and 2) decrypt the retrieved group data stored in cloud storages. The employee who leaves the group or is revoked cannot retrieve or decrypt the stored data in cloud storages. Moreover, we evaluate the computation and communication costs for the CSC in our design and conclude our design is effective and efficient for the enterprise users to share data and collaborate as a group.

Table 2: Computation Cost of the CSC

| Operation | Required Basic Operations |
|---|---|
| GrpStore | $(n + m' + 2l + 2)\ SclrMul_{G_1}$,<br>$(l - 1)\ Add_{G_1}$, $2l\ HashToPoint$ |
| Retrieve | $3\ SclrMul_{G_1}$, $(2m' - 1)\ Add_{G_1}$ |
| GrpDecrypt | $(m + 1)\ Pairing$, $m\ Mul_{G_2}$ |
| UpdateAU | $1\ Exp_{G_2}$ |
| UpdateSI | $1\ Add_{G_1}$, $2\ HashToPoint$ |

Table 3: Experimental Benchmark

| Basic Operation | Operation Description | Time |
|---|---|---|
| $Mul_{G_2}$ | *multiplication in* $G_2$ | $1\ \mu s$ |
| $Add_{G_1}$ | *addition in* $G_1$ | $9\ \mu s$ |
| $Exp_{G_2}$ | *exponentiation in* $G_2$ | $0.22\ ms$ |
| $Pairing$ | *bilinear pairing* | $1.79\ ms$ |
| $SclrMul_{G_1}$ | *scaler multiplication in* $G_1$ | $2.24\ ms$ |
| $HashToPoint$ | *hash to element in* $G_1$ | $5\ ms$ |

### 5.1 Security Analysis

The security of our protocol is based on the underlying *ABEar* [18] and *muPECK* [28]. On the one hand, the *ABEar* is proved to be semantically secure under *selective-id chosen plaintext attack (IND-sID-CPA)* assuming *decisional bilinear Diffie-Hellman (DBDH)* is hard. Based on these formal arguments, we can conclude that the unauthorized entity (either CSC or CSP) cannot forge searchable indexes and searchable trapdoors since these actions are involved in solving the hard problem.

On the other hand, the *muPECK* is proved to be semantically secure under *multi-user ciphertext from random against chosen keyword attacks (IND-mCR-CKA)* assuming *decision linear Diffie-Hellman (DLDH)* is hard. Therefore, the unauthorized entity cannot calculate attribute keys for decryption, either because these actions are involved in solving the *DLDH* hard problem.

As for data dynamics, the data is re-encrypted to the same ciphertext space. The re-generated key is also distributed uniformly in the key space. Any adversary cannot gain any more advantages since he/she has to deal with the same hard problems as the ones before data/key update. In addition, the user dynamics is handled by adding/removing one part of searchable index of that user and issuing/updating the attribute keys of that group. The secrecy of the data encrypted under specified access policy can be guaranteed when group members join or leave, while the access control of search capability of group members can be assured.

### 5.2 Performance Analysis

We analyze local computation, storage, and communication cost for the CSC. Suppose that our scheme supports maximum $n$ users. To add data with $l$ keywords and for $m'$ authorized users, the CSC uses *muPECK.mPECK* to generate indexes and *ABEar.Enc* to produce ciphertext. $(n + m' + 2l + 2)$ scalar multiplication in $G_1$ $(SclrMul_{G_1})$, $(l - 1)$ addition in $G_1$ $(Add_{G_1})$ and $2l\ Hash$ are required to prepare the data stored in the cloud, where $n$ is the maximum number of attribute supported. To retrieve data, Retrieve costs $3\ SclrMul_{G_1}$ and $(2m' - 1)\ Add_{G_1}$, while to decrypt the data, $m + 1\ Pairing$ and $m$ multiplication in $G_2$ $(Mul_{G_2})$ are needed. As for key update, the CSC is required to send its old keys to exchange for updated ones. UpdateAS uses $1$ scalar multiplication in $G_1$ for one authorized user, while UpdateSI depends on $2\ Hash$ and $1$ addition in $G_1$ for the inclusion/exclusion of one single searchable index. Please refer to Table. 2. As for communication cost, the CSC has to initiate a request for GrpStore, Retrieve, UpdateAK, UpdateCT, UpdateAU, and UpdateSI. Then the CSC receives the response from the CSP. Only one round of communication is required.

The experimental benchmark is conducted using local server with Intel Xeon processor E5620 at 2.40 GHz running Ubuntu 11.10. We use *GNU multiple precision arithmetic library (GMP)* [37] and *pairing-based cryptography library (PBC)* [38] libraries. We select one supersingular curve over one base field of size $512$ bits and the embedding degree is 2. Thus the security level is set to be ECC-160 bits. The size of one group element in $G_1$ is $1024$ bits. The cost of one addition in $G_1$ costs $9\ \mu s$, while one multiplication in $G_1$ requires $2.24\ ms$. One multiplication in $G_2$ requires $1$ $\mu s$, while one exponentiation in $G_2$ costs $0.22\ ms$. Finally, the bilinear pairing needs $1.79\ ms$, and hash to $G_1$ element consumes $5.00\ ms$. (See Table 3)

## 6. Conclusion

In this paper, we propose a novel cloud storage construction enabling the management of searchable dynamic data for group collaboration. Our contributions are summarized in the following three major features of our protocol: (1) Explicitly addressing enterprise application scenario of cloud storages in terms of system architecture and functionality. (2) A novel access-control scheme for the enterprise users to share the dynamic data and collaborate as a group, and (3) A cost-effective design in terms of the enterprise user's storage, computation and communication while (2) is achieved. For the future work, we would like to further integrate other important functionalities for the enterprise such as public auditing and secure cloud data computation, to enable a full-fledged cloud storage for future enterprise applications.

## Acknowledgment

This research was supported by National Science Council, Taiwan, under contract no. NSC98-2221-E-009-079-MY3.

## References

[1] P. Mell and T. Grance, "The nist definition of cloud computing (draft) recommendations of the national institute of standards and technology," *Nist Special Publication*, vol. 145, no. 6, p. 7, 2011.

[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, and M. Zaharia, “Above the clouds: A berkeley view of cloud computing,” EECS Department, University of California, Berkeley, Tech. Rep., Feb 2009.

[3] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, “Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility,” *Future Generation Computer Systems*, vol. 25, no. 6, pp. 599 – 616, 2009.

[4] Nist, “Fips pub 197: Announcing the advanced encryption standard (aes),” *NIST*, 2001.

[5] J. Jonsson and B. Kaliski, “Public-Key Cryptography Standards (PKCS) #1: RSA Cryptography Specifications Version 2.1,” no. 3, February 2003. [Online]. Available: http://www.ietf.org/rfc/rfc3447.txt

[6] D. Boneh and M. Franklin, “Identity-based encryption from the Weil pairing,” *SIAM J. of Computing*, vol. 32, no. 3, pp. 586–615, 2003, extended abstract in Crypto’01.

[7] N. Virvilis, S. Dritsas, and D. Gritzalis, “Secure cloud storage: Available infrastructures and architectures review and evaluation,” in *Trust, Privacy and Security in Digital Business*, ser. Lecture Notes in Computer Science, S. Furnell, C. Lambrinoudakis, and G. Pernul, Eds. Springer, 2011, vol. 6863, pp. 74–85.

[8] K. Yang and X. Jia, “Data storage auditing service in cloud computing: challenges, methods and opportunities,” *World Wide Web*, vol. 15, no. 4, pp. 409–428, 2012.

[9] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “A view of cloud computing,” *Commun. ACM*, vol. 53, pp. 50–58, Apr. 2010.

[10] A. Sahai and B. Waters, “Fuzzy identity-based encryption,” in *Advances in Cryptology - EUROCRYPT 2005*, ser. Lecture Notes in Computer Science, R. Cramer, Ed. Springer, 2005, vol. 3494, pp. 557–557.

[11] M. Pirretti, P. Traynor, P. McDaniel, and B. Waters, “Secure attribute-based systems,” in *Proceedings of the 13th ACM conference on Computer and communications security*, ser. CCS ’06. New York, NY, USA: ACM, 2006, pp. 99–112.

[12] V. Goyal, O. Pandey, A. Sahai, and B. Waters, “Attribute-based encryption for fine-grained access control of encrypted data,” in *Proceedings of the 13th ACM conference on Computer and communications security*, ser. CCS ’06. New York, NY, USA: ACM, 2006, pp. 89–98.

[13] J. Bethencourt, A. Sahai, and B. Waters, “Ciphertext-policy attribute-based encryption,” in *Proceedings of the 2007 IEEE Symposium on Security and Privacy*, ser. SP ’07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 321–334.

[14] R. Ostrovsky, A. Sahai, and B. Waters, “Attribute-based encryption with non-monotonic access structures,” in *Proceedings of the 14th ACM conference on Computer and communications security*, ser. CCS ’07. New York, NY, USA: ACM, 2007, pp. 195–203.

[15] B. Waters, “Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization,” in *Public Key Cryptography - PKC 2011*, ser. Lecture Notes in Computer Science, D. Catalano, N. Fazio, R. Gennaro, and A. Nicolosi, Eds. Springer, 2011, vol. 6571, pp. 53–70.

[16] A. Boldyreva, V. Goyal, and V. Kumar, “Identity-based encryption with efficient revocation,” in *Proceedings of the 15th ACM conference on Computer and communications security*, ser. CCS ’08. New York, NY, USA: ACM, 2008, pp. 417–426. [Online]. Available: http://doi.acm.org/10.1145/1455770.1455823

[17] R. Bobba, H. Khurana, and M. Prabhakaran, “Attribute-sets: A practically motivated enhancement to attribute-based encryption,” in *Computer Security âĂŞ ESORICS 2009*, ser. Lecture Notes in Computer Science, M. Backes and P. Ning, Eds. Springer, 2009, vol. 5789, pp. 587–604.

[18] S. Yu, C. Wang, K. Ren, and W. Lou, “Attribute based data sharing with attribute revocation,” in *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security*, ser. ASIACCS ’10. New York, NY, USA: ACM, 2010, pp. 261–270. [Online]. Available: http://doi.acm.org/10.1145/1755688.1755720

[19] D. X. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” in *Security and Privacy, 2000. S P 2000. Proceedings. 2000 IEEE Symposium on*, 2000, pp. 44 –55.

[20] E.-J. Goh, “Secure indexes,” *IACR Cryptology ePrint Archive*, vol. 2003, p. 216, 2003.

[21] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, “Public key encryption with keyword search,” in *Advances in Cryptology - EUROCRYPT 2004*, ser. Lecture Notes in Computer Science, C. Cachin and J. Camenisch, Eds. Springer, 2004, vol. 3027, pp. 506–522.

[22] M. Abdalla, M. Bellare, D. Catalano, E. Kiltz, T. Kohno, T. Lange, J. Malone-Lee, G. Neven, P. Paillier, and H. Shi, “Searchable encryption revisited: Consistency properties, relation to anonymous ibe, and extensions,” in *Advances in Cryptology - CRYPTO 2005*, ser. Lecture Notes in Computer Science, V. Shoup, Ed. Springer, 2005, vol. 3621, pp. 205–222.

[23] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, “Searchable symmetric encryption: improved definitions and efficient constructions,” in *Proceedings of the 13th ACM conference on Computer and communications security*, ser. CCS ’06. New York, NY, USA: ACM, 2006, pp. 79–88.

[24] M. Bellare, A. Boldyreva, and A. O’neill, “Deterministic and efficiently searchable encryption,” in *Advances in Cryptology - CRYPTO 2007*, ser. Lecture Notes in Computer Science, A. Menezes, Ed. Springer, 2007, vol. 4622, pp. 535–552.

[25] P. Golle, J. Staddon, and B. Waters, “Secure conjunctive keyword search over encrypted data,” in *Applied Cryptography and Network Security*, ser. Lecture Notes in Computer Science, M. Jakobsson, M. Yung, and J. Zhou, Eds. Springer, 2004, vol. 3089, pp. 31–45.

[26] L. Ballard, S. Kamara, and F. Monrose, “Achieving efficient conjunctive keyword searches over encrypted data,” in *Information and Communications Security*, ser. Lecture Notes in Computer Science, S. Qing, W. Mao, J. López, and G. Wang, Eds. Springer, 2005, vol. 3783, pp. 414–426.

[27] D. Park, K. Kim, and P. Lee, “Public key encryption with conjunctive field keyword search,” in *Information Security Applications*, ser. Lecture Notes in Computer Science, C. Lim and M. Yung, Eds. Springer, 2005, vol. 3325, pp. 73–86.

[28] Y. Hwang and P. Lee, “Public key encryption with conjunctive keyword search and its extension to a multi-user system,” in *Pairing-Based Cryptography âĂŞ Pairing 2007*, ser. Lecture Notes in Computer Science, T. Takagi, T. Okamoto, E. Okamoto, and T. Okamoto, Eds. Springer, 2007, vol. 4575, pp. 2–22.

[29] Y. Yang, H. Lu, and J. Weng, “Multi-user private keyword search for cloud computing,” in *Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on*, 29 2011-dec. 1 2011, pp. 264 –271.

[30] M. Li, S. Yu, N. Cao, and W. Lou, “Authorized private keyword search over encrypted data in cloud computing,” in *Distributed Computing Systems (ICDCS), 2011 31st International Conference on*, june 2011, pp. 383 –392.

[31] N. Koblitz, A. Menezes, and S. Vanstone, “The state of elliptic curve cryptography,” *Des. Codes Cryptography*, vol. 19, pp. 173–193, March 2000.

[32] V. S. Miller, “Short programs for functions on curves,” in *IBM Thomas J. Watson Research Center*, 1986.

[33] G. Frey, M. Muller, and H.-G. Ruck, “The tate pairing and the discrete logarithm applied to elliptic curve cryptosystems,” *Information Theory, IEEE Trans. on*, vol. 45, no. 5, pp. 1717 –1719, jul 1999.

[34] R. Sandhu, E. Coyne, H. Feinstein, and C. Youman, “Role-based access control models,” *Computer*, vol. 29, no. 2, pp. 38 –47, feb 1996.

[35] B. R. Waters, D. Balfanz, G. Durfee, and D. K. Smetters, “Building an encrypted and searchable audit log,” in *NDSS*, 2004.

[36] D. Boneh and B. Waters, “Conjunctive, subset, and range queries on encrypted data,” in *Theory of Cryptography*, ser. Lecture Notes in Computer Science, S. Vadhan, Ed. Springer, 2007, vol. 4392, pp. 535–554.

[37] *GMP: The GNU Multiple Precision Arithmetic Library*, Free Software Foundation, Inc, 2006, available at http://gmplib.org/.

[38] B. Lynn, *PBC: Pairing-Based Cryptography Library*, 2008, available at http://crypto.stanford.edu/pbc/.

# Signature as Public Key Structured Cryptosystem and its Application

**Kenichi Arai**[1] **and Hiroyuki Okazaki**[2]
[1]Tokyo University of Science, 2641 Yamazaki Noda-City, Chiba 278-8510, Japan
[2]Shinshu University, 4-17-1 Wakasato Nagano-city, Nagano 380-8553, Japan

**Abstract**—*In this paper, we present a ignature as Public Key tructured ( aPK ) Cryptosystem. he aPK Cryptosystem is a method of improving the function of certain cryptographic protocols. For improving a cryptographic protocol, we often change or replace a part of the protocol. However, it is difficult to analyze the security of the protocol improved in this manner. In the aPK Cryptosystem, we add a new function without modifying the original protocol. he main concept of our aPK Cryptosystem is to use a signature as the public key of another cryptographic protocol. oreover, we show the improvement of one of a ring signature scheme using the aPK Cryptosystem and present a method to reply to anonymous signers of a ring signature. ur method enables the generation of a ciphertext that only the signer of a ring signature can decrypt using the aPK Cryptosystem.*

**Keywords:** Using Digital Signature as Public Key, Anonymity, Ring Signature, Security of Improved Protocol

## 1. Introduction

In this paper, we present a Signature as Public Key Structured Cryptosystem (hereinafter called the SaPKS Cryptosystem) as a method of improving certain cryptographic protocols. For improving a cryptographic protocol, we often change or replace a part of the protocol. However, analyzing the security of such a protocol is problematic. In the SaPKS Cryptosystem, we add a new function without modifying the original protocol. In this paper, we describe a method to improve digital signature schemes, which can also be applied to other cryptographic protocols. The basic concept of the SaPKS Cryptosystem is using a signature as the public key of another cryptographic protocol.

A typical digital signature scheme consists of three algorithms: the key generation algorithm $KeyGen$, the signing algorithm $Sign$, and the verification algorithm $Ver$. First, the user generates a key pair, private key $x$, and public key $y$ using $KeyGen$. Second, the user generates signature $\sigma$ to sign a message $m$ using $Sign$ with the private key. Next, the validity of signature $\sigma$ can be verified using $Ver$. Now, suppose that a signature is used as a public key. In other words, a digital signature scheme works as the key generation algorithm of another cryptographic protocol, for example, as a public key cryptosystem or a digital signature scheme, then we can realize a "SaPKS Cryptosystem" or a "SaPKS Signature," respectively.

In the SaPKS Cryptosystem, the user generates a key pair of the SaPKS Cryptosystem using a digital signature scheme. The private key $x_{_{SaPK}}$ of the SaPKS Cryptosystem is the same as $x$, i.e., the private key of the signature scheme. The public key $y_{_{SaPK}}$ of the SaPKS Cryptosystem is the same as $\sigma$, i.e., the signature generated by the user to sign a message $m$ using $Sign$ with the private key $x$. We use the digital signature scheme as the key generation algorithm of the SaPKS Cryptosystem. Anyone having signature $\sigma$ can encrypt a message $m_{_{SaPK}}$ into the ciphertext $c$ using the encryption algorithm $Enc$ with the public key of the SaPKS Cryptosystem, $y_{_{SaPK}} = \sigma$. Then, the signer of $\sigma$ can decrypt $c$ using the decryption algorithm $Dec$ with the private key of the SaPKS Cryptosystem, $x_{_{SaPK}} = x$.

From a practical standpoint, the SaPKS Cryptosystem would be worthless if we employ an ordinary digital signature scheme as the key generation algorithm. This is because it is essential to verify signature $\sigma$ using the public key of the signature scheme. However, some well-known signature schemes such as Group Signature[1] and Ring Signature[2] preserve signer anonymity. Generally, in these signature schemes, any member of a group can generate a signature such that a public verifier can verify that the signature is generated by a group member; however, the verifier cannot identify the signer. If we use such a signature scheme as the key generation algorithm of the SaPKS Cryptosystem, we can generate a ciphertext that only the signer of $\sigma$ can decrypt without losing anonymity.

Moreover, we show the improvement of a signature scheme using the SaPKS Cryptosystem. We present a way to reply to the signer of the Linkable Ring Signature scheme, presented by Liu, Wei, and Wong[3]. In ring signature schemes, the signer of a signature cannot be identified. There are several Ring Signature applications including whistle-blowing and auctions. In these applications, it is useful if we can secretly send messages to the anonymous signer of the ring signature without losing signer anonymity. Then, we propose a method of generating ciphertexts that only the signer of the ring signature can decrypt using the SaPKS Cryptosystem.

The remainder of this paper is organized as follows. In Section 2, we present the definitions of the SaPKS

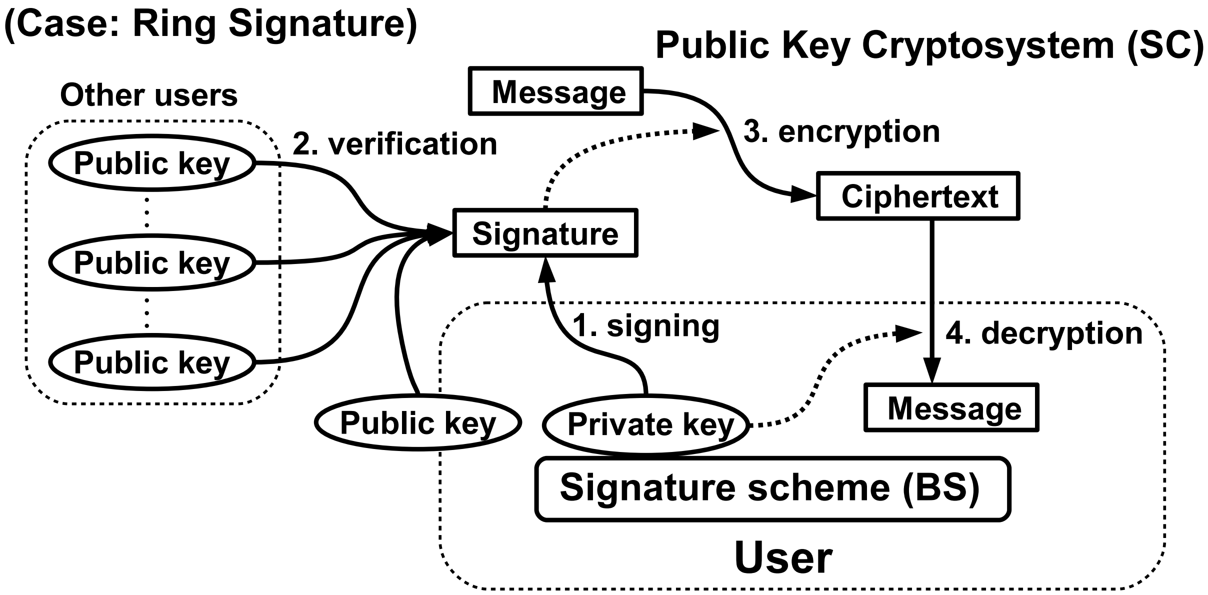


Figure 1: Concept of the SaPKS Cryptosystem

Cryptosystem. In Section 3, we present an application of the SaPKS Cryptosystem, i.e., a method to reply to the signer of LSAG signature. In Section 4, we discuss the security of the scheme presented in Section 3. We conclude the paper in Section 5.

## 2. Definitions of the SaPKS Cryptosystem

In this section, we present the definitions and notations of the SaPKS Cryptosystem.

In the SaPKS Cryptosystem, we employ two cryptographic schemes: the Base Signature scheme (BS) and the SaPKS public key Cryptosystem (SC). We consider BS as the key generation algorithm of SC.

First, the user generates a key pair of SC, $(x_{_{SaPK}}, y_{_{SaPK}})$. The private key $x_{_{SaPK}}$ of SC is the same as $x$ of the BS. The public key $y_{_{SaPK}}$ of SC is the same as $\sigma$, i.e., the signature of BS generated by the user to sign a message $m$ with the private key $x$. Once the user makes signature $\sigma$ public, everybody will be able to generate the ciphertext, which only the signer of $\sigma$ can decrypt. Figure 1 shows the concept of the SaPKS Cryptosystem.

BS and SC should have the same trapdoor. As long as this principle holds, we can separately consider BS and SC. We will originally propose cryptographic schemes for BS and SC and will employ known secure stand-alone schemes as BS and SC.

In this paper, we refer to the SaPKS Cryptosystem whose SC is A and BS is B as “SaPKS A on B.” For example, if SC is the ElGamal public key cryptosystem and BS is the Schnorr signature scheme, we refer to the SaPKS Cryotosystem as “SaPKS ElGamal public key cryptosystem on Schnorr signature.”

Note that it is not necessarily that we can employ any combination of a digital signature scheme and a cryptosystem. At least, the signature scheme and the cryptosystem should have the same trap door.

### 2.1 Base Signature as the Key Generation Algorithm of the SaPKS Cryptosystem

The BS signature scheme is composed of three algorithms: the key generation algorithm $Gen()$, the signing algorithm $Sig()$, and the verification algorithm $Ver()$. First, the user generates the private key $x$ and public key $y$ using $Gen()$. The user can generate a signature $\sigma = Sig(m, x)$ to sign a given message $m$. A public verifier verifies a given signature $\sigma$ with $Ver(\sigma, m, y)$. $Ver()$ outputs “Accept” if signature $\sigma$ is generated with the private key $x$ corresponding to the public key $y$; otherwise, $Ver()$ outputs “Reject”.

The private key $x_{_{SaPK}}$ of SC is the same as $x$ of BS. That is, the user decrypts ciphertexts of SC with $x_{_{SaPK}} = x$. The public key $y_{_{SaPK}}$ of SC is the same as $\sigma$, which the verification algorithm $Ver()$ accepts. Anybody who wants to communicate with the signer of $\sigma$ will be able to generate the ciphertexts of SC using $\sigma$ as the public key of SC.

Note that the SaPKS Cryptosystem would be worthless if we employ an ordinary digital signature scheme as BS. This is because we must use the public key of BS to verify $\sigma$. However, if we employ the signature schemes that preserve signer anonymity, such as Group Signature[1] and Ring Signature[2], we can verify $\sigma$ without losing signer anonymity. We present such a SaPKS Cryptosystem whose BS is a Linkable Ring Signature in Section 3.

### 2.2 Encryption of the SaPKS Cryptosystem

Let $Enc_{_{SaPK}}()$ be the encryption algorithm of SC. We can encrypt a message $m'$ into the ciphertext $c$ as follows:

$$c \;=\; Enc_{_{SaPK}}(m', y_{_{SaPK}}).$$

Here, the public key of SC $y_{_{SaPK}}$ is the same as signature $\sigma$ of BS.

### 2.3 Decryption of the SaPKS Cryptosystem

Let $Dec_{_{SaPK}}()$ be the decryption algorithm of SC.

The signer of $\sigma$ can decrypt the cipher text $c$ with the private key $x_{_{SaPK}}$ of SC as follows:

$$m' \;=\; Dec_{_{SaPK}}(c, x_{_{SaPK}}).$$

Here, the private key $x_{_{SaPK}}$ of SC is the same as $x$ of BS.

### 2.4 Security of the SaPKS Cryptosystem

In this section, for clarity, BSS refers to the scheme for BS and SCS refers to the scheme for SC. Both BSS and SCS are independent protocols. Therefore, we can separately analyze each protocol's security. In addition, we can employ known secure stand-alone schemes as BSS and SCS. However, it is unclear whether the SaPKS Cryptosystem composed of BS and SC is secure even if both BSS and SCS are secure when they are separately used. This problem can be resolved by employing UC–secure[4] cryptographic schemes as both BS and SC. However, it is interesting to discuss the security of the composition of non–UC–secure cryptographic schemes because they are widely used at present.

Now, we discuss the security of the SaPKS Cryptosystem composed of BS and SC. Let BSsec and SCsec be certain security requirements. We assume that stand-alone BSS and SCS satisfy BSsec and SCsec, respectively. Let BSAdv and SCAdv be polynomial time algorithms that can break BSsec and SCsec, respectively. Furthermore, let cSCAdv be the polynomial time algorithm that can break SCsec when BSS and SCS are used in the SaPKS Cryptosystem composed of BS and SC. Naturally, such BSAdv and SCAdv do not exist by the assumption mentioned above. If cSCAdv can break the SCsec that SCAdv cannot break in the stand-alone usage of SC, cSCAdv might obtain some information to break SCsec from BS. If so, we may break BSsec using cSCAdv. Thus, we analyze the security of the SaPKS Cryptosystem with the following strategy.

First, we show that we can reduce cSCAdv into BSAdv$'$, which breaks BSsec in polynomial time. Figure 2 and 3 show the model of adversaries.

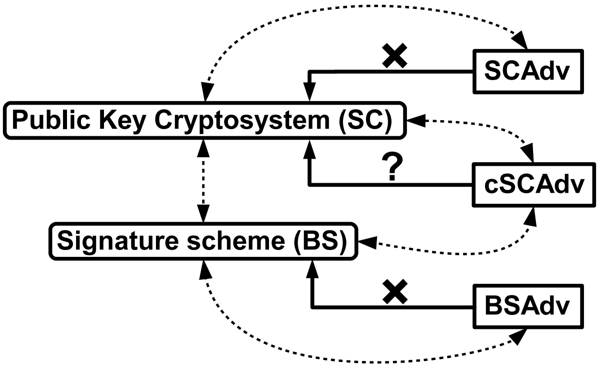


Figure 2: Model of Adversaries

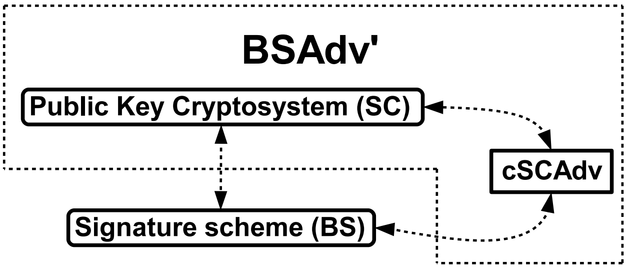


Figure 3: BSAdv$'$

Next, we evaluate BSAdv$'$. If equation (1) holds, BSAdv$'$ is indistinguishable from BSAdv in computational complexity:

$$\left| Prob\begin{pmatrix}\text{BSAdv}\\ breaks\\ \text{BSsec}\end{pmatrix} - Prob\begin{pmatrix}\text{BSAdv}'\\ breaks\\ \text{BSsec}\end{pmatrix}\right| \leq \frac{1}{\mu(k)}, \qquad (1)$$

where $k$ is a security parameter, $\mu$ is any polynomial, and $\frac{1}{\mu(k)}$ is negligible probability. Then, we can break BSsec using cSCAdv. Roughly speaking, we discuss the security of the SaPKS Cryptosystem by reducing it to that of BS as the difficult problem. Our strategy is similar to reducing the security of a general cryptosystem into difficult problems such as the Discrete Logarithm Problem (DLP) and Factoring.

## 3. Construction of a SaPKS Cryptosystem

Liu, Wei, and Wong presented, Linkable Spontaneous Anonymous Group (LSAG) signature, which is an interesting signature scheme[3]. The LSAG signature scheme is a Linkable Ring Signature scheme. In this section, we propose a method of generating ciphertexts that only the signer of the ring signature can decrypt using the SaPKS Cryptosystem.

## 3.1 Linkable Spontaneous Anonymous Group Signature

In this section, we briefly introduce Liu et al.'s LSAG signature[3]. LSAG signature satisfies the following security requirements.

**Existential Unforgeability:**
Given the public keys of all group members with no corresponding private keys, an adversary cannot forge a signature for any message.

**Signer Ambiguity:**
It is infeasible to identify which the private key was actually used in generating a given LSAG signature to find out which the member has generated a signature.

**Linkability:**
Two LSAG signatures with the same public key list $L$ are linked if they are generated with the same private key.

Let $G = \langle g \rangle$ be a group of prime order $q$. Let $H_1 : \{0,1\}^* \to \mathbb{Z}_q$ and $H_2 : \{0,1\}^* \to G$ be hash functions. For $i = 1, \ldots, n$, each user $\mathcal{U}_i$ picks the private key $x_i$ and then calculates the public key $y_i = g^{x_i}$. Furthermore, let $L = \{y_1, \ldots, y_n\}$ be the list of $n$ members public keys.

### 3.1.1 Signing

To sign a given message $m \in \{0,1\}^*$, the user $\mathcal{U}_\pi$ calculates the signature $\sigma = LsagSig(m, x_\pi, L)$. $LsagSig()$ is the signing procedure given as follows:

**Step 1 :**Calculate $h = H_2(L)$ and $\tilde{y} = h^{x_\pi}$.

**Step 2 :**Pick $u \in \mathbb{Z}_q$ at random, and calculate

$$c_{\pi+1} = H_1\left(L, \tilde{y}, m, g^u, h^u\right).$$

**Step 3 :**For $i = \pi+1, \ldots, n, 1, \ldots, \pi-1$, pick $s_i \in \mathbb{Z}_q$ at random, and calculate

$$c_{i+1} = H_1\left(L, \tilde{y}, m, g^{s_i} y_i^{c_i}, h^{s_i} \tilde{y}^{c_i}\right).$$

**Step 4 :**Calculate $s_\pi = u - x_\pi c_\pi \mod q$, and then output the signature

$$\sigma = \{c_1, s_1, \ldots, s_n, \tilde{y}\}.$$

### 3.1.2 Verification

A public verifier verifies a given signature $\sigma$ with the verification procedure $LsagVer(\sigma, m, L)$ as follows:

**Step 1 :**Calculate $h = H_2(L)$.

**Step 2 :**For $i = 1, \ldots, n$, calculate $z'_i = g^{s_i} y_i^{c_i}$, $z''_i = h^{s_i} \tilde{y}^{c_i}$, and

$$c_{i+1} = H_1\left(L, \tilde{y}, m, z'_i, z''_i\right).$$

**Step 3 :**Accept $\sigma$ if $c_1 = H_1\left(L, \tilde{y}, m, z'_n, z''_n\right)$, otherwise reject it.

### 3.1.3 Linking

Let $\sigma' = \{c'_1, s'_1, \ldots, s'_n, \tilde{y}'\}$ and $\sigma'' = \{c''_1, s''_1, \ldots, s''_n, \tilde{y}''\}$ be valid signatures signed with the same public key list $L$. If $\tilde{y}' = \tilde{y}''$, $\sigma'$ and $\sigma''$ are generated by the same signer. Otherwise, the two signatures are generated by two different signers.

## 3.2 SaPKS Cryptosystem whose BS is LSAG Signature

In this section, we show the way of constructing a SaPKS Cryptosystem on LSAG signature. We employ LSAG signature as BS. SC should have the same trap door as BS, i.e., DLP. Thus, the SC of this scheme is a public key cryptosystem based on DLP.

Let $\sigma(j) = \{c_1(j), s_1(j), \ldots, s_n(j), \tilde{y}(j)\}$ be the BS signature signed on to the message $m(j)$ by the signer $\mathcal{U}_j$. The private key of $\mathcal{U}_j$ is $x_j$. The public key of $\mathcal{U}_j$ is $y_j = g^{x_j} \in G$ that belongs to the public key list $L = \{y_1, \ldots, y_j, \ldots, y_n\}$. In our SaPKS Cryptosystem, we can compute a ciphertext that only $\mathcal{U}_j$, which is the signer of $\sigma(j)$, can decrypt. Although, we cannot know who had signed $\sigma(j)$. Here, we employ BS as the key generation algorithm of SC. Therefore, $\sigma(j)$ is the public key of SC and $x_j$ is the private key of SC.

Let $Enc_{_{SaPK}}()$ and $Dec_{_{SaPK}}()$ be the encryption and decryption algorithms of SC, respectively. The following relationship holds

$$\begin{aligned} c &= Enc_{_{SaPK}}(m, g, y), \\ m &= Dec_{_{SaPK}}(c, x), \end{aligned}$$

where $y = g^x$.

### 3.2.1 Encryption

We can encrypt a message $m'$ into the ciphertext $c$ as follows:

$$c = Enc_{_{SaPK}}(m', h, \tilde{y}(j)),$$

where $\tilde{y}(j)$ is the last term of the BS signature $\sigma(j) = \{c_1(j), s_1(j), \ldots, s_n(j), \tilde{y}(j)\}$. Note that $\tilde{y}(j) = h^{x_j}$.

### 3.2.2 Decryption

The signer of $\sigma(j)$, $\mathcal{U}_j$, can decrypt the ciphertext $c$ with the private key $x_j$ as follows:

$$m' = Dec_{_{SaPK}}(c, x_j).$$

Here, $\mathcal{U}_j$ used $x_j$ to generate the BS signature $\sigma(j)$.

## 3.3 Application of SaPKS Cryptosystem on LSAG Signature: How to Reply to an Anonymous Signer Secretly

We outline an application of the SaPKS Cryptosystem on LSAG signature. There are some interesting applications of LSAG signature, including whistle-blowing and auctions. In

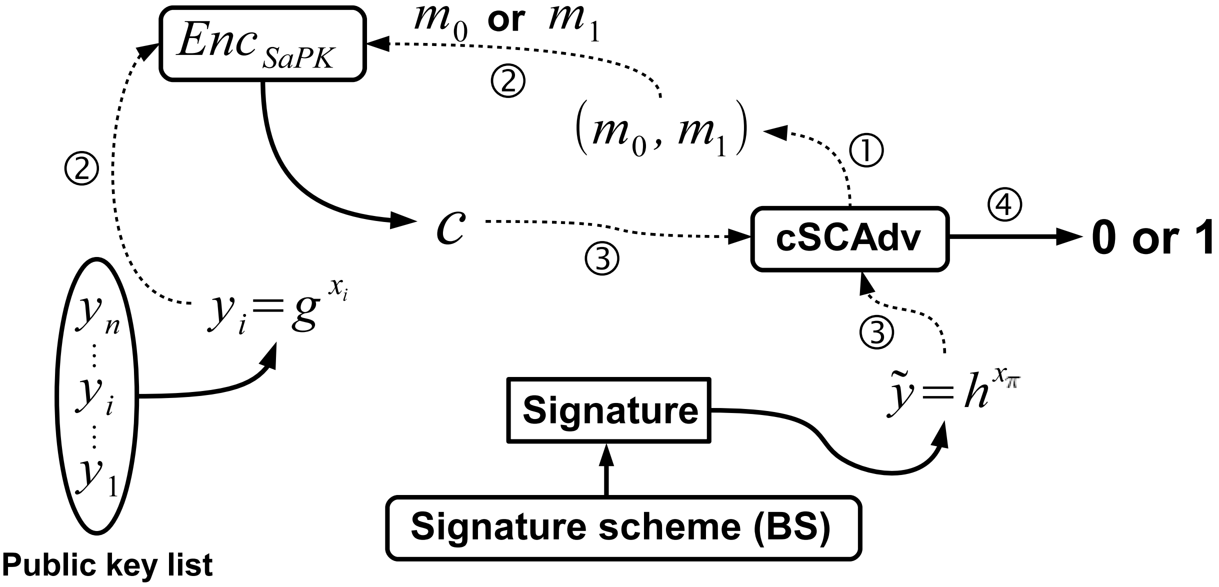


Figure 4: BSAdv′ of our SaPKS Cryptosystem on LSAG Signature

each application, LSAG signature achieves both anonymity of signers and linkability of signatures. Linkability prevents the injustice caused by the anonymity of the signers. Our SaPKS Cryptosystem enables to secretly communicate with the signer of LSAG signature, whistle-blower, and winning bidder. For example, we can achieve such goals by sending a ciphertext of SC to all users who might be the signers.

We can achieve a similar protocol without the SaPKS Cryptosystem if a public key is in the signed message. However, key management becomes easy in our SaPKS Cryptosystem because each user has to keep only the private key of BS secretly.

## 4. Security of the SaPKS Cryptosystem on LSAG Signature

In this section, we analyze the security of the SaPKS Cryptosystem on LSAG signature, presented in Section 3, on the basis of the discussion in Section 2.4. LSAG signature is *signer ambiguity* provided the Decisional Diffie-Hellman Problem (DDHP) is hard, in the random oracle model[5]. Now, we assume that the SCS of our SaPKS Cryptosystem, a public key cryptosystem scheme based on DLP, holds *indistinguishability* if we use stand-alone SCS. We then show that we can break the *signer ambiguity* of LSAG signature if the *indistinguishability* of SC is broken when we use SCS as the SC of our SaPKS Cryptosystem. In this section, we refer an adversary who can break the *signer ambiguity* of a stand-alone LSAG signature as BSAdv. We then refer to the adversary who can break the *indistinguishability* of the public key cryptosystem scheme of our SaPKS Cryptosystem's SC as cSCAdv.

First, we define the adversary cSCAdv that can break the *indistinguishability* of the public key cryptosystem scheme $(Enc_{_{SaPK}}(), Dec_{_{SaPK}}())$ as follows:

**Step 1 :**cSCAdv outputs a pair of messages $(m_0, m_1)$ at random.

**Step 2 :**We choose either $m_0$ or $m_1$ as the message $m_j$ at random, and encrypt $m_j$ into the ciphertext

$$c \ = \ Enc_{_{SaPK}}(m_j, g, y)$$

with the public key $y = g^x$.

**Step 3 :**We ask cSCAdv which $m_0$ or $m_1$ was encrypted into $c$. cSCAdv outputs the answer

$$\text{cSCAdv}(c, h, \tilde{y}) \rightarrow j$$

if $c = Enc_{_{SaPK}}(m_j, g, y)$, where $j = 0$ or 1; otherwise cSCAdv outputs 0 or 1 with a probability of $\frac{1}{2}$, respectively.

The latter case is, for example, when $c$ is not the ciphertext of either $m_0$ or $m_1$ or $c$ is encrypted with another public key.

Next, we define the adversary that can break the *signer ambiguity* of LSAG signature as follows:

$$\begin{aligned} \text{BSAdv}(\sigma, g, y_i) \ \ \rightarrow \ \ & 1 \quad if \quad i = \pi, \\ & 0 \quad if \quad i \neq \pi, \end{aligned}$$

where $\sigma = LsagSig(m, x_\pi, L)$ is the LSAG signature signed by the user $\mathcal{U}_\pi$, $x_i$ is the private key of the user

$\mathcal{U}_i$, $y_i = g^{x_i}$ is the public key of $\mathcal{U}_i$, and $L = \{y_1, \ldots, y_n\}$ is the public key list of $n$ users. LSAG signature is *signer ambiguity* provided DDHP is hard, in the random oracle model. Therefore, such a BSAdv does not exist under the assumption that DDHP is hard, in the random oracle model.

Now, we can achieve BSAdv$'$ using cSCAdv in the following game:

**Step 1 :**Given $\sigma$, a LSAG signature, and the public key list $L$.
**Step 2 :**Call cSCAdv, and receive $(m_0, m_1)$ from cSCAdv.
**Step 3 :**Choose either $m_0$ or $m_1$ as the message $m_j$ at random, and encrypt $m_j$ into the ciphertext $c = Enc_{_{SaPK}}(m_j, g, y_i)$ with the user $\mathcal{U}_i$'s public key $y_i$.
**Step 4 :**If the output of cSCAdv is correct, i.e., cSCAdv$(c, h, \tilde{y}) = j$, continue this game up to k times, otherwise output $0$ and terminate this game.
**Step 5 :**Output $1$ if all answers of $k$ times are correct.

Here, $\tilde{y}$ in the LSAG signature $\sigma$ holds $h \in G$, $\tilde{y} = h^{x_\pi} \in G$, and the $y_\pi$ in the public key list $L$. Moreover, the public key of the signer of $\sigma$ holds $y_\pi = g^{x_\pi} \in G$. Figure 4 shows BSAdv$'$.

The probability that BSAdv$'$ mistakes the judgment of the signer is $2^{-k}$. Thus, roughly estimating,

$$|Prob(\text{BSAdv}(\sigma, g, y_i)) - Prob(\text{BSAdv}'(\sigma, g, y_i))| \simeq 2^{-k}.$$

Therefore, any polynomial time algorithm can distinguish between BSAdv and BSAdv$'$ at negligible probability. Thus, we can break *signer ambiguity* of LSAG signature if the *indistinguishability* of SC is broken when we use SCS as the SC of our SaPKS Cryptosystem. Consequently, our SaPKS Cryptosystem holds *indistinguishability* if stand-alone LSAG signature holds *signer ambiguity*.

## 5. Conclusion

In this paper, we have presented the concept of a SaPKS Cryptosystem. In addition, we presented an application of the SaPKS Cryptosystem, a method to reply to the signer of LSAG signature, and analyzed the security of our SaPKS Cryptosystem. We believe that this SaPKS Cryptosystem can be an efficient way to enhance various signature schemes. In the future, we would like to discuss other applications based on a similar concept; for example, SaPKS Signature where the signer of a signature of BS can generate another signature to sign another message.

## References

[1] D.Chaum, E.van Heijst, *roup ignature*, Advances in Cryptology - EUROCRYPTO'91, pp.257–265,Springer-Verlag, 1991.

[2] R.Rivest, A.Shamir, and Y.Tauman, *How to leak a secret* ASIACRYPTO 2001, pp.552–565, LNCS Vol.2248, Springer-Verlag, 2001.

[3] J.K.Liu, V.K.Wei, and D. S.Wong, *inkable pontaneous Anonymous roup ignature for Ad Hoc roups*, ACISP 2004, pp.325–335, LNCS Vol. 3108, Springer-Verlag, 2004.

[4] R.Canetti, *niversally Composable ecurity, A ew Paradigm for Cryptographic Protocols*, Proceedings of the 42nd Foundations of Computer Science conference, pp.136–145, 2001.

[5] M.Bellare, P.Rogaway, *andom oracles are practical a paradigm for designing efficient protocols*, In First ACM Conference on Computer and Communications Security, pp.62–73, 1993.

# NOVEL WEIGHT BASED INTRA CLUSTER GROUP KEY MANAGEMENT APPLYING RSA ALGORITHM

**Dr.Meera Gandhi**[1] **Roop lata kumari**[2] **K.Gomathi**[3] **Dhiraj Kumar**[4]

Professor, Dept of CSE,
Sathyabama University,
Chennai, India

PG Student,
Sathyabama University,
Chennai, India

Research Scholar
Sathyabama University,
Chennai, India,

PG Student,
Sathyabama University,
Chennai, India

**Abstract**-*Key management is vital part of security; this issue is even bigger in wireless network compared to wired network. The distribution of keys in an authenticated manner is a difficult task in WMANETS and when a node leaves or joins, it need to generate new session key to maintain forward and backward secrecy. This paper reviewed technological solutions for managing keys by dividing the network into clusters. A clustering architecture increases network lifetime and fault tolerance results in more efficient use of network resources. Cluster Head (CH) will maintain the group key; it will also update the group key whenever there is a change in the membership. And the CH is responsible for intra-cluster communication. A Secondary Cluster Head (SCH) is also elected to avoid the CH from becoming a bottleneck, and also it acts a monitoring node for CH lifetime. The combination of Weight based Intra Cluster Routing Protocol (WICRP) and RSA has been proposed to secure multicast key distribution in which the source node uses Ad hoc On-Demand Distance Vector (AODV) routing protocol to reach its destination. The weight based clustering approach is based on combined weight metrics that takes into account of several system parameters like the degree difference, transmission range, battery power and mobility of each node. The performance of the system is evaluated based on the few metrics like Packet Delivery Ratio (PDR), and end to end delay. As demonstrated, our algorithm reduces frequent head election phenomena by having SCH, thus improves the overall performance and reduces energy utilization.*

**Keywords:** WMANETs**,** Cluster Head, Secondary Cluster Head, AODV, RSA, Group Key Security

## 1. Introduction

Wireless Mobile Ad hoc Networks (WMANETs) are deployed in difficult environments where interruption of connectivity is consistent on occurrences. Wireless Mobile Ad hoc Networks (WMANETs) [1] have less infrastructure, multi-hop, dynamic networks for a collection of heterogeneous mobile nodes. It consists of autonomous nodes that communicate with each other, most frequently using a multi-hop wireless network. Nodes do not necessarily know each other and come together to form an ad hoc group. Secure group communication (SGC) [16] is defined as the process by which members in a group can securely communicate with each other and the information being shared is inaccessible to outside members. In such a scenario, a group key is established among all the participating members and this key is used to encrypt all the messages destined to the group. As a result, only the group members can decrypt the messages. Due to dynamic behavior of the MANET, secret key used for communication is need to be updated whenever any node joins or leaves the network in order to maintain the forward and backward secrecy within the network. If the network is large and the mobility is higher, key updating is a frequent phenomenon. It also consumes more battery power. In order to manage the energy utilization, a network is divided into clusters by using Weight based Clustering and the keys are generated using well known RSA Algorithm. The rekeying process will be performed only if there is any mobility of nodes in the clusters. The rest of this paper is organized as follows. Section 2 presents related work done in Clustering and key management approaches. Section 3 presents the proposed Weight based Clustering and Group Key Agreement Section 4 presents performance evaluation and finally, Section 5 presents conclusions and the upcoming work.

## 2. Related work

Key management has hang around a challenging issue in wireless networks due to the constraints of node resources. Majority of research on security of ad hoc networks emphasize the secure routing protocols. There are some proposals on key generation and distribution issues. At the same time key management will be a tedious process when we have large network due to high mobility of nodes. To manage these issues, various cluster based routing schemes have been proposed in the literature namely low-maintenance clustering approach [8] mobility-based clustering approach, Weight-Based Clustering approach, Flooding-Based Clustering approach, Channel Based Clustering. Clustering protocols are categorized into different approaches based on its distinguished features. Tree structure of key management approaches is as follows.

## 1.2 Types of key management

The group key management protocols are typically categorized into three headings.

*Centralized Group Key Distribution (CGKD)*

In CGKD, there exists a central entity (i.e. a group controller (GC)) which is responsible for generating, distributing, and updating the group key e.g. Logical Key Hierarchy (LKH) and One Way Function (OWF) [4]

*De-Centralized Group Key Management (DGKM)*

The DGKM approach involves splitting a large group into small subgroups. Each subgroup has a subgroup controller which is responsible for the key management of its subgroup .e.g. IOLUS [4]

*Distributed/Contributory group Key Agreement (CGKA)*

The CGKA schemes involve the participation by all members of a group towards key management. Such schemes are characterized by the absence of the GC. The group key in such schemes is a function of the secret shares contributed by the members. Typical CGKA schemes include binary tree based ones [4] and n-party Diffie-Hellman key agreement [5, 6]. RSA is a group key management scheme proposed in [4]. The basic idea is to combine the efficiency of the tree structure with the contributory feature of DH.

## 1.3 Clustering protocols

Clustering protocols are categorized into different approaches based on its distinguished features.

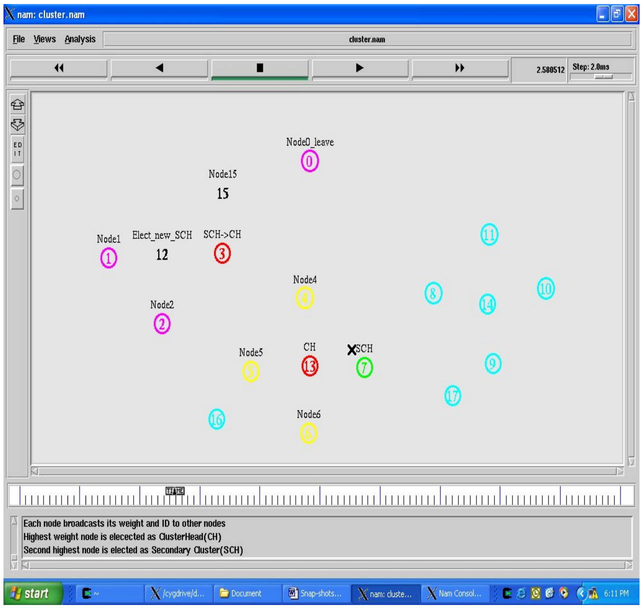
*Low-Maintenance Clustering* approach [9, 10, and 11] provides a stable cluster structure incurring less maintenance cost. *Lowest-ID*: Elect the node as a cluster head that has the lowest ID relative to its neighbors. *Maximum Connectivity Clustering* (MCC) [13] The MCC uses the degree of connectivity instead of the node ID in the cluster head election.

*Mobility-Based Clustering* approach considers mobility feature of the mobile nodes for cluster formation. It achieves maximum cluster stability by grouping mobile nodes of similar patterns into a single cluster. *MOBIC* uses the mobility metric as a basis of cluster formation and cluster head selection.

*Weight-Based Clustering* approach [2, 3] takes weight of the mobile nodes into consideration for the choice of the CH.

*On-Demand Weighted clustering algorithm (On-Demand WCA) :* The assignment of weight to a mobile node is the combined effect of several system parameters like ideal node degree, degree difference, transmission power, cluster head serving time and mobility. The advantage of this clustering scheme is the flexibility of adjusting the weighting factors for each system parameter to make it suitable for different scenarios.

*Flooding-Based Clustering* approach forms the cluster by disseminating information over the whole Network. *Max min D-cluster algorithm [7]: This* allows the control and flexibility in the determination of the CH density .*Channel Based Clustering* facilitates efficient utilization of channels by scheduling transmissions of the mobile nodes.

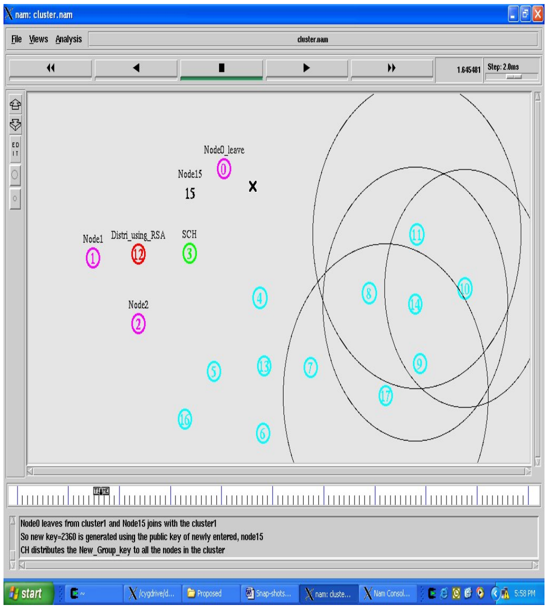


```
Each node broadcasts its weight and ID to other nodes
Highest weight node is elecected as ClusterHead(CH)
Second highest node is elected as Secondary Cluster(SCH)
```

Figure 1 Selection of Cluster Head

# 3. Proposed work

In our proposed work, the combination of eight based Clustering and RSA algorithm for secure multicast key distribution is applied, in which source node uses AODV routing protocol to reach its destination.



Node0 leaves from cluster1 and Node15 joins with the cluster1
So new key=2360 is generated using the public key of newly entered, node15
CH distributes the New_Group_key to all the nodes in the cluster

Figure 2 Selections of SCH and Intra Communication

The CH is responsible for cluster management, membership maintenance and Group Key Distribution. To begin with, all the nodes are assigned an id, its private key and public key.

The reason behind using WeightBased Clustering approach in this paper takes weight of the mobile nodes into consideration for the choice of the CH. The Weight based clustering approach is based on combined weight metric that takes into account of several system parameters like the degree difference of the node, transmission range, battery power and mobility of the node. But in existing clustering algorithms like MCC, MOBIC and lowest ID, any one of the weight metric is used for electing the CH. Rekeying is done by Cluster Head whenever any node joins or leaves the network to ensure backward secrecy (i.e., a new member should not know the previous information that was exchanged) and forward secrecy (i.e., an existing member should not receive the information exchanged after it leaves the network).Figure 1 shows selection of Cluster Head and Figure 2 depicts the selection of SCH and intra communication.

Key management is an essential cryptographic primitive upon which other security primitives such as privacy, authenticity and integrity are built. However, none of the existing key management schemes are suitable for ad hoc networks. The major limitation of these schemes is that most of them rely on a Trusted Third Party (TTP), thus not fulfilling the self-organization requirement of an ad hoc network. Special mechanisms and protocols though designed specifically for ad hoc network. For this reason, Distributed/Contributory group Key Agreement (CGKA) approach is used for establishing the group key by the contribution of cluster members.

## 3.1 Phases in Clustering and key management

The key management system consists of two phases:

*Initialization phase:* At first, weight will be computed for every node based on the factors like degree difference of the node, transmission range, and battery power and mobility of the sensor node. This module consists of following sub modules like Cluster Head election and cluster formation. Figure 3 depicts the flow Diagram of CH formation and Key Management using RSA.

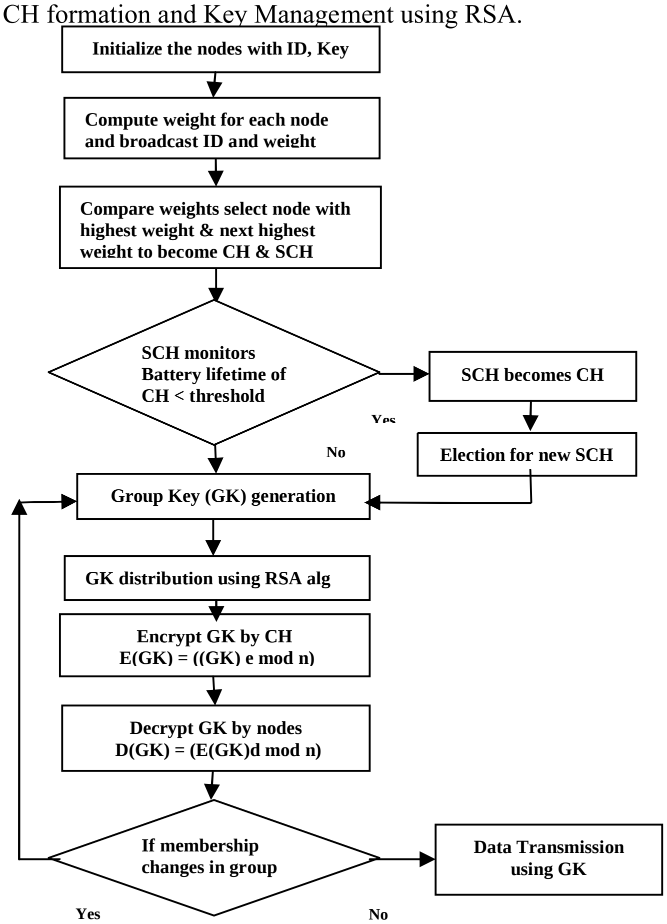

Initialize the nodes with ID, Key
Compute weight for each node and broadcast ID and weight
Compare weights select node with highest weight & next highest weight to become CH & SCH
SCH monitors Battery lifetime of CH < threshold
Yes
SCH becomes CH
Election for new SCH
No
Group Key (GK) generation
GK distribution using RSA alg
Encrypt GK by CH E(GK) = ((GK) e mod n)
Decrypt GK by nodes D(GK) = (E(GK)d mod n)
If membership changes in group
Data Transmission using GK
Yes
No

Figure 3 Flow chart for CH election and Group Key Management

*a) Weight Computation:* Each node is assigned a random ID value. It broadcasts its ID value to its neighbors and builds its neighborhood table. Each node calculates its own weight based on the following factors like Node degree difference, Energy sustainment, Mobility, distance from all other neighboring nodes. The distance between nodes and mobility is considered to keep the balance between clusters. The weight computation W for all the weights is given as follows in equation (1).

$$W_n = W_1 * \Delta_n + W_2 * E_n - W_3 * M_n + W_4 * D_n \quad \text{-----}\ \mathbf{(1)}$$

Where $\Delta_n -$ Degree Difference of node

$E_n$ -Energy in each node represented by Joules

$M_n$ - Mobility of each node. (Less mobility nodes have more probability to become a CH)

$D_n$ - Distance from all other neighboring nodes

The co-efficient used in weight calculations $W_1, W_2, W_3, W_4$ are assumed as follows. $W_1 = 0.5, W_2 = 0.35, W_3 = 0.05, W_4 = 0.1$. The sum of these co-efficient is 1. The factors degree difference and energy are given more importance and assumed higher co-efficient values 0.5 and 0.35. The combined weight is calculated by using the parameters of $\Delta_n, E_n, M_n, D_n$ from the equations 2,3,4,5 respectively. After finding its own weight, each node transmits its weight to its neighbors based on neighborhood table. The neighborhood table consists of one hop reachable nodes; its weights. It is maintained by the CH.

**Degree difference:**

It is defined as the difference between the cluster size N and the actual number of neighbours $d_n$.From the equation (2), it is known that $\Delta_n -$ Degree Difference of node 'n'. In order to find the Degree $d_n$ of the node 'n' by counting its neighbors. Compute the Degree difference for the node 'n', where N is a threshold for the cluster's size.

$$\Delta_n = \left| d_n - N \right| \quad \text{---------------}\ \mathbf{(2)}$$

**Energy:**

Energy in each node represented by Joules. It is represented by $E_n$ - Energy (Battery Power) of node 'n'. Energy $E_n$ is calculated as

$$E_n = E_0 - E_{residual} \quad \text{--------------}\ \mathbf{(3)}$$

$E_0$ and $E_{residual}$ are initial and remaining energy of node 'n'

**Mobility of each node:** Less mobility nodes have more probability to become a CH. It is represented by $M_n$ - Mobility (Speed) of each node. It is calculated as $M_n$ - Mobility speed of every node by following formula

$$M_n = \frac{1}{T}\sum_{t=1}^{T}\sqrt{(X_t - X_{t-1})^2 + (Y_t - Y_{t-1})^2} \quad \mathbf{-\ (4)}$$

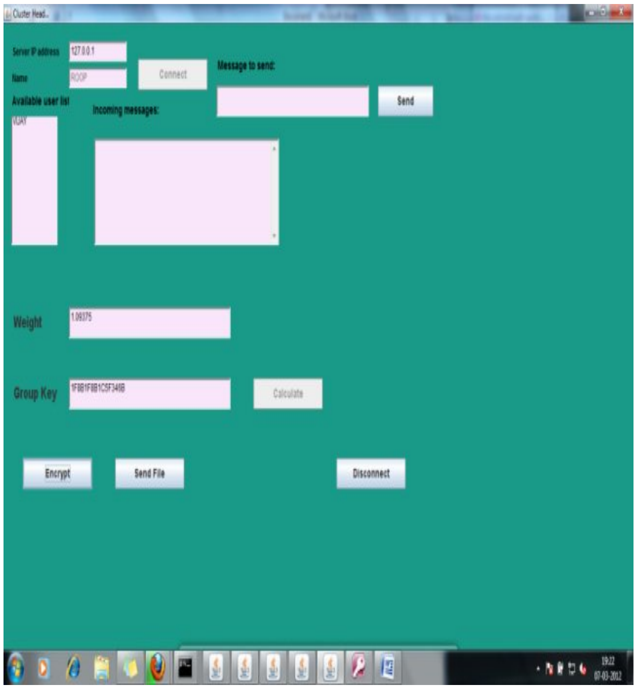
Where $(X_t, Y_t)$ and $(X_{t-1}, Y_{t-1})$ are the co-ordinate positions of node 'n' at time t and t-1, T= cumulative time.

**Distance:**

Distance from all other neighboring nodes is represented by $D_n$.Here; the sum of the distance between member nodes and its neighbors *is* defined by the equation (5). In order to find the neighbor $N(n)$ of each node 'n', the $D_n$ is calculated as

$$D_n = \sum_{n' \in N(n)} \{ distance\ (n, n') \} \quad \text{----}\ \mathbf{(5)}$$

$D_n$ - The sum of the distances between node 'n' with its entire neighbor.



Server IP address: 127.0.0.1

Name: ROOP

Available user list: VIJAY

Weight: 1.09375

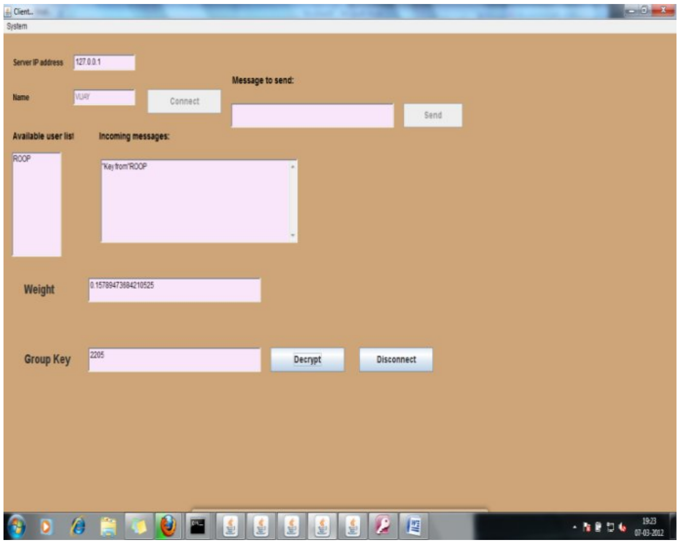
Group Key: 1F8B1F8B1C5F346B

Figure 4 Algorithm - Encryption for key distribution

As shown in the figure 4, CH is generating the group key and the key is encrypted by Cluster Head using RSA algorithm for secure communication.

## 3.2 Cluster Head Election and Cluster Formation - Here the Cluster Head election

and Cluster Head lifetime monitoring algorithm are discussed.



Server IP address: 127.0.0.1

Name: VIJAY

Available user list: ROOP

Incoming messages: "Key from"ROOP

Weight: 0.15789473684210525

Group Key: 2205

Figure 5 Decryption of Group Key

Decryption of group key is happened in nodes so that the nodes will be connected to CH and communication will be occurring between the nodes which have been depicted from the figure 5.

```
ALGORITHM FOR CLUSTER HEAD ELECTION
Step 1: Calculate weight for every node based on the
        Metrics like Node Degree, Mobility,
        Connectivity and Energy Remaining.
Step 2 : Broadcast Weight value and its Id to all its
        Neighboring nodes and the neighborhood
        table updated with weight value
Step 3: Cluster Head (CH) and Secondary Cluster
        Head (SCH) elected based on the weight
        value
        If (The Node with highest weight value)
            Elect that Node as a CH
        If (The Node with next highest weight)
        Elect that Node as a SCH
Else    Ordinary nodes send Join Request to CH to
form a Cluster
```

Figure 6 Algorithm for Cluster Head Election

```
ALGORITHM FOR CLUSTER HEAD LIFETIME
SCH monitor the battery level of CH for every 30s.
If (Battery level of CH< Minimum Threshold Level)
        SCH will become New CH,
        Send CH_LIFE DOWN Msg to all
        member nodes,
        Election procedure initiated to find
        new SCH
Else Re election not needed
```

Figure 7 Algorithm for CH life time

## 3.3 Group Key Agreement (Generation & Distribution)

Weight Computation and its algorithm for cluster formation has been depicted in figure 6 and figure 7 shows the CH life time. After Weight Computation and cluster formation, members in the clusters will send their id and public key. Cluster Head receive the message and initiate the group key calculation by using RSA. The algorithm for group key generation is shown in figure 8.From the equation (6), the member's Public Key is used to calculate group key as follows.

$$CH : GK =((\alpha)^{pk1+\ pk2+\ldots.pKn\ +CH_k} \bmod p) \text{ x } (R_v) \text{ -- } \mathbf{(6)}$$

GK- Group Key Where

$\alpha$ – primitive root of p

$CH_k$ – secret key of cluster head,

$pk_1,pk_2 \ldots pk_n$ – public keys of individual nodes w ithin the cluster,

p – prime number

$R_v$ – secret random value generated every time while re-keying.

```
ALGORITHM FOR THE GK GENERATION

Step 1 :
    If (Node present within the cluster)

        if CH gets public keys of all
        nodes
        Calculate group key as follows:
        $CH : GK =((\alpha)^{pk1+\ pk2+\ldots.pKn\ +CH_k}$
        $\bmod p) X (R_v)$
        End if
    End if
```

Figure 8 Algorithm for the GK generation

```
ALGORITHM - ENCRYPTION AND DECRYPTION FOR
KEY DISTRIBUTION

Step 1:
    CH → nodes within the cluster
    Encrypt using RSA algorithm
    E (GK)
    Where E (GK) = ((GK) e mod n)
     in which {e, n} are public key pair.

Step 2:
    Decrypt Nodes: D (GK) = (E (GK) d mod
    n) in which {d,n} are private key pair.
```

Figure 9 Algorithm for Encryption and Decryption in key distribution

The distribution of group key is initialized using RSA algorithm. Cluster Head having (e,n) public key and every node maintains (d,n) private key. Whenever any new node joins into the cluster, Cluster Head calculates a new group key and multicast to the already existing nodes. Cluster Head unicast [15] the group key to a new node along with private key for RSA algorithm. The algorithm for encryption and decryption in key distribution has shown in figure 9. The bit length used for RSA module is 1024 bits and p=47 and q=59 assigned values for prime generation and the Elliptic Curve Cryptography value will be 163 bits. Secret Random value, prime number selection will be taken care by the CH itself and not by the third party. If the group key is GK, an element in the group of integers modulo p, the process for encrypting/decrypting done by Cluster Head once the group key is generated.

## 4. Results and Discussion

The number of nodes used in the simulation results varies between 20 and 100. The simulations run for 300 seconds. The cluster size was fixed at 15. In the first set of simulations, the scalability of the algorithm is measured in terms of nodes density and transmission range. In this paper, the NS-2 simulator [14] is used for the simulation. The total energy consumption and Packet Delivery Ratio of total network are compared between with and without clusters.

*Packet Delivery Ratio (PDR):* Data Packet Delivery Ratio can be calculated as the ratio between the numbers of CBR (Constant Bit Rate) packets that are received by the destinations (sinks) over the total number of packets sent by all the sources within the period of simulation.

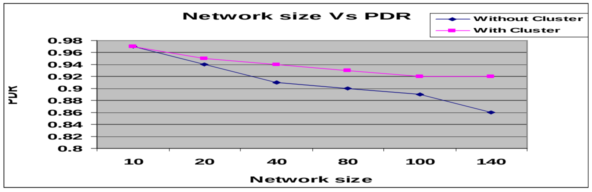


Figure 10 Network size vs PDR

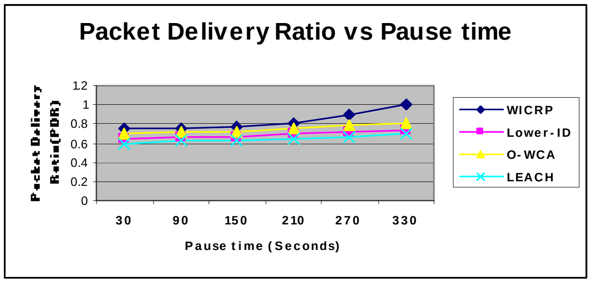


Figure 11 PDR vs. Pause time

The difference in the delivery ratios increases as the network's size increases, which show the performance gained by WICRP. Figure 10 describes the range of total PDR of entire Network nodes. Pause time is the time duration for which all nodes hold the same positions at random waypoints (mobility model). From the figure 11, it is known that the mobility model used is the random way point model which generates waypoints at random for WICRP, lowest-id, LEACH, O-WCA. Packet Delivery Ratio can be interpreted as the loss ratio that will be experienced at the routing layer for observing the performance of routing protocols which in turn has an impact on the overall throughput of network support.

A node moves at the given speed to a random waypoint and when it hits that, it chooses another waypoint at random and begins moving toward it. If the pause time is 100, there is no node mobility since the simulation time is also 100 seconds. If it is 10, then a node holds its position for 10 seconds whenever it hits each waypoint.

*Total Energy consumption:* Initially both the protocols consume energy almost the same but after a time of 400 seconds, there is a change in energy consumption of two protocols. Figure 12 shows the total energy consumption of entire network nodes vs time comparatively with and without clusters. So, the proposed protocol weight Clustering algorithm can save more energy better than other Clustering algorithms.

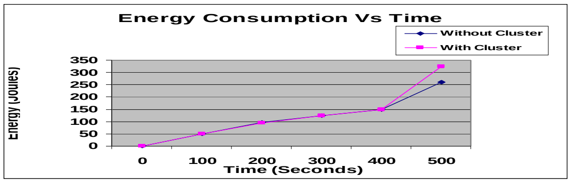


Figure 12 Energy Consumption vs Time

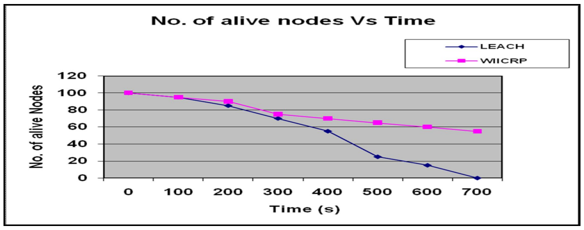


Figure 13 No. of live nodes vs Time

In LEACH, every node directly communicate with the BS, so it wastes its maximum energy for transmitting the data, but by applying weight based clustered hierarchy(WICRP) node directly send their data to its nearest CH and it consumes only less transmission power. CHs then cumulate these data's and send them to BS. So this clustered architecture results energy saving for nodes but at the same time, the energy of CHs is consumed. To solve this problem, there is a

Monitoring node for checking the energy level of CH, when it reaches minimum energy SCH will take part in the communication and it will act as a CH and CH algorithm will be initiated for selecting new SCH. In Figure 13, the proposed protocol (WICRP) has more number of alive nodes than the LEACH protocol.

# 5. Conclusion

In this work, Weight based Clustered Key Management scheme using RSA for Wireless Mobile Ad hoc Networks has been proposed. This approach is based on combined weight metric that takes into account of several system parameters like the degree difference of the node, transmission range, battery power and mobility of the sensor node. Since energy consumption is the most important criteria in cluster based routing schemes, our protocol provides better results than existing Lowest-id, WCA algorithm and LEACH algorithm. Performance metrics like network size with PDR, Energy Consumption have been evaluated between with clusters and without clusters. In the near future, some other performance metrics like fault tolerance can be taken for performance evaluation and this protocol can be extended to include group key management for inter clustering with multi-hops to provide secure transmission of collected data. Effective utilization of power, Bandwidth wastage helps in improving the Quality Of Service in WMANs by applying the Weighted Clustering Algorithm.

# 6. References

[1].Anitha, V.S., Sebastian, M.P.(2009), "SCAM: Scenario-based clustering algorithm for mobile *ad hoc* networks ", in Proceedings of the 13th IEEE/ACM International Symposium on Distributed Simulation and Real-Time Applications, pp. 97-104.

[2]. Basagni, S. et al. (2006), "Localized protocols for *ad hoc* Clustering and backbone formation: A performance comparison", IEEE Trans. Parall. Distrib. Sys. 2006, 17, pp. 292-306.

[3]. Dhurandher, S.K.; Singh, G.V (2005), "Weighted-based adaptive clustering algorithm in mobile ad hoc networks". in Proceedings of ICPWC'2005, New Delhi, India, pp. 96-100.

[4]. K. Akkaya and M. Younis, (2003) "A Survey of Routing Protocols in Wireless Sensor Networks", Elsevier Ad Hoc Network Journal, 3(3), pp .325-349.

[5]. LI Fangmin et al. (2008), "Power Control for Wireless Sensor Networks", Journal of Software Engineering, 19(3): pp. 716−732

[6].Zhang Jian-wu et al.(2008), "Weighted Clustering Algorithm Based Routing Protocol in Wireless mobile adhoc sensor networks (WMASNs)",

[7]. Sanjay kumar padhi et al. (2008), "Review of routing protocols in sensor and Adhoc networks ", International journal of reviews in computing

[8] Adeel Akhtar, Abid Ali Minhas, and Sohail Jabbar (2010),"Energy Aware Intra Cluster Routing for Wireless Mobile Adhoc Sensor Networks (WMASNs) ", International Journal of Hybrid Information Technology Vol.3, No.1, pp. 29-48

[9].Naveen chauhana (2011), "Distributed Weighted Cluster Based Routing Protocol for MANETS" w*ireless Sensor Network*, 2011, 3, pp. 54-60.

[10].Tzung-Pei Hong et al. (2011), "An Improved Weighted Clustering Algorithm for Determination of Application Nodes in Heterogeneous Sensor Networks", Journal of Information Hiding and Multimedia Signal Processing, Ubiquitous International, Volume 2, Number 2, pp. 173- 184.

[11]. R. Pandi Selvam et al. (2011), "Stable and Flexible Weight based Clustering Algorithm in Mobile Ad hoc Networks", International Journal of Computer Science and Information Technologies, Vol. 2 (2), pp. 824-828.

[12]. Jutao Hao et al. (2011), "Energy Efficient Clustering Algorithm for Data Gathering in Wireless Sensor Networks", Journal of Networks, Vol. 6, no. 3, pp. 490 - 497.

[13]. Huiheng Liu et al. (2011), "Cooperative Spectrum Sensing and Weighted-Clustering Algorithm for Cognitive Radio Network", I.J. Information Engineering and Electronic Business, vol. 2, pp. 20-27.

[14].NS-2 simulator. Available online: http://www.isi.edu/nanam/ns (17 May 2011)

[15].Y. Challal, H. Bettahar, and A. Bouabdallah, "SAKM: A scalable and adaptive key management approach for multicast communications," ACM SIG- COMM Computer Communications Review, vol. 34,no. 2, pp. 260-271, Apr. 2004.

[16]. Maneesha V.Ramesh ,Abishek T.K, Aparnadhusoodanan, Wireless Sensor Network Based Localization Scheme for Tracking Emergency Responders in Disaster Area ICWCA-2011 ,Aug 01-03,2011.

# End-to-End Encrypting Android Phone Calls

**I. Burns, K. Gabert, and J. Zheng**
Department of Computer Science and Engineering, New Mexico Institute of
Mining and Technology, Socorro, NM, USA
iburns@cs.nmt.edu, kasimir@cs.nmt.edu, zheng@nmt.edu

**Abstract**— *Cellular phones are regularly used to discuss sensitive personal and business information. However, little attention is given to the security of these conversations, which can be particularly important for international businessmen. Current encrypted phone call solutions for this problem require either an Internet connection for VOIP or a specialized handset. In this paper, we propose an architecture to encrypt phone calls as an addition to the Android smartphone operating system. The proposed architecture utilizes TLSPGP to provide strong encryption with a peer-to-peer web of trust.*

**Keywords:** Android operating system, mobile computing, computer security, encrypted communications.

## 1. Introduction

As cell phones are becoming more common, they are being used to communicate numerous types of sensitive information: personal banking details, business secrets, and personal health information are a few examples. Keeping this information private is extremely important. In the past, encrypted communications have only been accessible to those with vast resources. As technology has moved forward, the ability to communicate with high-level encryption has become a reality for more and more people. Currently, there are no means of having encrypted communications in consumer cell phones which cannot be monitored by the telephone service provider.

To compound this issue, numerous flaws have been presented in the current encryption used by the cellular telephone providers. These flaws allow any individual with a computer to quickly and easily monitor all cell-phone communications near them [1].

This paper presents a method for adding an encrypted communications stack to Android. Android-based smartphones are becoming more and more common [2] and present a unique opportunity to allow individuals and organizations without vast resources to communicate using encrypted communication.

Adding encryption into Android is not a new idea. RedPhone, from Whispersys [3], is a great example of recent Android-based products which provide encrypted phone calls. However, the current solutions, including RedPhone, rely on identity-based encryption. Identity-based encryption is not peer-to-peer and requires a trusted central server to manage the security of the communications. If this trusted server is the cell phone provider, then the communications will still not be private in our model.

These solutions also are not implemented in Android itself [4]. Instead, all of the current projects start from user space, as an application available for Android, and add encryption there. Starting from user space means that if Android is in control of another entity, such as the phone network provider, then these solutions will not provide any protection [5]. Having Android in control of another entity is a reasonable assumption since most phone providers will install a custom Android on each phone sold. Also, these solutions are vulnerable to user space malicious software (malware) [4].

The rest of this paper is organized as follows. In next section, we introduce some background information on security problems during an Andriod phone call, limitations of existing solutions and related security and privacy protocols. The Andriod radio subsystem is described in Section III. In Section IV, we present our solution for end-to-end encrypting Andriod phone calls and discuss some difficulties faced in implementation. Finally conclusions are drawn in Section V.

## 2. Background

### 2.1 Security Problems during an Andriod Phone Call

#### 2.1.1 Security Problems in the Android Operating System

The Android operating system takes a pervasive and multifaceted approach to security, incorporating mechanisms from Linux as well as introducing its own security features and environment. These features include POSIX security mechanisms, sandboxing through the Dalvik JVM, and restricted tasks that require user consent at application install-time. Finally, Android devices are distributed with the user (and any installed applications) limited to a non-administrative account (the user doesn't have root). Ideally, this prevents users from allowing software to make deep changes to the operating system and therefore limits the abilities of malware.

Even with this extensive security architecture, Android has a very large number of security-related exploits, including exploits granting root permissions [6]. The problem is exacerbated by nature of Android's distribution. Although Google is in charge of Android's creation, it is distributed by a multitude of phone manufacturers and wireless carriers, who each must provide users with a patch after Google creates it. This means the lead time between the discovery of an exploit and users receiving a fix can be very long, even if Google responds rapidly. Additionally, the vendors often abandon the patching of phones relatively quickly, leaving some users very vulnerable [6].

### 2.1.2 Security Problems in Cellular Networks

A secure phone does little to preserve privacy if it uses an insecure network. In an effort to preserve privacy, cellular radio standards such as GSM include encryption. Unfortunately, these encryption schemes have a long history of being inadequate for dependable privacy protection [7].

The most modern ciphers available for GSM networks are Kasumi-based encryption schemes (A5/3 for GSM, UEA1 for UMTS, GEA3 for GPRS [8]) are currently unbroken, but appear somewhat vulnerable. A practical time attack against Kasumi exists [9], but does not appear to be workable on the overall encryption schemes employed by GSM-based networks [8].

### 2.1.3 Security Problems in Telephone Networks

Outside of the handset-to-tower link, cellular networks revert to unencrypted traditional phone networks. If the infrastructure provider is trusted, this can be a minor problem. However, when the infrastructure provider is not trusted (such as during international business travel), this lack of security through the phone network can be a problem. Even when the nation controlling the phone network is trusted, the network may not be trustworthy.

In 2005, Vodafone Greece disclosed an attack that compromised two central office switches. The compromised switches were used to unlawfully wiretap 100 high-level officials in the Greek public and private sectors, including the Prime Minister and his wife. The attacking rootkit was not discovered until a software malfunction caused a deep investigation into the switch software [10].

In Italy, both official and unofficial phone intercepts have created numerous media scandals and resulted in arrests of high-profile officials. These scandals have spurred demand for end-to-end encryption for phones used to discuss sensitive matters there [11].

The market research firm ABI states that 79% of companies use mobile phones for communications that would be damaging if intercepted, and other studies have indicated that the average cost of significant data loss is up to $1.3 million per incident [12]. End-to-end encryption is the only way to fully address interception concerns [12].

## 2.2 Limitations of Existing Solutions

Encryption has been added to Android phone calls in the past. The most notable project to add encryption is a phone system denoted RedPhone, by Whispersys [3]. RedPhone is a product which provides identity based encryption in a userland application using Voice over Internet Protocol (VoIP). While it does support protection against eavesdropping and provides authentication for the communicating party, it is limited. First, it requires a trusted third party which will handle the encryption keys. This becomes a large problem if there is no third party which should be trusted. The second reason that RedPhone does not satisfy the conditions is because it is a user-space application. This means it is unable to provide any security assurances about the underlying OS, and it is limited to VOIP as opposed to phone network-based calls.

## 2.3 Related Security and Privacy Protocols

### 2.3.1 Pretty Good Privacy (PGP)

PGP is a form of public key cryptography originally developed for use in email [13]. The OpenPGP standard allows the implementation of various asymmetric and symmetric algorithms, including RSA and AES. PGP is structured around a message-packet format, with encryption for privacy and digital signatures for integrity [13].

Where PGP is special among modern encryption schemes is its trust model, called the Web of Trust. At its core, PGP relies of the user to determine whether they trust the identity of a key. The Web of Trust model is design to aid this process, without introducing a central authority that can be compromised. The concept of “trust" is deliberately vague so a user can determine what it means for them. An untrusted key still provides privacy, but is unable to provide identity verification [14].

### 2.3.2 Transport Layer Security (TLS)

TLS is a very popular cryptographic protocol in widespread use for numerous Internet applications. The most widely known of which is HTTPS. It provides symmetric key exchange and an integrity check using a keyed Message Authentication Code [15].

In traditional TLS, a trusted Certificate Authority verifies the identities of key holders. However, there have been several recent high profile attacks against the certificate authority system [16]. Once a certificate authority is compromised the entire system breaks down.

It is possible to use PGP instead of a certificate authority structure in the TLS Handshake Protocol. RFC 6091 implements such a standard [17]. It allows a Web of Trust authentication model to be used while providing the commonly available and well studied TLS protocols.

# 3. Android System Architecture

In this section we will discuss the Android radio subsystem and how it interacts with Android as a whole. The information in this section was gathered using the Ice Cream Sandwich source code.

Android is composed of five major components: Applications, Application Framework, Libraries, Android Runtime, and the Linux Kernel [18]. The cellular call system is implemented in Android as the Telephony Manager component, a subcomponent of the Application Framework.

## 3.1 Android Telephony

The Telephony Manager component is a platform-dependent component, meaning that for each platform (or different mobile phone / tablet) parts of it will have to be adjusted to work with the potentially proprietary Vendor Radio Interface Layer (RIL). Figure 1 contains a graphical representation of the various blocks that compose the telephony component.

The RIL interactions start right above the baseband, which is the firmware specific to the platform. The file `android/hardware/ril/reference-ril/reference-ril.c` contains functions which are the closest Android gets to interacting with the GSM network. This file interfaces with the baseband to perform various GSM actions such as dialing numbers, hanging up calls, accepting calls, and so on. The RIL library performs callbacks into this file. Those callbacks take the form

```
onRequest (int request, void *data,
    size_t datalen, RIL_Token t)
```

where `request` contains the action of the request, and `RIL_Token` allows for stateful operations. On the other end of the spectrum, the Android package `com.android.internal.telephony` contains various classes dealing with controlling the phone. Similar to the RIL commands, the commands exposed through these objects control the phone state, such as `acceptCall()`, `rejectCall()`, `clearDisconnected()`, etc.

## 3.2 Android Media

The Media components of the Application Framework as responsible for Android's multimedia support. The audio portion of this framework is of interest for phone calls.

The basic unit of Android audio is the stream. Streams can be input or output to/from any audio device and application. The routing of these streams to various available microphones and speakers is handled by AudioFlinger `android/frameworks/base/services/audioflinger`. Presumably to increase platform independence, AudioFlinger does little direct routing by default (although it retains the capability to route arbitrarily). Instead, it sets a system-wide routing mode such as `MODE_IN_CALL`, or `MODE_RINGTONE`. This routing mode is then passed down into vendor-specific code supporting the media system through the interface defined in `android/hardware/libhardware/include/hardware/audio.h`. The vendor code is then ultimately responsible for determining where the audio from a stream (such as the call audio stream) ends up being played.

AudioFlinger does not provide a user-friendly interface at the level Android applications are expected to function at. This interface is provided by `android.media.AudioSystem`, which bridges from the Java-based Application Framework to the native-code-based media system using `android/frameworks/base/media/libmedia`. The AudioSystem class provides constants for generic devices and streams, such as `DEVICE_OUT_EARPIECE` and `STREAM_VOICE_CALL` It also provides methods meaningful to applications that control the media system, such as `muteMicrophone` or `setRingerMode`.

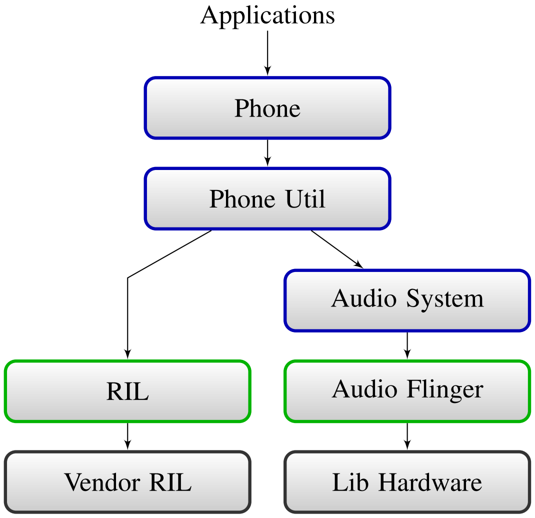


Fig. 1: The various blocks of the telephony component used in placing and communicating during a phone call. The blue outlined components are found in the Application Framework layer of Android, the green outlined components are found in the Library layer, and the black outlined components are found in the Vender layer.

## 3.3 Phone Calls in Android

Phone calls in Android are the end result of a collaboration between the telephony and media frameworks. They begin when the user interacts with the phone application. When the user has entered the number to dial and has placed the phone call, the phone application will issue a request for a phone call with the specified number. This request will propagate down into the vendor RIL. The vendor RIL will then send the request directly to the baseband on the phone. If the phone call request was issued correctly then an acknowledgment of the phone call will immediately be issued.

The baseband will then make the appropriate cell network request. Upon success, it will begin transmitting from the phone speaker and outputting the received voice transmission directly to the speaker. Finally, once a connection is established, the RIL in Android will receive an `onRequest` callback and will update the phone application interface appropriately. This interaction can be seen in Figure 2.

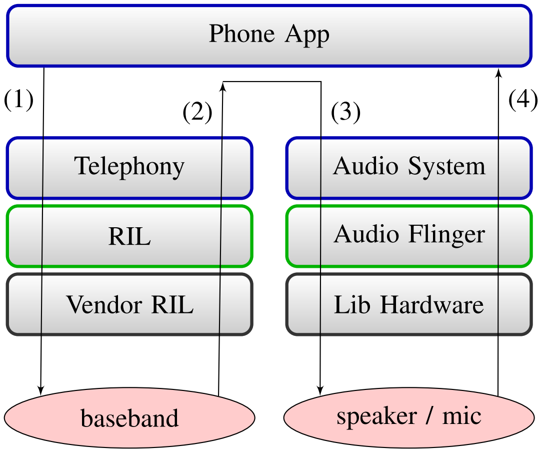


Fig. 2: The structure of placing a phone call in Android. Even though the call is placed in Android, the actual digital voice stream is set up by the low-level code provided by the vendor. (1): Placing a call. (2): Call in progress. (3): Set mode MODE_IN_CALL. (4): Returning back, displaying call in progress.

Once the call is established, the telephony system puts the phone's media system into `MODE_IN_CALL`. By default, this causes the vendor's media code to do something reasonable, like routing the call audio stream `STREAM_VOICE_CALL`. If a bluetooth or wired headset is connected, the vendor's code handles the change in audio routing.

Essentially, there is a "dumb phone" (as compared to a "smart phone") running alongside Android in an Android "smart phone". Android does not actually perform a phone call itself. Instead, it effectively dials the number through the baseband ("dumb phone") and then the baseband will alert Android whenever it is in a phone call. This leaves all of the implementation of the phone components itself up to baseband and outside of the reach of Android.

Such an architecture for sending and receiving calls complicates the matter of adding encryption. Most of the decision-making components lie in vendor code instead of the Android system. Any encryption scheme must work within the context of current decisions, routing around and through existing, proprietary code.

# 4. End-to-End Encrypting Phone Calls in Android

This section contains our proposed solution to this problem. First, we present the recommended encryption scheme for this project. Second, we will discuss where this encryption should be placed in the existing Android telephony stack to provide encrypted phone calls to users while preserving the integrity of current Android systems and applications. Finally, some difficulties in implementation the solution are discussed.

## 4.1 Encryption Scheme

To ensure privacy in the event of a malicious cellular network provider, we propose using a combination of PGP and an AES stream cipher. PGP provides a widely-used Web of Trust which will help authenticate users.

This implementation will follow RFC 6091 [17] which describes how to adjust the TLS Handshake Protocol to use PGP keys. The TLS protocol is a widely-implemented and well trusted protocol providing perfect forward secrecy. The GNU TLS implementation of the TLS protocol contains extensions to enabled the use of PGP keys. However, it does not implement any trust model for these PGP keys [19].

We performed some benchmarking to determine if a target Android phone could support AES at a speed fast enough for the purpose of encrypting phone calls. For these benchmarks, we used a Samsung Nexus S running the Android Open Source Project version of Ice Cream Sandwich (Android 4.0). This phone uses a 1 GHz ARM Cortex A8 based processor, and we believe it represents a fairly typical current Android device.

We measured performance of AES using 128-bit keys, no padding, and the cipher in ECB mode. We compare performance of the Java-based AES implementation ("managed") using the default provider for javax.crypto in Android and the C implementation of AES in the nettle library ("native"). The nettle library implements a variety of cryptography algorithms which is also used by GNU TLS. We measured the time required (as reported by the system clock) to encrypt 10 megabytes of random data. The results were then reported in megabytes encrypted per second. Each test was repeated 10 times. The first execution was discarded in which managed code was slower than subsequent executions.

We found the managed implementation of AES produced an average encryption rate of about 3.66 MB/s with a standard deviation of 0.02 and the native implementation produced an average rate of 8.24 MB/s with a standard deviation of 0.35. Either of these rates is completely sufficient for our purposes, where we anticipate the requirement of encrypting Andriod phone calls to be less than 100 kilobytes per second.

## 4.2 Details of Implementation

Implementation of the proposed solution in Android requires a few changes to the operating system. These changes form 3 new subsystems: the Cryptographic Control Module (CCM), the Cryptographic Module (CM), and the Cryptographic Interface (CI). These modules create a tweaked media stack for encrypted calls, and add a few additional libraries to support their functionality. Figure 3 contains an overview of the modules' interactions during a phone call process. This section will detail the changes required to Android in order to realize these modules.

The Cryptographic Module will actually perform the encryption on the audio stream. It will be added at the system Library level. This encryption module will act as a virtual audio input/output while performing the encryption of the phone call. Then the module will route the decrypted plaintext audio to the system's original call audio destination—which could be the phone's ear piece or a peripheral handsfree device. The incoming plaintext from the microphone will similarly be encrypted before being passed through to the baseband system.

The Cryptographic Module requires a library supporting TLSPGP protocol functionality must be added at the system Library level. We propose the use of GNU TLS for this purpose. To support the implementation of the web of trust, PGP will also be added to the system libraries. Java Native Interface (JNI) bindings will then be added for the relevant portions of each of these libraries to allow application-level code to interact with the call cryptography system.

The Cryptographic Control Module will consist primarily of adjustments to the AudioSystem and other Application Framework components of the media and telephony systems. These adjustments will allow the redirection of the call audio stream to the Cryptographic Module. The `setParameters` method of the `AudioSystem` class enables the CCM to redirect the call audio stream using high-level constructs understood at the application layer. That is, using constructs such as the `CALL_AUDIO_STREAM` constant instead of direct references to audio stream data.

The Cryptographic Interface exposes encrypted call and key management functionality to the user. It consists of modified Phone and Contacts apps. The Phone apps will expose enable/disable encryption functionality when a public key is available, provide indications that a secure or insecure call is in progress, and allow the user to place secure and insecure phone calls. The Contacts app will expose the web of trust functionality and key management, by providing keys associated with the user's contacts.

## 4.3 Implementation Difficulties

The implementation of this architecture faces a variety of challenges. The first is porting the required libraries. When they are fully implemented in C, this process is typically straightforward. However, it is difficult to predict if or where assembly code will be encountered for improved performance. This code is typically targeted at x86 machines, so switching to ARM-based Android devices can be challenging.

The implementation also depends on data being sent through the phone network in a manner that reproduces the original bit-for-bit. If lossy compression is employed below the encryption, the data will be irrecoverable. Any compression performed by the phone network needs to be evaluated. If it is performed in the RIL on the device, access to the RIL's code will be required for successful implementation.

In general, the architecture of the Android system has the potential to get in the way. In addition to compression concerns, odd behavior of vendor code may derail implementation. Correct permissions for all actions need to be located and assigned, and there is potential for unforeseen difficulties in implementation from the structure of the Android system.

The latency of the phone network provides another unknown. With cellular, wired, and international phone networks involved, there is the potential for significant latency in the transmission system. We expect that TLS will be able to tolerate whatever latency is present because of its common use with HTTP. A phone call, however, can be viewed as a real-time application where the encryption and transmission of data must be mroe timely than web browsing. The length of the handshake is also a concern, especially when the end-to-end latency of the phone network is high. In the worst case, we would expect that the proposed solution will suffer from latency less than satellite phones, which is tolerable for most users.

Finally, the battery life of the phone making the encrypted phone calls is a concern. The encryption process will result in higher CPU utilization and power consumption during a call. The exact effect of this increased power consumption on battery life needs to be investigated in future. In general, somewhat reduced battery life is a necessary trade-off for secure calls.

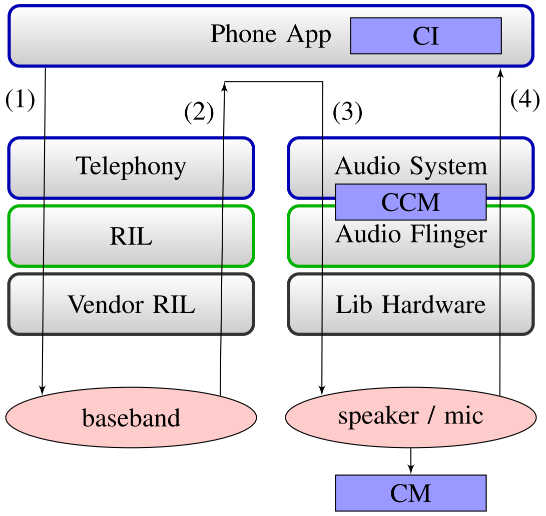


Fig. 3: The structure of placing an encrypted phone call in Android. In this structure, the phone call will have its voice routed through the encryption module in the Library layer.

## 5. Conclusion

This implementation will satisfy the conditions required to prevent the described adversary model. The encryption will be triggered when a GSM phone call is placed and a public key is known for the corresponding contact. Since the encryption will be implemented in Android, it will be resistant to user-land malware present on the phone. Since the encryption is a modification to Android itself, users will be motivated to place a custom OS on their phone which will help prevent attacks from the cell phone service provider. Finally, using a combination of PGP and AES will be sufficient to ensure privacy of phone calls without relying on a trusted central authority. Using TLS and PGP will help prevent an implementation lock-in for the cryptography and will allow multiple developers to independently implement similar systems interacting on the same network.

Adding this encrypted radio module and modifying the radio stack in Android will help provide privacy where it is desperately needed. By implementing such a system, attacks on core cellular or telephone infrastructure as well as specific attacks against individuals will be mitigated. With truly ubiquitous cell phone use, such a system becomes exceedingly important.

## References

[1] S. Gold, "Cracking GSM," *Network Security*, vol. 2011, no. 4, pp. 12–15, Apr. 2011. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1353485811700393

[2] Taylor Wimberly, "Android growth outpaces iPhone," http://androidandme.com/2009/08/news/android-growth-outpaces-iphone/.

[3] Whisper Systems, "Mobile Security for Android," http://www.whispersys.com/.

[4] A. Shabtai, Y. Fledel, U. Kanonov, Y. Elovici, S. Dolev, and C. Glezer, "Google Android: A comprehensive security assessment," *Security & Privacy, IEEE*, vol. 8, no. 2, pp. 35–44, 2010. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5396322

[5] S. Trapp, M. Wählisch, and J. Schiller, "Short Paper: Can Your Phone Trust Your Friend Selection?" in *CCS '11: Proceedings of the 18th ACM conference on Computer and communications security*, 2011. [Online]. Available: http://page.mi.fu-berlin.de/waehl/papers/tws-spypt-11.pdf

[6] T. Vidas, D. Votipka, and N. Christin, "All Your Droid Are Belong To Us : A Survey of Current Android Attacks," in *USENIX Workshop on Offensive Technologies August 2011*, 2011, pp. 1–10. [Online]. Available: http://mendeley.citizenlab.org/VidasVotipkaChristin2011.pdf

[7] E. Barkan, E. Biham, and N. Keller, "Instant Ciphertext-Only Cryptanalysis of GSM Encrypted Communication," *Journal of Cryptology*, vol. 21, no. 3, pp. 392–429, 2008. [Online]. Available: http://www.springerlink.com/index/ythkwv4gfq0fr5j4.pdf

[8] S. Bellec, "The attack against Kasumi cannot practically be used against the GSM," 2010. [Online]. Available: http://pro.01net.com/editorial/511353/the-attack-against-kasumi-cannot-practically-be-used-against-the-gsm/

[9] O. Dunkelman, N. Keller, and A. Shamir, "A practical-time attack on the A5/3 cryptosystem used in third generation GSM telephony," in *Proceedings of the 30th Annual Cryptology Conference (CRYPTO 2010)*, no. December 2009, 2010. [Online]. Available: http://cryptome.org/a5-3-attack.pdf

[10] V. Prevelakis and D. Spinellis, "The Athens Affair," *Ieee Spectrum*, vol. 44, no. 7, pp. 26–33, 2007. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4263124

[11] P. Kiefer, "Phone taps in Italy spark a rush for cellular encryption," 2007. [Online]. Available: https://www.nytimes.com/2007/04/29/technology/29iht-encrypt30.1.5487929.html

[12] S. Bransfield-Garth, "Mobile phone calls as a business risk," *Network Security*, vol. 2010, no. 9, pp. 4–11, Sep. 2010. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1353485810701148

[13] J. Callas, L. Donnerhacke, H. Finney, D. Shaw, and R. Thayer, "OpenPGP Message Format," *RFC*, vol. 4880, 2007. [Online]. Available: https://tools.ietf.org/html/rfc4880

[14] M. Lucas, *PGP and GPG*. No Starch Press, Incorporated, 2006.

[15] T. Dierks and E. Rescorla, "The Transport Layer Security (TLS) Protocol," *RFC*, vol. 5246, 2011. [Online]. Available: https://tools.ietf.org/html/rfc5246

[16] "GlobalSign stops secure certificates after hack claim," 2011. [Online]. Available: http://www.bbc.co.uk/news/technology-14819257

[17] N. Mavrogiannopoulos and D. Gillmor, "Using OpenPGP Keys for Transport Layer Security (TLS) Authentication," *RFC*, vol. 6091, 2011. [Online]. Available: https://tools.ietf.org/html/rfc6091

[18] Google, Inc., *Android Documentation*, http://developer.android.com/guide/.

[19] Free Software Foundation, Inc., "Gnu TLS," http://www.gnu.org/software/gnutls/.

# A Three-stage Phase Encoding Technique for Quantum Key Distribution

**F. Zamani, S. Mandal, and P. K.Verma**
School of Electrical and Computer Engineering, University of Oklahoma, Tulsa, Oklahoma, USA

**Abstract** - *The only known transmission technique that provides unconditional security is based on quantum cryptography. The current workhorse of quantum cryptography, more precisely, quantum key distribution (QKD), is based on the BB'84 protocol. Most commercial QKD implementations are based on phase-coding the BB'84 protocol, where the unbalanced Mach-Zehnder Interferometer (UMZI) is used as the information coder and decoder. This paper presents a three-stage phase encoding quantum key distribution protocol which is based only on one two-way quantum channel. This protocol does not need key sifting for key establishment. The proposed technique thus obviates the use of a classical communication channel for key sifting and key distillation which are subject to eavesdropping. The proposed technique thus increases the key efficiency and security compared with the current techniques.*

***Keywords:*** Quantum key distribution (QKD); sifted key; key reconciliation; Phase coding protocol; Plug and play system

## 1 Introduction

The aim of cryptography is to prevent a cryptanalyst from deciphering information sent by one party to another party. In classical cryptography, the process of encryption and decryption of information is based on mathematical algorithms and the security of the classical cryptography is based on the difficulty of these algorithms. Quantum cryptography is a good solution for providing key distribution.

Quantum key distribution (QKD) applies the quantum mechanics concept to guarantee security of a key exchange between two legitimate parties in the presence of cryptanalyst [1, 2, 3, 4]. According to the Heisenberg uncertainty principle and no-cloning theorem, measuring or copying the quantum state of the qubit alters the original information in the qubit [5]. In quantum key distribution, we are able to detect whether the eavesdropper has eavesdropped on the message or not, while the classical cryptography does not offer this feature.

The BB'84 is the most commonly used quantum key distribution protocol proposed by Bennett and Brassard in 1984 [5, 6]. It utilizes one quantum channel and one two-way classical channel. The classical channel is a regular telecommunication channel, over which Alice and Bob exchange information. The aim is establishing a secret key between two authorized parties, Alice and Bob, with the possibility that Eve can be present during the key establishment process. The original proposal of BB'84 relies on the polarization encoding of a photon. Since the alignment and the stabilization of polarization axes between Alice and Bob is practically difficult and depolarization happens due to medium transmission changes, the phase-coding technique for implementing BB'84 is used. Called the Plug and Play protocol, it is one of the common implementations of BB'84. The Plug and Play system utilizes one two-way quantum channel and one two-way classical channel which is basically an Ethernet channel. This classical channel is used for key sifting and key distillation. In this paper we propose a three stage QKD protocol which uses a single two-way quantum channel, thus removing the need for depending on a classical channel. The proposed mechanism will increase the security and key efficiency of the key establishment procedure and there will be no need to have the key sifting step which reduces the size of the final key.

The rest of the paper is organized as follows: the phase coding BB'84 protocol reviewed in Section II, Section III describes Plug and Play system, Section IV describes our proposed protocol, Section V illustrates the scenario in the presence of an eavesdropper and in Section VI the various advantages of the proposed protocol are described as compared with other phase coding protocols. Finally, in Section VII, we present our conclusions.

## 2 Phase coding BB'84 Protocol

As mentioned earlier, the BB'84 protocol was first implemented using polarization encoding. At present, phase encoding is used because using the polarization states causes some alignment and stabilization difficulties in practice. The basic configuration of phase coding is shown in Fig. 1 [7, 11]. The sender (Alice) transmits a photon through an asymmetric Mach–Zehnder interferometer. When the photon passes through the interferometer the phase difference $\theta_a$ between the two paths is randomly chosen from one of four values, namely {0, π} and {π/2, 3π/2} by PMA (phase modulator of Alice). After sending the photon to the receiver, Bob receives the photon and then passes it through his interferometer which is identical to Alice's. The phase difference $\theta_b$ on Bob's side is randomly chosen from {0, π/2} by PMB (phase modulator of Bob).The photon is then detected at one of the two interferometer's outputs, each of which has a detector.

Using this method, a secret key is generated by the protocol which is as follows:

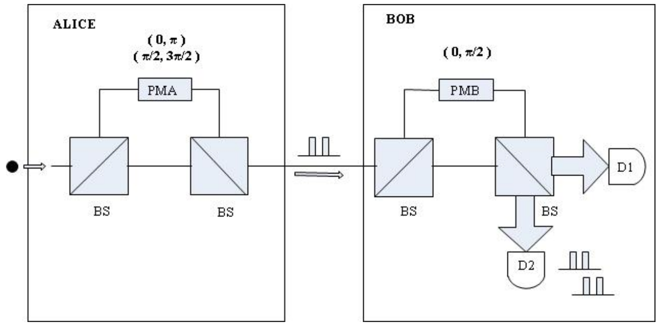


Figure 1. Schematic diagram of phase coding protocol

1) A number of photons are sequentially transmitted from Alice to Bob.
2) After the completion of the transmission through the quantum channel, Bob tells Alice which phase was chosen by him to detect the photon.
3) Alice lets Bob know using the classical channel, whether she chose $\theta_a$ from {0, π} or from {π/2, 3π/2) for the detected photon. This phase information enables Bob to know if the detection event was deterministic or probabilistic.
4) In case of deterministic detection events, Alice considers $\theta_a$ zero (X basis) axis or π/2 (Y basis) as bit "0" and $\theta_b$ equals π (X basis) or 3π/2 (Y basis) as bit "1", whereas, Bob considers the detector 1 click as bit "0" and the detector 2 click as bit "1". In case of probabilistic detection events, they ignore them. A bit string created in this step is the sifted key.
5) Finally, after obtaining the sifted bit string, the two parties should communicate with each other through the classical channel to apply error correction and privacy amplification to obtain the final secret key.

## 2.1 Plug and Play System

In this section, we describe a phase coding auto-compensating system which is part of the Plug and Play implementation [8, 9, 10, 11]. The structure of the phase coding auto-compensating system is as shown in Fig. 2. The steps under which the system operates are as follows:

Bob's Side:

1) Laser (L) produces strongly linearly polarized pulses of photons on Bob' side.
2) The beam is then separated into two equal parts at the 50/50 beam splitter (BS).
3) The long arm of the interferometer contains a delay line (DL) and the Bob's phase modulator (PMB) is not used at this step of transmission.
4) The linear polarization is rotated by 90 degrees in the shorter arm of the interferometer, not shown in Fig. 2.
5) Both the beams then recombine and come out one after another from Bob's side. (The first pulse passed the short and the second pulse passed the long arm of the interferometer).
6) The pulses are then transmitted to Alice through the optical fiber.

Alice's side:

7) Pulses that reach Alice, passing the BS 10/90 (90% of the intensity will be registered in the detector DA).
8) The other output of the beam splitter which is 10% is attenuated by the variable attenuator (VA) and reflected by a Faraday Mirror (FM), where the polarization states are reversed.
9) Alice applies a phase of 0 or π (bit 0 and 1 in the X-basis) and π/2 or 3π/2 (bit 0 and 1 in the Y-basis) on the second pulse for implementing the BB'84 protocol with her phase modulator (PMA).
10) The two pulses coming out as the output from Alice's side are orthogonal to each other but they have their polarizations interchanged because they have been reflected by the Faraday Mirror (FM). Thus, a compensation of all accumulated polarizations changes can take place on the way back from Alice to Bob.

Bob' side:

11) Two pulses arrive at Bob's interferometer and the first pulse now enters the long arm because of the changed polarization states.
12) Bob randomly chooses the measurement basis by applying a 0 or a π/2 phase shift on the first pulse by using his phase modulator (PMB).
13) The second pulse passes the short path in the interferometer.
14) Both pulses arrive at the same time at beam splitter (BS) and interfere with each other.
15) They are detected either in the detector 1 (D1) or after passing through the circulator (C) in the detector (D2).
16) The system is a usual QKD system which is using phase encoding between coherent pulses for transmitting a key from Alice to Bob.

If Bob's phase (PMB) = 0 and Alice's phase (PMA) = 0 or π, then measuring it in the X basis, one of Bob's detectors obtains a conclusive result, which thereby determines the bit to be a 0 or 1. On the other hand, when the phase of PMB is 0 and that of PMA is π/2 or 3π/2, either of the two detectors of Bob clicks with equal probability. This is because Alice chooses the Y-basis and Bob chooses the X-basis, which are different bases. A complementary process happens for PMB=π/2.

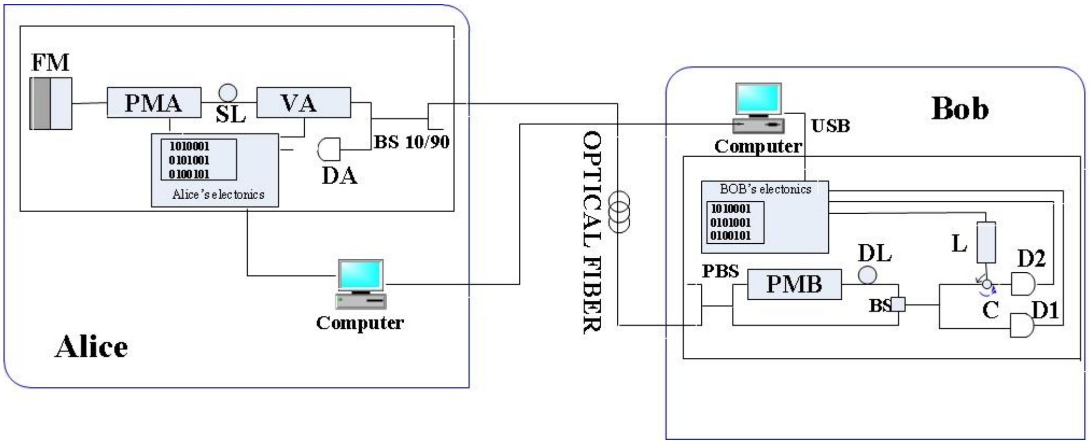


Figure 2. Schematic diagram of Plug and Play system, (FM: Faraday Mirror, BS: beam splitter, PBS: polarization beam splitter, PMA: Phase modulator of Alice, PMB: Phase modulator of Bob, DL: delay line, VA: variable attenuator, C: circulator, D1&D2&DA: detectors, L: laser, SL: storage line)

Alice and Bob after exchanging raw keys through the quantum channel need to communicate with each other through the classical channel to sift the raw key and discard Bob's random clicks. Also, after obtaining the sifted bit string, the two parties should communicate with each other through the classical channel to apply key distillation and Privacy Amplification to obtain the final secret key.

In this paper, we propose a new quantum key distribution protocol, which utilizes a two way quantum channel and does not need the two parties to communicate to get the sifted key. This three stage quantum key distribution protocol will be discussed in the next section in detail.

## 2.2 The Proposed Protocol

As we discussed in the previous section, the Plug and Play system is based on using a classical channel for key sifting and key reconciliation.

Our proposed protocol is based on encoding the qubits in a similar manner to the Plug and Play system, but we propose to use a two way quantum channel for the quantum key distribution. The idea for omitting the classical channel to sift the key comes from Kak's three stage polarization based protocol which is the only QKD protocol using three stages for key establishment [14,15].

The key establishment technique of our proposed protocol is described in Fig. 3. In the proposed protocol Alice has a bit string that she wants to share with Bob as the secret key. The key distribution procedure is achieved in the following way:

1- Strong linearly polarized pulses of photons are produced by a laser from Alice's side. The qubits are encoded in the relative phase between two subsequent pulses.

2- Alice applies an arbitrary phase of α1 to the first pulse and β1 to the second pulse and sends these two pulses through the quantum channel to Bob.

3- Bob receives these pulses and applies another arbitrary phase α2 to the first pulse and β2 to the second pulse and sends them back to Alice. It should be mentioned here that these arbitrary phases are known only to Alice and Bob respectively.

4- Alice again receives the two pulses from Bob and applies –α1 to the first pulse and –β1 to the second pulse for encoding bit "0". On the other hand, for encoding bit "1", Alice applies (–α1+π) and –β1 to the first and second pulses respectively and sends them back to Bob.

5- Bob receives the two pulses from Alice and applies –α2 and –β2 to the first and second pulses. These pulses after passing through the unbalanced interferometer are finally detected either in detector 1 (bit "0") or detector 2 (bit "1") according to their phase difference (0 or π).

The system structure of the proposed protocol is sketched in Fig. 4. In the proposed system, at Alice's side, a laser produces strong linearly polarized pulses of photons. The beam is then separated into two parts at the 50/50 beam splitter (BS) and enters into the unbalanced interferometer with a delay line DL in the long arm, which produces two pulses. Then with the help of the phase modulator (PMA1), Alice applies an arbitrary phase of α1 to the first pulse and β1 to the second pulse and sends these two pulses through the quantum channel to Bob. Now Bob uses his phase modulator (PMB1) to apply another arbitrary phase α2 to the first pulse and β2 to the second pulse and sends them back to Alice. Next, Alice encodes bit "0" and "1", according to the key establishment procedure described above using the phase modulator (PMA2) and sends these encoded qubits back through the quantum channel (optical fiber) to Bob.

Bob receives the two pulses from Alice and applies (–α2) and (–β2) to the first and second pulses using his phase modulator (PMB2). Bob then inverts the polarizations of the pulses and sends them to the interferometer. Due to this reversal of the polarizations of the pulses, the pulses which had travelled through the shorter arms of the interferometer at Alice's side will pass through the long arm of the interferometer at Bob's side and vice versa. As a result, the two pulses will reach the beam splitter where they interfere. Then they are detected either in detector 1(D1) or detector 2 (D2) according to their phase difference. Now Bob considers the detector 1 clicks as bit “0” and the detector 2 clicks as bit “1”.Thus, in this way, Alice shares her identical key with Bob without the process of key sifting and hence obviating the need for a classical channel in this step. We will consider the existence of Eve in our protocol in the next section.

## 2.3 Impact of an Intruder

One of the primary advantages of the proposed protocol is that Alice and Bob use arbitrary phases each time for the key establishment and thus an intruder Eve cannot guess the actual guess the actual phase which is applied to the pulses in each transmission. So, in this case, Eve cannot get any information from avesdropping on the channel and she can only apply some random phases and disturb the information which Bob receives.

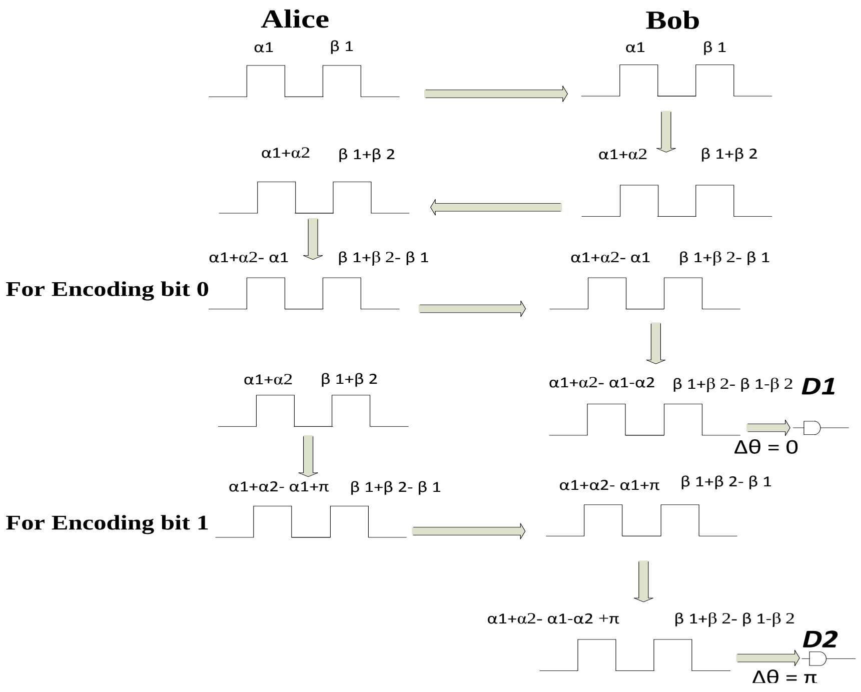


Fig.3. Schematic diagram of key establishment procedure of the proposed protocol

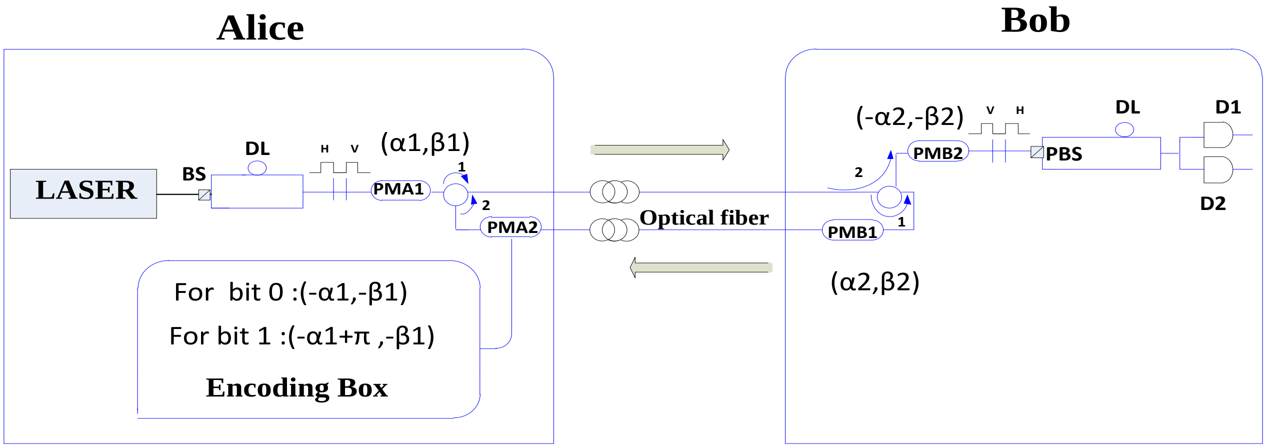


Figure 4. Schematic diagram of proposed protocol system, (BS: beam splitter, PBS: polarization beam splitter,PMA1&PMA2: Phase modulator of Alice, PMB1&PMB2 : Phase modulator of Bob, DL: delay line, C: circulator, D1&D2: detectors)

In the case where Eve applies some random phase to the pulses in order to disturb the information, the phase difference of the two pulses will be detected randomly by the two detectors.

To overcome this vulnerability, Alice and Bob put some portion of the previous established key as a test key and send it at the first of the transmission in order to check the presence of any intruder on the channel. If any of the two parties get some error in the test key, they know that there is an intruder present and they abort the transmission.

Moreover, for avoiding the man in the middle attack, Alice and Bob use this test key part for authentication and thus in this case there is no chance for Eve to attack the protocol by disguising to be the man-in-the-middle.

### 2.4 The Advantages of Proposed Protocol in Comparison with the Other Phase-Coding Protocols

1. Raw key efficiency is defined as the length of the raw key shared by Alice and Bob divided by the length of the random bits generated by Alice [10]. In the proposed protocol due to the usage of two-way quantum channel and three stages of transmission, the raw key efficiency of the protocol can reach 100 percent, which is twice as much as that of the other phase coding protocols [12,13], which is only 50 per cent.
2. The other phase coding protocols need both quantum channel and classical channel for raw key generation, which is basically a weak channel in the sense of security but the proposed protocol only utilizes a two-way quantum channel for raw key generation and this causes Eve's information reduction dramatically.
3. The proposed protocol is more secure according to other protocol because there is no need to transmit the phase information through the classical channel and Eve cannot get any information about the key so the privacy amplification step which is necessary for the other protocols is obviated. As a result the key establishment security and efficiency in our protocol is more than others.
4. In the proposed protocol, Alice can send the encoded key stream to Bob and because there is no key sifting in this method, an effective coding scheme based on forward error correction will reduce the probability of error in the final key to an arbitrary small value so there is no need to have classical channel for error correction. Proposed method uses a three-stage protocol. Fiber-based polarization changes are addressed through the use of phase encoding and an auto compensation technique used in the plug and play system.
5. The proposed protocol is robust against photon number splitting attack and it does not need to have single photon generator because if Eve intercept the photons cannot get any information without having the phase of pulses.

## 3 Conclusions

This paper has presented a novel means for exchanging quantum key between two entities, Alice and Bob. The technique is based on the use of a two-way quantum channel instead of one-way quantum channel as in BB'84. The two way quantum channel is used both for exchanging the quantum key as well as for replacing the conventional key sifting and key distillation on a classical channel. The proposed method thus obviates the need for a classical channel which is an additional security threat. The two-way quantum channel uses a three-stage protocol described in this paper.

## 4 References

[1] Lo HK, Chau HF. Unconditional security of quantum key distribution over arbitrary long distances. Science 1999; 283(5410): 2050-2056.

[2] V.L. Kurochkin, I.G. Neizvestny, "Quantum Cryptography", 10th International Conference and Seminar EDM'2009, Section III, 2009.

[3] V.Teja1, P. Banerjee2, N. N. Sharma3 and R. K. Mittal3," Quantum Cryptography: State-of-Art,Challenges and Future Perspectives", Proceedings of the 7th IEEE International Conference on Nanotechnology, 2007.

[4] M.S. Sharbaf, "Quantum Cryptography: A New Generation of Information Technology Sec urity System", Sixth International Conference on Information Technology: New Generations, 2009.

[5] G. Benenti, G. Casatti, and G. Strini, Principles of Quantum computation, vol. I: Basic Concepts, World Scientific Publishing, New Jersey, 2004.

[6] C. H. Bennett, and G. Brassard, "Quantum Cryptography: Public Key Distribution and Coin Tossing. In Proceedings of IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India, December 10-12, 1984, pp. 175-179.

[7] G. Massimo Palma, "Quantum Cryptography", in Handbook of Information Security, Volume II, Part 3, pp. 606-616, John Wiley and Sons Inc., New Jersey, 2006.

[8] K.Inoue, "Quantum Key Distribution Technologies", IEEE Journal of Selected Topics In Quantum Electronics, VOL. 12, NO. 4, JULY/AUGUST 2006.

[9] A.muller, T. herzog, B.Huttner, W.tittle and H.zbinden," Plug and Play Systems for quantum cryptography", Appl.Phys.Lett. 70, 793(1997) 98, 102

[10] G.Ribordy, J.-D.Gautier, N. Gisin, O. Guinnard and H. Zbinden, "Automated plug and play quantum key distribution", Electronics Letters, VOL.34 ,No.22 , 1998

[11] V.Scarani, H.Bechmann-Pasquinucci, N. J. Cerf, Dusek, N.Lutkenhaus, M. Peev," The Security of Practical Quantum Key Distribution", Foundations of Physics ver.3, 2009; arXiv: quant-ph/0802.4155

[12] K.Inoue, E.Waks, Y.Yamamoto, "Differential-Phase Shift Quantum Key Distribution"Physical Rev. Lett, Vol.89, 2002,037902.

[13] H.Takesue, T.Honjo, K.Tamaki, and Y.Tokura, "Differential Phase Shift-Quantum Key Distribution" IEEE Communications Magazine 0163-6804, 2009

[14] S. Kak, "A three-stage quantum cryptography protocol." Foundations of Physics Letters 19, 293, 2006; arXiv: quant-ph/0503027.

[15] P. Basuchowdhuri, "Comparing BB84 and Authentication-Aided Kak's Three-Stage Quantum Protocol",CITED AS: ARxIV:CS/0703092v1

# Optimization schemes for privacy key management protocol at WIMAX network

Masoud shabani[1], Marjan Abdeyazdan[2], Mohammad Reza Moini[3]

[1] Department of Computer Engineering, Mahshahr branch, Islamic Azad Universiy, mahshahr, Iran.
e-mail: shabani.masoud@yahoo.com
[2]Department of Computer Engineering, Mahshahr branch, Islamic Azad Universiy, mahshahr, Iran.
e-mail: abdeyazdan87@gmail.com
[3]Department of Computer Engineering, Mahshahr branch, Islamic Azad Universiy, mahshahr, Iran.
e-mail: rezamoini_it@yahoo.com

**Abstract:** Standard IEEE802.16 designed to establish relation in the WMAN network wireless portion and triggers data link layer and physical reference model OSI data link layer has a security sub layer at this standard that its role is operating safety in these networks. Security in this sub layer operates by use of privacy key management protocol (PKM) and a simple protocol that performs pictography data formats in the MAC layer . This article studied PKMv2 and PLMv1 and proposed schemes such as applying digital license OMAC and WTLS to increase and improve security and also as for applying ECC for security problems in wireless networks such as authentication , signature and key exchange that proposed use of ECC instead RSA .

**Index Terms:** 802.16 standard , WIMAX security , authentication , PKM , ECC.

## 1. Introduction

IEEE 802.16 is one of the new technologies in wireless communications that nowadays style with WIMAX commercial name . Wireless networks encounter to more vulnerability and threats and thus need to more serious studies in safety field [ 1,2] .Many security mechanisms defined to establish safety in this standard by WIMAX forum . OKM or privacy key management protocol is one of the main WIMAX security protocols that its role is performing authentication , Authorization key , distribution and key synchronization [3,4] . IEEE 802.16-2004 standard uses of initial version (PKMv1) that authentication process performs single –side and thus is mean man attack will be possible . due to existing defects in initial version, present optimized versions and presents second version ( PKMv2) but second version also has defects such as iterative attack and layering attack [5,6,7]. Due to many vulnerabilities and WIMAX threats, accrue among authentication and Authorization key [6,8] , because of authentication performs by PKM protocol , schemes and increasing their security.

### 1.1. article structure

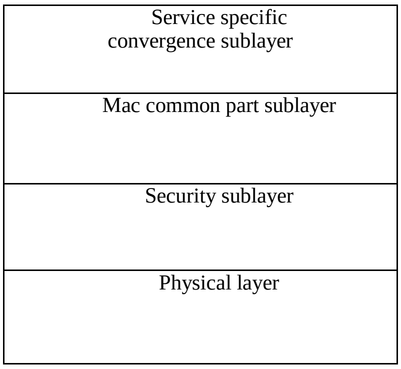
In the second section , we study IEEE 802.16 and security architecture and in the third section , we evaluate first and second version of privacy key management and study an attack to this protocol in fourth section and present schemes for increasing and improving security.

## 2. review on the standard IEEE 802.16

WIMAX born at 2001 of union derived of this name famous to WIMAX forum , that its objective was making development basis and developing and converging activities to trigger wireless metropolitan area networks IEEE 802.16 standard – based . due to fast development of this technology and various applications such as applications of light of sight and non(near)LOS , useful different versions of it . a cross WIMAX ordinary standards , we can suggest to IEEE 802.16 2004 and IEEE 802.16 e. because of WIMAX network is enable to offer wide bandwidth , so can presents services such ad high speed internet , VOIP , live images prevalence , conclave and pictorial communications and etc . to it self users . due to WIMAX gained popularity among it self users, can obtained good situation in new race networks [8].

### 2.1. IEEE 802.16 security sub layer architecture

IEEE 802.16 standard such as other IEEE 802.x standards operate data link layer and reference model physical layer. Layers and sub layers of this standard shown in figure 1 . MAC sub layer of data link layer in this standard has a security sub layer that its role is operating security features at network wireless section [4].



Service specific convergence sublayer

Mac common part sublayer

Security sublayer

Physical layer

Figure 1 . layers and sub layers of IEEE 802.16 standard

This sub layer responsible , encoding operations , key encoding and management . security operates in this layer by applying two protocol:

a) a simple protocol that encode data formats in MAC layer and then passes of network wireless section.

b) PKM protocol that distribution and

synchronizing adaptive keys perform safely from fixed station to client station .

PKM protocol follows of ' client / server" model and to exchange safe key between client station (ss) and fixed station (BS) use of digital Authorization X.509, common key encoding algorithm and power full encoding algorithms such as 3 DES and AES . each SS , apply privacy key management protocol until enable to gain Authorization key (AK) and " data current code key" (TEK) from fixed station. Originality office performs originality from SS to BS and forms a private key (AK) from common key encoding and then (SS) enrolls in network by use of key management protocol . during this process uses of AK to exchange (TEK) [6]

### 3. privacy key management protocol (PKM)

As represented in (2) section, PKM role is performing operations such as authenticator, authorization key and distribution and synchronizing TEK and AK keys. At this section discuss How to do them. Authentication shown in figure 2 briefly.

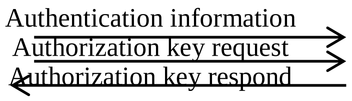

Authentication information
Authorization key request
Authorization key respond

Figure 2. Summary of authentication operations

Routine operation include:

- SS sent " licence issuance demand to BS until request licence of BS.
- BS check demand ant SS identity to response receiving these messages and if confirmed, send encoding set and creates authorization key and send to SS from Auth Rep message. Deviously authorization key coded set in Auth Rep message.

In IEEE 802.16e standard, PKM operates by uses two method. These two methods include: PKMv1 and PKMv2 first version of PKM similar to PKM that in (3) section discussed.

### 3.1. problems of PKMv1

In this protocol use of one – sided authentication and accordingly , man attack is possible other defects that can point of this protocol include role attack or iterative attack [7,6]. First version of this protocol improved by using corrected protocols that use one-sided authentication and protocols that use nance and temporal signet and finally introduced PKMv2 that survey follow.

### 3.2. second version of PKMv2

Support of extensible Authentication protocol (EAP)[10] and two – sided Authentication are advantages of this version on security case . in this version exits an addition message that adds to end of main authentication protocol PKM . authentication protocol of PKMv2 shown in figure.3

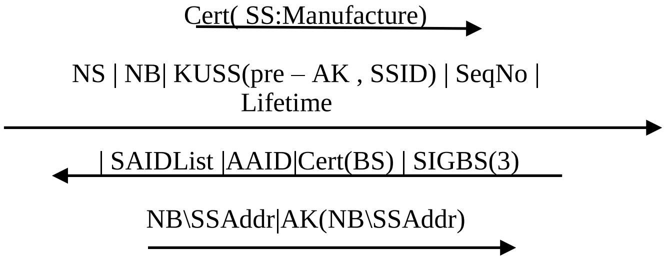

Cert( SS:Manufacture)
NS | NB| KUSS(pre – AK , SSID) | SeqNo | Lifetime
| SAIDList |AAID|Cert(BS) | SIGBS(3)
NB\SSAddr|AK(NB\SSAddr)

Figure -3- authentication protocol of PKMv2

Used characteristics in this protocol discussed in the table 1 .

Table (1) – used characteristics in PKMv2 protocol

| Discussion | Characteristic |
|---|---|
| X.509 certificate belonging to manufacturer | Cert(SS Manufacturer) |
| X.509 certificate belonging to SS that authorized by SS manufacturer | Cert(SS) |
| An identity or number consumes once and use of it to confirm ate new message | NS |
| List of cryptography and authentication set that supported from SS | Capabilities |
| Main connection identity or Basic connection (BC) | BCID |
| Authorization key encoded by common key (SS) | KUSS(AK) |
| 4- bit serial number for AK key | SeqNo |
| Comprise identities and features of initial SA and zero or multi dynamic SA | SAIDList |
| BS signature on third phase message | SAGBS(3) |
| Authorized Association (AA) | AAID |
| (SS's identifier) | SSID |
| MAC address of SS | SSAddr |

### 3.3 . new section of PKv2

Second version of privacy key management has new sections that incomed in table (2) comparatively.

Table (2) – comparing first version and second version of PKM

| PKMv2 | PKMv1 |
|---|---|
| Two- sided authentication | One - sided authentication |
| Support of various methods on Authentication by EAP structure. | RSA – based authentication |
| Types of security convent more over uncast connections of multi cast and over all cast that support by use of Group SA. | Two sided security convent for point to point connections protection . |
| Case machine TEK with cases, messages and new events (comprise 7 cases and 11 events). | Case machine TEK comprise 6 case and 8 event to establish and synchronize TEK key between connection two sides. |
| Of AES and DES algorithms and AES algorithm to exchanged data code. | DES and AES for exchanged data code. |

a- support of various methods on authentication by EAP structure: unlike key management protocol in first version that applys rife key encoding algorithm (RSA) to do authentication operations , in PKMv2 used different method to do authentication operations because of supporting EAP. EAP has series of messages that uses to start and to end a communication main advantage of EAP is allowing its user for applying each authentication protocol by two sides of communication .contents of this method don't define and cause that designers might use appropriative procedures and / or procedures that invent future.

b- Types of security convent: security convent is set of safety information comprising acceptive algorithms , respective parameters , keys , etc. that establish BS with one SS or several SS commonly until through it , can create a safe connection in IEEE 802.16 network . two types of SA presented in IEEE 802.16 version

- Bilinear security convent (SA) , use to protect of single – broad cast.
- Group SA (GSA) use to protect of multi – broad cast.

c- Case machine TEK : case machine TEK has cases, messages , and various events on establishment and synchronizing TEK key between two-sides of connection . case machine TEK in PKMv2 has states , messages and new events the version 2004 , and support of multi – broad cast messages.

d- Cryptography procedures: new cryptography procedures use to cryptography MAC format data and data current cryptography key in this privacy key management protocol . in the e . version of standard use AES and DEC algorithms (to CBC style) and AES algorithm (to CCM style) for exchanged data code between two sides [3]

## 4. attacks to privacy key management protocol (PKMv2)

First attack occurs when communication is without SS signature , in this state requested message is changeable or simply forgeable . even by presence of signature might to attack . in fact , presence of signature in versions that use nance won't helpful scheme on this attack shown in the fiure4 .

α.1 means that message 1 in α sample protocol and IR(SS) , is pervasive that replaced SS. When α executes, IR replace SS and send message that previously sent by SS , message 1 to BS to success in this authentication, IR must respond it with α.3 when receives α.2 from BS .

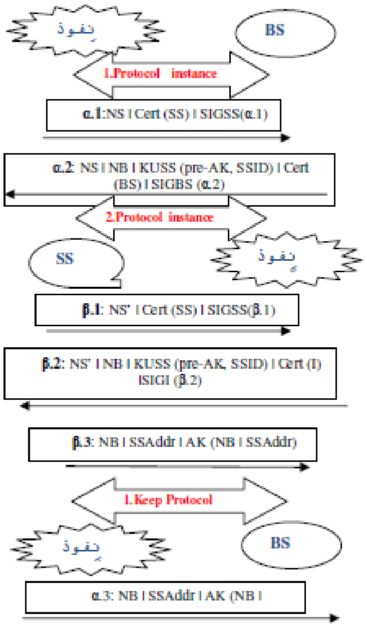

نفوذ
BS
1.Protocol instance
α.1:NS I Cert (SS) I SIGSS(α.1)
α.2: NS I NB I KUSS (pre-AK, SSID) I Cert (BS) I SIGBS (α.2)
2.Protocol instance
SS
نفوذ
β.1: NS' I Cert (SS) I SIGSS(β.1)
β.2: NS' I NB I KUSS (pre-AK, SSID) I Cert (I) ISIGI (β.2)
β.3: NB I SSAddr I AK (NB I SSAddr)
1.Keep Protocol
نفوذ
BS
α.3: NB I SSAddr I AK (NB I

figure 4- layering attack to PKMv2

to respond , it must has AK key until by it encodes nance challenge. But now IR isn't enable to send this respond, because of it can't decode coded message by common key SS and reaches to AK. Though IR can use SS as oracle to respond to this question . but maybe IR forces SS to execute a β protocol . IR send transmitted nance challenge from BS to SS . SS send β.3 massage to IR that includes coded nance and thus IR can send it to BS and end α.

For success in attack remain two problems . firstly, due to AK in PKMv2 obtained from primary authorization key with BSAddr and SSAddr and then to assimilate AKs in α and β , IR must forge BSAddr, that this work simply operates in wireless networks. Next problem is utilization AA by PKMv2 to establish a security session that this work also can operate by forging or iteration via IR to SS . to preven of this attack one can add BSs identity to message 3.

## 5. schemes to increase and improve security

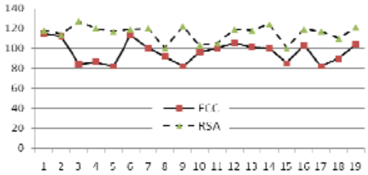
5.1. use Elliptic curve cryptography instead RSA because of most encoder systems actually have limitations such as bandwidth , processing capacity and memory content , applying one system that can provide needed security by smaller key than other systems will be important . small key that applied in ECC- based systems , causes decreasing calculations which increases system efficiency and responds to mentioned limitations above.

ECC can presents solutions for security problems in wireless networks such as authentication, signature and key exchange. A key with length 1024-bit in RSA creates uniform security level by key with length 163-bit in pictography algorithms on Elliptic carve cryptography , other security levels list in table 3 [11].

Table 3- comparison security levels at key length

| | | | |
|---|---|---|---|
| 409 | 283 | 163 | Key size ECC |
| 15360 | 3072 | 1024 | Key size RSA |

Amount of delay yield authentication phase during use of ECC and RSA shown in figure 5. Results from 19 experiments by Quaint 4.5 simulator shown that delay amount is Less than RSA [9].



Performed experiments times
Figure 5- delay yield authentication by applying ECC and RSA

### 5.1. USE OF WTLS AFFIDAVIT INSTEAD X.509 AFFIDAVIT

In the main model , digital licence X.509 send to BS by MS in [9] , proposed an model that in it use of WTLS licence why that occupied space in WTKS licence decreases because of small publisher's identity and serial number , that cause memory store and preventing of use addition memory. Result of [13] shown that ECC operation better than RSA in WTLS licence and moreover economy in memory time, low transmission time required for a handshake.

### 5.3. use OMAC instead HMAC

Mixer function used in main model of HMAC that in [9] propsed a model that instead it, use of one-key CBC MAC which has more simple algorithm and has complete safe stability versus iterative attack. This issue stated in [5]

## 6. result and discussions

New standards IEEE 802.16 support of two privacy key management protocol, OKMv2 and PKMv1 . Authentication was one-sided in PKMv1 and man attack in mean and iteration attack in it is possible, but Authentication is two – sided in PKMv2 and support of Extensible Authentication protocol (EAP). But still doesn't protect versus attacks. Schemes to increase and improve security presented that contain use of ECC instead RSA and OMAC instead HMAC this standard in various security scopes requires to serious surveys, particularly , in Authentication mechanism.

## 7. REFERENCE

[1] Ruixue LI , Zhiyi FANG ,Pang XU , Wei XIAO and Wei WANG, "Experimental Research on a New Authentication Protocol for Wireless Communication Network Based on WiMAX '2008 IEEE

[2] LUO Cuilan, "A Simple Encryption Scheme Based on WiMAX" IEEE, 2009

[3] Institute of Electrical and Electronics Engineers, "IEEE Standard for Local and metropolitan area networks", Part 16: Air Interface for Broadband Wireless Access Systems, IEEE Std 802.16$^{TM}$-2009, 29 May 2009

[4] Syed Ahson, Mohammad Ilyas , WiMAX Standards and Security. Boca Raton: CRC Press, 2008

[5] Michel Barbeau, "WiMax/802.16 Threat Analysis" ACM Int. Workshop on Quality of Service & Security in Wireless and Mobile Networks, Q2SWinet '05,October 13,2005

[6] S.Xu, M. Matthews, and C. –T . Huang, "Security Issues in Privacy and Key Management Protocols of IEEE 802.16", Proceedings of the 44th ACM Southeast Conference(ACMSE 2006), March 2006.

[7] S.Xu, and C. –T . Huang, " Attackes on PKM Protocols of IEEE 802.16 and its later versions" . In ISWCS '06: Proceedings of the 3rd International Sympsium on Wireless Communication Systems, September 2006.

[8] Masood Habib, Massod Ahmad , " A Rrview of Some Security Aspects of WiMAX & Converged Network" , Software and Networks, IEEE 2010

[9] Masood Habib, Tahir Mehmood , Fasee Ullah , Muhammad Ibrahim, "Performance of WiMAX Security Algorithm", IEEE 2009

[10] Aboba, B., Blunk, L. , Vollbrecht, J., Carlson, J ., and H. Levkowetz, "extensible authentication protocol (EAP)", RFC 3748,June 2004

[11] Jia Xiangyu, Wang Chao, 'The Application of Elliptic Curve Cryptosystem in Wireless Communication", IEEE 2005

[12] Albert Levi and Erkay Savas,u "Performance evaluation of public-key cryptosystem operations in WTLS protocol" Proceedings of the Eighth IEEE International Symposium on Computers and Communication, 2003

# SESSION

# SECURITY SYSTEMS AND MANAGEMENT

## Chair(s)

### TBA

# IRS: An Issue Resolution System for Cyber Attack Classification and Management

Chris B. Simmons, Sajjan Shiva, Vinhthuy Phan,
Vivek Shandilya
Department of Computer Science
University of Memphis
Memphis, TN, USA
{cbsmmons, sshiva, vphan, vmshndly}@memphis.edu

Lakisha Simmons
Department of Management of Information Systems
Indiana State University
Terre Haute, IN, USA
lakisha.simmons@indstate.edu

*Abstract*—**Cyber-attacks have greatly increased over the years, where the attackers have strategically improved in devising attacks toward a specific target. In order to correctly classify cyber-attacks there is a considerable need to neatly organize a representation scheme that is useful in an application setting. The classification of cyber-attacks within knowledge bodies, such as Computer Emergency Readiness Team (CERT) and Common Vulnerabilities and Exposures (CVE), are very useful for organizations gathering data as information is made available. However, there is substantial information to decipher when locating relevant details that are prevalent in local networks. We propose an issue resolution system (IRS) to detect and extract information from external vulnerability repositories and internal log files to assist with classifying and disseminating defenses. In this work we provide a frequent pattern classification algorithm that performs data mining techniques to classify attack vector information from the national vulnerability database (NVD). The results suggest the IRS presents a viable solution to correctly extract vulnerability information within a local knowledge base.**

***Keywords-Security; Security Management, Information Extraction; Algorithm; Taxonomy***

## I. Introduction

Cyber-attacks have received increased attention over the last decade, where researchers are investigating the relationships between attacks and the associated defenses. Organizations, particularly small-to-medium sized, lack the capacity to effectively capture cyber-attack related information and disseminate appropriate defenses. These particular organizations rely on a select set of knowledgeable security personnel to resolve cyber-attack related issues. With cyber threats on the rise, it is necessary to correctly identify the suspected threat in a timely manner. Frequent pattern analysis has been used consistently within data mining through the ability to relate patterns.

Frequent patterns are defined as itemsets, subsequences, or subsets that appear in a data set with a certain level of frequency [1]. Sequential pattern mining is the discovery of frequent subsequences [2]. Both frequent and sequential analysis types are beneficial in correlating attack vector information using frequencies and sequences. Correlation is typically used with machine learning approaches or pattern detection algorithms. Lee et al. [3] used timed signatures to tag signatures in discovering database intrusions. We extend this approach to a repository of common vulnerabilities and exposures (CVE), where the vulnerabilities are associated to a specific application setting.

Han et al. [1] highlighted the importance of frequent pattern analysis for data indexing, classification, clustering, etc. Accurate cyber-attack classification is pertinent for suitable damage assessment and recovery. Web logs are used locate potential attack vectors within a particular application. Parsing web log files and database insertions enable an analysis of the current state and transitions of the application. Information stored within the repository can be easily queried to retrieve frequent and/or sequential data items. Although beneficial, using this method can become expensive when querying a large dataset. Pei et al. [2] highlighted the need for extending sequential patterns towards considering time constraints, time windows, and/or a taxonomy. We extend this concept by utilizing a cyber-attack taxonomy, consisting of Attack Vector, Operational Impact, Defense, Informational Impact, and Target (AVOIDIT). AVOIDIT facilitates a classification mechanism that increases the efficiency of correlating attack vector information. From a cursory scan of literature, there is a lack of research focused on correlating external knowledge bodies, such as Computer Emergency Readiness Team (CERT), with internal extracted information, such as web log files, to produce a knowledge base containing a sequential order of attack steps. One of the problems faced by research pertaining to attack classification is how to classify the vast nature of attacks and their potential to polymorph. Understanding cyber-attack defense is chess and not checkers, we can provide a way to capture the wide array of moves an attacker may take through appropriate classification and response.

In this paper, we propose initial work of an Issue Resolution System (IRS) for extracting and dessiminating attack vector information in a local application setting. The IRS uses a classification algorithm which consists of two major methods, a classification method and a decision tree method. The classification algorithm uses AVOIDIT to identify the related attack vector information for classification. Once classified, IRS presents the information via a SilverStripe knowledge base for analysis. We demonstrate the IRS applicability through mining and extracting 160 CVE descriptions from the National Vulnerability Database.

This paper is organized as follows. In section II we highlight the literature involving correlation and frequent pattern algorithms. In section III we propose the issue resolution system's architecture and describe its components in detail. In section IV we provide a preliminary experiment and results of our issue resolution system and section V we conclude our work with insight into future work.

## II. Related Work

There are several recent efforts regarding techniques to automate correlating attacks, where defenders can view a collection of data seamlessly. However, this task creates a massive amount of data that defenders are unable to decipher in a reasonable amount of time. There is a wide array of pattern analysis techniques. In this section we provide a review of literature relating the frequent pattern analysis.

Hu and Panda [4] proposed a data mining approach for detection intrusion alerts targeted towards data corruption. Their approach concentrates classification rules to mine the database for dependencies between two or more data items. Hu and Panda [4] used the database logs to deduce data dependencies. Data dependencies that are not compliant are flagged as anomalous. The result performance increased where dependencies were stronger amongst data items.

Han et al. [1] proposed a frequent-pattern tree approach to mining large amounts of frequent patterns in a transactional database. Han et al. provides an extension to the Apriori algorithm through the use of partitions, divide-and-conquer growth patterns. This approach utilizes solutions to smaller tasks. The approach scans the database twice, one to get frequencies and another to construct the frequency tree. Efficiency is achieved using a three techniques, a large database is compressed into smaller data structure, a fp-tree-based mining using pattern-fragment growth to avoid costly generation, and partitions-based divide and conquer method.

Leung, et al. [5] proposed a canonical-order tree algorithm that captures the content of the transaction database and orders tree nodes, called CanTree. This work provided an extension to the FP-tree algorithm for incremental mining. Leung, et al. uses CanTree to efficiently arrange tree nodes according to canonical order, which are unaffected by frequency item changes. This provides easy maintenance when transactions are modified within a database.

Zaki [6] presented a TreeMiner algorithm which discovers all frequent subtrees in a forest. This novel algorithm performed a depth first search for frequent subtrees using a tree representation called scope list. The use of scope list improved the ability for fast support counting of candidate trees. Cheung and Zaiane [7] proposed a FELINE algorithm, which is tree based incremental mining algorithm, containing a CATS tree.

Pei et al. [8] proposed a prefix-projected sequential pattern mining (Prefix-Span) that explores prefix-projection in sequential patterns. Prefix-Span was developed to reduce the time of subsequence generation while mining the complete set of patterns. The goal of PrefixSpan is to examine the prefix subsequences which allow a representation of the postfix subsequences in the database. They presented valuable information needed to successfully gather the prefix attack information from various input sources in order to successfully disseminate the post attack information. This provides insight into the proposed IRS, as attack vector information becomes available it can be used to retrieve potential defenses regarding suspecting attacks.

Ning et al. [9] proposed three utilities to facilitate correlating a large dataset of IDS related alerts. These utilities are adjustable graph reduction, focused analysis, and graph decomposition. This resulted in the correlation of using consequences of earlier events with prerequisites later events. Ning et al. [9] presented an interesting approach to navigate through the enormous amounts of data captured from an IDS. Ning et al. [10] follow-up work is an extension which focused on correlation to construct attack scenarios using hyper-alert type representing prerequisite and consequences of each alert type of an attack.

We propose IRS, which suggests a new correlation algorithm that uses a cyber-attack taxonomy towards a classification of attack vector information. The correlation algorithm uses the discovery of new attack vectors in aspiration of establishing a relationship between attack vectors based on frequency of sequential events. We further propose a tree based algorithm to be used within a knowledge repository to use the classified attack vectors information with assisting defenders to view the complete path of an attack.

## III. Issue Resolution System Architecture

The issue resolution system enables an organization to use a formalized apparatus to communicate seamlessly regarding the discovery of attacks and defenses. It supports security awareness within organizations by offering attack identification, attack classification, and assist with attack resolution to the identified attack. IRS provides seamless communication by using the following five components: (i) an Ontology, (ii) a Cyber-attack Taxonomy, (iii) a Classification algorithm, (iv) a Log Parser, (v) and a Knowledge Base. Figure 1 depicts each component in a diagram of the issue resolution system.

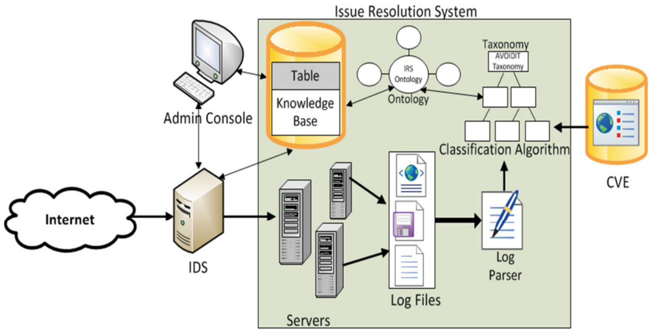


Figure 1. Issue Resolution System Architecture

The ontology is representative of a global communication scheme which consists of the cyber-attack taxonomy. The ontology uses input from the knowledge base and the classification algorithm to assist with facilitating identification, response, and resolution.

The taxonomy is based on a cyber-attack taxonomy called AVOIDIT [11]. Based on the parameters provided as input from external repositories and internal log files, the taxonomy uses the classification algorithm to identify the characteristics of an attack.

The classification algorithm uses frequent and sequential pattern analysis to classify attack vector information. This information is then stored in the repository to identify relevant attacks and potential defenses.

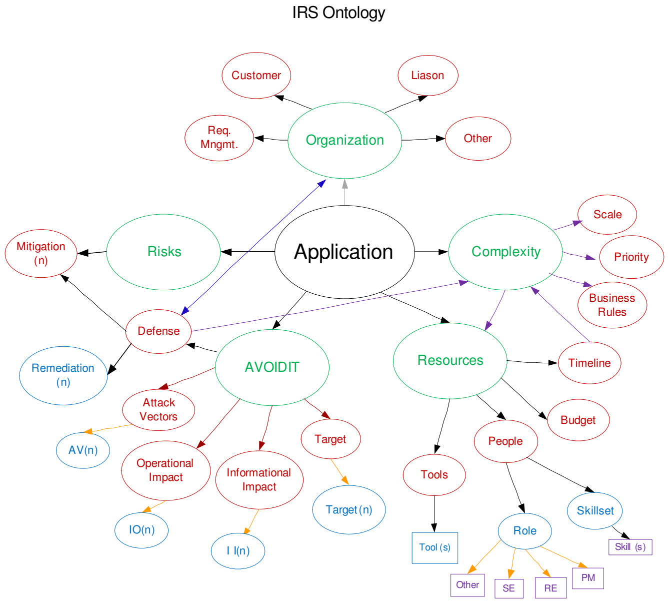
The Log Parsers are scripts written as sensors to identify potential intrusions. These scripts are written in Microsoft Log Parser and provide the ability to parse various log files for suspicious activity. In this paper we limit the discussion to web server log files.

Once an attack is identified, the knowledge base acts as the repository and searches for additional attack vector information. The knowledge base stores information related to the sequence of an attack and provides potential defenses to mitigate and/or remediate the damage to a system.

## A. *IRS Ontology*

An ontology is an explicit specification of concepts related to a specific domain and the relationships amongst those concepts to create an organized knowledge base. Ontologies are a common way to organize knowledge and involves the description of objects and relationships [14]. Cyber-attack management is critical in the application of the IRS. The ontology processing will capture attack details from the knowledge base to begin attack analysis.

Figure 2 highlights the IRS ontology to support the communication flow within an organization upon attack discovery. The objective of the security ontology is to provide knowledge representation of the most relevant security concepts within an organization. The **ovals** refer to major concepts that are needed to successfully communicate an incident within the IRS. The **boxes** refer to the terminal entities that provide specifics of the superclass concepts. The **arrows** refer to the relationships between concepts relevant to incident management.



→ **associatedWith** – who assigns a task

→ **has/have** – information needed to complete superclass. The direction is one-way.

→ **is** – an option associated with a higher level concept.

→ **consistOf** – is the superclass concept that contains vital subclass relationship.

→ **approvedBy** – who approves

→ **influencedBy** – the associated impact

Figure 2. IRS Ontology

## B. *Cyber Attack Taxonomy*

This section describes the cyber-attack taxonomy used in an application setting to classify attacks. Simmons, et al. [11] provided the AVOIDIT cyber-attack taxonomy to support comprehending each attack classification and how a variety of attacks are represented in each category. Using AVOIDIT we classify attacks via attack vectors where the repository disseminates the potential attack for an appropriate defense selection. Due to space constraints, Figure 3 highlights a representative subset of AVOIDIT, where each classification is defined in the following sections. Our approach follows this pattern:

- **Classification by Attack Vector**
  When an attack takes place, there is a possibility that it uses several vectors as a path to a complete cyber-attack. An attack vector is defined as a path by which an attacker can gain access to a host.

This definition includes vulnerabilities, as it may require several vulnerabilities to launch a successful attack.

- **Classification by Operational Impact**
  Classification by Operational Impact involves the ability for an attack to culminate and provide high level information known by security experts, as well those less familiar with cyber-attacks. We provide a mutually exclusive list of operational impacts that can be categorized and concisely presented to the public.

- **Classification by Defense**
  We extend previous attack taxonomy research to include a defense classification. We provide the possibility of using both mitigation and remediation when classifying attack defenses, as an attack could be first mitigated before a remediation can occur.

- **Classification by Informational Impact**
  An attack on a targeted system has potential to impact sensitive information in various ways. A committed resource must be able defend information warfare strategies in an effort to protect themselves against theft, disruption, distortion, denial of service, or destruction of sensitive information assets [15].

- **Classification by Target**
  Attacks target a variety of hosts, leaving the defender unknowingly susceptible to the next attack. This section is used to classify targets an attack uses to perform unauthorized privileges.

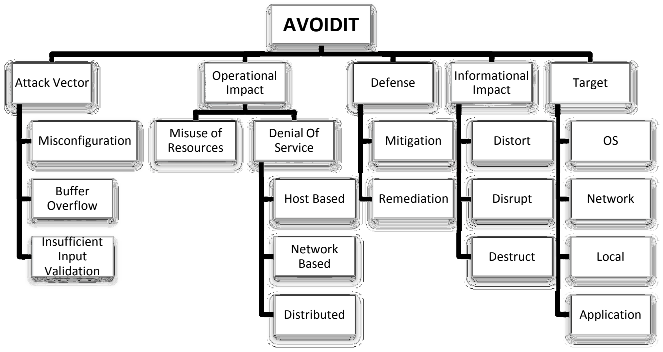


Figure 3. AVOIDIT: A Cyber Attack Taxonomy

### C. *Classification Algorithm*

The classification algorithm is the functioning mechanism behind the issue resolution system, which takes the keywords as input, matches them with those that are accounted for in the AVOIDIT taxonomy to traverse the tree to reach a leaf node. This provides the ability of identifying the attack described by the keywords. The algorithm is implemented in a program written in PHP. The MySQL database is used for the representing and analysis of attack vectors.

Each set of input strings the application retrieves the associated attack vectors using the classification algorithm (Algorithm 1) and AVOIDIT to classify input description against keywords within the repository. If the input description has at least one match, then we select all the classifications of attack vectors associated to the input description. Using each classification correlated with the input description, the frequency is retrieved. The classification algorithm retrieves the attack vector that contains the highest probability of the attack vector being associated to the description received from the input. Once the attack vectors are returned, the attack vector is passed to the tree algorithm (Algorithm 2) to retrieve the associated attack tree. This provides discovery of the initial and later parts of the identified part of the attack to complete the attack. The algorithm is an intuitive depth search algorithm searching all attack trees to find a match. This information can be used by the defender to assist with devising defense strategies. The application continuously receives input description from the vulnerability repository to update information related to attack vectors.

Table 1. Notations used in AVOIDIT correlation algorithm

| **Notation** | **Definition** |
|---|---|
| $n$ | Extracted string from a log |
| $cav_i$ | Classifications of previous attack vectors |
| $T$ | Attack tree |
| $v$ | Set of Vertices |
| $x$ | Parent node |

Algorithm 1. Classification Algorithm (A data retrieval algorithm, which is retrieved directly from repository)

**Input:** a set of strings $N$.
**Output:** A set of attack vectors $cav$ names associated to set $N$ and $freq$

**The** *AVOIDIT Classification* **Algorithm**

```
for each $n \in N$
    if $n_i$ is at least one match
        select all classifications $cav$ including frequency
    else if
        for each variation of $n_i$ //uses php pecl function
            select all classifications $cav$ including frequency
        end for
    else
        set $n_i \rightarrow$ unclassified
    end if
end for
retrun classifications $cav$ and $freq$
```

The CVE descriptions or application log files are parsed for keywords, which are described in Algorithm 1. Algorithm 1 accepts an input file containing a set of strings. The algorithm parses the data and uses the keyword matches within the repository to classify new attack vectors and their associated target. The algorithm retrieves the associated classifications for each keyword recognized within the repository.

The premise behind the attack tree algorithm is to provide a database scan of the classified attack vectors to retrieve the associated trees. If no tree exists, the attack vector becomes the root level for future related attack tree retrieval. If a tree exists for the classified attack vector, the child nodes are compared with other existing nodes. If a tree exists where two or more attack vectors contain the same parent the attack trees are merged into one tree. This assists an organization with capturing knowledge to itemized phases of a complete attack.

Algorithm 2. Attack Tree Algorithm (A depth first search to retrieve attack tree)

**Input:** a set of classified attack vectors *CAV*.
**Output:** A set of attack trees $T_{av}$ that simulates the complete path of an attack.

**The** *AVOIDIT Tree* **Algorithm**

```
for each $cav_i \in CAV$
  find corresponding attack Trees $T_n$
  if $cav_i$ does not have a corresponding $T_n$
    $cav_i \rightarrow T_{av}$
  else
    Given a set of tree $T_{1\ to\ n}$
    for each $T_n$
      for all $v \in T_n$
        visited $(v) = false$
      for all $v \in T_n$
        if not visited $(v)$: $explore\ (T_n)$
      construct a tree $T_{av}$ s.t. all $v$ are covered for $cav_i$
      set parent in $T_{av}$ of $cav_i \rightarrow x_i$
    end for
  end if
end for
for all $x_{1\ to\ i}$ are equal
  union $T_{av}\ (x_i, x_{i+1})$
return a set of attack trees $T_{av}$.
```

## D. Log Parser

Log data is capable of recording important events, which should be analyzed on an ongoing basis, consistent with the monitoring of other key centralized security controls [12]. MS Log Parser [13] tool was developed to parse recorded events that have occurred in a system and/or application. It contains a core SQL engine facilitating the use of data repository for further analysis.

IRS uses MS Log Parser to turn large amounts unstructured text into a form (in a MySQL database) that can be manipulated to understand patterns, relationships, and meanings by using sensors. This enables IRS to retrieve various log data via a local network to correlate with pre-existing attack vectors. LogParser Query 1 highlights example queries used to retrieve pertinent events within a Windows registry or from an IIS web server log file.

LogParser Query 1. Windows Registry Events and Web Log File Events

**Input:** Windows registry location.
**Output:** A set of registry keys that have been modified within the past 24hrs.

**Example** *Log Parser* **Query**

```
logparser SELECT path, valuename from hklm\software where lastwritetime >= sub(system_timestamp(), timestamp('0000-01-02', 'yyyy-MM-dd')
```

**Input:** IIS Web log file.
**Output:** A set of status code changes for a selected files.

**Example** *Log Parser* **Query**

```
logparser SELECT DISTINCT date, time, c-ip, cs-uri-stem, sc-status from %web.log% WHERE c-ip IN (SELECT DISTINCT c-ip FROM %web.log% WHERE sc-status = 304) AND (sc-status=200 OR sc-status=304) ORDER BY date, c-ip
```

## E. Knowledge Base

The KB is a component within the IRS that facilitates the storage of various attack related information. Once the IRS identifies an attack vector, this information is forwarded to the KB which then extracts further information related to the attack vector to ensure appropriate classification and retrieve potential solutions.

The attack related information consists of external vulnerability descriptions and internal log data. We utilize information from various sources depicting the complete path to an attack. These various sources highlight information which can be used to correlate disparate and unstructured data within the KB.

We envision the KB to work in either an offline or online mode. The KB can operate in online mode by as described above. The KB operates in offline mode by updating itself semi-autonomously by using its access to online vulnerability databases and security expert intervention.

# IV. Experiment and Results

This work emphasizes an issue resolution system appropriate for auditing external repositories and internal web log files to correlate attack vector information. In this paper focus was placed on the classification algorithm for searching and classifying attack vectors information within a repository. Experiments show the classification algorithm is practical within the issue resolution system to proceed with further development.

### A. *Methodology*

The use of the common vulnerabilities and exposures (CVE) database was used to classify pre-existing attack vectors. In conducting this preliminary experiment the CVEs from the National Vulnerability Database (NVD) were used for classification. Considering the massive number of CVEs, focus was placed on a real world scenario, where an organization uses Joomla! as a web based content management system.

The methodology for our experiment involved a training set of 60 positive CVEs associated with Joomla! for learning. Table 2 depicts a CVE description specific to Joomla!, which highlights the vulnerability information. The algorithm was trained using a standard unigram bag of words approach. In each CVE description, experiential knowledge was used to classify the concepts of interests from the text relative to the attack vector, operational impact, defense, informational impact, and target.

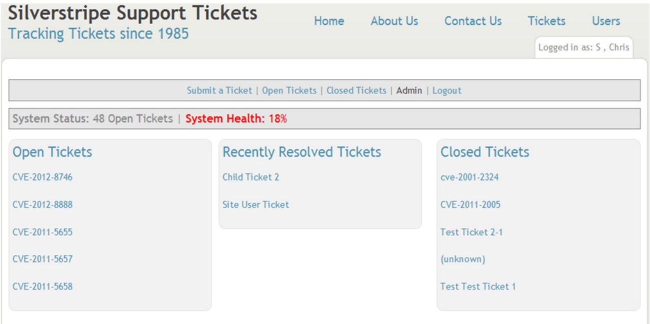
Table 2. CVE Description for Joomla!

| **CVE -2011-4808** | |
|---|---|
| **Summary** | SQL injection vulnerability in the HM Community (com_hmcommunity) component before 1.01 for Joomla! allows remote attackers to execute arbitrary SQL commands via the id parameters in a fnd_home action to index.php. |
| **Published** | 12/14/2011 |
| **CVSS Severity** | 7.5 (High) |

The second step involved a test set of 100 unique CVEs associated with Joomla! and 50 random CVEs (for noise) associated with various open and closed source software. The goal is to ensure Joomla! related CVEs are correctly classified and distinguished from irrelevant CVEs. The irrelevant CVEs were discarded by the IRS. Performing this step simulates data being pushed to the user via the IRS providing needs specific attack vector information. Preliminary evaluation on this minimal dataset highlighted the algorithms ability to correctly classify application pertinent incidents specific to an organization.

### B. *Results*

In this section we provide preliminary results of the IRS prototype giving insight to the potential success of our concept. Figure 4 highlights the Silverstripe knowledge base used for our experiment.



Silverstripe Support Tickets
Tracking Tickets since 1985

Home About Us Contact Us Tickets Users

Logged in as: S , Chris

Submit a Ticket | Open Tickets | Closed Tickets | Admin | Logout

System Status: 48 Open Tickets | System Health: 18%

| Open Tickets | Recently Resolved Tickets | Closed Tickets |
|---|---|---|
| CVE-2012-8746 | Child Ticket 2 | cve-2001-2324 |
| CVE-2012-8888 | Site User Ticket | CVE-2011-2005 |
| CVE-2011-5655 | | Test Ticket 2-1 |
| CVE-2011-5657 | | (unknown) |
| CVE-2011-5658 | | Test Test Ticket 1 |

Figure 4. Silverstripe knowledge base interface

Precision was used to measure the accuracy in which the classification algorithm was able to correctly classify attack vector information. The use of precision was measured by dividing the total number of correctly classified items by the total number of extracted items, as provided in Equation 1. Recall was also used to measure the percentage of available correct information extracted, which highlights the algorithms ability to extract relevant information. Recall is the number of correctly classified by the total possible correct classifications, as provided in Equation 2. Table 3 displays the actual precision and recall computed for both the training set and test set using the classification algorithm.

$$Precision = \frac{Total\ Correctly\ Classified}{Total\ Extracted} \qquad (1)$$

$$Recall = \frac{Total\ Correctly\ Classified}{Total\ Possible\ Correct\ Classifications} \qquad (2)$$

Table 3. Algorithm Precision and Recall Computation

| | CVEs Correctly Classified | Precision | Recall |
|---|---|---|---|
| Training Set | 47/60 | 78.3% | 78.3% |
| Test Set | 92/100 | 92% | 92% |

Based on the results we can derive the following facts. The percentage of correctly classified CVEs are relatively high. This may be due to the use of only Joomla! as the test application for this experiment. It further seems a lot of CVEs mainly deal with input validation vulnerabilities, which allowed our classification algorithm to remain somewhat consistent.

Although, the classification algorithm is preliminary, it presents an approach of consideration using machine learning and good information extraction to increase efficiency. The time required for processing attack vector information was reasonably small satisfying any stringent timing requirement.

## V. Conclusion

It is imperative the attack vector information is correctly classified and disseminated throughout an organization. In this paper we presented a new approach to attack vulnerability classification that locates keyword matching from the NVD repository containing CVEs. Such keyword matching is correlated to the local network for issue resolution of the current system.

Our preliminary experiment indicated a promising solution so as to move forward with more detailed research. We assume the collected keywords were pertinent to the type of attack vector information within the repository and can be used generally. We believe providing a mechanism to distribute public information and correlate within a local network will assist organizations with identifying the underling details of an attack. Although our approach is preliminary, we believe the approach provides a direction for information extraction and machine learning in security management.

Future work will investigate an elaborate machine learning technique through the use of natural language processing that will allow the classification algorithm to learn various descriptions within the CVE for use within a local setting. We will use attack test beds such as metasploit [17] to explore the ability of the classification algorithm to identify the correct attack.

We further investigate a technique intended to use the AVOIDIT algorithms in a game inspired defense architecture with aim to extend the functionality of previously proposed game models addressing a broader array of cyber and data engineering problems [16]. Using the AVOIDIT algorithms we intend to build a Game Theoretic Defense System, which will investigate the applicability of AVOIDIT in determining the action space of the attacker and defender.

## REFERENCES

[1] Han, J., Pei, J., Yin, Y.,  and Mao, R. Mining frequent patterns without candidate generation: a frequent-pattern tree approach. Data Mining and Knowledge Discovery, 8(1), pp. 53–87, Jan. 2004.

[2] Pei, J., Han, J., Asi, B.M., Pino, H. PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth, in: Proc. The Seventeenth International Conference on Data Engineering, April 2001, pp. 215–224.

[3] Lee, V. C.S., Stankovic, J. A., Son, S. H. Intrusion Detection in Real-time Database Systems Via Time Signatures. In Proceedings of the Sixth IEEE Real Time Technology and Applications Symposium, 2000.

[4] Hu, Y., & Panda, B. A Data Mining Approach for Database Intrusion Detection. Proceedings of the 19th ACM Symposium on Applied Computing, Nicosia, Cyprus, 711-716, 2004.

[5] Leung, C.K., Khan, Q.I., Li, Z., Hoque, T. CanTree: a canonical-order tree for incremental frequent-pattern mining, Knowledge and Information Systems 11(3) (2007) 287–311.

[6] ZAKI, M. J. Efficiently mining frequent trees in a forest. Inf. Syst. 17, 8, 1021 – 1035, 2005.

[7] Cheung, W. and Zaiane, O.R. “Incremental mining of frequent patterns without candidate generation or support constraint,” In Proc. IDEAS 3003, pp. 111–116.

[8] Pei, J., Han, J., Asi, B.M., Pino, H. PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth, in: Proc. The Seventeenth International Conference on Data Engineering, April 2001, pp. 215–224.

[9] Ning, P., Cui, Y., and Reeves, D. S. Analyzing intensive intrusion alerts via correlation. In Proc. of the 5th Int'l Symposium on Recent Advances in Intrusion Detection, October 2002.

[10] Ning, P., Cui, Y., Reeves, D.: Constructing attack scenarios trough correlation of intrusion alerts. In: CCS '02: Proc. 9th ACM Conference on Computer and Communication Security, ACM Press (2002) 245-254.

[11] Simmons, C., Shiva , S., Dasgupta , D., and Wu, Q., “AVOIDIT: A cyber attack taxonomy,”, Technical Report: CS-09-003, University of Memphis, August 2009.

[12] Scarfone, K., Grance, T., Masone, K. “Computer Security Incident Handling Guide,” NIST SP 800-61, *2008*.

[13] Giuseppini G, Burnett M, Faircloth J, Kleiman D. Microsoft Log Parser toolkit. Syngress; 2005.

[14] Tran, Quynh-Nhu Numi, and Graham Low. "MOBMAS: A methodology for ontology-based multi-agent systems development." *Information & Software Technology* 50, no. 7/8 (June 2008): 697-722.

[15] Cronin, B. and Crawford, H., "Information warfare: Its Application in military and civilian contexts", Information Society, volume 15, pp. 257-263, 1999.

[16] Shiva, S., Dasgupta, D., Wu, Q. “Game Theoretic Approaches to Protect Cyberspace,” Office of Naval Research, Grant Number N00014-09-1-0752, 2009.

[17] http://www.metasploit.com/, retrieved on March 3, 2012.

# CamInSens - An Intelligent in-situ Security System for Public Spaces

**David d'Angelo[1], Carsten Grenz[2], Colin Kuntzsch[2], and Manfred Bogen[1]**
[1]Fraunhofer IAIS, Schloss Birlinghoven, Sankt Augustin, Germany
[2]Leibniz Universität Hannover, Hannover, Germany

**Abstract**— *In this paper we present a novel third-generation surveillance system for public spaces. In contrast to analyzing a threat after its occurrence, this systems aims on* in-situ *detection of salient trajectories and events. We use a network of smart camera nodes, which is capable of detecting and tracking people across an entire camera network. In addition, we developed an analysis framework which uses the extracted trajectory and event information from the smart camera network for autonomous online detection of potential threats in real-time. Finally, we present new visualization and interaction techniques for security control rooms and mobile devices based on an evaluation of existing infrastructures and components.*

**Keywords:** Surveillance Technologies, Security Operations, Event Detection, Trajectory Analysis, Human Computer Interaction

## 1. Introduction

Research on *Preventive Security* is concerned with looking for and implementing intelligent strategies and processes prior to security incidents happening and so minimizing security risks across all sectors. *In situ* event detection is the overall challenge. Security risks in this sense are process breaches, technical defects, and targeted attacks such as asymmetrical threats, cyber crime or human error. The objective is to develop concepts and technologies aimed at recognizing these security threats at an early stage. The disaster of the love parade in Duisburg, Germany, is an example where the application of preventive security strategies could have saved the life of many humans.

The CamInSens project [2] sponsored by the German Federal Ministry of Education and Research explores new ways of recognizing danger situations induced by people themselves. It is about the protection of public areas which differs from the protection of critical infrastructures because humans and the public become the center of interest. In order to protect public areas from dangers caused by people themselves, technologies are needed which can find the balance between increased levels of security and existing surveillance regulations.

The focus of the CamInSens project is to explore the possibilities of a practical and legally compliant surveillance system based on a network of smart video cameras and automatic event detection in order to increase the public safety without compromising data and privacy rules. We expect the output of the system to be highly relevant to operators and security staff to detect and assess critical situations in good time and to provide the information base for resolving them. One of the key components we developed is a new camera system which can acquire and process visual data cooperatively using intelligent camera nodes. This allows the detection and tracking of people across an entire smart camera network. In order to minimize the manual interaction in the data analyzing process, we further developed algorithms to automatically analyze and classify trajectories and other sensor events. This creates the possibility for an autonomous preemptive detection of potential threats and it decreases workload for the security personnel.

All information and classification results are transmitted to a security control room as well as to mobile devices of security personnel on patrol in real-time, where it is visualized and evaluated by the security personnel in a convenient way. We also examine questions of technical and operational reliability such as safeguarding against failure or failure tolerance. Another focus is the analysis of economic viability and the regulatory framework [6].

In the remainder of this paper we discuss related work, then we explain the overall system setup followed by a detailed description of the single components. Finally, we draw a conclusion and give an outlook of future work.

## 2. Related Work

Surveillance systems are technological tools that assist humans by augmenting their limited (physical) capabilities regarding data collection, storage and reasoning about situations of interest occurring within the monitored environment [19]. Over the last 20 years, much effort has been put into researching methods of automated monitoring, in particular video surveillance, in order to satisfy the ever-increasing need for safety and security within society [22]. During this time, video surveillance systems have evolved into three generations of surveillance systems ( [7], [18]) by gradually replacing analogous hardware with digital equivalents.

The first generation of video surveillance systems, which still represents the most common practical application, is centered around human personnel performing visual inspection of video data. Security personnel typically monitors multiple locations simultaneously, which results in significant stress on the security personnel involved [8]. Later, second-generation surveillance systems tried to resolve the burden on security personnel by (partially) automatizing the monitoring task by using automatic object detection/tracking within sequences of images and automatic event detection. Most recently, the third generation of surveillance systems utilized distributed networks of smart cameras/sensors with fully digital back-ends allowing for integrated data collection, storage and analysis processes. For comprehensive reports on recent developments in object detection, tracking and classification, please refer to e.g. [4], [13].

The surveillance system evolution process is driven by the continuous increase of cost-efficient hardware availability. Contrary, these surveillance systems are restricted by real-time performance needs and limitations in terms of available resources. Several automated surveillance systems have been proposed in recent years; a comprehensive review of several video surveillance systems can be found in [22].

## 3. System Description

This section describes the overall *CamInSens* system architecture, followed by a detailed description of the single components.

Apart from its legal compliance, the system under development had to fulfill several requirements in order to meet today's standards in terms of third generation video surveillance systems. This includes an in-situ event detection, an autonomous online detection of potential threats, autonomous and smart cameras, an automatic tracking of persons over multiple cameras, an automatic classification and analysis of trajectories and a strong coupling of the security control center with mobile operators on patrol. These requirements will be described in this chapter in great detail.

### 3.1 System Design

The two main input data sources of the system are a smart camera network and a set of additional sensors, which are combined into a smart sensor network (see 3.2). These autonomous sensors can be of arbitrary nature, such as theft, glass breakage or fire sensors and every state change of interest is directly reported to the central CamInSens gateway server. Each smart camera runs a *Multi-Object Tracker* instance, which accesses a camera's image data and extracts partial human movement trajectories. These results are passed to the *Multi-Camera Tracker* in the central control room, which generates global trajectories. This mechanism allows the tracking of persons over multiple camera viewports and represents the basis data set for subsequent trajectory analysis (see 3.3).

On demand, the video data acquired by the camera network can be sent to a streaming server, where it is accessible to other components of the system.

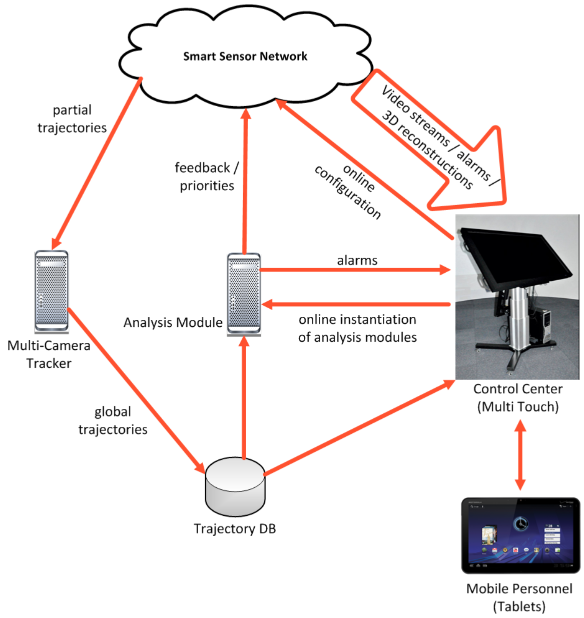


Fig. 1: Schematic overview of the *CamInSens* system setup

Furthermore, the smart camera system provides a 3D reconstruction component, which can be used to generate a detailed 3D model of a potential offender. To meet existing legal regulations [21], this component needs to be started by a human operator.

The complete data is automatically sent to and visualized in the connected security control room in real-time (see 3.4). This gives the security personnel the possibility to quickly assess the current situation and take counter measures if needed, e.g. give orders to the mobile personnel (see 3.5). Those are equipped with custom tablets, which gives them the possibility to access approved information and also to input data into the *CamInSens* system. Figure 1 shows a simplified overview of the system architecture.

### 3.2 Smart Sensor Network

The input source for the CamInSens system is a sensor network consisting of smart cameras and different additional sensors, which make up a *Smart Sensor Network*. One of the main challenges faced in the CamInSens system is the integration of a variety of highly heterogeneous hardware platforms into a self-organizing distributed system. Because of the huge differences in performance (i.e. CPU speed, RAM, battery lifetime, network speed, etc.) the authors chose to extend two smart sensor platforms to be part of one large distributed system.

The main goal is a fully decentralized sensor system with a high level of abstraction towards users and connected systems (e.g. computing components as the central control room and trajectory analysis systems). This high level of abstraction enables the system to serve surveillance requests coming either from a human operator or other connected

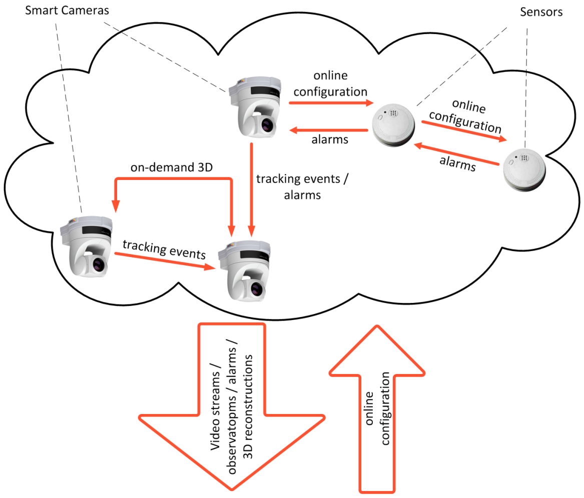


Fig. 2: Data Flow in the Smart Sensor Network

systems in a self-organizing way. Furthermore, the system contains some new highly-developed modules, i.e. a self-organizing field of view calibration, in-network event processing and dynamic on-demand 3D-offender reconstruction. Therefore, new distributed algorithms have been developed to build a reliable and responsive system.

### 3.2.1 Smart Cameras

Each *Smart Camera (SC)* consists of an Axis 214 PTZ IP-camera, local processing capabilities and two or more communication interfaces. At least one interface is used for SC-to-SC communication, while another one connects a SC to a subset of the *additional sensor motes* (see 3.2.2).

Smart cameras distinguish themselves from “classical” sensor nodes (e.g. temperature sensors) in their primary sensor being a photosensitive chip. This leads to two main differences: Smart cameras generate a high volume of data which is magnitudes higher than that of other sensor nodes. Furthermore, the directional characteristic makes it infeasible to cover a whole region of interest all at once [9]. To overcome the latter limitation, distributed algorithms have been developed, which make use of the pan, tilt and zoom capabilities of the camera to reorganize its field of view (FoV) in a distributed fashion to suit the current task [12].

To lessen the interconnection network’s burden of all the image data generated, many efforts have been taken to analyze the pictures inside the SC itself or at least near their point of creation. This means, that only high-level event information representing events in the observed environment is delivered to a sink. The image streams of cameras are only sent on demand, i.e. when the security personnel actually want to look at them.

The concepts for the *Smart Camera Software* itself, which had been introduced in earlier works [11], have been extended in different ways. The most important extension is a dynamic plug-in structure to enable the concurrent use of different control and analysis modules. The modules can make use of the local camera control, consume image data and communicate to module instances on remote SCs.

To make the usage of arbitrary modules in different configurations feasible, new challenges arise (see sections 3.2.4 and 3.2.5).

### 3.2.2 Additional Sensor Motes

Our additional sensor motes are build upon standard Mica2 motes used in wireless sensor networks [5]. The Mica2 motes can be connected to various types of sensors, e.g. temperature and humidity sensors, simple switches, photoelectric barriers, etc. Due to their main power supply being a battery and their small size, they have a much smaller performance profile than SCs considering their computational power and (wireless) connectivity.

### 3.2.3 Interconnection Networks

In classical WSNs based on MicaMotes, each packet intended for a sink (e.g. the central control room) has to travel through the network. That is why sensor nodes in a geospatial neighborhood of a sink may have to communicate exceedingly just to forward packets of other nodes, leading to so-called *routing hotspots*. The routing algorithms developed for the CamInSens system have been designed with the heterogeneity of the system nodes in mind and making use of the higher performance profile of SCs. The sensor nodes organize themselves in different ad-hoc networks depending on their configuration. While the Smart Cameras auto-configure their IP-based communication, the sensor motes self-organize themselves in different routing trees, each one rooted at a Smart Camera. To meet the requirements of a robust and self-healing system, the nodes react on changes in their environment and they reconfigure to overcome any disturbances.

### 3.2.4 Task-Oriented Abstraction of Surveillance Tasks

The Smart Sensor Network is able to fulfill a variety of different surveillance tasks without overwhelming a user with configuration issues. This is done by using self-organizing software-agents that occupy cameras and work on their computational resources. These agents can execute object tracking, pattern recognition and camera alignment. [14]

While some modules run most of the time (e.g. person tracking), others are triggered by human operators or other systems.

The following modules have been implemented:

**Distributed Tracker Component** The *distributed tracker component* is a single-camera multi-object tracker. It generates as long trajectories as possible out of the local cameras’ image stream. These trajectory parts are sent towards the *Multi-Tracker* component, which combines these parts of the same person to global tracks spanning the field of view of many cameras (see section 3.3.1 ff.).

**3D Online FOV Partitioning** The goal of the online partitioning agent is "an overlap-free monitoring of the observation area, considering the distinct priority of an area element" [15]. This enables the tracker component to collect as many trajectories as possible while also ensuring to meet the user's demands of seeing video streams from a certain area, which can be expressed by such priorities. Another system, which makes use of this module, is the analysis module: Depending on the current trajectory analysis, certain regions may be assigned with a priority.

**Dynamic 3D Reconstruction** The user can trigger a *dynamic 3D reconstruction* of a selected person. This leads to the instantiation of a reconstruction agent taking care of acquiring the needed data, i.e. selecting two cameras that are able to generate a multi-view of the person, taking two timely synchronized pictures (one from each camera) and generating the 3D model on one of the participating cameras.

### 3.2.5 On-Demand Reorganization of Data Storage and Forwarding

As said before, the CamInSens sensor network consists of different ad-hoc networks, which are build using wired as well as wireless network technologies. Despite the ongoing advances in computer architecture concerning computational power and communication interfaces, energy consumption is still a major concern while designing sensor networks. This is why, the system's design takes advantage of the heterogeneity present in the following ways:

Smart cameras and the additional sensor motes do not form disjoint ad-hoc networks, but are integrated into one large sensor network with the smart cameras acting as access points for the sensor motes. This is achieved by extending the smart cameras with another communication interface that acts as a base station towards the sensor motes. This way, the set of sensor motes becomes partitioned into different ad-hoc networks with each one connecting to another smart camera.

Since it is not a given, that each sensor node is in a spatial proximity to a smart camera to be able to connect to it directly, the sensor nodes still need to build up routing trees to make use of multi-hop communication towards the next smart camera.

## 3.3 Trajectory Analysis and Event Detection

The presented surveillance system is designed to perform automated recognition of safety critical behavior of observed persons (carried out by individuals or groups of persons) within the scene. To this end, a centralized trajectory analysis module assists the security personnel in detecting indications for such behavior, which can be roughly classified into criminal acts against other persons (e.g. assault or theft) or inanimate objects (vandalism) as well as self-endangerment (intentional or unintentional). There are two mechanisms for feedback based on analysis results: notification of the security personnel (raising an alarm), detailing on the origin of the detection as well as feedback towards the self-organizing camera network, introducing priorities within the distributed tracking task based on observed uncommon movement behavior.

The analysis module is physically linked to the output of tracking processes within the self-organizing camera network, using the extracted movement trajectory information within the observed scene as a basis for further analyses in real-time. The analysis module, however, does not perform image processing on image sequences in order to reduce data transfer volume and allow for better scalability of the system in terms of performance with increasing numbers of sensors/cameras.

The module does not provide a fixed set of rules for safety critical behavior, as the definition of safety relevant behavior may vary in different application scenarios. Instead, the module provides highly configurable analysis techniques. Positive responses to an analysis for a given set of observations do not automatically trigger counteractive measures but notify the security personnel providing details of the detected (potential) incident.

### 3.3.1 Characterization of Input Data

Results from the person observation/tracking task performed by the distributed smart camera network are camera-local current observations, i.e. short trajectory pieces observed by the individual cameras. Camera-local identities are maintained as long as an object is successfully tracked within the respective camera's local field of view. This (local) identity is communicated using unique LocalIDs. Assigned IDs are maintained as long as the observed person remains visible within the camera's local field of view (FoV). If a previously tracked person re-enters the FoV of a camera, a new ID is assigned, as, for legal reasons, no identifying features of persons are communicated/stored within the system.

Local observations within individual camera images are aggregated for a brief period of time (up to about 1 second) and transmitted towards the analysis module as short trajectory segments, consisting of a list of tuples (CameraID, LocalID, position, timestamp). Discrete timestamps of observation are synchronized over all cameras within the network. Positions (using the bottom center point of the object's bounding box within the image plane as point of reference) of observed/tracked entities within the image sequence are transformed into a global coordinate system using the current local camera calibrations.

### 3.3.2 Preprocessing Steps

In a first step, those local observations need to be integrated with each other in order to obtain a global view on all currently observed pedestrian trajectories, performed by a multi-camera tracker, including an assignment of unique GlobalIDs. Data integration is based only on the trajectory geometries, as image features are not communicated. The multi-camera tracker needs to deal with local positioning errors resulting from different angles of view and different qualities of occlusion which may lead to different trajectory locations (and thus differing trajectory geometries). Simultaneous observations (in overlapping areas of view within the camera network) need to be matched in terms of person identity, as no identifying features are used/communicated within the system. Global representations of trajectories are then stored/updated within a central trajectory database for retrieval by the security central within a limited period of time after the observation, after which data needs to be deleted for legal reasons.

#### 3.3.3 Conceptual View on the Analysis Module

Due to the self-organization capabilities of the camera network, the observed trajectories feature a number of additional properties: as one of the goals of the system is to demonstrate resource-efficient observation of a large scale scene with a relatively low number of cameras, the coverage of surveillance is incomplete. This can lead to an incomplete and/or fragmented knowledge about the people within the scene. Persons will usually be tracked for only a limited amount of time, before resources within the network are redistributed in order to maximize coverage. This requires the trajectory analysis module to also work on short episodes of movement behavior in order to quickly decide whether a currently observed movement requires further tracking. Positive analysis results based on short trajectories trigger a priority change (degree of priority changed based on the severity of the corresponding alarm) for continuation of tracking for the currently observed individual. This guarantees a continuous tracking until the alarm has been resolved by the security personnel.

The analysis software itself provides a number of analysis modules for different purposes that offer a set of configurable types of analyses for, e.g., single trajectory analysis, analysis of trajectories within the spatio-temporal context of the system installation or group pattern analysis. Typical single trajectory analyses deal with primary spatio-temporal properties of the observed movement and geometric characteristics of trajectories like speed, heading, curvature as well as detection of changes of the aforementioned parameters, i.e. acceleration and stops or sudden turns. In preparation for analyses in relation to spatio-temporal context, the analysis module continuously performs on-line learning techniques in order to create spatio-temporal models of typical movement behavior within the scene. This allows anomaly detection on trajectories within the learned context. For group pattern analysis, several algorithms for group movement patterns have been implemented, including well-known examples like leader-follower, flock/convoy, divergence/convergence or avoidance. All analysis techniques are implemented with regard to real-time capabilities.

#### 3.3.4 Interaction with Adjacent Modules

All provided analyses are accessible by the security control room at runtime by defining analysis tasks providing values for required parameters, a unique analysis task ID and a degree of severity for the alarm associated with a positive detection. Each type of analysis may be defined with multiple different parameters sets in order to create multiple instances of similar analyses. Any previously defined analysis can be deactivated by the security personnel at a later point in time. Within the analysis module, a cascade of currently active analysis tasks is performed for incoming (current) observations, triggering individual alarms (according to the pre-defined degrees of severity) whenever the defined criteria for a positive detection are met. As analyses are independent from each other, this process is highly parallelizable.

## 3.4 Control Center

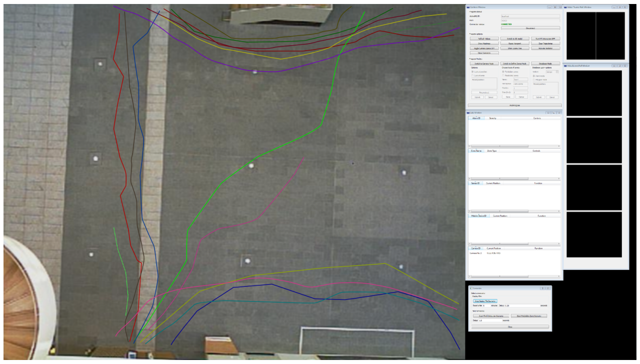


Fig. 3: Screenshot of the control center trajectory visualization module.

All data acquired by the automatic trajectory analysis and event detection system is transferred to the security control room in real-time for human evaluation and further processing. This puts special requirements not only on the used hardware but also on the visualization and interaction techniques.

At the heart of the CamInSens security control room is a 56" multi-touch enabled quad HD display (resolution: 3840x2160 pixels) from Barco [1]. The display has been specially designed for use in dedicated professional applications and delivers crisp, clear and color-accurate images on a large display size. The touch functionality is based on the *dreaMTouch* overlay manufactured by the German company Citron [3]. It differs from other multi-touch overlays for flat panel displays (e.g. [17]) in the choice and distribution of IR sensor elements and provides robust detection of 32 touch points at 50Hz.

Research on multi-touch started already in the early 1980ies (see [16]) and with the more recent advent of touch enabled smartphones and tablets it made its transition from a research prototypes into the mass market. We decided to use a fully multi-touch based interaction for the control center, since this allows a more parallel interaction, reduces task complexity and increases efficiency over standard WIMP (Windows, Icons, Menu, Pointer) interaction, which is based on serial discrete events.

The display area of the screen is split into a large map view of the surveillance area and into panels for different system control tasks. In the online mode, all global trajectories (a global trajectory can consist of multiple local trajectories detected by different cameras) are shown and updated in real-time (see figure 3). The system also has a trajectory database, which supports geometric queries, such as selecting all trajectories which passed a specific area in a certain time interval.

All alarms, automatically generated by the trajectory analysis module, or the sensor network, are received and displayed at their respective location on the main map. The module responsible for generating the alarm also automatically associates it with a severity level. This allows a priority sorting of the alarms in order to avoid an information overflow. In case an alarm cannot be resolved from the control center personnel directly, it can be assigned to available mobile personnel for further handling.

The CamInSens system allows the dynamic creation of different surveillance zones. Typical types are restricted zones, which people are not allowed to enter, or zones where people are not supposed to stay for longer periods of time. The security control room operators can create such zones by directly drawing polygonal shapes on the map and the trajectory analysis module updates itself.

In certain cases it is useful to generate a 3D model of a suspicious person or object. The operator can then generate such a task and the smart camera network will automatically reconfigure itself to process the request. Please note that 2 cameras with overlapping field-of-view are needed to successfully calculate the 3D reconstructions, and depending on the system configuration this may not be possible for all locations. In addition, the system assigns all task requests with a certain priority which is used to avoid contradicting tasks, or deadlocks.

## 3.5 Mobile Personnel

In current security systems, mobile personnel typically have only very limited data access. This is mainly due to the lack of sensors and connectivity functionality of exiting devices. E.g. most devices are limited to voice communication between the mobile security personnel and the security control room. In the interviews, which we conducted with the mobile personnel of large train stations and a state prison, it became clear, that this restriction is seen as one of the major shortcomings of the currently used devices. To overcome this limitation, and to allow a tighter coupling of the security control room and the mobile personnel we developed a novel *Mobile Control System* (MCS), using tablets running the *Android* operating system as hardware devices.

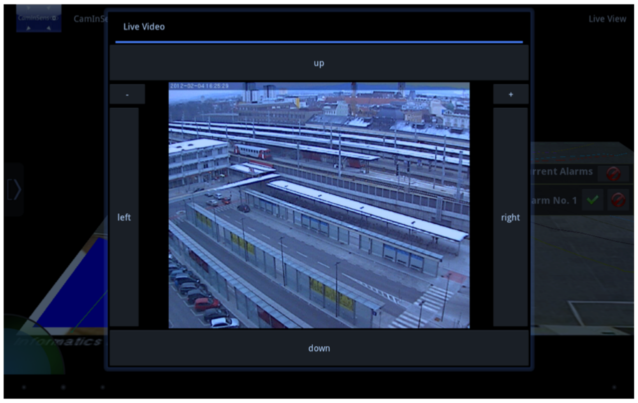


Fig. 4: Screenshot of the mobile user interface. In this case, a synthetic alarm was triggered and the system automatically shows a video stream containing the area. Note that the camera network is automatically reconfigured in case not camera monitoring this area.

The core of the MCS visualization is a geo-referenced map visualization module to display the monitored area and its surroundings. This provides the personnel with a reference for additional data, such as global trajectories, events, alarms, or locations of mobile personnel, which are received and displayed in real-time. In most cases, the system simultaneously detects a large number of trajectories, so it is crucial to provide mechanisms to cluster groups of similar trajectories and quickly find abnormal trajectories, which may be of interest. To achieve this on the limited computation resources available on the tablets, we implemented a simplified version of the algorithm proposed by Hoeferlin et al. [10]. All alarms, automatically generated by the trajectory analysis module, or explicitly by the personnel in the security control room, are always visualized in the foreground and also trigger haptic and audio feedback (see figure 4). An informal user study about the multi-modal alarm feedback showed, that the visual and audio feedback were rated as most valuable [20]. Another important functionality of the MCS is the possibility to access the video streams of the entire camera network.

In addition to the data visualization, it is crucial to provide an efficient interaction concept, taking the special requirements for mobile security personnel, e.g. the system is most often used while standing or walking, into account.

Since we are using multi-touch tablets as hardware basis of our mobile devices, an important part is the selection of natural gestures, which can be learned by the security personnel without the need of extensive training. Because of this, our gestures are based on the work of Wobbrock et al. [23] who analyzed a large number of gestures, provided by users who didn't have any prior experience with multi-touch devices in order to find and classify sets of natural and intuitive gestures.

In addition, all navigation and system control tasks can either either be triggered via a menu system or by drawing gestures on the device. The system furthermore provides functionality to create and map custom gestures to all system actions.

## 4. Conclusion and Future Work

This paper proposed a novel system design for a third-generation surveillance system. During the first chapters, we motivated the need for such systems and derived both mandatory and optional system requirements. Design decisions and interdependencies of these requirements are illustrated for both the overall system and the main system modules, i.e. the self-organizing smart camera network, the trajectory analysis module, and the visualization and interaction concept for both the security control room and the mobile personnel.

The proposed design has been implemented and tested in a first prototype, which demonstrates the feasibility of the chosen approach. A second demonstration of the complete system is planed for late 2012. The system's promising approach of integrating different types of sensors into self-organizing networks, that can be controlled from unified interfaces, makes way to support even more types of sensors, e.g. mobile camera nodes based on quadcopters. Future work will also include the evaluation of 3D visualization and its requirements on new gesture based interaction models for security control rooms.

## Acknowledgements

The authors want to thank the CamInSens project members and all participants of the on-site interviews. The CamInSens project is funded by the German Federal Ministry of Education and Research as part of the "Research for Civil Security Programme" of the German Federal Government under the project number 13N10809 - 13N10814.

## References

[1] Barco website. http://www.barco.com, Mar. 2012.

[2] Caminsens website. www.caminsens.org, Mar. 2012.

[3] Citron website. http://www.citron.de, Mar. 2012.

[4] J. Aggarwal and Q. Cai. Human motion analysis. *Computer vision and image understanding*, 73(3):428–440, 1999.

[5] I. F. Akyildiz and M. C. Vuran. *Wireless Sensor Networks*. Wiley, 2010.

[6] K. Berkler, K. Ammann, H. Bendel, and R. Müller. 2010 - 2011 Annual report of Fraunhofer Institut for Intelligent Analysis and Information Systems. Technical report, Fraunhofer IAIS, Sep 2011.

[7] M. Bramberger, A. Doblander, A. Maier, B. Rinner, and H. Schwabach. Distributed embedded smart cameras for surveillance applications. *Computer*, 39(2):68–75, 2006.

[8] A. C. M. Fong and S. C. Hui. Web-based intelligent surveillance system for detection of criminal activities. *Comput. Control Eng. J.*, 12(6):263–270, 2001.

[9] J. Hähner, U. Jänen, C. Grenz, and M. Hoffmann. Selbstorganisierende Smart-Kamera-systeme. *Informatik-Spektrum*, pages 1–9, Feb. 2012.

[10] M. Höferlin, B. Höferlin, D. Weiskopf, and G. Heidemann. Interactive schematic summaries for exploration of surveillance video. In *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, ICMR '11, pages 9:1–9:8, New York, NY, USA, 2011. ACM.

[11] M. Hoffmann, J. Hähner, and M.-S. Christian. Towards self-organising smart camera systems. In *Architecture of Computing Systems – ARCS 2008*, volume 4934 of *Lecture Notes in Computer Science*, pages 220–231. Springer Berlin / Heidelberg, 2008.

[12] M. Hoffmann, M. Wittke, J. Hähner, and M.-S. Christian. Spatial partitioning in self-organising camera systems. In *IEEE Journal of Selected Topics in Signal Processing*, volume 2. IEEE, aug 2008.

[13] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics*, 34(3):334–352, 2004.

[14] U. Jänen, J. Hähner, and C. Müller-Schloer. Competing agents for distributed object-tracking in smart camera networks. In *Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC), PhD forum*, 2009.

[15] U. Jänen, M. Huy, C. Grenz, and J. Hähner. Distributed three-dimensional camera alignment in highly-dynamical prioritized observation areas. In *Fifth ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*, 2011.

[16] S. Lee, W. Buxton, and K. C. Smith. A multi-touch three dimensional touch-sensitive tablet. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '85, pages 21–25, New York, NY, USA, 1985. ACM.

[17] G. D. Morrison. A cmos camera-based man-machine input device for large-format interactive displays. In *ACM SIGGRAPH 2007 courses*, SIGGRAPH '07, pages 65–74, New York, NY, USA, 2007. ACM.

[18] T. D. Räty. Survey on contemporary remote surveillance systems for public safety. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40(5):1–23, 2010.

[19] C. S. Regazzoni, V. Ramesh, and G. L. Foresti. Scanning the issue/technology special issue on video communications, processing, and understanding for third generation surveillance systems. In *Proc. IEEE*, volume 89, pages 1355–1367, 2001.

[20] M. Roedder. Visualisierung und interaktion von trajektorien und assoziierter multidimensionaler daten mittels multi-touch tablets. Master's thesis, Koblenz-Landau University, 2012.

[21] A. Roßnagel, M. Desoi, and G. Hornung. Gestufte kontrolle bei videoüberwachungsanlagen. *Datenschutz und Datensicherheit - DuD*, 35:694–701, 2011. 10.1007/s11623-011-0166-z.

[22] M. Valera and S. A. Velastin. Intelligent distributed surveillance systems: A review. In *IEE Proc.-Vis. Image Signal Process.*, volume 152, pages 192–204, 2005.

[23] J. O. Wobbrock, M. R. Morris, and A. D. Wilson. User-defined gestures for surface computing. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 1083–1092, New York, NY, USA, 2009. ACM.

# Vulnerability Assessment In Cloud Computing

**Srujan Kotikela[1,a], Krishna Kavi[2,a], and Mahadevan Gomathisankaran[3,a]**
[a]Department of Computer Science and Engineering, University of North Texas, Denton, Texas, USA

**Abstract**—*As vulnerabilities keep increasing exponentially every year, the need to efficiently classify, manage, and analyse them also increases. Many of the previous attempts at managing vulnerabilities have not been so successful because of the use of taxonomy approach. Few of the recent approaches have used ontologies for vulnerability management. Ontologies are real world concepts that are modelled using an ontology language. Ontologies are more appropriate for vulnerabilities as vulnerabilities can not be strictly classified into hierarchies (taxonomies) and tend to overlap. Ontologies support both these characteristics of vulnerabilities. Cloud computing is redefining the way computers are used. As more and more users, applications and businesses move to cloud it becomes very important to have proper vulnerability management in cloud. In this paper we present a vulnerability management framework for cloud computing.*

**Keywords:** security; vulnerability; ontology; cloud computing

## 1. Introduction

Security vulnerabilities are prevalent across all facets of software. The vulnerabilities are increasing every year at an exponential rate. Our experience with software engineering shows it is very difficult, even impossible to build software without vulnerabilities, because of the complexity of modern software systems. So the only way to deal with vulnerabilities is find them and patch them. Discovering and patching vulnerabilities is not an easy task. To deal with this complex vulnerability management we need standard and efficient methods and tools.

The first step to deal with vulnerabilities is classifying them. Vulnerability classification is a well-studied area in computer security. Many vulnerability classifications have been proposed and devised. Most of them have chosen the taxonomy [1] approach to classify vulnerabilities. However many of these classifications have proven to be inefficient, incomplete or erroneous. In taxonomy based classification the elements being classified are divided into groups and sub-groups (hierarchy). Hence the taxonomy approach requires assigning vulnerabilities to one and only one sub-group. But many times vulnerability would be present in more than one sub group. This could be due to incomplete and/or incorrect definition of the vulnerability or the subgroup. It has been observed that this situation arises due to the nature of vulnerabilities themselves [2] [3].

Vulnerabilities are concepts, not entities themselves. It is natural for them to overlap across different groups. Ontologies are better suited than taxonomies to model concepts. Ontology [4] is a knowledge representation technique which is used to model real-world concepts and their relationships [5]. It is one of the prominent techniques used to model and share a domain specific knowledge in the field of information science. Ontologies are widely used in artificial intelligence, semantic web, and library science where classification of concepts is very essential. These properties of ontologies make them perfect candidate for vulnerability classification. A rich collection of existing tools and frameworks will make creating ontology based vulnerability classification easy and efficient. The structured nature of ontologies makes it easy to reason, query and infer. These features of ontologies have led to adoption of ontologies in many security solutions such as [6] [7] [8] [9].

As Cloud Computing [10] continues to expand and evolve it is influencing the way we think about computing. Every aspect of computing is now connected to cloud computing. It is a big game changer across all verticals of computing. This demands a lot of attention and research for cloud computing. The Cloud Security Alliance had mentioned that, security is one of the biggest roadblocks in adopting cloud computing. As many businesses and users are adopting and using cloud, there will be lot of software running in the cloud. Vulnerability management is still relatively new. This makes the problem even more interesting [11] with respect to cloud computing.

In this paper we present a solution for vulnerability management in cloud environments. Our solution uses well defined ontologies. The proposed framework consists of Ontological Vulnerability Database, Semantic Natural Language Processor and Attack Code Database. We designed an ontology by extending the Ontology for Vulnerability Management (OVM). Then we designed a framework around the ontology and created an Ontological Vulnerability Database (OVDB) which has semantic collection of vulnerabilities. The OVDB is linked to an attack script database in which there is a many-to-many mapping between vulnerabilities of the OVDB and scripts of the attack script database. The attack script database is a collection of attack scripts which will invoke runnable attack code from the attack code database. The attack code database is compilation of attack

[1] SrujanKotikela@my.unt.edu
[2] krishna.kavi@unt.edu
[3] mgomathi@unt.edu

codes from popular attack codebase like Metasploit. This attack database can be used to launch attacks on applications to test for the associated vulnerability. A natural language processor will facilitate natural language and keyword search on the OVDB. The semantic nature of the ontologies will facilitate the reasoning and inferences on the OVDB. The framework facilitates vulnerability scanning and vulnerability assessment of an application. This work can be further expanded to assess the runtime environment by extending the ontology to include configurations of the environment.

The rest of the paper is organized as follow. Section 2 outlays some background concepts related to our work. Section 3 describes the Related Work. Section 5.3 explains the various Ontologies. Section 4 explains the architecture of our proposed framework. Section 5 presents the Implementation of our framework followed by the Future Work in Section 6 and Conclusion in Section 7.

## 2. Background

*Common Vulnerabilities and Exposures (CVE)*: CVE [12] is a publicly available listing of vulnerabilities and exposures in software. This project is initiated and maintained by MITRE organization. CVE doesn't attempt to classify the vulnerabilities. It just enumerates all the vulnerabilities. Every vulnerability in CVE has a unique identifier, description and list of software systems along with corresponding versions that are affected by this vulnerability. This public repository helps many other vulnerability research projects. The CVE project started by the MITRE organization now lies at the core of many security/vulnerability research projects. Our framework also depends directly on CVE repository at its lowest level. However there are many refined layers available on top of CVE, such as NVD. We will be using them than the raw CVE format.

### 2.1 Ontologies

Ontologies are at the heart of our research and many vulnerability assessment projects. in this section we will provide brief introduction about ontologies. Ontology is defined as "A formal explicit description of concepts in a domain of discourse, properties of each concept describing various features and attributes of the concept, and restrictions on properties. Ontology is a conceptualization of a domain of interest". It consists of concepts, relationships between these concepts and rules specifying the limitations of these relationships. The concepts from the real world are modelled as classes in ontology. The members of these classes can be individuals (real-world-objects) or other classes or a combination of both. The properties model various attributes of the individuals or the properties of the classes in general. Properties are also used to model relationships between two individuals or classes.

Ontologies are expressed in ontology languages. OWL [13] is the World Wide Web Consortium (W3C) standard for representing ontology. OWL stands for Web Ontology Language. There are 3 sublanguages for OWL: OWL-Lite, OWL-DL and OWL-Full. The three languages differ in their expressiveness.

*OWL-Lite* is the simplest of the three. It is used where simple hierarchy and simple constraints are sufficient. It is easy to build ontology in OWL-lite and it is best choice to migrate an existing taxonomy/hierarchy to an ontology using OWL-Lite.

*OWL-DL* is based on Description Logics (DL) and is more expressive than the OWL-Lite. The inclusion of DL in OWL can be exploited for automated reasoning due to the First Order Logic properties. It is also the most used OWL variant by many researchers.

*OWL-Full* is required for situations where high expressiveness is desired. It is to be chosen when high expressiveness is more essential than decidability or computational completeness of the language. OWL-Full cannot be used for automated reasoning.

## 3. Related Work

In this section we will present major components of our framework followed by an algorithm.

*Security Content Automation Protocol (SCAP)*: SCAP [14] is a suite of interoperable specifications for automating security management. SCAP is a standard developed by NIST along with community participation. By using SCAP protocol to build a security solution will ensure that a security solution will be interoperable with other related security solutions. Our proposed OVDB framework is SCAP compliant. SCAP is essential for bringing automation, standardization, and regularity to many security related initiatives. It is the defacto standard for achieving inter-operability between various security automation projects. Hence we also align our ontology with SCAP so that we can leverage existing fine works that are SCAP compliant and also ensure that our framework interoperable with other similar initiatives and projects.

*National Vulnerability Database (NVD)*: NVD [15] is a SCAP compliant vulnerability database maintained by NIST. It is essentially the SCAP compliant version of the CVE enumeration. The NVD database is released as NVD feeds in XML format. This NVD database is used as an input for the OVDB creation. NVD is a refined version of CVE. It has all the CVE data and in a SCAP compliant format. Hence, using NVD makes security solutions more robust and more interoperable. In the same light we also use the NVD database as a source for our Ontological Vulnerability Database.

*Ontology for Vulnerability Management (OVM)*: OVM [16] is ontology for managing vulnerabilities. Ontologies are more suitable to model vulnerabilities than taxonomies [2]. OVM uses ontology to store and refer to vulnerabilities mapped from the NVD vulnerabilities list. OVM can be

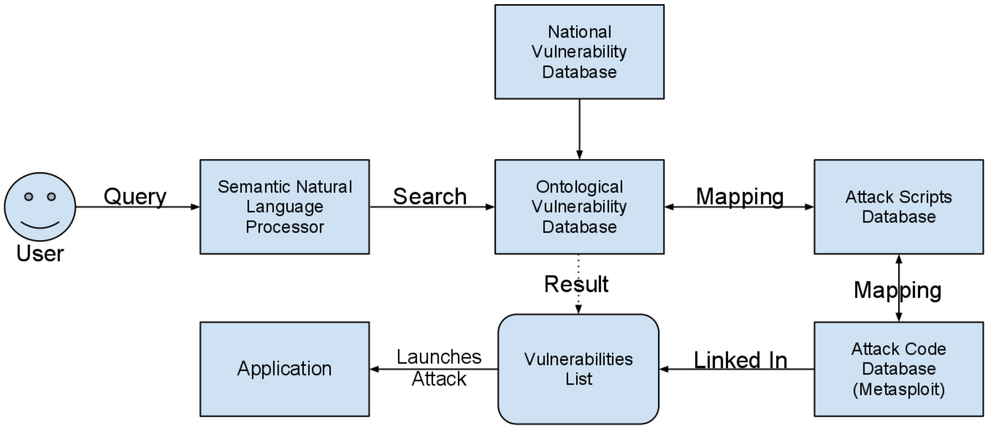

National Vulnerability Database
User
Query
Semantic Natural Language Processor
Search
Ontological Vulnerability Database
Mapping
Attack Scripts Database
Result
Mapping
Application
Launches Attack
Vulnerabilities List
Linked In
Attack Code Database (Metasploit)

Fig. 1: Vulnerability Assessment Framework

used to store and retrieve vulnerabilities. It can be queried using SWRL (Semantic Web Rule Language) through which we can perform semantic comparisons between two related products. OVM is the precursor for OVDB. Though both the databases share many similarities, OVDB is modified to consist only concrete and dis-ambiguous components and is intended to apply for cloud computing use-cases also.

*OVM Software Assessment Tool (OSAT)*: OSAT [17] is an ontology based software assessment tool. It is built on top of OVM. It uses all the vulnerability information present in OVM and tries to measure the security of software applications. It uses the CVSS scores of each vulnerability present in NVD and computes total security measure for particular software using a formula which sums up the Common Vulnerability Scoring System (CVSS) scores of all the vulnerabilities present in that software. OSAT is really useful tool and one of the first of its kind. It brings quantification for vulnerability assessment. Our Vulnerability Assessment framework is also similar to OSAT and has some interesting improvements like more user-friendly search.

*Ontology Of Cybersecurity Operational Information*: Is an ontology [18] developed for identifying cybersecurity information in cloud computing. The basis of the ontology is derived by applying cybersecurity operations that are prevalent in regular non cloud computing environments and applying them to cloud computing. The set of operations identified are generalized out of the cybersecurity operations performed by various cybersecurity practitioners in USA, Japan and Korea. Ontology of Cybersecurity Operational Information is Cloud agnostic and aims at assessing the cloud environment for vulnerability assessment. In future OVDB is going to combine the OVDB (which targets application vulnerability) with the Ontology of Cybersecurity Operational Information to create a complete end-to-end cloud vulnerability assessment.

## 4. Vulnerability Analysis Framework

Our Vulnerability Analysis Framework consists of *Semantic Natural Language Processor* (SNLP), *Ontological Vulnerability Database* (OVDB), *Attack Scripts Database*, and *Attack Code Database*.

*Semantic Natural Language Processor (SNLP)*: The semantic capabilities of OWL ontology aids in performing semantic reasoning on the ontological vulnerability database (OVDB). The user enters generic or specific information about his application and the SNLP is responsible to search through the OVDB and pull out the vulnerabilities that match user's keywords. Certain keywords by the user can be used to reason semantically than just perform a keyword search/match. The SNLP is capable of performing both keyword search as well as semantic search.

*Ontological Vulnerability Database (OVDB)*: OVDB is ontology database of vulnerabilities listed in the National Vulnerability Database. The OVDB includes lot of additional information about vulnerabilities like consequences, countermeasures, attacks that reveal a particular vulnerability etc. The ontology for OVDB is a modified version of the ontology found in OVM. There is a one-to-one mapping between OVDB and Attack Scripts database.

*Attack Scripts Database*: Attack Scripts Database is a collection of scripts which can invoke attacks from the attack code database. The scripts are customized for each attack individually as the parameters required can vary greatly for each attack. The scripts are mapped and a link to the script is stored along with associated vulnerability in the OVDB.

*Attack Code Database*: Attack Code Database is a database of attack codes primarily taken from Metasploit. The scripts in the attacks scripts database invoke the code in this database. This code will receive the parameters from the attack script and launch attacks on the application.

The usage of the framework has been explained in Algorithm 1. The vulnerability assessment starts by user typing

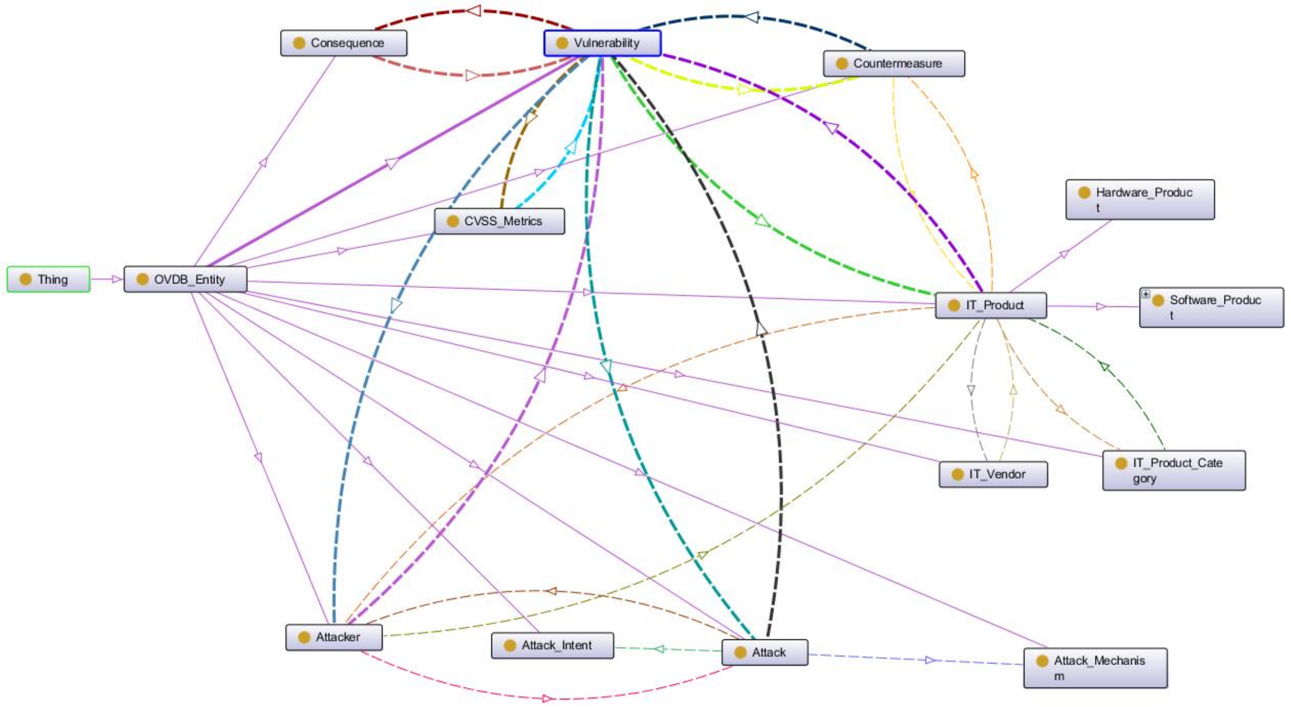


Fig. 2: OVDB Ontology

**Algorithm 1** Working algorithm of OVDB framework

```
1: User enters keywords for the search
2: SNLP processes the user query and displays list of
   related vulnerabilities
3: User selects the vulnerabilities he wants to test the
   application for
4: User launches associated attacks for the vulnerabilities
   selected
5: Attacks are performed on user's application
6: A summary of attack results is posted for the user
```

in the keywords which describe the application that is to be tested. This user query is submitted to the SNLP module. SNLP dissects the query and fetches various vulnerabilities related to the given keywords. These vulnerabilities will have a unique identifier (CVE-ID), brief description, impact score and a check box and launch attack button. After the user selected all the vulnerabilities he want to test, he can click Launch Attack button. This will invoke the associated attack script(s) from the attack-script database. The attack scripts will invoke necessary attack code from the attack code database. After all the selected vulnerabilities are tested, the user is presented with an analysis of what vulnerabilities have been tested positive and what have been tested negative.

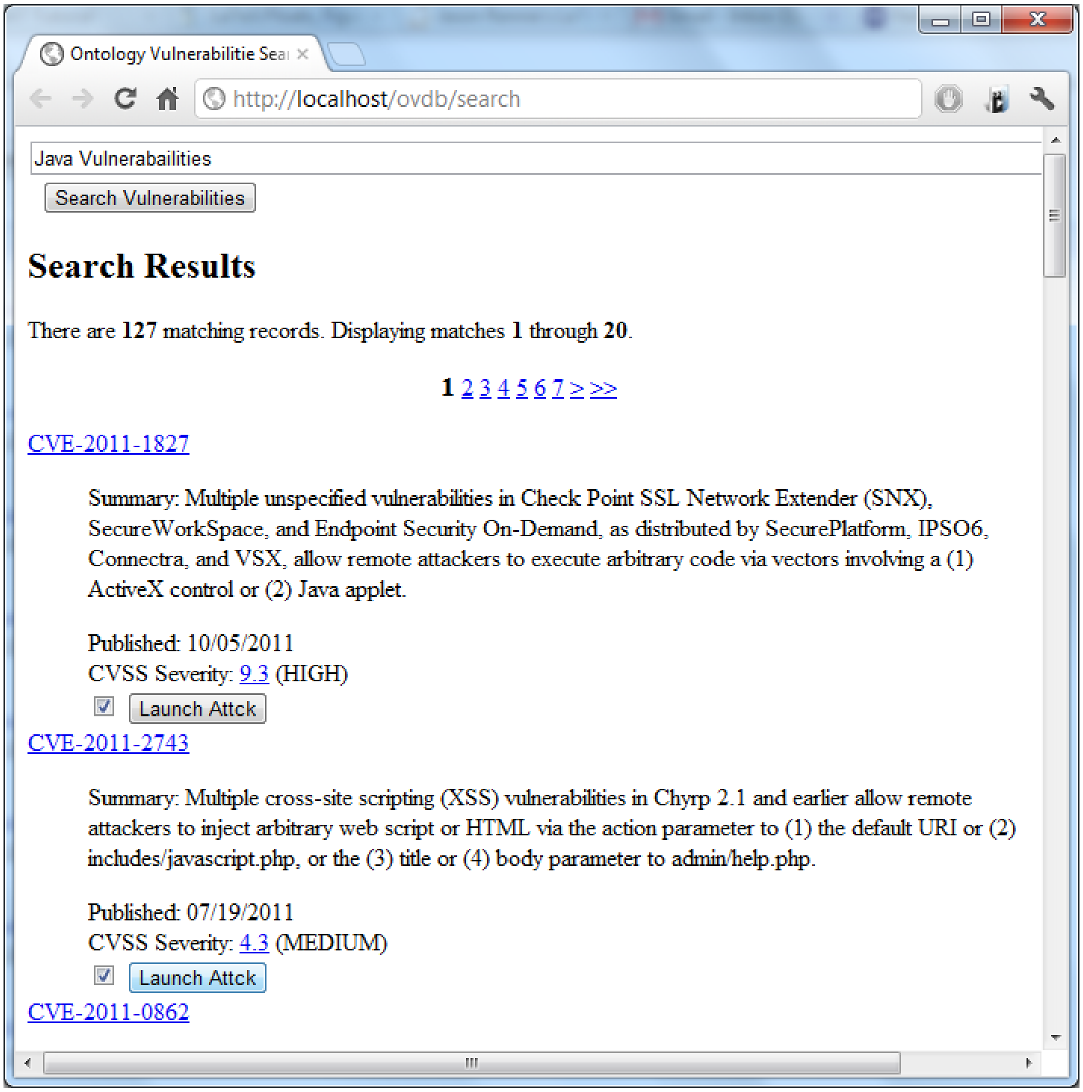
## 5. Framework Implementation

The framework we propose is built on top of the existing state-of-art vulnerability assessment solutions such as OVM and OSAT and extends them with subtle modifications. Hence, to understand our framework, one will need a good understanding of OVM and OSAT.

### 5.1 OVM and OSAT

OVM is a vulnerability database (populated using NVD) which has a query interface. OVM can be queried using standard query language SWRL [19]. These queries go through the OVM and pulls out vulnerability information. A user can write queries and infer results very efficiently with SWRL. For example, if a user is looking for vulnerabilities in browsers, he doesn't have to perform his search for each browser individually. Instead, he can issue search terms querying to look up for vulnerabilities associated with applications like Firefox. The reasoner will automatically infer which applications in the database fall in the category (browser) as Firefox and pulls out all the vulnerabilities in those applications. SWRL is a robust and expressive language which allows users to perform customized and efficient queries according to their needs.

OSAT is a Security Assessment tool built on top of OVM. OSAT takes advantage of all the ontological properties of OVM and reports comprehensive and qualitative measure-



Java Vulnerabailities

## Search Results

There are **127** matching records. Displaying matches **1** through **20**.

**1** 2 3 4 5 6 7 > >>

CVE-2011-1827

Summary: Multiple unspecified vulnerabilities in Check Point SSL Network Extender (SNX), SecureWorkSpace, and Endpoint Security On-Demand, as distributed by SecurePlatform, IPSO6, Connectra, and VSX, allow remote attackers to execute arbitrary code via vectors involving a (1) ActiveX control or (2) Java applet.

Published: 10/05/2011
CVSS Severity: 9.3 (HIGH)
☑ Launch Attck

CVE-2011-2743

Summary: Multiple cross-site scripting (XSS) vulnerabilities in Chyrp 2.1 and earlier allow remote attackers to inject arbitrary web script or HTML via the action parameter to (1) the default URI or (2) includes/javascript.php, or the (3) title or (4) body parameter to admin/help.php.

Published: 07/19/2011
CVSS Severity: 4.3 (MEDIUM)
☑ Launch Attck

CVE-2011-0862

Fig. 3: OVDB Framework Search Page (Mock)

ments of security. As the OVM, OSAT also follows the SCAP protocol. OSAT populates its reports from the OVM data. With OSAT we can enter a software product and ask the tool to list it's vulnerabilities. Alternatively we can provide input such as type of vulnerability, scope of the effect and nature of the vulnerability etc. Depending on user's input the OSAT infers the OVM database and builds a reports the vulnerability information. We can also ask the OSAT to find similar software along with security scores. This feature will allow us to compare which software product is more secure in a given product line.

### 5.2 OVDB Framework

Both OVM and OSAT are pioneering projects which have shown the power of using ontologies for vulnerability management. We build our OVDB vulnerability assessment framework as an extension to the ideas of OVM and OSAT. OVM and OSAT have reasoning and reporting which is limited to the applications in the database. They can not be used for user created applications. OVM and OSAT reports details from the database depending on the user query. These two tools report the vulnerabilities listed in the database. They can not analyze the application and tell us what vulnerabilities are present in the application right now. This makes these tools static in nature, where we can only look up existing information. The OVDB framework aims at solving this problem. The framework includes a attack code database which is mapped to the vulnerabilities in the OVDB. Whenever a user wants to analyze his application, he will use the framework to search for vulnerabilities. The search module will report possible vulnerabilities. User can select and launch attacks to test corresponding vulnerabilities. This is explained in more detail below.

### 5.3 Ontology

We have developed an ontology for implementing OVDB. It is a modified and extended version of OVM. Figure 2 shows the various entities and their relationships between them. We have developed this ontology in Protégé(ontology editor) [20] . All the concepts in the ontology are derived from Thing (a generic entity signifying every child entity is thing). At the top we have a wrapper entity for our ontology, called the OVDB_Entity which signifies that every child entity belongs to OVDB framework. Vulnerability is at the centre of the ontology. It has relations with other entities like IT_Product, Countermeasure, Consequence, CVSS_Metrics, Attack and Attacker. The relationship of the Vulnerability with these entities is described below:

*Consequence* signifies that every vulnerability has a consequence. Having this information associated with the vulnerability helps us to search vulnerabilities using their consequence. Many normal users may not technically classify a vulnerability, but they can identify the consequence and use it for searching the vulnerabilities.

*Countermeasure* entity contains the countermeasure for a vulnerability. It gives information on how to patch the associated vulnerability. This information will help users to patch their software and get rid of the vulnerability.

*CVSS_Metrics* is the set of CVSS metrics for a particular vulnerability. Which is a standard measurement for the severity of the vulnerability. It also tells which of the security properties of the information (confidentiality, integrity, availability) is being effected by the vulnerability.

*IT_Product* is the class of IT Products which have a particular vulnerability. This relation helps us to find vulnerabilities not only within the application but also the complete application stack. For e.g. if we are testing a Java Enterprise Application running in an application server, we can give the details of the application server to get the list of possible vulnerabilities in the application stack.

*Attacker* is the entity which is interested in exploiting the associated vulnerability. Having information about attacker will help to protect the application more efficiently.

*Attack* is the type of the attacks that can exploit a particular vulnerability. This allows user the flexibility to search if a particular attack is possible on his application. This relation will allow a quick evaluation of the application against dangerous attacks.

### 5.4 Working

Figure 3 shows a sample search results page. The user performs a search query by giving keywords describing the application such as technology, framework, language etc. (Java Vulnerabilities in the above example). The SNLP searches for the keywords in the OVDB and reports a list of vulnerabilities that are matching the user's keywords. User's keywords will be used for semantic search. After the search is done, the SNLP presents user with a list of vulnerabilities. These results are pulled out of OVDB. There is a check box after every vulnerability. The user can either choose some or all of the vulnerabilities and launch attacks corresponding to these vulnerabilities. User can click on launch attack button for every attack he wants to be performed upon the application. After the attacks are performed on the application a detailed report is generated on the security status of the application.

## 6. Future Work

In future we are planning to combine our ontology with the Ontology Of Cybersecurity Operational Information to provide a more robust and complete security in the cloud. The OVDB ontology primarily targets vulnerability of applications where as Ontology Of Cybersecurity Operational Information targets vulnerabilities of the cloud environment itself. Therefore, by combining these two ontologies we can achieve ontology for vulnerability assessment of the entire cloud infrastructure (application and environment). Since these two ontologies are cloud platform and application agnostic, we can perform vulnerability assessment for any application in any cloud.

## 7. Conclusions

In this paper we have proposed and successfully implemented an Ontological Framework for Vulnerability assessment in cloud. The framework is capable of assessing the vulnerabilities in popular software as well as software created by users. The framework can be installed in any cloud platform and used for assessing any technology applications. The framework allows security professionals and as well normal users to search through the database and assess the software. The framework is equipped with a nice user interface which makes the searching of vulnerabilities very easy. The framework makes the tedious task of vulnerability management and assessment easy and effective. With vulnerabilities growing exponentially everyday, this framework will have a great use in present and future. As the framework is built with the state-of-art security automation protocols, it is both automotive and interoperable with other applications.

## Acknowledgment

This work is supported in part by a grant from NSF (#1128344) and the Net-centric Industry/University Cooperative Research Center.

## References

[1] Matt Bishop, David Bailey. A Critical Analysis of Vulnerability Taxonomies). http://www.cs.ucdavis.edu/research/tech-reports/1996/CSE-96-11.pdf).

[2] Pascal Meunier. Classes of Vulnerabilities and Attacks. Technical Article, 2009.

[3] Simon Hansman and Ray Hunt. A taxonomy of network and computer attacks. *Computers & Security*, 24(1):31 – 43, 2005.

[4] www.wikipedia.org. Ontology (information science). http://en.wikipedia.org/wiki/Ontology_(information_science).

[5] B. Chandrasekaran, John R. Josephson, and V. Richard Benjamins. What are ontologies, and why do we need them? *IEEE Intelligent Systems*, 14(1):20–26, January 1999.

[6] Jeffrey Undercoffer, Anupam Joshi, and John Pinkston. Modeling computer attacks: An ontology for intrusion detection. In Giovanni Vigna, Christopher Kruegel, and Erland Jonsson, editors, *Recent Advances in Intrusion Detection*, volume 2820 of *Lecture Notes in Computer Science*, pages 113–135. Springer Berlin Heidelberg, 2003.

[7] Yanxiang He, Wei Chen, Min Yang, and Wenling Peng. Ontology based cooperative intrusion detection system. In Hai Jin, Guang Gao, Zhiwei Xu, and Hao Chen, editors, *Network and Parallel Computing*, volume 3222 of *Lecture Notes in Computer Science*, pages 419–426. Springer Berlin / Heidelberg, 2004. 10.1007/978-3-540-30141-7_59.

[8] F. Abdoli and M. Kahani. Ontology-based distributed intrusion detection system. In *Computer Conference, 2009. CSICC 2009. 14th International CSI*, pages 65 –70, oct. 2009.

[9] Andrew Simmonds Peter, Peter S, and Louis Van Ekert. An ontology for network security attacks. In *In Proceedings of the 2nd Asian Applied Computing Conference (AACC04), LNCS 3285*, pages 317–323. Springer-Verlag, 2004.

[10] Cloud computing: An overview. *Queue*, 7:2:3–2:4, June 2009.

[11] Timothy Grance Wayne Jansen. Guidelines on security and privacy in public cloud computing, Jan 2011.

[12] FIRST. Common Vulnerability Scoring System. http://www.first.org/cvss.

[13] W3C. OWL Web Ontology Language . http://www.w3.org/TR/owl-ref/.

[14] NIST. The Security Content Automation Protocol. http://scap.nist.gov.

[15] NIST. National Vulnerability Database. http://nvd.nist.gov.

[16] Ju An Wang and Minzhe Guo. Ovm: an ontology for vulnerability management. In *Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research: Cyber Security and Information Intelligence Challenges and Strategies*, CSIIRW ’09, pages 34:1–34:4, New York, NY, USA, 2009. ACM.

[17] Ju An Wang, Minzhe Guo, Hao Wang, Min Xia, and Linfeng Zhou. Ontology-based security assessment for software products. In *Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research: Cyber Security and Information Intelligence Challenges and Strategies*, CSIIRW ’09, pages 15:1–15:4, New York, NY, USA, 2009. ACM.

[18] Takeshi Takahashi, Youki Kadobayashi, and Hiroyuki Fujiwara. Ontological approach toward cybersecurity in cloud computing. In *Proceedings of the 3rd international conference on Security of information and networks*, SIN ’10, pages 100–109, New York, NY, USA, 2010. ACM.

[19] W3C. SWRL, 2004. http://www.w3.org/Submission/SWRL/.

[20] ProtÃľgÃľ Team. What is ProtÃľgÃľ-owl? http://protege.stanford.edu/overview/protege-owl.html.

# A Customer Perspective Investigation on Internet Banking Security of Licensed Banks in Hong Kong

**S. Limwiriyakul**[1] **and P. Subsorn**[2]
[1] SECAU, Edith Cowan University, Joondalup, Western Australia
[2] Bangkok, Thailand

**Abstract -** *Currently, internet banking security is a critical issue when implementing an internet banking service. Confidentiality, integrity, and privacy are serious concerns for both the banking industry and the internet banking customers due to the potential harm inherent when these principles are breached. The subject of this paper was the comparative investigation of internet banking security features as offered by 19 selected licensed banks in Hong Kong. The aim of the comparative analysis of the selected banks was to construct realistic and comprehensive guideline that incorporated a usable weight rating scale of security features for the banking industry in Hong Kong. The findings from this paper exposed the lack of related internet banking security information on the websites in all 19 of the selected banks in Hong Kong. The lack of provision of information of this kind can negatively impact the confidentiality of the banks as well as influence the decision of existing personal and potential internet banking customers to adopt this type of banking service.*

**Keywords:** Adoption; customer perspective; Hong Kong; internet banking security; licensed banks; personal internet banking

## 1 Introduction

Hong Kong is one of the core financial centers in the world for both international and local banking institutions. It is banking institutions offer a wide variety of products and services to both international and local customers [1]. The three-tier banking system as utilized in Hong Kong is comprised of licensed banks, restricted licence banks and deposit-taking companies which are all generally termed authorized institutions [2]. These banking institutions have been supervised by the Hong Kong Monetary Authority (HKMA).

Since the dawn of the cyber era, the Internet has become an influential factor in which the banking industry operates [3]. It provides a new banking service delivery medium and communication channel between the banks and their customers. Banking operations through this medium allows the banking industry to reduce administration and operation costs in terms of staff and physical branches, while at the same time, customers enjoy the benefits of convenience, 24/7 availability, reliability, and rapidity. These advantages can enhance internet banking customer satisfaction when using the internet banking system [4]. Although, internet banking systems have several advantages, the service comes with some inherent disadvantages in the form of information security threats and risks [5]-[8]. These information security threats and risks are the main concern for both the banking industry and internet banking customers as they relate to confidentiality, privacy, and integrity of online transactions and associated private information [5]-[7], [9]-[10].

The focus and emphasis of this paper was therefore to examine the security of the information and banking transacted on the internet through the selected licensed Hong Kong banks particularly on the personal internet banking service. The aim was to deploy a refined internet banking security checklist that described the information security weaknesses associated with the website information of the specified banks.

The rest of the paper is structured into four major sections: research methodology, results, comparison between the selected licensed banks in Hong Kong, and conclusions and recommendations.

## 2 Research methodology

The comparative analysis was compiled into two major sections: (1) the accessibility of internet banking security features of the selected licensed banks in Hong Kong and (2) the results and findings from the selected licensed banks in Hong Kong. This comparative analysis method scrutinized the dissimilarity in the internet banking security features between the selected licensed banks in Hong Kong.

### 2.1 Data sample and collection

Out of the total of 23 licensed banks incorporated in Hong Kong, 19 were chosen to accomplish the goal of this paper [11]. These selected licensed banks had similarities in providing banking websites, internet banking websites, and English version websites for their customers as claimed in order to be meet appropriate conditions for this analysis. Secondary data sources which were obtained freely from websites of these selected banks were used in this analysis. See Table 1 for full details on the list of the selected licensed banks in Hong Kong. All the data pertaining to the internet banking websites were collected and analyzed during the period between February and March 2012.

## 2.2 The refined internet banking security checklist

Table 1 provided full details on a refined internet banking security checklist (Version 1.04) deployed in this particular analysis. It provided detailed background and information of internet banking security features to the banks' existing personal and potential internet banking customers. Furthermore, it was grouped into six main security feature categories for these selected licensed banks in Hong Kong.

## 2.3 The scoring technique

Each of the six main categories in this refined checklist was integrated with a weight rating score aimed to provide an extra useful and inclusive guideline for the selected licensed banks in Hong Kong and the banks' existing personal and potential internet banking customers. A highest possible score of 10 value points had been delivered in each of the sub categories except Sections 5.1 and 5.5 which have a highest possible score of 15 value points. Both sections were considered to be very sensitive and significant in terms of the internet banking security approach. The value points allocated in the sub-points in each of the sub categories were derived from item's significance based on present knowledge.

# 3 Results

Table 1 displays and concludes the analysis and results findings for the reader to assess the refined internet banking security checklist. A coding technique was deployed to this refined checklist with the discussions as follows.

| | Represents | | Represents | | Represents |
|---|---|---|---|---|---|
| ✓ | Yes | * | Optional | D | 3DES-EDE-CBC 168-bit SSL encryption |
| NI | No information | A | AES 256-bit SSL encryption | E | Entrust Secure Server CA |
| R | RC4 128-bit SSL encryption | C | Condition | V | VeriSign Authentication |
| K | Akamai Subordinate CA | T | Thawte | | |

TABLE 1. A SUMMARY OF THE REFINED INTERNET BANKING SECURITY CHECKLIST

| Internet banking information security checklist (Version 1.04) | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Security feature categories | Licensed Hong Kong banks | | | | | | | | | | | | | | | | | | | Weights |
| | | BOCHK | HKBEA | CCB ASIA | CHIYU | CHONG HING | CITIBANK | CITIC | DAH SING | DBS | FUBON | HANG SENG | HSBC | ICBC ASIA | NCB | PUBLIC HK | SHANGHAI | STANDARD CHARTERED | WING HANG | WING LUNG | |
| **1. General online security and privacy information available to internet banking customers (40)** | | | | | | | | | | | | | | | | | | | | | |
| **1.1** | **Account aggregation or privacy and confidentiality** | | | | | | | | | | | | | | | | | | | | **10** |
| | Compliance with the National Privacy Principles and the Privacy Act | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 10 |
| **1.2** | **Losses compensation guarantee** | | | | | | | | | | | | | | | | | | | | **10** |
| 1.2.1 | 100% | C | C | C | C | C | C | C | C | C | C | C | C | C | C | C | C | C | C | C | 10 |
| 1.2.2 | No information | | | | | | | | | | | | | | | | | | | | 0 |
| **1.3** | **Online/internet banking security information** | | | | | | | | | | | | | | | | | | | | **10** |
| 1.3.1 | Threats: Hoax email, scam, phishing, spyware | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | ✓ | 1.5 |
| 1.3.2 | Trojan and virus/malicious programs | ✓ | | ✓ | ✓ | | | ✓ | | ✓ | | ✓ | ✓ | | ✓ | | | ✓ | ✓ | ✓ | 1.5 |
| 1.3.3 | Keyloggers | | | | | | | | | | | ✓ | ✓ | | | | | | | | 1 |
| 1.3.4 | General online security guidelines | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 2 |
| 1.3.5 | Security alert/up-to-date issue | | | | | | | | | | | | | | | | | ✓ | | | 1 |
| 1.3.6 | Provision of password security tips | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 2 |
| 1.3.7 | Others (e.g. wireless) | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | 1 |
| **1.4** | **Bank security mechanism system** | | | | | | | | | | | | | | | | | | | | **10** |
| 1.4.1 | Antivirus protection | | | | | | | | | | | | ✓ | | | | | | | | 2.5 |
| 1.4.2 | Firewall(s) | ✓ | | | ✓ | ✓ | | | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | 2.5 |
| 1.4.3 | IDS/alert/monitoring system | | | | | | | | ✓ | ✓ | | | ✓ | | | | | | | ✓ | 2.5 |
| 1.4.4 | Others (e.g. data encryption, password protected, physical security, regular audit) | | | | | | ✓ | | | ✓ | ✓ | | ✓ | ✓ | | | | | | ✓ | 2.5 |
| 1.4.5 | No information | | ✓ | ✓ | | | | ✓ | | | | ✓ | | | | | | ✓ | ✓ | | 0 |
| **2. IT assistance, monitoring, and support (20)** | | | | | | | | | | | | | | | | | | | | | |
| **2.1** | **Hotline/helpdesk service availability for personal internet banking customers** | | | | | | | | | | | | | | | | | | | | **10** |
| 2.1.1 | 24/7 customer contact center by phone **OR** | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | | | | ✓ | ✓ | ✓ | ✓ | 5 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2.1.2 | No 24/7 customer contact center by phone | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | | | | | | ✓ | ✓ | ✓ | | | | | 3 |
| 2.1.3 | Secured email/message box | | | ✓ | | | ✓ | | | | | | ✓ | | | | | ✓ | ✓ | ✓ | 2 |
| 2.1.4 | Email | ✓ | | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | 1 |
| 2.1.5 | Demo/FAQ/online support form | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 2 |
| 2.1.6 | No information | | | | | | | | | | | | | | | | | | | | 0 |
| **2.2** | **Internet banking transaction monitoring by the banks** | | | | | | | | | | | | | | | | | | | | **10** |
| 2.2.1 | Provides dedicated team and technology for monitoring all transactions | ✓ | | ✓ | ✓ | | | | | | | | | | ✓ | | | | | | 10 |
| 2.2.2 | No information | | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | 0 |
| **3. Software and system requirements and settings information based on bank website information (30)** | | | | | | | | | | | | | | | | | | | | | |
| **3.1** | **Compatibility “best” with the popular internet browsers** | | | | | | | | | | | | | | | | | | | | **10** |
| 3.1.1 | Google Chrome | ✓ | ✓ | | ✓ | | ✓ | | | | | | ✓ | | ✓ | | | | | ✓ | 2 |
| 3.1.2 | Firefox | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | | | | ✓ | | ✓ | | | ✓ | | ✓ | 2 |
| 3.1.3 | Internet Explorer | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 2 |
| 3.1.4 | Netscape | | | | | | | | | | | | | | | | | | | | 2 |
| 3.1.5 | Opera | | | | | | | | | | | | | | | | | | | | 1 |
| 3.1.6 | Safari | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | | | | ✓ | | ✓ | | | ✓ | | ✓ | 1 |
| 3.1.7 | No information | | | | | | | | | | | | | | | | | | | | 0 |
| **3.2** | **Internet banking user device system and browser setting requirement** | | | | | | | | | | | | | | | | | | | | **10** |
| 3.2.1 | Hardware device | | ✓ | | | ✓ | | | ✓ | ✓ | ✓ | | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | 2 |
| 3.2.2 | Operating system | | ✓ | ✓ | | ✓ | | | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | 2 |
| 3.2.3 | Type of browser and setting (e.g. cookie, java, certificate) | | ✓ | | | ✓ | | | | | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | 2 |
| 3.2.4 | Screen resolution | | ✓ | | | ✓ | | ✓ | ✓ | | ✓ | | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | 2 |
| 3.2.5 | Browser automatic or manual test feature available | | | | | | | | | | | | | | | | | | | | 2 |
| 3.2.6 | No information | ✓ | | ✓ | ✓ | | ✓ | | | | | | | | ✓ | | | | | | 0 |
| **3.3** | **Free/paid security software/tool/information available to personal internet banking customers** | | | | | | | | | | | | | | | | | | | | **10** |
| 3.3.1 | Antivirus/anti-spyware | | | | | | | | | | | | | | | | | | | ✓ | 2 |
| 3.3.2 | Internet security suite | | | | | | | | | | | | | | | | | | | ✓ | 3 |
| 3.3.3 | Provides internet information/links to security software vendor(s) (e.g. firewall) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 2 |
| 3.3.4 | Other services (e.g. automated tool, online security scanning) | | | | | | | | | | | | | | | | | | | | 3 |
| 3.3.5 | No service and/or information | | | | | | | | | | | | | | | | | | | | 0 |
| **4. Bank site authentication technology (10)** | | | | | | | | | | | | | | | | | | | | | |
| **4.1** | **Employed encryption and digital certificate technologies** | | | | | | | | | | | | | | | | | | | | **10** |
| 4.1.1 | 128/168-bit SSL encryption **OR** | R | R | R | R | | R | R | R | D | | R | R | | R | | R | | R | R | 5 |
| 4.1.2 | 256-bit SSL encryption | | | | | A | | | | | A | | | A | | A | | A | | | 6 |
| 4.1.3 | Extended validation SSL certificates | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | ✓ | | ✓ | ✓ | 3 |
| 4.1.4 | Signing CA | V | V | V | V | E | V | V | V | V | T | V | V | V | V | V | V | K | V | V | 1 |
| **5. User site authentication technology (60)** | | | | | | | | | | | | | | | | | | | | | |
| **5.1** | **Logon requirement** | | | | | | | | | | | | | | | | | | | | **15** |
| 5.1.1 | Bank/credit card/customer ID/email **OR** | | ✓ | ✓ | | ✓ | | | ✓ | | | | | | ✓ | ✓ | | | | | 2 |
| 5.1.2 | Registered bank username (characters) **OR** | 8<br>-<br>1<br>6 | ✓ | | 8<br>-<br>1<br>6 | 8<br>-<br>1<br>6 | 6<br>-<br>5<br>0 | 6<br>-<br>1<br>5 | | 8<br>-<br>1<br>2 | 9<br>-<br>1<br>6 | ✓ | ✓ | 6<br>-<br>2<br>0 | ✓ | | ✓ | 8<br>-<br>1<br>6 | 8 | ✓ | 3 |
| 5.1.3 | Smart ID card with digital certificate embedded | | | | | | | | ✓<br>* | | | | | | | | | | | | 4 |
| 5.1.4 | Password/pin/security no. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 3 |
| 5.1.5 | Additional password or secret question **OR** | | | | | | | | | | | ✓ | ✓ | | | | | | | | 3 |
| 5.1.6 | Others (e.g. CAPTCHA) | | | | | | | | | | | | | | | | | | | | 2 |
| 5.1.7 | Two-factor authentication/USB key digital cert./dynamic password card (min. 6 digit pins) | | | | | | | | | ✓ | | | 6<br>,<br>8 | | | | | | | 6 | 5 |
| **5.2** | **Logon failure limitation** | | | | | | | | | | | | | | | | | | | | **10** |
| 5.2.1 | Standard max. (3 times) **OR** | | | | | | | | | ✓ | | | | | | ✓ | | | | | 10 |
| 5.2.2 | Max. more than 3 times **OR** | | | | | 5 | 5 | 4 | 4 | | | | | 4 | | | | | 5 | | 8 |

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5.2.3 | In use but does not specific max. number of failure allowed | | | | | | | | | | | | ✓ | | | | | | | | 5 |
| 5.2.4 | No information | ✓ | ✓ | ✓ | ✓ | | | | | | ✓ | ✓ | | | ✓ | | ✓ | ✓ | | ✓ | 0 |
| **5.3** | **Logon user input type** | | | | | | | | | | | | | | | | | | | | **10** |
| 5.3.1 | Keyboard only **OR** | ✓ | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 7 |
| 5.3.2 | Combination of keypad and keyboard **OR** | | | | | | | | | | | | | | | | | | | | 8 |
| 5.3.3 | Scrambled keypad with/without keyboard | | ✓ | | | ✓ | | | | | | | | | | | | | | | 10 |
| **5.4** | **Password restriction/requirement** | | | | | | | | | | | | | | | | | | | | **10** |
| 5.4.1 | Enforce good password practice | ✓ | | | ✓ | ✓ | ✓ | ✓ | | | ✓ | | ✓ | ✓ | ✓ | | | | ✓ | ✓ | 2 |
| 5.4.2 | Password/pin length (min. 8 characters length) | NI | NI | | NI | 8-12 | 6-50 | 6-8 | | 8 | 8-16 | 8 | | | NI | | 6 | 8-16 | 8 | | 1 |
| 5.4.3 | Numbers only **OR** | | | | | | | ✓ | | | | ✓ | | | | | | ✓ | | | 0 |
| 5.4.4 | Combination of numbers and letters | | ✓ | | | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | | | ✓ | ✓ | ✓ | 1 |
| 5.4.5 | Combination of upper and lower cases | | | | | | | ✓ | | | ✓ | | ✓ | ✓ | | | | ✓ | | ✓ | 1 |
| 5.4.6 | Special characters or symbols | | | | | | ✓ | | | | C | | | | | | | | | | 1 |
| 5.4.7 | Different passwords as compared to any of three previous used passwords | | | | | | 1 | | | | | 1 | | | | | | | | | 1 |
| 5.4.8 | No two or more consecutive identical characters (e.g. aa, 11) | 3 | | | 3 | 2 | 3 | ✓ | | | | | ✓ | | 3 | | | 3 | 6 | | 1 |
| 5.4.9 | No three or more consecutive characters (e.g. abc, 123) | ✓ | | | ✓ | ✓ | 3 | ✓ | | | | | ✓ | | ✓ | | | 4 | 6 | | 1 |
| 5.4.10 | Automatic password strength check on creation or change of password | | | | | | | | | | | | | | | | | | | | 1 |
| 5.4.11 | No information | | | ✓ | | | | | ✓ | | | | | | | ✓ | | | | | 0 |
| **5.5** | **Transaction verification for external or sensitive transaction (e.g. unregistered 3rd party account, BPAY)** | | | | | | | | | | | | | | | | | | | | **15** |
| 5.5.1 | Token device/dynamic password card **OR** | | | ✓ | | ✓ | | | | | | ✓ | ✓ | | | | | | | 6 | 15 |
| 5.5.2 | SMS (no. of digit pins) **OR** | ✓ | ✓ | | ✓ | ✓ | | ✓ | | ✓ | ✓ | | | 8 | ✓ | 6 | | 6 | | | 15 |
| 5.5.3 | Others (e.g. USB key digital certificate) **OR** | ✓ | ✓ | | ✓ | ✓ | | | | | | | | ✓ | ✓ | | ✓ | | | | 15 |
| 5.5.4 | Extra password/reserved verification info. | | | | | | | | | | | | | ✓ | | | | ✓ | | | 10 |
| 5.5.5 | No requirement | | | | | | ✓ | | ✓ | | | | | | | | | | ✓ | | 0 |
| **6. Internet banking application security features (40)** | | | | | | | | | | | | | | | | | | | | | |
| **6.1** | **Automatic timeout feature for inactivity** | | | | | | | | | | | | | | | | | | | | **10** |
| 6.1.1 | Max. (mins) **OR** | | | 10 | | | 15 | | 10 | 10 | 30 | 20 | | 15 | | 20 | | 15 | | 5 | 10 |
| 6.1.2 | In use but does not specify timeout length | ✓ | | | ✓ | ✓ | | ✓ | | | | | ✓ | | ✓ | | ✓ | | | | 8 |
| 6.1.3 | No information | | ✓ | | | | | | | | | | | | | | | | ✓ | | 0 |
| **6.2** | **Limited default daily transfer amount to sensitive transaction (e.g. unregistered 3rd party and international accounts)** | | | | | | | | | | | | | | | | | | | | **10** |
| 6.2.1 | Less or up to HKD $50,000 | | | ✓ | | ✓ | | | | | ✓ | | | | | ✓ | ✓ | | ✓ | ✓ | 6 |
| 6.2.2 | More than HKD $50,000 | | | | | | | ✓ | ✓ | | | | | ✓ | | | | | | | 4 |
| 6.2.3 | The default maximum daily transfer limit is variable dependant on the type of the internet banking customers | | ✓ | | | | | | | | | | | | | | | | | | 2 |
| 6.2.4 | The maximum daily transfer limit may be increased with the approval by the banks | | | | | ✓ | | ✓ | | | ✓ | | | ✓ | | | | ✓ | ✓ | ✓ | 2 |
| 6.2.5 | No information | ✓ | | | ✓ | | ✓ | | | ✓ | | ✓ | ✓ | | ✓ | | | | | | 0 |
| **6.3** | **Logging information and alert** | | | | | | | | | | | | | | | | | | | | **10** |
| 6.3.1 | Last login | ✓ | | | ✓ | ✓ | | ✓ | ✓ | ✓ | | | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 4 |
| 6.3.2 | Activity log/transaction history | ✓ | ✓ | | ✓ | ✓ | ✓ | | | ✓ | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 3 |
| 6.3.3 | Alert available via email and/or SMS | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | | | ✓ | | | | | 3 |
| 6.3.4 | No information | | | | | | | | | | | ✓ | | | | | | | | | 0 |
| **6.4** | **Session management** | | | | | | | | | | | | | | | | | | | | **10** |
| 6.4.1 | Use of cookie technology **OR** | | ✓ | ✓ | | | ✓ | ✓ | | | ✓ | | | ✓ | | | ✓ | ✓ | ✓ | ✓ | 7 |
| 6.4.2 | Use of page tokens **OR** | | | | | | | | | | | | | | | | | | | | 10 |
| 6.4.3 | Use of session tokens | | | | | | | | ✓ | | | | | | | | | | | | 10 |
| 6.4.4 | Use of cookies for other purposes (e.g. marketing, research, and/or statistics) | ✓ | | | ✓ | ✓ | ✓ | | ✓ | | | ✓ | ✓ | | ✓ | ✓ | | | | ✓ | 0 |
| 6.4.5 | Capture of other information (e.g. IP | | | | | | | | | | ✓ | | | | | ✓ | | | | | 0 |

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | address) | | | | | | | | | | | | | | | | | | | | |
| 6.4.6 | No information | | | | | | | | | ✓ | | | | | | | | | | | |

## 3.1 General online security and privacy information to the internet banking customers

Majority of the 19 banks provided useful general internet security information on their websites. On the other hand, only four out of the 19 banks did not provide any information about the bank's security mechanism system on their websites. Provision of such information can increase internet banking security awareness and confidentiality assurance to internet banking customers as shown in Table 1 in Sections 1.3 and 1.4.

All of the 19 banks had a limitation of liability for any claim, loss or damage in relation to the use of the internet banking service. Most of them only take responsibility if the fraud, omission, negligence, or willful default was proven to have been caused by the bank. Liability was limited to the amount of the relevant transaction, the amount of the direct loss, or actual damage (whichever was less). Otherwise, the bank will not accept any liability or responsibility for any instruction/transaction consequences acted by the internet banking customers and/or any third-party [12]-[13].

## 3.2 IT assistance, monitoring, and support

Less than half (9) of the 19 banks provided 24/7 telephone support. On the other hand, majority (18) of the 19 banks provided several types of IT assistance and support such as demo, FAQ, and online support through their internet websites. In relation to internet banking transaction monitoring by the banks, only four out of the 19 banks provided this on their websites. See more details in Table 1 in Section 2.2.

## 3.3 Software and system requirements and settings information based on the website information of the bank

All of the 19 banks declared that their internet banking website compatible was best with the Internet Explorer (V. 6+). Moreover, some of these 19 banks also stated compatibility with other well-known internet browsers such as Firefox (V.3+), Google Chrome, and Safari (V.3.2+). Providing compatibility to the several popular and well-known internet browsers can increase a bank's internet banking flexibility to existing personal and potential internet banking customers.

In terms of internet banking user device system and browser setting requirement, five of the 19 banks provided no information on their websites. Provision of such information can increase flexibility as well as security to internet banking customers and is in the best interests of the banks. See Table 1 in Section 3.2 for more details.

In addition, all of the 19 banks provided internet security information in relation to antivirus and personal firewall software to their internet banking customers. However, only one (Wing Lung Bank) of the 19 banks had a partnership arrangement with a third-party internet security software vendor whereby limited time discounts were offered on the purchase of antivirus, personal firewall, or encryption software [14]. See more details in Table 1 in Section 3.3.

## 3.4 Bank site authentication technology

There were 13 of the 19 banks which employed 128-bit SSL encryption and five which employed the highest available SSL encryption technology of 256-bit SSL encryption. Only one bank (DBS Bank) used 168-bit SSL encryption. Furthermore, 14 out of the 19 banks employed the extended validation SSL certificate. The use of both 256-bit and extended validation SSL certificate can provide better security, confidentiality, and identity of the banks' websites to both existing personal and potential internet banking customers.

## 3.5 User site authentication technology

There were 17 of the 19 banks which allowed personal internet banking customers to register their own login IDs instead of using the bank account ID. This method can increase the level of login security as well as provide convenience to users in terms of remembering the login ID. However, in terms of best practice guidelines, the length of a login ID should be at least 8 characters. In addition, Dah Sing Bank employed an authentication technology which allows a smart ID card with embedded digital certificate to use as an alternative login ID. Honk Kong ID is one of the classic examples of the smart ID card types [15].

In terms of logon input type, majority of the 19 banks used a keyboard for the input of username and password. However, the other two banks were HKBEA and Chong Hing Banks used a scrambled keypad password input type for its internet banking customers. Use of scrambled keypad technology can reduce potential risk of keylogger attacks against a personal internet banking customer's computer. Section 5.3 in Table 1 elaborates on this.

In relation to password restriction/requirement, nine out of the 19 banks displayed a password length requirement on their websites. Two out of the nine banks set the password length requirement to a minimum of six digits whereas one of these banks (Shanghai Bank) required a maximum of six digits password length only. In general the minimum password length should be any combination of eight alphanumeric characters. Obviously, a longer length will provide better security and is strongly recommended. See Section 5.4 in Table 1 for further details.

Additionally, there were three out of 19 banks which did not require any type of verification for external or sensitive transactions. These three banks were Citibank, Dah Sing Bank, and Wing Hang Bank. We recommend that a verification feature be provided to internet banking customers. This will provide a second layer of protection to the system authentication. Two factor authentication technologies such as Short Message Service (SMS), token device, and Universal Serial Bus (USB) digital certificate are good examples of current technologies which should be considered for internet banking transaction verification. Moreover, additional passwords can be another cost effective and basic solution for transaction verification process.

### 3.6 Internet banking application security features

In terms of the automatic timeout feature for inactivity, 17 out of the 19 banks declared that their internet banking system had this feature in place while the remaining two banks (HKBEA and Wing Hang Banks) provided no such information on their websites. Of the 17 banks with the timeout feature, 10 provided specific inactivity time information ranging in duration from five to 30 minutes. Our considered opinion is that five minutes may be too short while 30 minutes is too long for the inactivity period. Furthermore, Hang Seng Bank provided its customers the option to continue the internet banking session after a period of 20 minutes inactivity. The continuation without another login is effected by clicking on the button of a warning message window. See more details in Table 1 in Section 6.1.

Seven out of the 19 banks did not provide any information regarding the limited default daily transfer amount to sensitive transactions such as non-registered third-party accounts, BPAY, and international transactions. The provision of such information can increase information security confidentiality of existing personal and potential internet banking customers.

Only one bank (Hang Seng Bank) did not provide information on its website about logging information including any alert features. Activity log or transaction history was provided by 12 of the 19 banks. Furthermore, 10 out of the 19 banks had alert features such as SMS and/or email communications automatically triggered to be sent to their internet banking customers. The provision of both activity logs and alert features is both handy and convenient and also serves to enhance the security of the internet banking system. See Table 1 in Section 6.3 for additional details.

In addition, 18 out of the 19 banks used cookie technology for their internet banking accounts. Only one bank (DBS Bank) had no information about cookies being available on its website. Information about the requirement for the cookie feature to be enabled on a customer's web browser for the proper operation of the internet banking was provided by 10 out of the 18 banks. All of these 10 banks had disclosures about not keeping a customer's sensitive information during the internet banking session. Rather, all the cookie information (cache) was automatically deleted after the end of the session. The avoidance of not using session management cookie for internet banking process except for statistical purposes can reduce the potential risk of a cross-site scripting attack [16].

Note that there are no points were awarded to any features not meeting the minimum requirements. For example, the password/pin length (min. 8 characters) has a required minimum of eight characters for a one point entitlement on the rating scale. Consequently, any banks that had a minimum password of six characters were not awarded a point as per the rating scale metric.

## 4 Comparison between the selected licensed banks in Hong Kong

The following table (Table 2) displays the details of the overall comparative analysis for each of the six main categories based on the previous section. The checked marks from all sub categories and sub sections in each sub category were used to calculate both the total mark and the percentage. Only two banks (Citibank and Hang Seng Bank) out of the 19 banks performed poorly with less than 50 percent. More than a half (10) of the banks scored between 50 and 59 percent. The remaining seven banks scored a reasonable 60 to 70 percent.

In terms of the loss compensation guarantee, all of the selected 19 licensed banks in Hong Kong did not provide solid or obvious messages to their internet banking customers. The provision of such messages can enhance the confidentiality of both existing personal and potential internet banking customers. Only four banks had a good score in Category 2 whereas 14 out of the remaining 15 banks performed poorly with less than 50 percent in the same category. Moreover, these 14 licensed banks did not receive any scores in Section 2.2 as they did not provide any information on internet banking transaction monitoring by their banks.

TABLE 2. A SUMMARY COMPARISON INFORMATION BETWEEN THE 19 SELECTED LICENSED BANKS IN HONG KONG

| **Banks** | **Categories** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **1**<br>40 marks | **2**<br>20 marks | **3**<br>30 marks | **4**<br>10 marks | **5**<br>60 marks | **6**<br>40 marks | **Total**<br>200 | **%**<br>100 |
| BOCHK | 30.5 | 16 | 9 | 9 | 32.5 | 15 | 111.5 | 55.75 |
| HKBEA | 25.5 | 5 | 17 | 9 | 34 | 15 | 105.5 | 52.75 |
| CCB ASIA | 28 | 17 | 9 | 6 | 27 | 26 | 113 | 56.5 |
| CHIYU | 30.5 | 16 | 9 | 9 | 32 | 15 | 111.5 | 55.75 |
| CHONG HING | 29 | 6 | 12 | 7 | 46 | 26 | 126 | 63 |
| CITIBANK | 25.5 | 4 | 9 | 9 | 27 | 23 | 97.5 | 48.75 |
| CITIC | 26 | 6 | 9 | 9 | 42 | 28 | 120 | 60 |
| DAH SING | 29 | 8 | 10 | 9 | 22 | 31 | 109 | 54.5 |
| DBS | 35.5 | 7 | 8 | 9 | 45 | 20 | 124.5 | 62.25 |
| FUBON | 32.5 | 8 | 12 | 7 | 34 | 28 | 121.5 | 60.75 |
| HANG SENG | 29 | 8 | 8 | 9 | 33 | 10 | 97 | 48.5 |
| HSBC | 38 | 8 | 11 | 9 | 47 | 11 | 124 | 62 |
| ICBC ASIA | 29 | 6 | 12 | 7 | 40 | 27 | 121 | 60.5 |
| NCB | 30.5 | 16 | 9 | 9 | 34 | 15 | 113.5 | 56.75 |
| PUBLIC HK | 29 | 6 | 12 | 10 | 37 | 26 | 120 | 60 |
| SHANGHAI | 26.5 | 8 | 12 | 9 | 28 | 28 | 111.5 | 55.75 |
| STANDARD CHARTERED | 29 | 9 | 15 | 7 | 33 | 26 | 119 | 59.5 |
| WING HANG | 28 | 9 | 12 | 9 | 27 | 18 | 103 | 51.5 |
| WING LUNG | 34.5 | 10 | 22 | 9 | 37 | 28 | 140.5 | 70.25 |

In Category 3, 11 out of the 19 banks scored over 50 percent and only Wing Lung Bank had the highest score in this category. This was due to the fact that Wing Lung Bank had a deal with a third-party vendor software to provide special discount offers for antivirus and personal firewall software to its internet banking customers.

The majority of the selected licensed banks (14) scored highly (over 7 out of 10 points) in Category 4. The remaining five selected banks scored average as their web sites did not deploy extended validation SSL certificates for their bank site authentication technology.

In Category 5, most of the selected licensed banks (15) scored reasonably. The remaining four banks scored below 50 percent as there was little or no information on password restriction and/or logon failure limitation provided on their websites.

Seven out of the selected 19 licensed banks scored below 50 percent in Category 6. These seven banks should consider providing more Category 6 information on their websites such as automatic timeout feature for inactivity as well as limited default daily transfer amount to sensitive transactions. This can increase security awareness and confidentiality to both existing personal and potential internet banking customers.

# 5 Conclusions and recommendations

Providing full details of internet banking information to cover Categories 1, 2, 3, 5, and 6 is strongly recommended as this practice enhances security awareness in internet banking usage as well as online payments for both existing personal and potential internet banking customers. In terms of bank site authentication technologies (Category 4), changing to 256-bit SSL encryption and extended validation SSL certificate are strongly recommended to the licensed Hong Kong banks which have not done so. Furthermore, providing special deals for antivirus or related software to internet banking customers is also recommended as such an incentive measure can provide added convenience and cost savings to internet banking customers.

By universally deploying the checklist recommendations to their internet banking systems, all of the selected 19 licensed banks in Hong Kong can standardize information security and usability as well as enhancing confidentiality for their existing personal and potential internet banking customers. The refined internet banking security checklist (Version 1.04) can also be employed or personalized in other similar organizations apart from the banking industry. For instance, this refined checklist can be adapted and adopted to validate an online payment system in the local government sector.

# 6 Acknowledgment

We would like to express our sincere thanks and appreciation to Mr. Yunous Vagh for his generous support in bringing our paper to a publishing standard.

# 7 References

[1] Financial Services and the Treasury Bureau, "Hong Kong: The facts - Financial services," the Information Services Department, Hong Kong Special Administrative Region Government (GovHK), 2011.

[2] Hong Kong Monetary Authority (HKMA), "The three-tier banking system," 2012.

[3] T. C. E. Cheng, D. Y. C. Lam, and A. C. L. Yeung, "Adoption of internet banking: An empirical study in Hong Kong," Decision Support Systems vol. 42 pp. 1558–1572, 2006.

[4] C. S. Yiu, K. Grant, and D. Edgar, "Factors affecting the adoption of Internet Banking in Hong Kong—implications for the banking sector," International Journal of Information Management vol. 27, pp. 336–351, 2007.

[5] P. Subsorn and S. Limwiriyakul, "A comparative analysis of the security of Internet banking in Australia: A customer perspective," presented at the 2nd International Cyber Resilience Conference (ICR2011), Perth, Western Australia: Edith Cowan University, 2011a.

[6] P. Subsorn and S. Limwiriyakul, "A comparative analysis of internet banking security in Thailand: A customer perspective," presented at the 3rd International Social Science, Engineering and Energy Conference 2011 (I-SEEC2011), Nakhon Pathom, Thailand, 2011b.

[7] P. Subsorn and S. Limwiriyakul, "An analysis of internet banking security of foreign subsidiary banks in Australia: A customer perspective," International Journal of Computer Science Issues (IJCSI), vol. 9, 2012.

[8] Usonlinebiz, "Types of Internet banking and security threats ", vol. 2011: Usonlinebiz, 2008.

[9] D. Hutchinson and M. Warren, "A framework of security authentication for internet banking," presented at the International We-B Conference (2nd), Perth, Western Australia, 2001.

[10] D. Hutchinson and M. Warren, "Security for Internet banking: A framework," Logistics Information Management, vol. 16, pp. 64 -73, 2003.

[11] Hong Kong Monetary Authority (HKMA), "List of licensed banks," 2012.

[12] Bank of China (Hong kong) Limited, "Conditions for services," n.d.

[13] China Construction Bank (Asia) Corporation Limited, "Terms and conditions for online banking services: Liabilities of the Bank," n.d.

[14] Wing Lung Bank, "Anti-virus, encryption software and emergency charger special offer," 2012.

[15] The Hong Kong Special Administrative Region Government (HKSAR), "The smart identity card," 2010.

[16] The Open Web Application Security Project (OWASP), "Cross-site Scripting (XSS)," 2011.

# Implementation of a Web Application for Evaluation of Web Application Security Scanners

L. Ertaul, Y.Martirosyan,
Mathematics and Computer Science, CSU East Bay, Hayward, CA, USA

*Abstract*-**With more and more people becoming Internet users there have been great increase in using Web in all areas of life, including communication, education and shopping. And as a result of these changes the security concerns have also grown. The web application vulnerability scanners help reduce these security concerns in Web-based applications. In today's market a large number of web application-scanning tools are available, e.g. QualysGuard WAS, Acunetix, Hailstorm, Appscan, WebInspect, etc. Although these tools are available in the market but question becomes how efficient they are to address security concerns in WEB applications? To compare vulnerability detection rate of different scanners, it is important to have an independent test suite. This paper describes a web application, which is intended to be used to evaluate the efficiency of QualysGuard WAS and Acunetix web application vulnerability scanners. The application implements real life scenarios for imitation of OWASPs Top Ten Security Risks that are presented in the wild. For several vulnerabilities presented in this application, we also explain defense measures, which secure the application significantly. The results of web application evaluation identifies the most challenging vulnerabilities for scanner to detect, and compare the effectiveness of scanners as penetration testing tools for exploiting OWASP Top Ten vulnerabilities. The assessment results can suggest areas that require further research to improve scanner's detection rate.**

*Index Terms*—**Black box testing, web security scanners, web security, web security vulnerabilities.**

## I. Introduction

In today's world many of the most dangerous security risks are based on vulnerabilities in web applications. ISO 27005 defines vulnerability as "*a weakness of an asset or group of assets that can be exploited by one or more threats where an asset is anything that can has value to the organization, its business operations and their continuity, including information resources that support the organization's mission*" [1]. According to National Vulnerability Database (NVD) [2] the number of vulnerabilities has become lower since 2009, which means that security measures has been incremented over last year. This is shown in Figure1.

In spite of this fact, percentage likelihood that at least one vulnerability will appear in a website remains very high. During 2010 every day almost every websites were exposed to at least one of high, critical, or urgent severity vulnerability, 64% of which had at least one Information Leakage vulnerability [2]. These web application vulnerabilities may cause attacks to exploit weaknesses on any tier or layer of web-based applications.

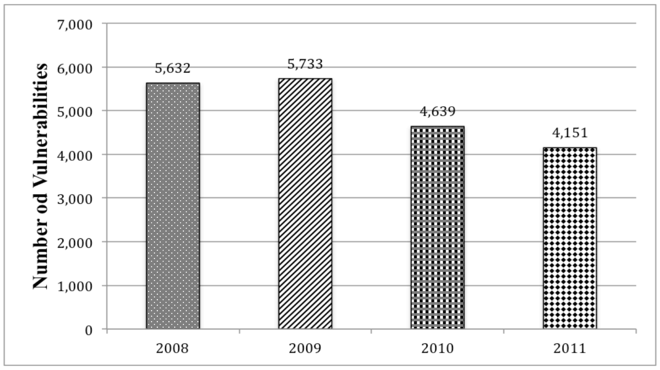


Fig. 1. Vulnerability distribution over years (2008-2011)

Most applications deployed on the Web implement a 3-tier architecture: presentation tier, business tier and data tier. Presentation tier is a web browser and dynamic web pages containing various types of markup language; business tier is a web application server; data tier is a database server. All tiers communicate with each other using strings to process input data. Web Application Server processes the inputs it receives from the clients and interacts with the database as shown in Figure 2.

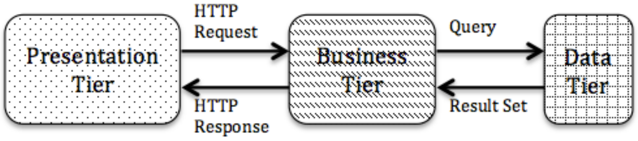


Fig. 2. The interaction between Client Tier (Web Client), Application Tier (Web Application) and Database Tier (Database Server)

Because web application server must validate and/or modify incoming strings before processing them or passing to database tier, in this paper we discuss input validation from client tier problem along with other most popular security flaws. Client Tier technologies include HyperText Markup Language (HTML) [53], Extensible Markup Language (XML) [57], JavaServer Pages (JSP) [58], JavaScript [55], and web applications continue to become more feature-rich and more dynamic, in particular with the advent of Asynchronous JavaScript and XML (AJAX) client tier technology. In the Web Application used to evaluate web application scanners we implement modern features such as JavaScript and AJAX to present more complex tasks for security scanners [3]. Another challenge that can result in limitations for security scanners presented in the Web Application is the difference of vulnerabilities within one class in terms of types of attacks vector. For example, exploiting Persistent XSS is more complex task than Non-Persistent XSS vulnerability [26]. Our goal is to assess the strengths and limitations of QualysGuard WAS [47] and Acunetix [56] tools and to report the test results. In the first part of our experiments we create a tested, the Web Application (MusicStore) that contains The Open Web Application Security Project (OWASP) Top Ten [4] most critical security risks. In the second phase we test QualysGuard WAS and Acunetix security scanners for vulnerability detection.

In Section II we present OWASP Top Ten web application security risks of 2010. Section III describes the technical characteristics, functionality and vulnerabilities of Web Application, which is implemented as a test suit for assessment of scanners. We explain defense mechanisms against web application attacks in Section IV. Section V contains web application assessment results. In Section VI we present conclusions.

## II. OWASP Web Application Security Risks

The OWASP security community has released its annual report in 2010 capturing the top risks in web application development as a combination of the probability of an event and its consequence. Following is the list of the top risks in web applications:

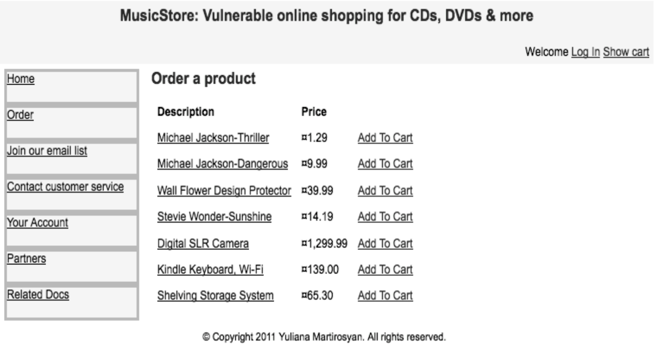
1. Injection
2. Cross-Site Scripting (XSS)
3. Broken Authentication and Session Management
4. Insecure Direct Object References
5. Cross-Site Request Forgery (CSRF)
6. Security Misconfiguration
7. Insecure Cryptographic Storage
8. Failure to Restrict URL Access
9. Insufficient Transport Layer Protection
10. Unvalidated Redirect and Forward

In web application described in this paper, we implement vulnerabilities 1 to 10, presenting them as real-life scenarios.

## III. Web Application (MusicStore)

There are several existing web applications to demonstrate common web application vulnerabilities such as "HacMe" series [41]. Those applications are well known by users and scanner developers. These applications may be used by scanner developers to optimize their performance. Other concern is the unavailability of the source code to estimate the rate of positive and false negative results of security scanner's findings. In addition to that these applications do not implement all the vulnerabilities from OWASP Top Ten report. Another well-known application is "WebGoat"[42] which is very complex web application. It is mainly used in educational purposes and not all of its test cases replicate real-life scenarios. Because of these drawbacks of available applications, there is a need to have an independent Web Application, which has real life scenarios and implements OWASP Top Ten vulnerabilities, to be used to test these web scanners. The Web Application (MusicStore) we present in this paper is designed to realistically simulate the steps a regular user goes through while using a dynamic web page and replicates the behavior of online store. The availability of source code and the control over server results in better evaluation of web application scanners.

Now let's have a look at functionality of the application. First a user creates an account, providing his/her personal data, including credit card number and shipping address. Second he/she selects the product and stores his selection in personal shopping cart. Later when the user decides to make the purchase an invoice is placed in queue for further processing. In addition to that the user can add reviews to products and read other customers' opinions, check partners' newsletters and subscribe to mailing list. Figure 3 illustrates the interface of the web application.



MusicStore: Vulnerable online shopping for CDs, DVDs & more

Welcome Log In Show cart

Order a product

| Description | Price | |
|---|---|---|
| Michael Jackson-Thriller | ¤1.29 | Add To Cart |
| Michael Jackson-Dangerous | ¤9.99 | Add To Cart |
| Wall Flower Design Protector | ¤39.99 | Add To Cart |
| Stevie Wonder-Sunshine | ¤14.19 | Add To Cart |
| Digital SLR Camera | ¤1,299.99 | Add To Cart |
| Kindle Keyboard, Wi-Fi | ¤139.00 | Add To Cart |
| Shelving Storage System | ¤65.30 | Add To Cart |

© Copyright 2011 Yuliana Martirosyan. All rights reserved.

Fig. 3. Web Application User Interface

The MusicStore Web Application is Java [50] based application, which is deployed on Tomcat Server [51]. It uses database on Oracle database management server [52] to store the data for the web site in its tables. Apache is a web server with Tomcat servlet/JSP engine. The application uses JSPs to present the user interface. It also uses HTML, CSS [54], JavaScript, and AJAX technologies. The presence of such technologies as AJAX and JavaScript in our web application gives additional opportunities. JavaScript is widely used in modern web applications and it is important to analyze the behavior of tools and their ability to parse JavaScript code.

The web application was developed based on OWASP Top Ten report of 2010. In this section we go over the characteristic vulnerabilities presented in the Web Application. The full list of the flaws designed in the project is available in Vulnerability Report [43]. As seen in the report we implement fifty-five variations of OWASP Top Ten Security Risks (see Table1 'Total' column).

*A. First Order SQL Injection:* ***recoverPassword*** function is intended to recover user's password based on her answer to security question.

```
String query = "SELECT Password FROM v_UserPass WHERE
(v_UserPass.EmailAddress = '" + emailAddress + "' AND
v_UserPass.Answer = '" + answer + "') ";
```

```
Payload:
emailAddress=test%40test.com%27%29--&answer=anycolor
```

*In* ***recoverPassword*** function concatenation is used to create dynamic SQL query. Attacker can easily impersonate site user and recover victims password by commenting out the part of the query using '--' single-line comment indicator [6].

*B. Blind SQL Injection*: ***updatePassword*** function is intended to update user's password based on her emailAddress.

```
    String query = "UPDATE    v_UserPass SET Password = ?, Answer =
'"+ answer+ "' WHERE EmailAddress = '"+ emailAddress + "'";
```

Manipulating 'answer' query parameter attacker can verify if email address he is interested in is stored in application database.

```
    True payload:
password=test11&answer=red%27+WHERE+EmailAddress%3D%28%27e
xistedEmail%40test.com%27%29--
```

If there is user with existedEmail@test.com email address in application database then query will be executed.

```
    False payload:
password=test11&answer=red%27+WHERE+EmailAddress%3D%28%27n
otExistedEmail%40test.com%27%29--
```

If there is not any user with notExistedEmail@test.com email address in application database then query will fail.

*C. SQL Injection Using Database constant*: ***insertReview*** function adds customer product reviews database in online store.

```
    String query = "INSERT INTO v_Reviews (Title, Message) VALUES
('"+title + "', '"+ message+ "' )";
    Payload:
title=%27%7C%7CSYSDATE%7C%7C%27&message=%27%7C%7CSYS
DATE%7C%7C%27
```

SYSDATE is Oracle function that returns date and time on a local database. This way attacker receives additional information about SQL Server.

*D. Non-Persistent XSS:* In this JSP Expression Language and Java example user registration information is stored in online store database after creditCardNumber parameter is validated on server side. No input inspection for firstName parameter is performed.

```
<form action="registrationServlet" method=post>
 First Name <input type="text" name="firstName"
 value="${newUser.firstName}">
 Card number <input type="text" name="creditCardNumber">
<input type="button" value="Continue">
</form>
Payload:
firstName=John"'><script>alert ("firstName parameter is
vulnerable")</script>&creditCardNumber=1234
```

If credit card number is incorrect firstName value be reflected on web page.

*E. Persistent XSS:* ***insertReview*** function adds customer product reviews database in online store.

```
    String query = "INSERT INTO v_Reviews (Message) VALUES ("'"+
message+ "'')";
    Payload:
message=message+%3Cscript%3Ealert%280%29%3C%2Fscript%3E&SU
BMIT=Submit
```

*F. DOM Based XSS:* web page uses firstName parameter in URL to greet the user. Web browser parses this HTML, which is received from server, into DOM. Parser executes the JavaScript code and as a result the XSS vulnerability is exploited.

```
<div id="greeting">
Hello
<SCRIPT>
var url = window.location.href;
var pos = url.indexOf("firstName=") + 10;
var firstName_string = url.substring(pos);
document.write(unescape(firstName_string));
</SCRIPT>
</div>
Payload in URL:
http://www.vulnerablewebapp.com/join_email_list.jsp?firstName=%3Cscri
pt%3Ealert%28%22DOM%20XSS%22%29%3C/script%3E
```

*G. Broken Authentication:* web application uses password recovery function, when you need to answer the security question. Using social engineering attacker can guess the country. Then using brute force dictionary method attacker can find the city and obtain victim's credentials [7], [8], [9], [10].

```
    Question: Where were you born?
    Payload is list of cities.
```

*H. Insecure Direct Object Reference:* web application receives reference to a file as form parameter 'letter', reads and displays the text. An attacker manipulates 'letter' parameter to access other objects.

```
Form parameter:    letter=SomePartner.html&SUBMIT=View+Letter
Java code:
File f = new File(path + "/" + request.getParameter("letter"));
String text = getFileText (new BufferedReader (new FileReader (f)), false);
Payload: ../../../../../../apps/java/apache-tomcat-6.0.16/conf/server.xmls
```

*I. CSRF:* the victim is authenticated at vulnerable online store. Attacker has placed malicious CSRF code on a web site. The browser will submit the request to vulnerable online store.

```
    Malicious CSRF code:
<img src=
"http://www.vulnerablewebapp.com/updateUserPassword?password=false
pass" width="1" height="1" border="0">
```

*J. Security Misconfiguration:* Web application server is vulnerable to slow HTTP headers DDoS attack. Using slowhttptest [11] tool attacker can get denial of service by slowing down requests.

*K. Failure to Restrict URL:* web application protects all data under "/user" directory. After user is authenticated web application makes possible to access /userAccess.jsp link. But it is not under /user directory and attacker can guess that hidden link and take advantage of it.

```
<% if (request.isUserInRole("user")) {%>
<a  href= "https://www.vulnerablewebapp.com/userAccess.jsp">User
Only</a>
```

*L. Insufficient Transport Layer Protection:* any data under "/user" directory should be protected using SSL.

```
    https://www.vulnerablewebapp.com/user
```

*M. Unvaidated Redirect and Forward:* Redirect and Forward functionality is very common in many web applications. But insecure implementation of it can result in tricking the user by an attacked into clicking the link that will navigate to unsafe destination. This is an example of Java code that demonstrates implementation of redirect function where site parameter value is the URL.

```
String site = request.getParameter("site");
if(site!=null && site!=""){
response.setStatus(response.SC_MOVED_TEMPORARILY);
response.setHeader("Location", site); }
    Payload: ="http://www.vulnerablewebapp.com
/partners/displayParnerLetter?site=http://www.attackerDestination.com
```

With all these threats widely available in web it is important to secure web application against them. In the next section we explain defense mechanisms and we show the implementation of several most important techniques in our web application.

## IV. Defense Mechanisms

Preventing vulnerabilities in applications extremely important due to high number of attacks (see Fig 1). In this section we describe several defense techniques against web application attacks used in our application.

- *SQL Injection and Cross-Site Scripting (XSS) Defense.* Server side defense using Prepared Statement [12] is the most effective way to protect from SQL Injection, because it ensures that intent of query is not changed. It is very important to lockdown database server and to follow the Principle of Least Privilege [13], [14]. Modern web applications also rely heavily on caching and database schema design to improve performance [40].

For prevention code injection attacks, including SQL Injection and XSS all user data should be validated. Input validation can be performed client side using JavaScript, but from security prospective it is not effective, because it doesn't provide protection for server-side code.

```
JavaScript Example:
    var emailexp = /^([A-Za-z0-9_\-\.])+\@([A-Za-z0-9_\-\.])+\.([A-Za-
z]{2,4})$/
    if (!isValid(emailexp,form.emailAddress.value)){
    return false}
```

Despite rule that input must be validated server-side sometimes validation should be performed client-side [15][16]. Web frameworks and filters that offer automate sanitization to prevent XSS in web applications are gaining popularity, because manual implementation of input sanitization in web application is prone to errors [17-25]. Unfortunately input filters can be circumvented with various attack vectors [26] [27].

- *Broken Authentication Defense, Session Management and Transport Layer Protection.* Authentication and session security is critically important because compromised credentials leads to impersonation and loss of confidentiality. To protect user's session ID strong efforts should be made to avoid XSS flaws as described in Injection Defense Section. Authentication key points are Password Strength and Password Use, including number of possible attempts and storage; and Password Recovery mechanism [28].

```
  Example:
  <security-constraint>
      <web-resource-collection>
        <web-resource-name>User</web-resource-name>
        <url-pattern>/user/*</url-pattern>
      </web-resource-collection>
      <auth-constraint>
        <role-name>user</role-name>
      </auth-constraint>
     <user-data-constraint>
        <transport-guarantee> CONFIDENTIAL
       </transport-guarantee>
      </user-data-constraint>
   </security-constraint>
```

Authentication relies on secure communication, so it is important to maintain Transport Layer Protection [29].

In this example data under /user/ directory will be transferred using secure connection. Also session cookie used to identify authenticated user should contain the "secure" or "HTTPOnly" attribute.

- *Insecure Direct Object Reference Defense.* This attack represents a serious threat to parameter-driven site if parameter is modified to point to a local file on the Web server. It is a good practice to use a reference map to prevent parameter manipulation.
- *Cross-Site Request Forgery Defense.* Main defense technique is using authorization token, generated web application on server side. The Anti-CSRF token should be a randomly generated value, specific to the user's current session [29-32].
- *Security Misconfiguration Defense.* Maintaining security settings of the application, frameworks, application server, web server, database server, and platform is very complex problem. Web servers are frequent target of attacks so trying to secure web servers the following aspects should be taken into account: Configuration, Web content and server-side applications, Operating System, Documentation [33].

*Example:*
*HTTP server is subject to Slow type HTTP Attack [34].*
*There is number of steps to protect against this attack [35]. The RequestReadTimeout directive value should be set to limit the time a client may take to send the request [36].*

- *Insecure Cryptographic Storage.* Sensitive data should not be displayed in clear form. The data should be stored encrypted with strong encryption algorithms, such as AES [44], RSA [45], and SHA-256 [46], [37] in database and decrypt it on server side upon request, or store hash of the data.
- *Failure to Restrict URL access.* Hidden pages are difficult to find, but sometimes it is possible to guess the URL, which is not intended for presence to unauthorized users. It is important to use an effective and trusted access control mechanism [38] and access control matrix that is carefully planned [39].
- *Unvalidated Redirect and Forward.* As for many previous discussed attacks parameter value validation should be performed before redirection. It can be done by ensure that the URL parameter is indeed a valid URL.

With all described flaws and defense mechanisms we need to find out whether the Web Application presented in this paper is useful to identify weak and strong points of a security scanner. In next section we examine the experimental results of running web application vulnerability scanners against our MusicStore web application.

## V. Evaluation of Web Application Vulnerability Scanners

Two web application vulnerability scanners, QualysGuard WAS (scanner Q) and Acunetix (scanner A), were tested using our MusicStore web application in order to find out whether those tools are actually successful at finding existent vulnerabilities. The results discuss the challenging vulnerabilities to detect, the possibility of false positive reports and the variation of vulnerabilities detection between different types. Both scanners support identification of web application vulnerabilities in the OWASP Top Ten approach, including dynamic and static search lists, links crawling, brute force and authentication.

Before the testing procedure Web Application is restored to original state. The setup consist of following steps:

1) Count and classify vulnerabilities in web application before the initial test.

2) The database server and web server are put in an initial state. This state includes seven products, two regular users and one administrator user in database and seven images for each product on web server.

3) Run web application scanner in initial mode.

4) Count the vulnerabilities found by web application scanner and compare to actual vulnerabilities report in step 1. The details of analysis presented next in this section.

5) Count False Positive/Maybe/Duplicate results.

The results of running Scanner A and Scanner Q against web application are shown in Table 1. The Table contains the following data:

- First column represents the vulnerabilities presented in the test suit. (Top 10 OWASP Vulnerabilities)
- Second column shows the different types of a vulnerability presented in first column.
- Third column contains the total number of vulnerabilities of each type existing in the web application MusicStore.
- Forth column contains the number of vulnerabilities detected by scanners.
- Fifth column is named False Positive (FP) results, which are reported by scanners but are not actually presented in the Web Application. The list included the findings of vulnerabilities marked as 'possible', which we will consider as 'maybe'; or vulnerabilities, which were reported, previously in the same type but with different description.
- The last column represents False Negative (FN) results, those are the vulnerabilities missed by the scanners.

Full report of running QualysGuard WAS and Acunetix against Web Application can be found in QualysGuard WAS Evaluation [48] and Acunetix Evaluation [59].

The Table 1 reports the vulnerabilities that were detected by web application scanners. As seen from the Table 1 both tools missed some weaknesses. Here we present the analysis of why the scanners missed certain vulnerabilities.

*1) SQL Injection.* Scanner A was able to discover all First Order SQL Injection vulnerabilities. But both scanners failed to find second order SQL Injection vulnerabilities, which are not executed immediately. The result of the injection is displayed on a page that should be navigated by user after the payload was submitted. Scanners fails to follow this logic thus interprets it as a negative response.

*2) Cross-Site Scripting.* Scanner Q discovered all Non-Persistent XSS vulnerabilities. Scanner A's results were very impressive too, but as a group most Persistent multi-step XSS and DOM XSS vulnerabilities were missed by both scanners.

TABLE 1
RESULTS OF WEB APPLICATION VULNARABILITY SCANNERS ASSESSMENT

| Vulnerabilities | Vuln. Type | Total | Detected | | FP | | FN | |
|---|---|---|---|---|---|---|---|---|
| | | | A | Q | A | Q | A | Q |
| SQL Injection | First Order | 2 | 0 | 2 | 1 | 0 | 2 | 0 |
| | Second Order | 4 | 0 | 0 | 0 | 0 | 4 | 4 |
| XSS | Non-Persistent XSS | 10 | 9 | 10 | 36 | 10 | 1 | 0 |
| | Persistent XSS | 4 | 1 | 3 | 1 | 3 | 3 | 1 |
| | DOM XSS | 4 | 3 | 1 | 0 | 0 | 1 | 3 |
| Broken Authentication | | 2 | 1 | 1 | 0 | 0 | 1 | 1 |
| Insecure Direct Obj. Ref. | | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| CSRF | | 11 | 0 | 4 | 0 | 8 | 11 | 7 |
| Security Misconfiguration | Password sent via GET method | 2 | 0 | 0 | 0 | 0 | 2 | 2 |
| | Web Server DDoS | 2 | 0 | 2 | 0 | 2 | 2 | 0 |
| | Sensitive Data display | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| Insecure Cryptographic Storage | | 7 | 2 | 4 | 0 | 0 | 5 | 3 |
| Failure to Restrict URL Access | | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| Insufficient Transport Layer Protection | Insecure session cookie | 2 | 2 | 2 | 0 | 0 | 0 | 0 |
| | Insecure Login (no SSL) | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Unvalidated Redirect and Forward | | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| | | 55 | | | | | | |

*3) Broken Authentication and Session Management.* In our web application we present two vulnerabilities of this type. The first one is vulnerability with weak password recovery model. The weakness is easily exploited by guessing. So scanners were not able to find the flow, which is not surprising. Both scanners easily discovered the second vulnerability because it had plain brute force attack possibility.

*4) Insecure Direct Object Reference.* Both security scanners were able to detect this type of vulnerability.

*5) CSRF.* Scanner Q found only 4 CSRF vulnerable links. Scanner A didn't show any results for this type of vulnerability. We relate this to the fact that during the information gathering phase the link crawling did not enumerate all the reachable pages. For those links presented in crawling report CSRF vulnerability was detected. For full information on links presented in our web application see Full Crawling Report [49].

*6) Security Misconfiguration.* The 2 vulnerabilities missed by the tool Q in this type are based on insecure data handling by web server, which is able to process requests sent by GET method. Scanners missed this vulnerability because the form with sensitive data was submitted by POST method although it was possible to send the request by adding the parameters in URL and process it as GET method. Scanner A didn't find any of the presented flows.

*7) Insecure Cryptographic Storage.* Both scanners discovered all session flaws. Although Scanner Q tested the possibility of sending credit card information securely, but it missed the same type of vulnerability: secure processing password and the answer to secret question. Those are application specific vulnerability.

*8) Failure to Restrict URL Access.* Both scanners did not detect the hidden link. The link is accessible by registered user only. Another way to reach the hidden link is force browsing which has failed for scanner specific testing.

*9) Insufficient Transport Layer Protection.* The scanners were able to detect all insecure cookie and session processing vulnerabilities.

10) *Unvalidated Redirect and Forward.* Scanner Q detected this vulnerability, while Scanner A didn't report any findings.

In Figure4 and Figure 5 we present the following details on our findings:

- False Negative Rate (FN)- The rate is calculated as the number of FN vulnerabilities of each type over total number of vulnerabilities of each type.
- False Positive/Duplicate/Maybe Rate (FP) – percentage of vulnerabilities reported by scanner, but not the actual weaknesses. The rate is calculated as the number of FP vulnerabilities of each type over total number of vulnerabilities of each type.

The interesting result for Scanner Q was found for CSRF vulnerability type as shown in Fig. 4. False Positive rate is higher than false Negative. This means that despite the fact that scanner is very attentive to this type of weaknesses and suspected many web pages to be vulnerable it wasn't able to reach all possible web pages to try there the attacks as a result of complex multi-step application design.

Fig. 5 shows that Scanner A has very high FP results for XSS vulnerability. Almost all FP reports were duplicates.

## VI. Conclusion

We described OWASP Top 10 Security Risks implemented in the independent web application, which was designed and used as a testbed for evaluation of effectiveness of QualysGuard WAS and Acunetix web application vulnerability scanners. Each vulnerability type, presented in the web application was implemented as separate real life scenarios, including the popular coding mistakes and possible defense mechanisms.

Web Application vulnerability scanners failed to crawl the entire web application, which resulted in missing vulnerabilities. The other challenge was the difficulty to exploit stored and multi-step vulnerabilities. This also resulted in high rate of False Negative results. False Positive report was mostly the result of duplicates and 'possible' vulnerabilities. The tools showed very good results on detecting straightforward vulnerabilities as Non-Persistent XSS, Transport Layer Protection and Insecure Direct Object Reference.

Our plans for future work include evaluation of another two well-known web application vulnerabilities scanners using MusicStore web application with a purpose to get more extensive independent scanner evaluation report.

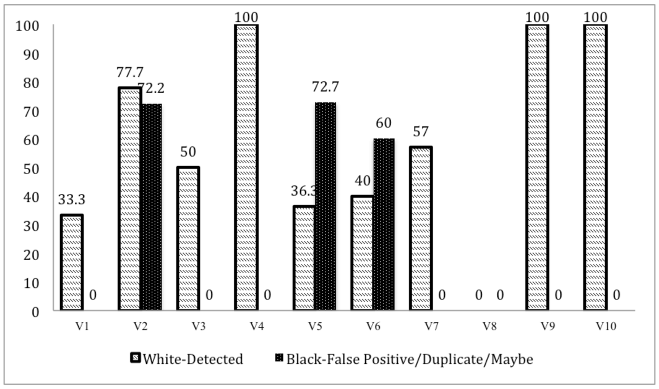


Fig. 4. QualysGuard. False Negative and False Positive/Duplicate/Maybe.
V1- SQL Injection, V2-Cross-Site Scripting, V3-Broken Authentication, V4-Insecure Direct Object Reference, V5-Cross-Site Request Forgery, V6-Security Misconfiguration, V7-Insecure Cryptographic Storage, V8- Failure to Restrict URL Access, V9- Insufficient Transport Layer Protection, V10-Unvalidated Redirect and Forward.

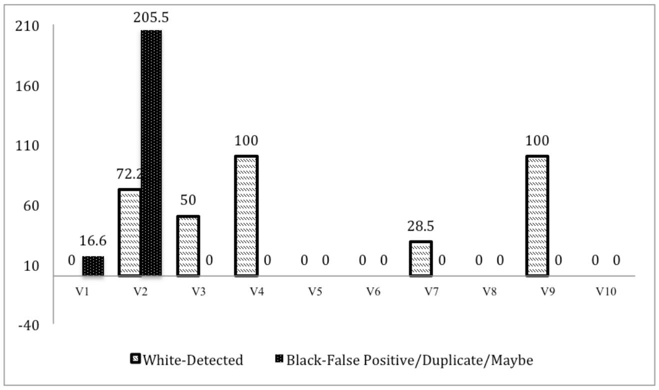


Fig. 5. Acunetix. False Negative and False Positive/Duplicate/Maybe.
V1- SQL Injection, V2-Cross-Site Scripting, V3-Broken Authentication, V4-Insecure Direct Object Reference, V5-Cross-Site Request Forgery, V6-Security Misconfiguration, V7-Insecure Cryptographic Storage, V8- Failure to Restrict URL Access, V9- Insufficient Transport Layer Protection, V10-Unvalidated Redirect and Forward.

## References

[1] International Organization for Standardization and International Electrotechnical Commission. ISO/IEC 27001:2005, Information technology – security techniques – information security management systems – requirements, 2005.

[2] National Vulnerability Database, http://nvd.nist.gov.

[3] Anthony T. Holdener III, "Ajax: The Definitive Guide Interactive Applications for the Web", O'Reilly Media, 2008.

[4] The OWASP Foundation, "OWASP Top Ten Web Application Security Risks", http://www.owasp.org/index.php/Category:OWASP_Top_Ten_Project, 2011.

[5] Oracle Learning Library. Defending Against SQL Injection Attacks, http://apex.oracle.com/pls/apex/f?p=44785:1:4073230388602787::NO

[6] Oracle PL/SQL Tutorial. http://www.java2s.com/Tutorial/Oracle/CatalogOracle.htm.

[7] THC Hydra, thc releases, thc-hydra v. 7.1, http://www.thc.org/thc-hydra/, 2011.

[8] John the Ripper password cracker, http://www.openwall.com, 2011.

[9] B. -S. Huang. "Brutus Project Groups Technical Report", Brutus Project. http://www.hoobie.net/brutus/.

[10] Massimiliano Montoro, Cain& Abel, http://www.oxid.it/cain.html.

[11] slowhttptest. Application Layer DoS attack simulator. http://code.google.com/p/slowhttptest, 2011.

[12] Java SE Technical Documentation. JDBC(TM) Database Access, Using Prepared Statements, http://docs.oracle.com/javase/tutorial/jdbc/basics/prepared.html, 2011.

[13] Jerome H. Saltzer and Michael D. Schroeder, The protection of information in computer systems. Proceedings of the IEEE, 63(9): 1278-1308, 1975.

[14] Fred B. Schneider, Least Privilege and More. IEEE Security & Privacy, pp. 55-59, September 2003.

[15] Matt Johansen and Kyle Osborn, "Hacking Google Chrome OS". Black Hat USA, Briefings and Trainings, August 2011

[16] Jeremiah Grossman, "Sometimes Input MUST be Validated Client-Side". WhiteHat Security. https://blog.whitehatsec.com/sometimes-input-must-be-validated-client-side-o_o/, September 1, 2011.

[17] E. Athanasopoulos, V.Pappas, A. Krithinakis, S.Ligouras, E. P. Markatos, "xJS: practical XSS prevention for web application development", Proceedings of the 2010 USENIX Conference on Web Application Development, 2010.

[18] P. Bisht, V. Venkatakrishnan, "XSS-GUARD: precise dynamic prevention of cross-site scripting attacks", Detection of Intrusions and Malware, and Vulnerability Assessment, pp. 23–43, 2008.

[19] W. Robertson, G. Vigna, "Static enforcement of web application integrity through strong typing", Proceedings of the 18th Conference on USENIX Security Symposium, SSYM 2009. USENIX Association, Berkeley (2009).

[20] P. Saxena, D. Akhawe, S. Hanna, F. Mao, S. McCamant, D. Song, "A symbolic execution framework for JavaScript", Proceedings of the 2010 IEEE Symposium on Security and Privacy, SP 2010. IEEE Computer Society, Washington, DC, USA 2010.

[21] P. Saxena, D. Molnar, B. Livshits, "Scriptgard: Preventing script injection attacks in legacy web applications with automatic sanitization", Tech. rep., Microsoft Research. September 2010.

[22] Zend Framework. Zend Filter. http://framework.zend.com/manual/en/zend.filter.set.html.

[23] Yii Framework. Special Topics. Security. http://www.yiiframework.com/doc/guide/1.1/en/, 2010.

[24] Template Toolkit.Manual. http://template-toolkit.org/docs/manual/Filters, January 2012.

[25] P. Saxena, S. Hanna, P. Poosankam, D. Song, "FLAX: Systematic discovery of client-side validation vulnerabilities in rich web applications", 17th Annual Network & Distributed System Security Symposium NDSS, 2010.

[26] K. K. Mookhey, Nilesh Burghate, Detection of SQL Injection and Cross-site Scripting Attacks, Symantec Connect Community, 02 November 2010.

[27] J. Weinberger, P. Saxena, D. Akhawe, M. Finifter, R. Shin, and D. Song, "A Systematic Analysis of XSS Sanitization in Web Application Frameworks", University of California, Berkeley, 2011.

[28] C. Miller, "Password Recovery". http://fishbowl.pastiche.org/archives/docs/PasswordRecovery.pdf, October 20 2002.

[29] K. Jaggi, "Securing Web Apps on Tomcat with SSL". Sun Developer Network, August 2006.

[30] Sun Microsystems. Mojarra Project. Mojarra JavaServerTM Faces JSF 2.0, 2011.

[31] Apache Struts. Class Token. http://struts.apache.org/2.0.14/struts2-core/apidocs/org/apache/struts2/components/Token.html.

[32] E. Sheridan. OWASP CSRFGuard Project.https://www.owasp.org/index.php/CSRF_Guard, 2010.

[33] N. Mendes, A. A. Neto, J. a. Dura˜es, M. Vieira, and H. Madeira, "Assessing and Comparing Security of Web Servers," Proceedings of the 2008 14th IEEE Pacific International Symposium on Dependable Computing. IEEE Computer Society, 2008.

[34] S. Shekyan. "Identifying Slow HTTP Attack Vulnerabilities on Web Applications". Qualys Community, July 7, 2011.

[35] S. Shekyan, "How to Protect Against Slow HTTP Attacks". Qualys Community, November 2, 2011.

[36] The Apache Software Foundation. Security Tips. Apache HTTP Server Version 2.5. http://httpd.apache.org/docs/2.3/misc/security_tips.html, 2012.

[37] Oracle Database Documentation Library. Developing Applications Using Data Encryption. Oracle® Database Security Guide 10g Release 1 (10.1). Part Number B10773-0.

[38] Vincent C. Hu David F. Ferraiolo D. Rick Kuhn, "Assessment of Access Control Systems". National Institute for Standards and Technology (NIST), September 2006.

[39] The Apache Software Foundation. The Apache Tomcat 5.5 Servlet/JSP Container Realm Configuration HOW-TO. http://tomcat.apache.org/tomcat-5.5-doc/realm-howto.html.

[40] M.Shema. "Seven Deadliest Web Application Attacks", Syngress, 2010.

[41] Foundstone Hacme Series. McAfee Corp.

[42] WebGoat Project. OWASP. http://www. owasp.org/index.php/Category:OWASP WebGoat Project .

[43] L.Ertaul, Y.Martirosyan, *"Vulnerability Report",* http://www.mcs.csueastbay.edu/~lertaul/WEBSEC/VulnerabilityReport.pdf, January 2012.

[44] NIST, "Advanced encryption standard (AES)," Nov. 2001, http://csrc.nist.gov/publications/fips/fips197/fips- 197.pdf.

[45] R. L. Rivest, A. Shamir, L. Adleman, "A Method for Obtaining Digital Signatures and Public-key Cryptosystems,"Commu- nications of the ACM, vol. 21, pp. 120-126, 1978

[46] NIST/NSA, "FIPS 180-2: Secure Hash Standard (SHS)", August 2002 (change notice: February 2004).

[47] QualysGuard Web Application Scanning (WAS), Qualys Inc., http://www.qualys.com/products/qg_suite/was/

[48] L.Ertaul, Y.Martirosyan, *"QualysGuard WAS Evaluation",* http://www.mcs.csueastbay.edu/~lertaul/WEBSEC/QualysGuardWASEvaluation.pdf, January 2012.

[49] L.Ertaul, Y.Martirosyan, "*Full Crawling Report*", http://www.mcs.csueastbay.edu/~lertaul/WEBSEC/FullCrawlingReport.pdf, January 2012.

[50] Java. Oracle Corporation, 1995.

[51] Tomcat Server. Apache Software Foundation.

[52] Oracle database management Server, Oracle Corporation.

[53] HyperText Markup Language (HTML), World Wide Web Consortium and Web Hypertext Application Technology Working Group (WHATWG).

[54] Cascading Style Sheets (CSS). World Wide Web Consortium.

[55] JavaScript. Brendan Eich. Netscape Communications Corporation. Mozilla Foundation.

[56] Acunetix Web Vulnerability Scanner. Acunetix. http://www.acunetix.com/vulnerability-scanner/

[57] Extensible Markup Language (XML), World Wide Web Consortium.

[58] JavaServer Pages Technologies (JSP), Sun Microsystems.

[59] L.Ertaul, Y.Martirosyan, *"Acunetix Evaluation",* http://www.mcs.csueastbay.edu/~lertaul/WEBSEC/AcunetixEvaluation.pdf, February 2012.

# Towards Design an Interoperability Framework for Security Policy Languages

Amir Aryanpour[1], Song Y. Yan[1], Scott Davies[2], Andrew Harmel-law[2]
[1]Department of Computer Science and Technology, University of Bedfordshire, UK
[2]Technology Services, Capgemini UK Plc

*Abstract* -- **Security policy languages ranged from a single role based access control (RBAC) to a highly sophisticated policy language which is capable of negotiating across a network, all aim to express clearly and concisely what the protection mechanisms are to achieve. However, these policy languages have often been designed independently and hence translation between these languages is difficult. The challenge of representing policies in different languages on a distributed network where there are multiple heterogeneous policy languages in use, affects the main benefit of policy-based security management, namely the enablement of resources and services be controlled and managed at a high level regardless of the adopted underlying policy language.**

**Through our research we have focused on this inability of transforming and translating policy languages. We intend to provide a framework which will facilitate the representation of security policy languages. However as a first step towards our goal, this paper details the steps which have been taken to theoretically prove that policy languages can be interoperable.**

## I. INTRODUCTION

The notation of protecting networked resources came to life at the very same moment computer networking was introduced. Many access control model and policy languages have been proposed in order to address the abovementioned concern. These languages, which have undergone a revolution during the last decade, usually come with different specifications, all of which aim to tackle different business requirements. When it comes to distributed networks, which includes virtual environments, different policies possibly written in heterogeneous policy languages and from different domains, could relate to many different resources. On the other hand, resources are often shared amongst different domains. In such a scenario, various policies must be integrated into a single policy in order to govern the access to the jointly owned resources. This has already been investigated and researched by others. Bearing in mind the fact that policy languages are not easily capable of being put into another language, this research however is intended to investigate the management of such an environment and to study the security policies at a more understandable and abstract level.

The lack of an interoperability framework for policy languages has received limited attention on previous occasions. Clemente et al presented a solution for this business requirement in [1]. In addition Basile et al proposed a system to address the problem of policy translation in [2]. Through our research we have re-examined these papers from different points of views. The following facts are applicable to both of them.

The proposed solutions on these papers focus on ontology based policy languages. The *extensibility at runtime* and *adaptability* of semantic policy languages convinced the authors of the abovementioned articles to choose semantically rich languages such as OWL over traditional languages. However, brief research shows that only a small set of available policy languages and frameworks which are widely used, are based on semantic languages. Hence, by imposing such a constraint on the framework, a large number of policy languages will not be able to use the solutions. In addition, despite the fact that semantic policy languages usually come with strong underlying formalism, both reports fail to demonstrate the possibility of cross-language policy translations using formal specifications. Undoubtedly we have been inspired by these papers in our research but have aimed to improve their research accordingly.

Figure (1) which serves crucial illustration purposes, displays a typical distributed network which utilises heterogeneous policy languages across different domains. In such an environment the management of domains at an abstract level is challenging if not impossible unless we introduce an abstract management policy framework to control the environment from a single point.

In essence, the policy framework as illustrated below is responsible for providing a platform to policy languages in order to make them interoperable which would undoubtedly be the main contribution of this research. However, before we architect the system and as a prerequisite to our framework's design, we would like to theoretically prove that such a translation is possible. The four steps which must be considered on route to this goal can be summarised as:

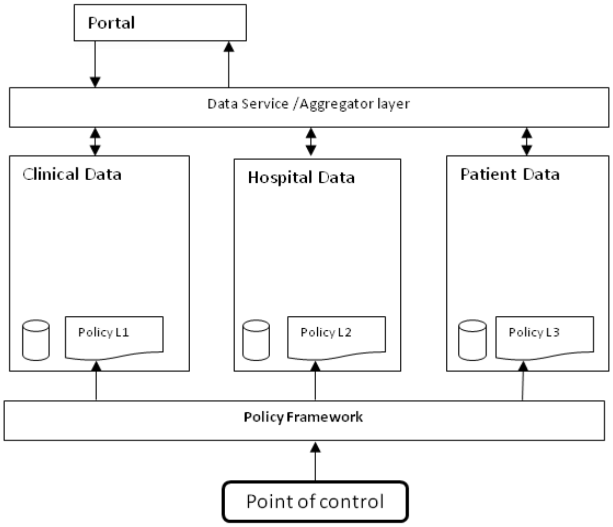


Figure1. A typical secure distributed environment

1. Policy Language Candidates
   Due to the fact that there are too many policy languages to choose from [3], we must classify policy languages into different categories. We must then select a policy language candidate from each individual group which represents the characteristics of that specific set. The result of this step is a set of policy language candidates which will then proceed on to the next step.

2. Algebra Candidate
   Logically, the most appropriate way to translate many-to-many languages is to translate them to and from an abstract language. We choose algebra as the abstract language throughout our PoC (Proof of Concept) hence the research must identify and utilise an appropriate algebra for policy languages.

3. Evaluation of Selected Algebra
   The algebra selected should be capable of expressing different scenarios written in possibly heterogeneous policy languages. Hence it should be evaluated against the policy language candidates chosen in step1.

4. Limitation and Solution
   During the final step, research must identify any existing weakness of the algebra, and formulate a solution to address that accordingly. Although the main contribution of the research is the design and development of the framework itself, but this step of the PoC can be presented as the contribution of this paper.

The rest of this paper is organised into three parts as follows: The first part details a comparison of policy languages and lists a set of policy language candidates. The second part describes the algebra we have used to theoretically prove that the translation of policy languages is achievable. Finally, the last part describes the limitation of the selected algebra and proposes a solution for this.

## II. COMPARISON OF POLICY LANGUAGES

Security policy languages have occasionally been compared over time by researchers in order to pave the way for future research or, to help security architects customise their security infrastructures. However, these comparisons have often focused on a small number of policy languages, or are already out-dated due to the fact that, the IT industry is continuously evolving at a rapid pace. From the vast array surveys regarding security policy languages reviewed by this research, we have used the following reports which compared these languages from different perspectives:

### A. *Scenario-based Comparison*

The ideology of Duma et al report is based on scenarios [4]. To be more precise, in order to compare policy languages the report defines sets of criteria. Following this, for each individual criterion which emerges from real user needs, the report presents a scenario to evaluate the language. If the scenario can be expressed and encoded in a language then the language fulfils the corresponding criterion. This point of view on the comparison of policy languages makes this report unique in its category.

### B. *Criteria-based Comparison*

De Coi et al. rigorously analysed twelve policy languages over a three years periods, with their report being concluded in 2008 [5]. The first criterion which they considered was that the language selected must be popular and widely used. The languages which they reviewed were Cassandra, EPAL , KAoS , PeerTrust , Ponder , Protune , PSPL , Rei , RT , TPL , WSPL and XACML. De Coi et al. evaluated these languages by comparing two sets of criteria namely, core policy properties and contextual properties. Each of these sets were further broken down into sub-categories.

As noted above, policy languages can be evaluated and classified from different perspectives. However we believe the top level classification of policy languages, as described in [5] and has restated here, fits perfectly in the context of this research.

Policy languages can be classified into one of three groups namely *standard-oriented*, r*esearch-oriented or "in between"* of these two mentioned groups. *Standard-oriented* policy languages are well defined and widely shared within the industry. However, they come with a restricted/minimal set of features. *Research-oriented* policy languages are popular amongst academics. They usually provide advanced features to their users and go beyond the boundaries which have been put in place by standard-oriented policy languages. There is also a third group of policy languages which are neither sufficiently advanced nor fully compatible with standardisation rules to be considered in either of the abovementioned categories. These languages are grouped in the third category referred to as "in between". We choose *XACML* from the standard-oriented group, *Ponder* from "in between" and *Protune* from the research-oriented group for consideration.

## III. ALGEBRA FOR POLICY LANGUAGES

As soon as a combination of heterogeneous security policies became vital, researcher realised that an independent nontrivial combination process, namely an algebra for policy languages, had to be introduced. As a result, a number of policy language algebras have been brought forth. Apart from the fact that using algebra minimises misunderstanding and ambiguity of policies when different parties refer to the same policy [6], algebra can be used for the decentralisation of policy descriptions where complicated and sophisticated policies get broken down into smaller manageable polices [7]. In the context of our research, it was almost obvious from the beginning that using algebra is inevitable as we needed it to describe all policies at an abstract level. As the result we reviewed and critiqued a number of policy language algebras in detail.

We believe the algebra introduced by Rao et al [8] paved the way and could be extended in order to cover all the policy languages we selected in section II. In the following section, we first discuss the policy semantics and then detail the limitation of the algebra before providing an appropriate extension to overcome the restriction.

### A. Policy Semantics

Rao et al. presented a simple yet powerful algorithm to describe *XACML* policy languages in [8]. This was then extended by Zhao et al. in [9]. In their notation, *A* is a finite set of *a* which characterises an object, subject or an environment. In the same sense, a domain defines a set of possible values and is denoted by *dom(a)*.

**Definition 1** *Let* $a_1, a_2, ..., a_k$ *be attribute names, and let* $v_i \in dom(a_i)$ *(*$1 \leq i \leq k$). $r \equiv \{(a_1, v_1), (a_2,v_2), \cdot\cdot\cdot, (a_k, v_k)\}$ *is a request.*

**Definition 2** A security policy is defined as a request evaluation function P : $ST \times A \rightarrow D$, where *ST* is the set of system states, *A* represents a finite set of actions and *D* denotes the set of decision tuple for authorisation and obligation associated with : $P\{Da,Do\} = \{(Y,Y),(Y,NA),(N,NA),(NA,NA)\}$. The function *P* takes a system state $st \in ST$ and an action $a \in A$ as input, and returns a decision tuple *(da,do)* determining whether a is authorised and obliged to execute in state *st*. As it is obvious from the decision tuple set, it does not include *(N,Y),(NA,Y)* as the obligation state is satisfied with positive authorisation.

**Definition 3** Let S be the set of subjects, T be the set of targets, E be the set of event triggers and C be the set of conditional constraints. System state is defined as $ST = E \times C \times S \times T$. This definition allows a system state to be described as $st = st(e, c, s, t)$ consisting of an event trigger $e \in E$, the conditional constraint $c \in C$, subject $s \in S$ and target $t \in T$.

### B. Policy Constants and operators

In addition the algebra defines its policy constants as follows:

**Permit policy:** $P_+ : ST\times A \rightarrow \{(Y, NA)\}$. $P_+$ authorises all authorisation requests in any state without considering any obligation.

**Deny policy:** $P\_ : ST \times A \rightarrow \{(N, NA)\}$. $P\_$ denies all authorisation requests in any state without considering any obligation.

The algebra also comes with its operators which are:

**Addition (+).** Integrated *policy* $P_I$ would be union of $P_1$ and $P_{2.}$

**Intersection (&).** Given $P_1$ and $P_2$, $P_I$ is defined as the intersection of these two polices if $P_I$ returns the same decision which is agreed by the two policies.

**Negation (**$\neg_a$ **,**$\neg_o$**).** $\neg P_1(st,a)$, return $P_I$ which effectively denies/permits all requests which $P_1$ permits/denies.

**Subtraction (-).** $P_I$ which denotes the result of $P_1(st,a) - P_2(st,a)$ is defined as a policy which allows all the requests which are authorised/obliged by $P_1(st,a)$ and are not applicable by $P_2(st,a)$.

**Projection (Π).** Taking into account the fact that the state of an environment is determined by events, constraints, subjects and targets, as described in Definition 3, and assuming c is a computable subset of ($ST \times A$), the projection operator restricts the policy *P* to requests which are satisfied by *c*.

$$PI\ (st,a)\ = \Pi_{c(ST,A)}^{(do,da)} =$$

$$\begin{cases} (da,do) & if(st,a) \in c(ST,A)\ and \\ & P(st,a) = (da,do) \\ (NA,NA) & otherwise \end{cases} \qquad (1)$$

## IV. ALGEBRA LIMITATION

The algebra, as briefly explained above and detailed in [8] and [9], is undoubtedly capable of expressing policy languages which are defined within a domain using a set of ground, i.e. variable free, authorisation and obligation terms. However in this section we would challenge the algebra with more sophisticated policy languages in order to extend the algebra which was presented by Rao et al and introduce *a fine grained algebra for policy languages with negotiation capability*. As mentioned earlier, extending the algebra presented by Rao can be presented as the contribution of this paper.

### A. *A simple negotiation scenario*

The following example, which is widely shared amongst policy languages with negotiation capabilities such as Protune, goes beyond the concept of unilateral negotiation as we know it in the traditional distributed system. By adopting the approach used by Duma et al in their report [4], we re-evaluate the defined algebra using the following scenario. In the following scenario, Alice, who is a police officer, would like to apply for a free Spanish course online. She does not mind providing information as long as it is not sensitive.

**Step 1** Alice requests to access eLearn's free Spanish course.

**Step 2** eLearn replies by requesting that Alice show a police badge issued by the State Police to prove that she is a police officer, and her driver's licence to prove that she is living in the same state.

**Step 3** Alice is willing to disclose her driver's licence to anyone, so she sends it to eLearn. However, she considers her police badge to contain sensitive information. She tells eLearn that in order to see her police badge, eLearn must prove that they belong to the Better Business Bureau (BBB).

**Step 4** Fortunately, eLearn has a Better Business Bureau membership card. The card contains no sensitive information, so eLearn discloses it to Alice.

**Step 5** Alice now believes that she can trust eLearn and discloses her police badge to eLearn.

**Step 6** After verifying that the badge is valid and that Alice owns it as well as the driver's licence, eLearn gives Alice the special discount for this transaction [10].

As is apparent from the above scenario, in simple terms negotiations can be divided into a series of steps. A message is usually exchanged in each step whilst the state is partially evaluated. Each evaluation leads the negotiation to the next level. The messages exchanged at each step are of different types, such as "Query message", for example, *Is Alice entitled to a discounted course?*, "Policy sets" messages which contain credentials, or simple decision messages which indicate the end of negotiation, possibly with a decision [11].

### B. *Limitation and the solution*

The policy algebra as been described so far is incapable of expressing the above mentioned scenario simply because we have added another dimension to our scenario; *negotiation*. In the above example, the negotiation process between eLearn and Alice and of course the final decision are unpredictable because Alice is located outside the boundaries of eLearn's environment and therefore her behaviour is beyond the visibility of eLearn. Hence, the algebra, as defined above, is not capable of formulating the state of environment as stated in Definition 3.

We are now ready to introduce the extended version of the algebra. In order to achieve this goal and formulate negotiation process two steps must be taken.

First, the definition of a security policy i.e. definition 2 needs to be finely modified as follow:

**Definition 2** A security policy is defined as a request evaluation function $P : \sum \times ST \times A \rightarrow D$, where $\sum$ denotes the finite set of non-ground literal, *ST* is the set of system states, *A* represents a finite set of actions and *D* denotes the set of decision tuple for authorisation and obligation associated with: *P{Da,Do} = {(Y,Y),(Y,NA),(N,NA),(NA,NA)}*. The function *P* takes a system state $st \in ST$, an action $a \in A$ and non-ground state literals $\sum$ (that intuitively specifies which state literals must be used) as input, and returns a decision tuple *(da,do)* determining whether a is authorised and obliged to execute in state *st*. As it is obvious from decision tuple set, it does not include *(N,Y),(NA,Y)* as obligation state is satisfied with positive authorisation.

Second, we must introduce another operator which utilises both sets of ground and non-ground literal states held at any given time. Assuming that *ST* denotes the ground literal states and refers to the set of literals which are held at the current state of environment, and $\sum$ denotes the set of non-ground literals, the *Trace* operator for policy languages with negotiation capability which provides partial evaluation with regards to $\sum \times ST$ can be defined as follows:

**Trace (H):** a Trace for policy *P*, which is a converging and non-ambiguous process, is defined as a set of policy sequences:

$$Pol_1 \xrightarrow{ST,\sum} Pol_2 \xrightarrow{ST,\sum} \ldots \xrightarrow{ST,\sum} Pol_{i-1} \xrightarrow{ST,\sum} Pol_i \ldots \qquad (2)$$

A trace is complete if for the last element $Pol_n$ in the sequence, there exists no policy $Pol_0$ such that $Pol_n$ $Pol_n \xrightarrow{ST,\sum} Pol_o$ . In simple words if criteria of policy *P*, denoted by $Pol_i$ remains unchanged (with regards to *ST and* $\sum$) then the trace is complete.

**Theorem 1:** For all policies *P*

1. In relation to *ST and* $\sum$ , policy *P* has no infinite complete traces.
2. All complete traces of policy *P* (which are defined as finite sequences of policies with an end policy element of $Pol_n$ with regards to *ST and* $\sum$) have the same final element, that is, policy *P's* decision.

**Proof:**

1. We must prove A) policy *P* cannot have a trace with infinite end elements and B) Policy *P* cannot have infinite traces with finite sets of end elements.

   A) The term *complete trace* used in theorem implies that the trace must come to an end that is, $policy_n$ (which denotes the final decision of the Policy *P*). In contrast, definition 2 clearly introduces a finite set of decision tuples for any Policy *P*. In other words Policy *P* cannot have an infinite set of complete traces with an infinite set of decisions/final elements.

   B) Arguably we could have an infinite number of scenarios with a finite number of final elements. However using the term *in relation to ST and* $\sum$ within theorem narrows down the number of scenarios and distinctly specifies which set of ground and non-ground literals is used.

   Taking the above facts into account the first part of the theorem proves itself because based on definition 2, policy P cannot have complete traces with infinite end elements $policy_n$ such that n→∞ and at the same time utilising the set of *ST and* $\sum$ narrows down the number of scenarios, hence policy P cannot have infinite complete traces with regards to *ST and* $\sum$.

2. Again using the terms *ST and* $\sum$ implies that the second part of the theorem is referring to a specific scenario. To prove this, we must refer to definition 2 which specifies that the final decision of a policy is determined by three inputs as $\sum \times ST \times A \rightarrow D$. In other words, the expression can be read as: As long as combination of $\sum$ , *ST* and A satisfies the policy P, it will make a decision. The way in which the policy collects this information has no effect on the decision which is made. Considering that trace is a sequence of policies which individually come to a decision and considering the fact that the order of evaluating the sub-policies has no effect on policy *P*'s decision (with regards to $\sum \times ST$) proves that different complete traces must have the same final element.

To make this part of the proof more tangible, let us take the abovementioned scenario when Alice asks for a course discount. If we keep the ground and non-ground literals of the environment the same, any alteration to the sequences of the event does not change the policy's decision. In other words if Alice asks eLearn to disclose their BBB membership number first and then discloses her driving licence and her police badge number, she would still be eligible for a discount.

### C. *Algebra completeness*

Proof of completeness of extended algebra is the next task on our list. We adopt the same approach described in [8] and [9] . In their approach they used a 2-dimensional combination matrix as shown in Figure (2). Bearing in mind the fact that each cell can potentially have four different values namely *(Y,Y),(Y,NA),(N,NA),(NA,NA)* there will be $4^{16}$ different combination matrices for a 2-dimensional matrix. The authors then proved there is an algebra expression to describe each and every individual policy which is presented by the combination matrix.

| P1 \ P2 | {N,NA} | {N,NA} | {Y,Y} | {NA,NA} |
|---|---|---|---|---|
| {Y,Y} | $e_1$ | $e_2$ | $e_3$ | $e_4$ |
| {Y,NA} | $e_5$ | $e_6$ | $e_7$ | $e_8$ |
| {N,NA} | $e_9$ | $e_{10}$ | $e_{11}$ | $e_{12}$ |
| {NA,NA} | $e_{13}$ | $e_{14}$ | $e_{15}$ | $e_{16}$ |

Figure2. A 2-dimentional combination matrix

**Theorem 2:** Let $M_i$ denote combination matrix in sequence *i* of a complete trace of a given policy *P* which requires partial evaluation with regards to *ST and* $\sum$ , then an algebraic expression exists which describes policy *P*.

**Proof:** The complete trace of policy P implies that the number of sequences in the trace are finite, ($1 \leq i < \infty$). Theorem 3 of [8] also proves that for any combination matrix there exists an algebra expression which contains algebra operators. Hence by adding up all expressions which describe the combination matrices of different individual sequences, we would be able to present policy P using algebra expressions.

What we must consider is the fact that for most cases the minimal set of operators which are needed to describe and formulate polices are {$P_{+}$ , $P_{-,}+$ , -, &, $\neg_{a}$ , $\neg_{o}$ , $\Pi$ }. The operator H is needed to describe and divide complex policies such as polices with negotiation capabilities to a finite sequence of sub-policies, in this case the abovementioned operators would be sufficient to describe them in detail.

### D. Algebra expressions

In practice, expressing a policy using proposed algebra requires multiple operators which must be used simultaneously, and as such we must define the algebra expression. An expression consists of a left associative, an operator and a right associative. Trace has the highest precedence, with negation and projection having the same priority, followed by intersection and addition respectively.

Assuming $P_1 = P_1(st,a)$ and $P2=P_2(st,a)$, the two algebra expressions which extend the expressions already defined in [8] can be presented as:

$$H\,(P_1 + P_2) = (HP_1)+(HP_2) \qquad (3)$$
$$H\,(P_1\ \&\ P_2) = (HP_1)\&(HP_2)$$

## V. CONCLUSION AND FUTURE WORK

In this report we have outlined the challenge of translating policies between different policy languages and have illustrated how the management of a distributed network which utilises heterogeneous policy languages is affected by that. Furthermore, we have described an interoperability framework for policy languages to address the issue. It is evident that our proposed framework, as a generic platform, would allow policy languages to be translated to and from each other and studying all available policy languages is unlikely to fit into the context of this research. In light of these facts, we have categorised policy languages into three groups according to their capabilities and have nominated one candidate from each set.

As these policy languages should eventually be mapped onto an abstract layer in order to be translated, we have illustrated how by using algebra at an abstract level we could prove that translation of policy languages to/from each other is possible. Finally, we have also summarised our study on available algebras for policy languages and suggested the one which is most suited for our research. In addition, we have pointed out the limitation of the algebra for policy language and provided our solution for overcoming the restriction.

There are many aspects of future work which have resulted from this paper. We are keen to design and develop a DSL (Domain Specific Language) for policy languages based on the facts we have proved and presented in this paper. The DSL at the core of our framework will facilitate the translation of policy languages to each other. We have already considered appropriate approaches and the development of the DSL is under way.

## REFERENCES

[1] Clemente, F., P´erez, G., Blaya, J., Skarmeta, A. (2005). Representing Security Policies in Web Information Systems. *In:* Proceedings of The International Workshop on Policy Management for the Web PM4W2005 The Fourteen International World Wide Web Conference WWW2005 Chiba Japan (2005),

[2] C.Basile, A.Lioy, S.Scozzi, M.Vallini, "*Ontology-based Security Policy Translation*", Journal of Information Assurance and Security, 2010, Vol.5, No.1, pp.437-445

[3] Review of Policy Languages and Frameworks, (Visited Jan 2012) http://www.w3.org/Policy/pling/wiki/PolicyLangReview.

[4] Claudiu Duma, Almut Herzog, and Nahid Shahmehri. Privacy in the Semantic Web: What Policy Languages have to Offer. Department of Computer and Information Science, Linköping University, Sweden. June 2007.

[5] Juri Luca De Coi, and Daniel Olmedilla. A Review of Trust Management, Security and Privacy Policy Languages. *International Conference on Security and Cryptography (SECRYPT 2008)*. INSTICC Press, July 2008.

[6] Ninghui Li and Qihua Wang. Beyond Separation of Duty: An Algebra for Specifying High-level Security Policies. Proceedings of the 13th ACM Conference on Computer and Communications Security. 2006.

[7] D. Wijesekera and S. Jajodia. A Propositional Policy Algebra for Access Control. ACM Transactions on Information and System Security. 2003.

[8] Prathima Rao, Dan Lin, Elisa Bertino, Ninghui Li, and Jorge Lobo. An Algebra for Fine-Grained Integration of XACML Policies. ACM Transactions on Information and System Security. May 2003.

[9] Hang Zhao, Jorge Lobo, and Steven M. Bellovin. An Algebra for Integration and Analysis of Ponder2 Policies. Proceedings of the 2008 IEEE Workshop on Policies for Distributed Systems and Networks. 2008

[10] Wolfgang Nejdl, Daniel Olmedilla, and Marianne Winslett. PeerTrust: Automated Trust Negotiation for Peers on the Semantic Web. In Workshop on Secure Data Management in a Connected World, Toronto, Aug. 2004

[11] P. A. Bonatti and D. Olmedilla. Reasoning on the Web with Rules and Semantics. Policy Language Specification. 2005.

# A Deception Framework for Survivability Against Next Generation Cyber Attacks

**Ruchika Mehresh**[1], **and Shambhu Upadhyaya**[2]
[1]Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY 14260, USA
[2]Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY 14260, USA

**Abstract**— *Over the years, malicious entities in cyber-space have grown smarter and resourceful. For defenders to stay abreast of the increasingly sophisticated attacks, the need is to understand these attacks. In this paper, we study the current trends in security attacks and present a threat model that encapsulates their sophistication.*

*Survivability is difficult to achieve because of its contradictory requirements. It requires that a critical system survives all attacks (including zero-day attacks), while still conserving the timeliness property of its mission. We recognize deception as an important tool to resolve this conflict. The proposed deception-based framework predicts an attacker's intent in order to design a stronger and more effective recovery; hence strengthening system survivability. Each design choice is supported by evidence and a detailed review of existing literature. Finally, we discuss the challenges in implementing such a framework and the directions that can be taken to overcome them.*

**Keywords:** Deception, mission critical systems, recovery, security, survivability

## 1. Introduction

This is the era of cyber-warfare and it is no longer limited to military domain. Knapp and Boulton [12] have reviewed information warfare literature from 1990 to mid-2005 and made a strong case for how cyber warfare has extended to other domains outside military. Baskerville [3] has discussed how the asymmetric warfare theory applies to information warfare and how it has expanded to the electronic business domain. According to the asymmetric warfare theory, attackers have the advantage of time and stealth over defenders. Thus, in order to counter this imbalance, defense needs to be "agile and adaptive."

Owing to this increasing hostility, critical systems in cyber space need to be protected. This need for protection extends beyond the routine fault tolerance and security into the domain of survivability. Ellison et al. [8] describe survivability as "the capability of a system to fulfill its mission in a timely manner in the presence of attacks, failures and accidents." Survivability focuses on continuity of a mission (set of essential services) without relying on the guarantee that precautionary measures will always succeed. It concentrates on the impact of an event rather than its cause. There are four basic layers of protection in a survivable system: Prevention or resistance against faults/attacks; Detection of faults/attacks; Full recovery of the essential services (mission) after the fault/attack and; Adaptation or evolution to reduce the possibility or effectiveness of future faults/attacks.

While the first two layers, prevention and detection already provide strong defense, recovery is the fallback option should these layers fail to protect the system. However, recovery being the last phase, needs protection (or a fallback). Mehresh et al. [20] discuss the possible attacks on the recovery phase of a critical system. Because adaptation/evolution mechanisms are generally activated during or after recovery, they are rarely effective if recovery fails. Therefore, recovery phase needs further protection to assure mission survivability.

One of the major challenges of designing a survivable system is to ensure that all the precedented or unprecedented threats are dealt with, while conserving the timeliness property of the mission. Since dealing with unprecedented attacks (zero-day attacks) requires monitoring the entire traffic, it becomes difficult to ensure timeliness property. Hence, these two requirements of surviving all kinds of threats and conserving the timeliness property are contradictory in nature. We propose deception as a tool to handle this conflict and even out the asymmetry in cyber warfare. Defensive deception is an act of intentional misrepresentation of facts to make an attacker take actions in defender's favor [7]. In this work, we make the following contributions:

- Study current trends in sophisticated attacks against mission critical cyber systems and present a next generation threat analysis/model (Section 3).
- Derive formal set of requirements for a survivable system to defend against such attacks (Section 4).
- Transform these requirements into a survivability framework where each design choice is supported with evidence and detailed reasoning (Section 5).

We review the related work in Section 2. In Section 3, we present an assessment of next generation threats. Formal requirements for the survivability framework are laid out in Section 4 and Section 5 presents the framework in detail. Section 6 concludes the paper by discussing the challenges involved in implementing this framework.

## 2. Related Work

The static nature of today's networks presents a sitting and vulnerable target. Patch development time for most exploits is much higher than the exploit development time. Repik [29] documents a summary of internal discussions held by Air Force Cyber Command staff in 2008. His work makes a strong argument in favor of using deception as a tool of defense. He discusses why planned actions taken to mislead hackers have merit as a strategy and should be pursued further.

Deception itself in warfare is not new [33], [6]. However, deception has several associated legal and moral issues with its usage in today's society. Cohen [6], the author of deception toolkit [5] discusses moral issues associated with the use of deception throughout his work. Lakhani [13] discusses the possible legal issues involved in the use of deception-based honeypots.

Deception aims to influence an adversary's observables by concealing or tampering with the information. Murphy [21] discusses the techniques of deception like, fingerprint scrubbing, obfuscation, etc. Her work is based on the principle of holding back important information from the attacker to render the attack weak. There is vast literature and taxonomies [6], [30], [29] on the use of deception to secure computer systems and information in general. Our framework essentially builds on these principles to handle sophisticated attacks.

## 3. Next generation threat assessment

In this section, we discuss the latest trends in sophisticated attacks on critical systems. This analysis helps us derive the attack patterns required to design a secure solution for the next-generation of critical systems.

Today's market forces and easy access to high-end technology have changed the cyber attack landscape considerably. As reported by Washington Post [22], malicious sleeper code is known to be left behind in the U.S. critical infrastructure by state-sponsored attackers. This sleeper code can be activated anytime to alter or destroy information. Similar stealth methodologies are also employed during multi-stage delivery of malware discussed in [28] and the botnet's stealthy command and control execution model in [11]. We already see a rising trend of stealthy smart malware all around [10]. Stuxnet, for instance, sniffs for a specific configuration and remains inactive if it does not find it. "Stuxnet is the new face of 21st-century warfare: invisible, anonymous, and devastating" [9]. Another instance of smart malware is 'Operation Aurora' that received wide publicity in 2009-10. The most highlighted feature of Aurora is its complexity, sophistication and stealth [16]. It includes numerous steps to gain and maintain access to privileged systems until the attacker's goals are met. The installation and working of this malware is completely hidden from the user.

A recently published report by McAfee surveyed 200 IT executives from critical infrastructure enterprises in 14 countries [2]. The report documents cyber-security experts expressing concerns about the surveillance of U.S. critical infrastructure by other nation-states.

Considering these trends, we set our focus on addressing attacks from a resourceful, adaptive and stealthy adversary. Note that aggressive attackers are easier to spot and hence the routine security measures generally take care of them. However, multi-shot, stealthy attackers rely on techniques that are difficult to detect and thus need innovative defense [28].

An attacker can cause maximum damage to a mission critical system during its crucial stage (like, the final stage). Multi-shot attackers are stealthy. They sniff around the system, install backdoors, place sleeper code, fragmented malware, etc., while evading detection. Thus, these stealthy attackers have a long time to infect the system. If discovered at a late stage, sometimes the only way left to recover is a system-wide sanitation which may disrupt the mission. Such drastic measures can cause a huge financial loss due to the heavy investments made during the course of a mission. At other times, the defender may not even get an opportunity to react. Thus, the need is to make these stealthy attackers manifest an easily detectable pattern at as early a stage as possible. Additionally, some stealth is required on defense system's part if it aims for a no-loss recovery from the attack. Most smart attackers and malware come with a contingency plan (to destroy or steal information on discovery). Thus, spooking an attacker without being prepared for the consequent contingency plan can be catastrophic. Thus, detection needs to be stealthy too, until the defender comes up with a plan to deal with a spooked attacker.

Based on the discussion above, we design a perceived smart attack flow. It is an extension of the basic one presented by Repik [29]. The attack flow is described in Algo. 1. Let $\phi$ be the set of exploitable vulnerabilities for a system with state s(t), where t is time. For each vulnerability $\nu$ in $\phi$, the amount of resources required to exploit it is represented by r[$\nu$]. Total resources available to an attacker is $\hat{r}$. Risk associated with exploiting each vulnerability $\nu$ is $\rho[\nu]$. Maximum risk that the attacker can afford is $\hat{\rho}$.

A sophisticated attack usually starts with intelligence gathering and initial planning. Based on the available resources, an attacker decides either to exploit a currently known vulnerability or keep searching for more. Attack occurs in multiple stages involving installing backdoors, rootkits, etc., until a crucial stage is reached. An attack during crucial stage has the maximum pay-off for the attacker. If discovered, most attackers have a contingency plan that may involve deleting or destroying information.

**Algorithm 1** Attack pattern for sophisticated attacks

```
while TRUE do
    while $\phi$ = NULL AND $\forall\nu$,$\rho[\nu]\geq\hat{\rho}$ do
        Gather intelligence
        Develop exploits
        Perform network reconnaissance
        Update vulnerability set $\phi$
    end while
    if $\exists\nu$, (r[$\nu$]$\leq\hat{r}$ AND $\rho[\nu]\leq\hat{\rho}$) then
        Install backdoors; Update $\hat{r}$
        while s(t) $\neq$ ATTACK_DISCOVERED do
            if s(t) $\neq$ CRUCIAL_STAGE then
                WAIT
            else if $\exists\nu$, (r[$\nu$]$\leq\hat{r}$ AND $\rho[\nu]\leq\hat{\rho}$) then
                Attack and exploit $\nu$; Update $\hat{r}$; Assess damage
                if s(t)=COMPROMISED then
                    Operation successful and Exit
                end if
            else
                Terminate operation
            end if
        end while
        if Contingency plan exists then
            Execute contingency plan
        else
            Terminate operation
        end if
    else
        Terminate operation
    end if
end while
```

## 4. Formal requirements

In light of the threat assessment presented in the previous section, we now list down requirements for a state-of-the-art deception-based security framework for mission survivability.

1) **Prevention**: It is generally the first step towards developing any effective solution in dealing with security threats. Prevention not only attempts to stop the attacks from succeeding, but also dissuades attackers with limited resources.
2) **Detecting the smart adversary**: We identify two main challenges for developing a security solution. First, it should force or manipulate a stealthy attacker into leaving a discernible and traceable pattern. Second, detection of such a pattern should be hidden lest the attacker should get spooked and execute a contingency plan for which the defender is not likely to be prepared. In addition to that, the solution should provide basic prevention, detection and recovery while conserving the timeliness property of the mission.

   For a given system state $s_1(t)$, there is a set $\phi_1$ of suspicious actions (for instance, a possible exploit attempt). A user that chooses an action from this set is malicious with a probability p. However, he could be benign with a probability 1-p. Let system states $s_1(t)$, $s_2(t)$,....,$s_n(t)$ (where, n is the total number of system configurations) have $\phi_1$, $\phi_2$,....,$\phi_n$ as their respective sets of suspicious actions. For some system states, this set of actions can be more clearly categorized as malicious with higher probabilities $p_i$ where, $1\leq i\leq n$. Choosing such states more frequently helps the defender to come up with a clear user profile in a shorter time. In honeypots, a defender can choose states with higher $p_i$'s, which means that if an attacker keeps choosing the actions from the set $\phi$, his probability of being malicious ($p_1.p_2.p_3....p_n$) will cross the threshold in a shorter time. Thus, choosing and controlling these states is crucial in determining if an attacker is malicious with a higher probability in a shorter time.
3) **Effective recovery with adaptation**: If the attacker has penetrated the system via existing vulnerabilities, recovering the system to the same old state does not remove these vulnerabilities. Therefore, the need is to ensure that during each recovery, vulnerabilities that are being exploited are patched. In this paper, we assume proactive recoveries that are periodically scheduled. It is much easier to predict the timing impact of proactive recoveries and hence conserve the timeliness property of a survivable system. Reactive recoveries, if evoked excessively, can harm system's performance and mission's survivability.
4) **Zero-day attacks**: Any good survivability solution must deal with zero-day attacks. Several anomaly-based detection systems have been proposed in order to detect such attacks [4]. However, Liu et al. [15] describe the big challenge, "how to make correct proactive (especially predictive) real-time defense decisions during an earlier stage of the attack in such a way that much less harm will be caused without consuming a lot of resources?" Schemes that attempt to recognize a zero-day stealth attack usually take two approaches: predictive and reactive. Under the predictive approach, all the suspected policy violations are taken as a sign of intrusion. This results in higher rate of false alarms and hence service degradation. Under the reactive approach, defender takes an action only when he is somewhat sure of a foul play. Generally, it is difficult to know when to react. If the system waits unless a complete attack profile emerges, it may be too late to react. A good trade-off is offered by honeypots (a form of deception). The defender redirects all the suspicious traffic through honeypots which is responsible for blacklisting/whitelisting the traffic flows [25], [24].

Authors in [14], [13] introduced methodologies for employing honeynet in a production-based environment.

5) **Conserving timeliness property**: Timeliness property describes the capability of a mission survivable system to stick to its originally planned schedule. This being said, a schedule can account for periodic recoveries and some unexpected delays due to miscellaneous factors. In order to conserve this property, it is essential that all the indeterministically time-consuming operations be moved out of the mission's critical path. Thus, it is essential to redirect the suspicious traffic to a separate entity for further examination.
6) **Non-verifiable deception**: A good deception should be non-verifiable [23]. Deception is difficult to create but easier to verify. For instance, an attacker attempts to delete a file. Even if a deceptive interface gives a positive confirmation, the attacker can always verify if the file exits.

   For a state s(t), an action $\chi$ is expected to have an effect $\omega$. Generally, deception (like in honeypots) involves confirming that $\chi$ has been performed but the effect $\omega$ is never reflected in the system. If the attacker has a feedback loop to verify $\omega$, a deception can be easily identified. Therefore, either the feedback loop needs to be controlled so as to give the impression that $\omega$ exists, or the feedback loop should be blocked for all regular users. An open and honest feedback loop will help attacker in figuring out ways around deception by trial-and-error.

# 5. The Framework

## 5.1 Basics

**Preventive deception** is the first step in mission survivability. Traditionally, measures like firewall, encryption techniques, access control, etc. have been used as preventive measures. These measures have proved to be very successful in deterring weak adversaries. However, strong and determined adversaries are always known to find their way around these. McGill [17] suggests that the appearance of a system being an easy or a hard target determines the probability of attacks on it. Based on similar literature, we categorize deception-based prevention methodologies under following four headings:

- *Hiding*: Hiding is the most basic form of deception. One could use schemes like fingerprint scrubbing, protocol scrubbing, etc. to hide information from an attacker [34], [31]. Similarly, these schemes could also be used to feed false information to the attacker. Yuill et al. [35] have developed a model for understanding, comparing, and developing methods of deceptive hiding.
- *Distraction*: McGill [17] demonstrates that given two targets of equal value, an attacker is more likely to attack the target with lesser protection. However, Sandler and Harvey analytically prove that this tendency continues only till a threshold. If more vulnerabilities are introduced to a system, an attacker's preference for attacking that system does not increase beyond a certain threshold.

  System observables that attackers rely on can be manipulated to feed misinformation or hide information from attackers. Thus, strategies can be devised to affect an attacker's perception about the system. Studies like [18] model threat scenarios based on target's susceptibility and attacker's tendencies. Such models can be used to assess the attractiveness of a target to an attacker if its apparent susceptibility is manipulated via its observables.

**Axiom 1**: Adding more vulnerabilities to one of the two equal-value systems increases the likeliness (till a threshold) of an attack on the one with more vulnerabilities.

- *Dissuasion*: Dissuasion describes the steps taken by a defender to influence an attacker's behavior in his favor. It involves manipulating system observables to make it look like it has stronger security than it actually does. This may discourage attackers from attacking it. As shown in Algo. 1, if the estimated resources for exploiting the system go over $\hat{r}$ or the estimated risk goes over $\hat{\rho}$, the attacker will be dissuaded from attacking the system. Dissuasion is generally implemented as deterrence or devaluation. Deterrence involves a false display of greater strength. Devaluation, on the other hand, involves manipulating observables to lessen the perceived value that comes out of compromising a system. McGill [17] develops a probabilistic framework around the use of defensive dissuasion as a defensive measure.

Deception techniques are complementary to conventional prevention techniques rather than a replacement.

**Axiom 2**: False display of strength dissuades an attacker from attacking the system.

**Axiom 3**: Increasing or decreasing the perceived value of a system affects the attacker's preference of attacking the system favorably or adversely, respectively.

**Honeypot** is a tool of deception. It generally comes across as a system capable of a low-resource compromise with high perceived gains. Honeypot not only distracts an attacker from attacking the main system, but also logs attacker's activity heavily. Studying these logs can help the defender to gauge an attacker's capability and come up with a good strategy to ward off any future attacks. Spitzner describes honeypot as "a security deception resource whose value lies in being probed, attacked, or compromised" [32]. Honeypots are generally classified into two categories: Physical and Virtual. Physical honeypots are when real computer systems are used to create each honeypot. Virtual honeypots use

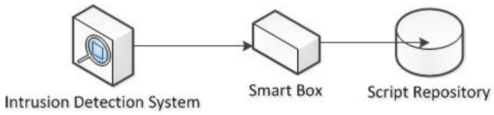


Fig. 1: Smart-Box

software to the workings of a real honeypot and the connecting network. They are cheaper to create and maintain and hence are used in the production environments more often. Virtual honeypots are further divided into high interactive and low interactive honeypots. Qasswawi et al. [27] provide a good overview of the deception techniques used in virtual honeypots.

High interactive honeypots provide an emulation for a real operating system. Thus, the attacker can interact with the operating system and completely compromise the system. Some examples are User Mode Linux (UML), VMware, Argos, etc. Low-interaction honeypots simulate limited network services and vulnerabilities. They can not be completely exploited. Examples are LaBrea, Honeyd, Nepenthes, etc. [26], [27].

Cohen's Deception Toolkit (DTK) laid the groundwork for low-interaction honeypots [5]. It led to the development of advanced products like Honeyd. Honeyd [32] simulates services at TCP/IP level in order to deceive tools like Nmap and Xprobe. Though it does not emulate the entire operating system, its observables are modified to give the impression that it does.

**Honeypot farm** is a cluster comprised of honeypots of the same or different kinds. Hybrid honeypot farms consist of a mixture of low and high interactive honeypots.

**Smart-box** is a module that we proposed to help figure out the best deception for a specific suspicious traffic flow. Conceptually, a smart-box works as shown in Fig. 1. It takes input from the IDS about the suspected traffic flow. The logic in smart-box then decides Attackers Intent, Objectives and Strategies (AIOS) based on this information [15]. Then it maps the AIOS to deception scripts. These scripts are stored in the script repository.

## 5.2 Design

Building up from the concepts discussed in the previous subsection, we extend the model presented in [13] to design our deception-based survivability framework.

As shown in Fig. 2, the mission survivable (production) system runs behind several layers of protection including firewalls, deception, etc. The first layer of proxy servers uses Axioms 1, 2 and 3 to mislead attackers into choosing systems that will re-route their traffic to honeypot farm via the smart-box. Rest of the unsuspected traffic goes through the main server, the firewall and the intrusion detection system. Intrusion detection system is another layer of defense which re-routes any suspicious traffic to the honeypot farm for further analysis.

Suspicious traffic is generally sieved out based on two criteria: either the intrusion detection system identifies an attack pattern in the traffic flow or the traffic originates to/from dark address space. Dark address space is the set of Internet's routable address reserved for future network expansion. These two criteria worked just fine until cloud computing came along. Now attackers can launch their attacks from behind the cloud using valid IP addresses and evade detection. Therefore, in addition to employing the above-mentioned two methods, we introduce a layer of distraction proxy servers. This layer contains a main server which is widely publicized to legitimate clients. This main server is extremely secure and its observed security is further enhanced (deception/deterrence). Thus, amateur attackers are dissuaded from attacking it. Other proxy servers expose a specific set of non-essential, vulnerable services. For instance, one server can keep the *ssh* port open to accept the traffic, while the other can mislead the attacker into thinking that it is running a vulnerable version of Windows operating system. These servers not only distract the malicious traffic away from the main server but, also inform the smart-box about attackers' intentions (based on their preference of proxy servers and vulnerabilities that they try to exploit).

In this design, we use smart-box to optimize resource allocation in hybrid honeypot farms. These honeypots should be assigned to the traffic flows based on the assessment about each flow's AIOS. This is because low-interaction honeypots can be easily verified if attacker suspects deception and tries to go in deeper. Use of high-interactive honeypots for deception is more fool-proof but consumes more computing and memory resources. Thus, smart-box helps in smart allocation of these honeypot resources by assessing the nature of an attack and re-routing the traffic to appropriate honeypots (similar to loading the deception scripts).

Logging tools and analyzer in the honeypot farm recognize an attack and create a complete attack profile. Based on this attack profile, the flow is whitelisted and forwarded to the production server or blacklisted. If blacklisted, either automated patches, if available, are executed in the next recovery cycle, or a system administrator is alerted. This is the step where the attack profile helps the defender to develop an effective patch for the next recovery cycle, while the unsuspected malicious actor stays busy playing with the honeypot. Thus, deception buys defender the time to design an effective recovery.

Since a system is "as secure as its weakest point", we need to make sure that this framework not only provides good security but is tamper-proof at all times. Since all the modules in this design like, proxy servers, the traffic redirection module, intrusion detection systems, etc. are connected to the same network, they are always susceptible to intrusions. Therefore, these modules need to be tamper-proof in order for the entire design to be tamper-proof.

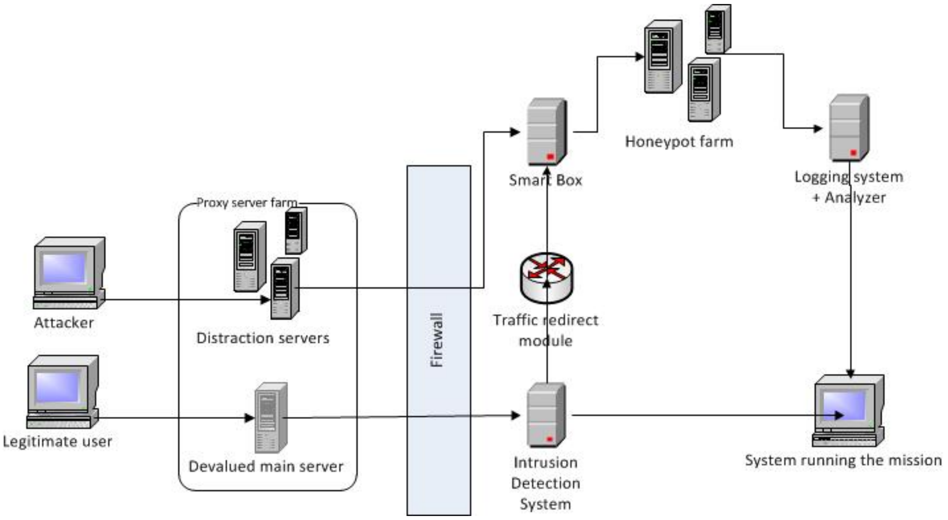


Fig. 2: Deception framework for mission survivability

We can use techniques like lightweight cyclic monitoring in order to make sure that IDS on all these modules (like proxy servers, IDS, etc.) stay tamper-proof [19]. Then using the scheme described by Mehresh et. al in [20], the integrity signature of each module is surreptitiously detected and sent to the production system for verification. The detection is secret so the attacker is not spooked. This arrangement introduces a cyclic integrity-check. All modules make sure that production system works tamper-free at all times, while the production server takes care of the integrity-check for all modules.

Anagnostakis et al. [1] proposed shadow honeypots as an effective solution to deploy honeypots in a production environment. Shadow honeypots use a combination of anomaly intrusion detection systems and shadow honeypots. A variety of anomaly detectors monitor traffic in the network and the suspected traffic is forwarded to a shadow honeypot. Shadow honeypot is an identical copy of the production server but instrumented to detect potential attacks. Misclassified traffic is verified by the shadow and transparently handled correctly. We see many challenges in this approach. First, predictive anomaly detectors (higher sensitivity) will have more false positives and will direct more misclassified traffic to the shadow honeypot, creating unnecessary delays and overhead. Reactive anomaly detectors (lower sensitivity) will take more time to create a complete profile and may miss a lot of malicious traffic before identifying a problem with the flow. Moreover, identifying zero-day attacks ask for a higher sensitivity intrusion detection. Additionally, each suspected traffic flow may need separate copies of shadow honeypot (else an attacker can verify deception by initiating two parallel malicious flows). This further increases the overhead.

## 6. Discussion and Conclusion

The most important factor to consider while designing any aspect of this deception framework is to remember that nothing will remain a secret if it is widely deployed. Hence, an effective deception must assume that an attacker knows about its existence with some probability. That's why all deceptions should be non-verifiable. In this case, when an attacker sees several proxy servers with vulnerabilities, he sends traffic flows to all the servers. The flow that gets through the fastest is the main server (under the safe assumption that going through honeypot farm adds to the delay). That's why, any feedback loop for the attacker must also be controlled with deception. Discussing deception in feedback loop is beyond the scope of this paper.

Another major challenge is the design of the smart-box. A smart-box performs two major functions: assess the nature of the traffic flow and, map the AIOS to a honeypot. Designing an implementation of both these functions is a major challenge and will benefit excessively from the use of machine learning algorithms. Deceptions in honeypots can also be made customizable based on the parameters provided by the smart-box. Other challenges like designing proxy servers, re-routing, choosing the IDS, etc. depends on the system that the framework is used for and the designer.

In this paper, we focus on designing survivable mission critical systems. We began with analyzing the current cyber security attacks to derive the next generation threat assessment. Multi-shot, stealthy attacks came out as a major threat in this assessment. We then defined a set of requirements around this threat for a survivability framework. Based on evidence and existing literature, we identified deception as an important tool of defense and designed a deception framework for survivable systems. This framework deals with zero-day attacks while still conserving the timeliness property of a mission. It uses concepts of deception to introduce a preventive layer of proxy servers that helps system to further narrow down the suspicious traffic. This traffic then gets rerouted to a smart-box that selects the honeypot this traffic is forwarded to. Honeypots provide an important functionality of uncovering the stealthy patterns in these traffic flows with a higher probability in a shorter time. This way, the framework helps in identifying and rooting out the stealth attacks at an early stage.

A major advantage of this framework is the strong recovery that it provides. It buys defender more time to analyze the suspected traffic flow without spooking the adversary. The analyzer and log modules help in designing a secure and more effective recovery patch. Hence, this framework ensures system survivability equipped with a strong recovery phase.

In future, we plan to address the challenges we discussed above and work towards implementing a prototype of this framework.

## 7. Acknowledgments

This research is supported in part by DoD Grant No. H98230-11-1-0463. Usual disclaimers apply.

## References

[1] K. G. Anagnostakis, S. Sidiroglou, P. Akritidis, K. Xinidis, E. Markatos, and A. D. Keromytis. Detecting Targeted Attacks Using Shadow Honeypots. *Proceedings of the 14th conference on USENIX Security Symposium*, page 9, 2005.

[2] S. Bake, N. Filipiak, and K. Timli. In the Dark: Crucial Industries Confront Cyberattacks. *McAfee second annual critical infrastructure protection report*, 2011.

[3] R. Baskerville. Information Warfare Action Plans for e-Business. *3rd European Conference on Information Warfare and Security*, pages 15–20, 2004.

[4] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41:15:1–15:58, 2009.

[5] F. Cohen. Deception Toolkit, 2001.

[6] F. Cohen, D. Lambert, C. Preston, N. Berry, C. Stewart, and E. Thomas. A Framework for Deception. *IFIP-TC11, Computers and Security*, 2001.

[7] D. C. Daniel and K. L. Herbig. *Strategic Military Deception*. Pergamon Press, 1982.

[8] R. J. Ellison, D. A. Fisher, R. C. Linger, H. F. Lipson, T. A. Longstaff, and N. R. Mead. Survivability: protecting your critical systems. *IEEE Internet Computing*, 3:55–63, 1999.

[9] M. J. Gross. A Declaration of Cyber-War, April 2011.

[10] A. Kapoor and R. Mathur. Predicting the future of stealth attacks. *Virus Bulletin Conference*, 2011.

[11] E. J. Kartaltepe, J. A. Morales, S. Xu, and R. Sandhu. Social network-based botnet command-and-control: emerging threats and countermeasures. *Proceedings of the 8th international conference on Applied cryptography and network security (ACNS)*, pages 511–528, 2010.

[12] K. J. Knappa and W. R. Boulton. Cyber-Warfare Threatens Corporations: Expansion into Commercial Environments. *Information Systems Management*, 23:76–87, 2006.

[13] A. D. Lakhani. Deception techniques using Honeypots. *MSc Thesis, ISG, Royal Holloway, University of London*, 2003.

[14] J. G. Levine, J. B. Grizzard, and H. L. Owen. Using honeynets to protect large enterprise networks. *IEEE Security and Privacy*, 2:73–75, 2004.

[15] P. Liu, W. Zang, and M. Yu. Incentive-based modeling and inference of attacker intent, objectives, and strategies. *ACM Transactions on Information and System Security (TISSEC)*, 8, 2005.

[16] McAfee Labs and McAfee Foundstone Professional Services. Protecting your critical assets, lessons learned from "Operation Aurora". *Technical report*, 2010.

[17] W. L. McGill. Defensive dissuasion in security risk management. In *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2009.

[18] W. L. McGill, B. M. Ayyub, and M. Kaminskiy. Risk Analysis for Critical Asset Protection. *Blackwell Publishing Inc*, 27:1265–1281, 2007.

[19] R. Mehresh, J. J. Rao, S. J. Upadhyaya, S. Natarajan, and K. Kwiat. Tamper-resistant Monitoring for Securing Multi-core Environments. *International Conference on Security and Management (SAM)*, 2011.

[20] R. Mehresh, S. J. Upadhyaya, and K. Kwiat. Secure Proactive Recovery - A Hardware Based Mission Assurance Scheme. *Journal of Network Forensics*, 3:32–48, 2011.

[21] B. S. Murphy. Deceiving Adversary Network Scanning Efforts Using Host-Based Deception. Technical report, Air Force Institute of Technology, Wright-Patterson Air Force Base, 2009.

[22] E. Nakashima and J. Pomfret. China proves to be an aggressive foe in cyberspace, November 2009.

[23] V. Neagoe and M. Bishop. Inconsistency in deception for defense. In *Proceedings of the 2006 workshop on New security paradigms*, 2007.

[24] R. R. Patel and C. S. Thaker. Zero-Day Attack Signatures Detection Using Honeypot. *International Conference on Computer Communication and Networks (CSI- COMNET)*, 2011.

[25] G. Portokalidis and H. Bos. SweetBait: Zero-Hour Worm Detection and Containment Using Low- and High-Interaction Honeypots. *Science Direct*, 51:1256–1274, 2007.

[26] N. Provos and T. Holz. *Virtual Honeypots: From Botnet Tracking to Intrusion Detection*. Addison-Wesley, 2008.

[27] M. T. Qassrawi and H. Zhang. Deception Methodology in Virtual Honeypots. *Second International Conference on Networks Security Wireless Communications and Trusted Computing (NSWCTC)*, 2:462–467, 24–25, 2010.

[28] M. Ramilli and M. Bishop. Multi-Stage Delivery of Malware. *5th International Conference on Malicious and Unwanted Software (MALWARE)*, 2010.

[29] K. A. Repik. Defeating adversary network intelligence efforts with active cyber defense techniques. Master's thesis, Graduate School of Engineering and Management, Air Force Institute of Technology, 2008.

[30] N. C. Rowe and H. S. Rothstein. Two Taxononmies of Deception for Attacks on Information Systems. *Journal of Information Warfare*, 3:27–39, 2004.

[31] M. Smart, G. R. Malan, and F. Jahanian. Defeating TCP/IP stack fingerprinting. *Proceedings of the 9th conference on USENIX Security Symposium*, 9:17–17, 2000.

[32] L. Spitzner. Honeynet Project, Know Your Enemy: Defining Virtual Honey-nets, 2008.

[33] S. Tzu. *The Art of War (Translated by James Clavell)*. Dell Publishing, New York, NY, 1983.

[34] D. Watson, M. Smart, G. R. Malan, and F. Jahanian. Protocol Scrubbing: Network Security Through Transparent Flow Modification. *IEEE/ACM Transactions on Networking*, 12:261–273, 2004.

[35] J. Yuill, D. Denning, and F. Feer. Using Deception to Hide Things from Hackers: Processes, Principles, and Techniques. *Journal of Information Warfare*, 5:26–40, 2006.

# Extension of Account Management Method with Blind Signature Scheme

**Ryu Watanabe and Yutaka Miyake**
KDDI R&D Laboratories, Inc., Ohara 2-1-15, Fujimino, Saitama, Japan

**Abstract**—*In order to reduce the risk of user information leaks, Dey et al. have proposed an account management method for a federated login system using a blind signature scheme. However their scheme includes a defect that allows a malicious user to generate multiple accounts without limitation. To address this problem, we have modified and extended the scheme. By introducing a token management scheme for user account generation at the user registration authority (a blind signature service, or BSS), user account generation and deletion on an IDP can be controlled. However, our proposed scheme has a limitation in terms of the number of IDPs and the number of user accounts on an IDP for each user. The BSS can handle only one IDP and one account for each user. In order to remove this limitation, we have extended our scheme by using modified token management between the BSS and IDP. In this paper, we describe the extension.*

**Keywords:** security, identity management (IdM), Blind Signature Scheme

## 1. Introduction

Recently, innovations in information technology (IT) and improved infrastructure for high-speed communication have led to the development of more flexible communication on the Internet. Various WEB services are available, such as on-line shopping, VoD services, and social network services (SNS). In this environment, in order to enhance both the security and usability of WEB services, there is increasing use of identity management technology for single sign-on (SSO) techniques. In addition, many different specifications and implementations are available[2][3]. By using an SSO technique, a single authentication from an identity provider (IDP) can substitute for authentication by each service provider (SP). Relevant issues in this area have been well studied under the designation 'identity management technique'.[7][8][9]

Once an IDP authenticates a user, it retains the user authentication status. When the user wants to use a service, the service provider delegates the user's authentication to the IDP. Then the IDP can inform the service of the user's authentication status. The SP receives the notification from the IDP and checks it, completing user authentication on the SP. If the user wants to use different services, each SP authenticates the user in the same manner. Therefore, users do not have to input each ID/PW pair and are freed from the nuisance of having to manage numerous ID/PWs.

In this SSO technique, only IDPs know the user identity for authentication, therefore, they have to strictly manage and maintain such important information. One of the issues of this form of account management is a privacy problem. Usually, in order to avoid a linkability problem, the accounts on an IDP and an SP are federated using a pseudonym. However, if an IDP is cracked, the relationship between the accounts at the IDP and the pseudonym is revealed. In this case, from the relationship, the user identity held by an IDP can be linked to user activities on each service.

To deal with this account problem, Dey et al. proposed an account management method called PseudoID, a federated login system that protects users from disclosure of private login data held by identity providers. In PseudoID, the user authentication function is separated from the IDP. The function is passed on to an authority called a blind signature service (BSS). The BSS independently deals with user authentication and generates tokens for account generation for an IDP. Therefore, even though an IDP is cracked, the user identity behind an authentication cannot be revealed, thereby protecting user privacy. However, the PseudoID focuses on user privacy protection, so the scheme has the drawback of multiple user account generation.

To deal with this drawback, we have proposed a modified scheme, which provides an account generation and deletion control function to IDPs by introducing token management on the BSS[1]. Checking the user's token status, the BSS can avoid issuing multiple tokens for a user account generation on an IDP. In addition, on account termination, the IDP generates a deletion token and the user submits the token to the BSS. By validating this deletion token, BSS regains the ability to reassign the user token for account generation.

However, our previously proposed scheme was limited with respect to the number of IDPs and accounts. In that scheme, the BSS handled only one IDP. Moreover, the scheme handled only one account for each user on an IDP. This limitation is not appropriate for current federated login systems because a user can use only one ID provider. To remove this limitation, we have extended our scheme so that the token for account generation is modified to manage multiple IDPs. In addition, multiple account management for a single user on one IDP is also introduced. With these extensions, more flexible account management can easily be

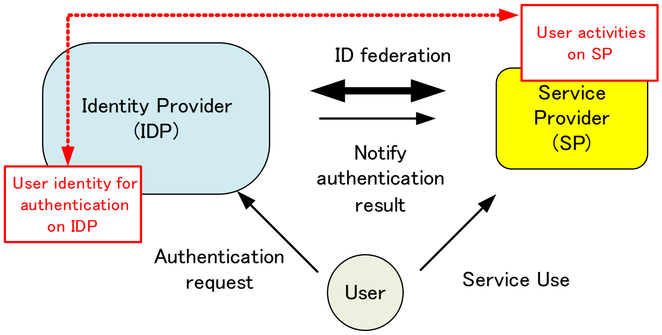


Fig. 1: Privacy problem on SSO with ID federation

provided.

## 2. Related work

### 2.1 Blind signature

Here we describe the blind signature scheme, which is a key component for the PseudoID. The blind signature scheme is an extension of the digital signature scheme, a public key cryptosystem. The digital signature scheme consists of the signing algorithm $S()$ with a private key, which belongs to the signer, and the verification algorithm $V()$ with a public key, which is open to the public and paired with the private key. The signer appends his/her signature to a message with his/her private key, which only he/she knows. The verifier verifies the signed message with the corresponding public key. In the verification algorithm, using the message $M$, which is the signing target and the signed message $S(M)$, the $V(M, S(M))$ is calculated for verification.

The blind signature scheme applies a blind messaging process to an ordinary signature scheme. The blinding function $B()$ is introduced and is used to make messages unreadable. The blinding function $B()$ and the signature function $S()$ satisfy the equation given below. In addition, $B^{-1}()$ is the inverse function of $B()$.

$$B^{-1}(S(B(m))) = B^{-1}(B(S(m))) = S(m) \qquad (1)$$

In a blind signature scheme, users want a signer to generate a signature without revealing his/her message to the signer. Therefore, the user encrypts the message using a blinding function and sends it to a signer. Then the signer generates a signature on the blinded message with its private key and returns it to the user. The user can unblind the signed message by using the relationship between the blinding function and signing function denoted in equation (1) and obtain the signature for the message. Then the user sends the signature to a verifier. The verifier can verify the signed message with the signer's public key.

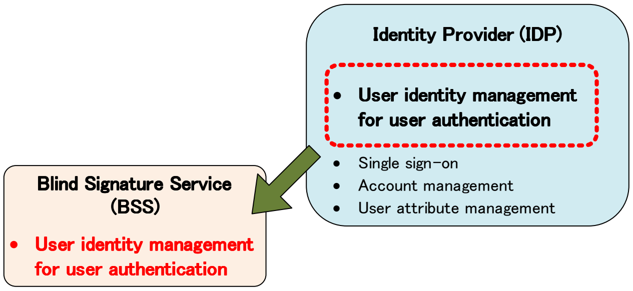

Identity Provider (IDP)
User identity management for user authentication
Single sign-on
Account management
User attribute management
Blind Signature Service (BSS)
User identity management for user authentication

Fig. 2: Concept of PseudoID account management

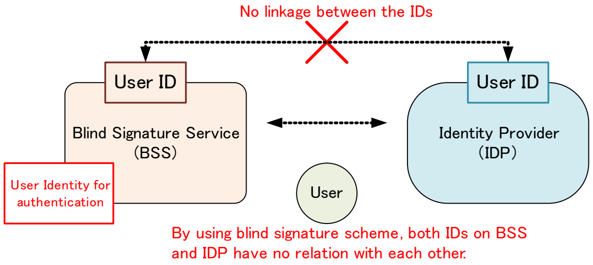

No linkage between the IDs
User ID
Blind Signature Service (BSS)
User Identity for authentication
User ID
Identity Provider (IDP)
User
By using blind signature scheme, both IDs on BSS and IDP have no relation with each other.

Fig. 3: Privacy protection system with PseudoID

### 2.2 PseudoID

In the current single sign-on scheme, an IDP manages user identity for account generation and the IDs on both the IDP and SPs are linked via an ID federation technique. Therefore, there is a security concern about user privacy. For example, if an IDP is cracked by a malicious party, the user's identity held on the IDP and the activities on services are linked and revealed through this connection between IDs on both the IDP and SPs. A conceptual representation of this problem is shown in Fig. 1

To resolve this privacy problem, Dey et al. proposed an account management method referred to as "PseudoID". In the PseudoID method, the user identification function is delegated to a dedicated party called a blind signature service from an IDP (Fig. 2). Between this dedicated party and the IDP, the IDs of both parties are not linked (Fig. 3). Therefore, if the IDP or BSS is cracked, the risk of an identity leak is reduced. In order to perform this function, the PseudoID applies the blind signature scheme. Before explaining the PseudoID scheme in more detail, the scheme's assumptions are summarized. The BSS can make user accounts for itself by identifying users' identities and retaining them for user management. The BSS opens its public key and other parties can confirm its legitimacy through a defined operation.

On PseudoID, first, the user prepares an ID/PW pair for an IDP and the pair is used as token between the BSS and the IDP. Then the user blinds the token and sends it to the BSS. The BSS authenticates the user and then generates a signature against the token. The token is then returned to

Table 1: Definition of symbols

| Symbol | Definition |
|---|---|
| $ID_X$ | User ID on entity X |
| $PW_X$ | User password on entity X |
| $B()$ | Blind function |
| $S_X(M)$ | Signature for message M by entity X |
| $V(M, S_X(M))$ | Verification function for Signature |
| $R_X$ | Random number generated by entity X |
| $\|\|$ | concatenation |

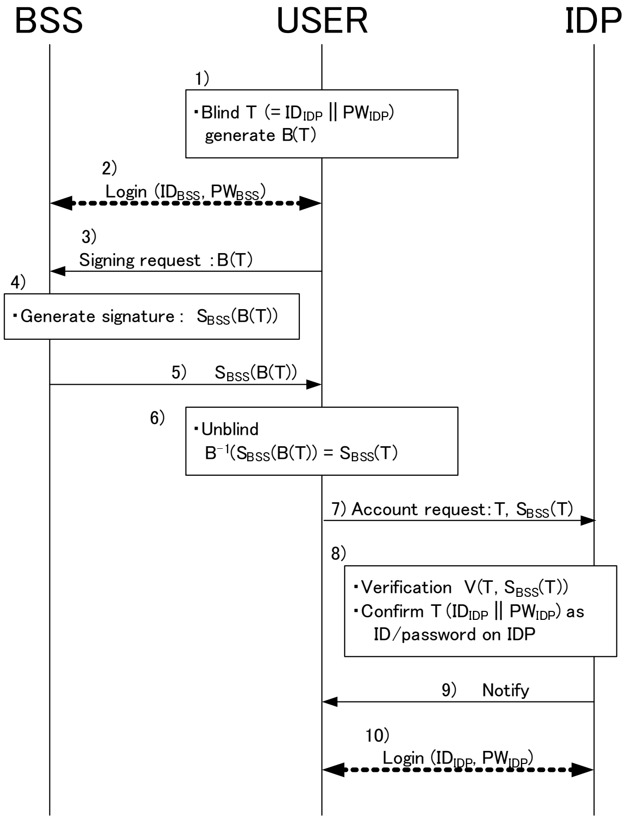

BSS
USER
IDP
1)
・Blind T (= $ID_{IDP}$ || $PW_{IDP}$)
generate B(T)
2)
Login ($ID_{BSS}$, $PW_{BSS}$)
3)
Signing request : B(T)
4)
・Generate signature : $S_{BSS}$(B(T))
5) $S_{BSS}$(B(T))
6)
・Unblind
$B^{-1}$($S_{BSS}$(B(T)) = $S_{BSS}$(T)
7) Account request: T, $S_{BSS}$(T)
8)
・Verification V(T, $S_{BSS}$(T))
・Confirm T ($ID_{IDP}$ || $PW_{IDP}$) as
ID/password on IDP
9) Notify
10)
Login ($ID_{IDP}$, $PW_{IDP}$)

Fig. 4: Sequence of PseudoID account management

the user. The user unblinds the token and submits the signed token to the IDP. If the IDP can verify the signature for the token, the IDP generates a user account and accompanying password for the user based on the contents of the token.

The sequence of account generation in PseudoID is denoted in Fig. 4. The definition of symbols in Fig. 4 is summarized in table 1.

In the PseudoID scheme, the ID/PW pair on an IDP is hidden from the BSS using a blinding function. Therefore, if a user requests the BSS to sign the tokens, which each have a different ID/PW pair, the user can generate multiple accounts on the targeted IDP without limitation(Fig. 5). This action is undesirable from the IDP's viewpoint. So the number of user accounts for one individual user on one IDP should be limited by BSS and SP.

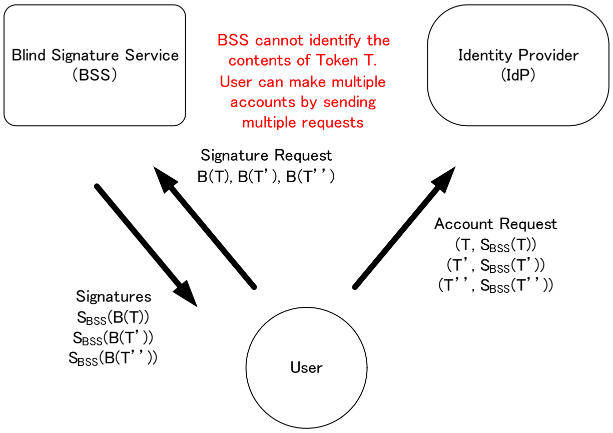


Fig. 5: Problem of multiple account generation on PseudoID

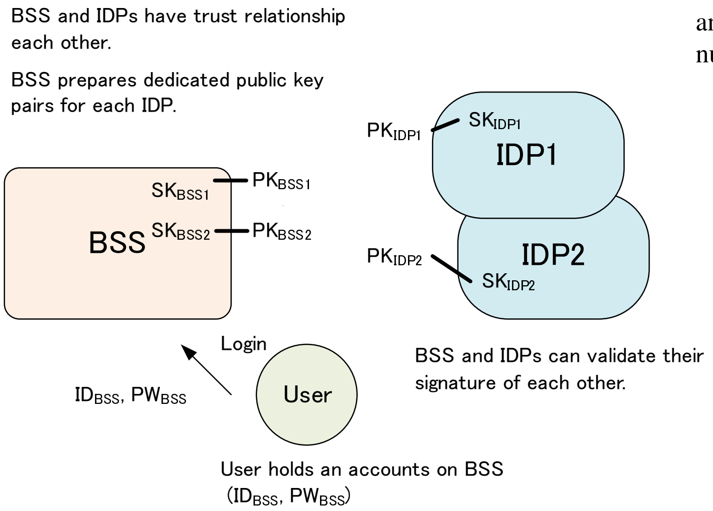
## 2.3 Our previous proposal

In order to avoid the multiple account problem in PseudoID, the authors have already proposed a modified scheme[1]. In our previous proposal, we introduced token generation management in the BSS, which is used for account generation in the IDP as an additional function. In the account deletion procedure, a blind signature scheme is used and prevents any linkage or relationship between accounts on both the BSS and IDP. Thus, the risk of identity leak is reduced. However, our previous scheme also has a limitation. The BSS handles only one IDP and cannot handle multiple accounts for one user on one IDP. This limits users' flexibility. Some use cases are assumed. For instance, currently, there are many IDPs. IDPs vary, as do the corresponding service providers. So, in this situation, users want to (or must) use multiple IDPs. In another situation, some users want to use two or more accounts (identities). Usually a person has two aspects, one private and the other public. In this case, a user wants to use two or more accounts corresponding to those roles. Therefore, we decided to extend our scheme and relax the limitation.

# 3. Our new proposal

Below, we lay out the requirements for our new proposal. For this extension of our scheme, we also took security and usability into account.

1) The accounts on a BSS and each IDP must have no relationship.
2) Users can regenerate their accounts on IDPs.
3) An IDP and a BSS cooperate to control the number of user accounts for the same user for one IDP.
4) The accounts on each IDP must have no relationship. So the accounts on different IDPs for the same user are independent.



BSS and IDPs have trust relationship each other.

BSS prepares dedicated public key pairs for each IDP.

BSS and IDPs can validate their signature of each other.

User holds an accounts on BSS (IDBSS, PWBSS)

Fig. 6: The assumptions of our new proposal

The first two requirements are already present in our old scheme, while the third and fourth requirements are new. The third is for usability. To provide users with more flexibility, we include a function that controls the number of IDP accounts. The upper limit on the number of user accounts for one user is determined by the IDP and the number is registered on the BSS in advance. The fourth requirement is for security reasons. User privacy must be protected in a multiple IDP situation.

### 3.1 Assumptions

The assumptions for our proposed scheme are described below. Fig. 6 illustrates the assumptions. In the figure, $SK_{BSS_1}$ is the private key of the BSS for IDP1. $PK_{BSS_1}$ is the corresponding public key of $SK_{BSS_1}$. $SK_{IDP_n}$ and $PK_{IDP_n}$ are public key pair of $IDP_n$ (n = 1, 2).

1) Users have user accounts on a BSS and login to the BSS with their ID and PW pairs.
2) Before making user's accounts on the BSS, the BSS checks each user credential. Therefore, the BSS knows the user's real identity.
3) The BSS and each IDP have a trust relationship with each other, so the BSS can validate the IDPs' signatures signed with their own private keys. In the other direction, IDPs can validate BSS's signatures.
4) The BSS prepares its own public key pairs for each IDP independently.

### 3.2 Account generation

To generate a user account on an IDP, the sequence is quite similar to that in the previous version. Initially, a user requests the IDP to create an account. Then the IDP generates a random number as the identifier for the account and passes it to the user; this is something like a sequence number. Fig. 7 shows the sequence for account generation.

1) A user requests $IDP_1$ to generate a user account.
2) IDP1 generates a random number $R_{IDP_1}$ and records its generation time to determine its subsequent lifetime. Then IDP1 sends the $R_{IDP_1}$.
3) The user generates a token ($B_{UK_{1n}}(T)$) with $R_{IDP_1}$ using blind function $B_{UK_{1n}}()$. (Whereas $T = R_{IDP_1}$, $UK_{1n}$ means the user's blind key for the n-th account on $IDP_1$)
4) The user logs in to the BSS with his / her user account on the BSS ($ID_{BSS}, PW_{BSS}$), informing the BSS that the user wants to make an account on $IDP_1$ and passes the token $B_{UK_{1n}}(T)$ to the BSS.
5) The BSS checks the token status of the user for IDP1. If the user's number of tokens on IDP1 does not exceed the generation limit, then the BSS generates a signature for the token. $S_{SK_{BSS_1}}(B_{UK_{1n}}(T))$ ($SK_{BSS}$ means the private key of $BSS_1$ for $IDP_1$). The BSS records the relationship among $ID_{BSS}, B_{UK_{1n}}(T)$ and $IDP(IDP_1)$. The BSS decreases the number of remaining token for $IDP_1$ and sends the signature $S_{SK_{BSS_1}}(B_{UK_{1n}}(T))$ to the user.
6) The user unblinds the token with function $B^{-1}_{UK_{1n}}$ and generates signed token $S_{SK_{BSS_1}}(T)$ and sends it to $IDP_1$.
7) IDP1 checks the status of random number $R_{IDP_1}$. If it is not expired, then the IDP validates the signature $V(T, S_{SK_{BSS_1}}(T))$. If it is valid, $IDP_1$ permits the user to create an account. The status of $R_{IDP_1}$ is changed to "used" to avoid the reuse of the random number.
8) The user sets his / her account on $IDP_1$.

### 3.3 Account deletion

The deletion procedure is also similar to that in the previous version of our proposed scheme. In the deletion procedure, the BSS generates a unique random number for identification. The account, which was generated by using the token B(T), is deleted, and the number of token generations that can be maintained on the BSS is again increased. The sequence of the deletion procedure is shown in Fig. 8.

1) The user logs in to $IDP_1$ with his / her account and password ($ID_{IDP_1}$ / $PW_{IDP_1}$) and requests account deletion from IDP1.
2) IDP1 searches for the random number $R_{IDP_1}$ in the data recorded from user ID ($ID_{IDP_1}$).
3) Then the IDP1 sends the $R_{IDP_1}$ to user. (Instead of doing so, the user can retain the random number.)
4) The user regenerates the token $B_{UK_{1n}}(T)$, which is used at account generation, by using blind function $B_{UK_{1n}}()$.

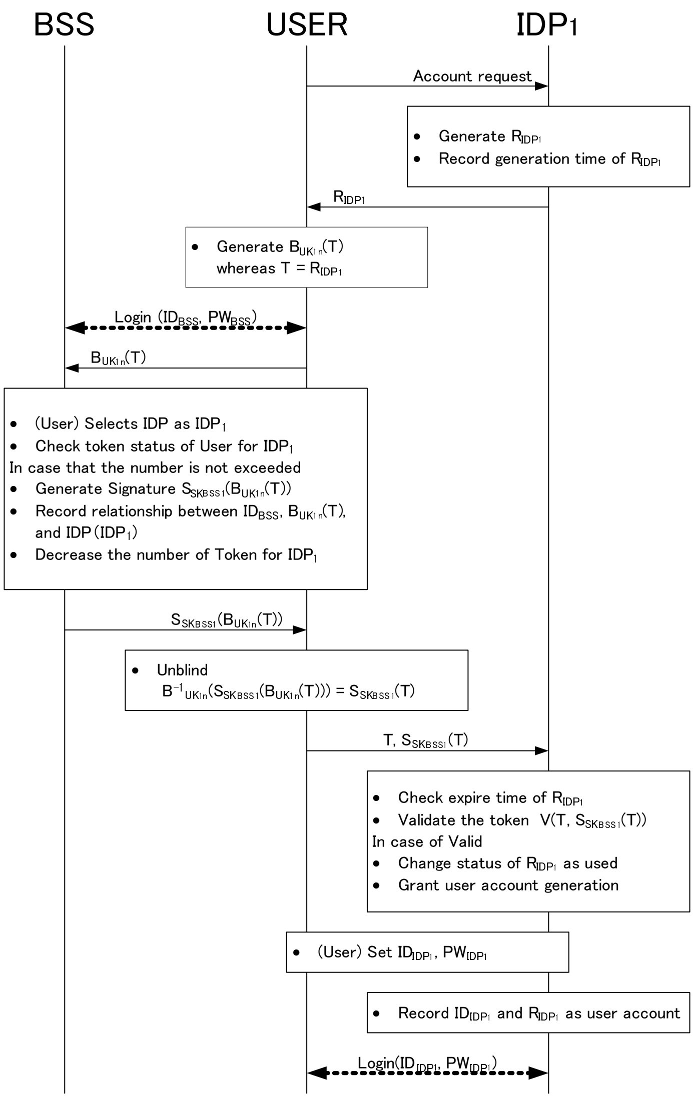

BSS
USER
IDP1
Account request
Generate $R_{IDP1}$
Record generation time of $R_{IDP1}$
$R_{IDP1}$
Generate $B_{UK1n}(T)$ whereas $T = R_{IDP1}$
Login ($ID_{BSS}$, $PW_{BSS}$)
$B_{UK1n}(T)$
(User) Selects IDP as $IDP_1$
Check token status of User for $IDP_1$
In case that the number is not exceeded
Generate Signature $S_{SKBSS1}(B_{UK1n}(T))$
Record relationship between $ID_{BSS}$, $B_{UK1n}(T)$, and IDP ($IDP_1$)
Decrease the number of Token for $IDP_1$
$S_{SKBSS1}(B_{UK1n}(T))$
Unblind $B^{-1}_{UK1n}(S_{SKBSS1}(B_{UK1n}(T))) = S_{SKBSS1}(T)$
T, $S_{SKBSS1}(T)$
Check expire time of $R_{IDP1}$
Validate the token V(T, $S_{SKBSS1}(T)$)
In case of Valid
Change status of $R_{IDP1}$ as used
Grant user account generation
(User) Set $ID_{IDP1}$, $PW_{IDP1}$
Record $ID_{IDP1}$ and $R_{IDP1}$ as user account
Login($ID_{IDP1}$, $PW_{IDP1}$)

Fig. 7: Account generation sequence in the proposed method

5) The user logs in to the BSS and requests the deletion of the account on IDP1. Then the user passes the blinded token $B_{UK_{1n}}(T)$ to the BSS1.
6) The BSS checks the status of the token, confirming whether the original sender of the token is the user and the token is already used. Then the BSS generates a random number $R_{BSS_1}$ and records the relationship among $ID_{BSS}$, $B_{UK_{1n}}(T)$, $R_{BSS_1}$, and $IDP_1$. Finally, the BSS sends $R_{BSS1}$ to the user.
7) The user generates a token for account deletion by using blind function $B_{UK'_{1n}}(D)$. (Where D = $R_{BSS_1}$). The user sends the deletion token $B_{UK'_{1n}}(D)$ to the IDP1. IDP1 checks the status of $R_{IDP_1}$ from user ID $ID_{IDP_1}$. If the status is "issued", then the IDP1 generates signature $S_{SK_{IDP_1}}(B_{UK'_{1n}}(D))$ and deletes user account $ID_{IDP_1}$. Then IDP1 changes the status of $R_{IDP_1}$ to "deleted" and records the relationship between $R_{IDP_1}$ and $B_{UK'_{1n}}(D)$. Finally, IDP1 sends the signed token $S_{SK_{IDP_1}}(B_{UK'_{1n}}(D))$ ($SK_{IDP_1}$ is $IDP_1$'s private key) to the user.

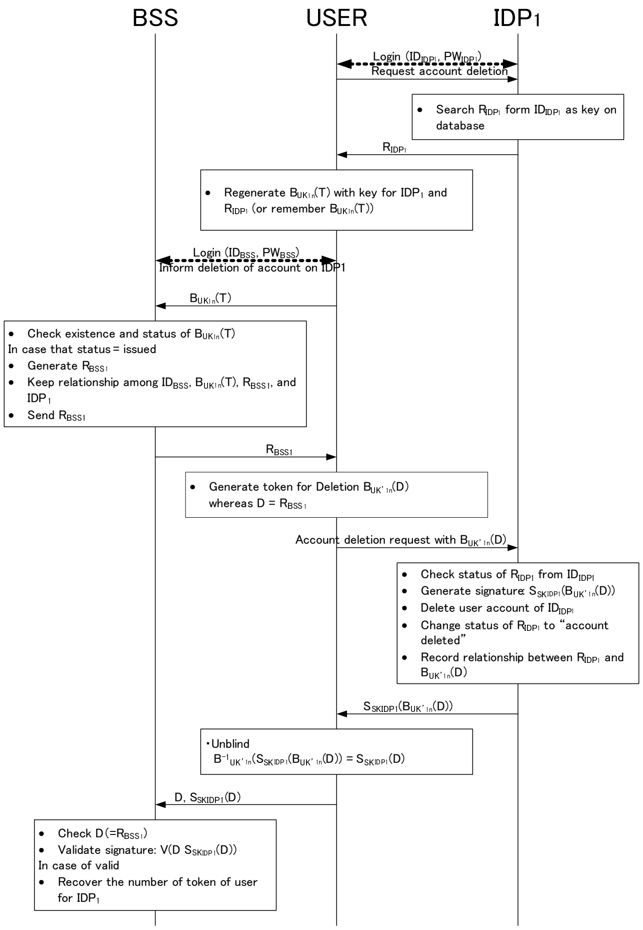

BSS
USER
IDP1
Login ($ID_{IDP1}$, $PW_{IDP1}$)
Request account deletion
Search $R_{IDP1}$ form $ID_{IDP1}$ as key on database
$R_{IDP1}$
Regenerate $B_{UK1n}(T)$ with key for $IDP_1$ and $R_{IDP1}$ (or remember $B_{UK1n}(T)$)
Login ($ID_{BSS}$, $PW_{BSS}$)
Inform deletion of account on IDP1
$B_{UK1n}(T)$
Check existence and status of $B_{UK1n}(T)$
In case that status = issued
Generate $R_{BSS1}$
Keep relationship among $ID_{BSS}$, $B_{UK1n}(T)$, $R_{BSS1}$, and $IDP_1$
Send $R_{BSS1}$
$R_{BSS1}$
Generate token for Deletion $B_{UK'1n}(D)$ whereas $D = R_{BSS1}$
Account deletion request with $B_{UK'1n}(D)$
Check status of $R_{IDP1}$ from $ID_{IDP1}$
Generate signature: $S_{SKIDP1}(B_{UK'1n}(D))$
Delete user account of $ID_{IDP1}$
Change status of $R_{IDP1}$ to "account deleted"
Record relationship between $R_{IDP1}$ and $B_{UK'1n}(D)$
$S_{SKIDP1}(B_{UK'1n}(D))$
Unblind $B^{-1}_{UK'1n}(S_{SKIDP1}(B_{UK'1n}(D)) = S_{SKIDP1}(D)$
D, $S_{SKIDP1}(D)$
Check D(=$R_{BSS1}$)
Validate signature: V(D $S_{SKIDP1}(D)$)
In case of valid
Recover the number of token of user for $IDP_1$

Fig. 8: Account deletion sequence in the proposed method

8) The user unblinds the token with the reverse function of $B^{-1}_{UK'_{1n}}()$ and sends the deletion token $S_{SK_{IDP_1}}(D)$ with $D$.
9) The BSS checks the $R_{BSS1}$ and validates the signature using IDP1's public key. If it is valid, then the BSS adds one to the number of user tokens for IDP1.

# 4. Discussion

Our new proposed scheme introduces more complex token management than the previous one, but the basic idea is same, so the first and second requirements described in third section are satisfied. In addition, the BSS holds and manages isolated public key pairs for each IDP. Therefore, nothing can link user accounts on different IDPs. As a result, the fourth requirement is also satisfied.

In our new proposed scheme, a user can request multiple accounts for one IDP. For this purpose, the BSS also maintains a number of tokens for account generation and counts the accounts that the BSS has already issued. When a user asks the BSS to generate a new token, the BSS checks to see that the number of issued tokens does not exceed the limit set by the relevant IDP. In that case, even though a user wants to obtain more accounts on an IDP, the user's request is refused by the BSS. Thus, the third requirement is also satisfied.

The main purpose of our proposed scheme is to limit the number of user accounts on one IDP, so the procedures have been designed for this purpose. Also, the requirement to protect user privacy is quite the same as in our previous scheme. In addition, in our scheme, in order to recover the token condition in the BSS, the BSS knows which IDP the users have set up their accounts in.

# 5. Conclusion

In this paper, the authors proposed an account management scheme that uses a blind signature method to protect user privacy. We modified our previous scheme and relaxed the limitation on the number of IDPs and the number of accounts on each IDP for each user. To provide this flexibility, we extended our scheme by introducing individual token management for each user and IDP in a BSS. By using a blind signature scheme, the accounts on the BSS and IDPs have no discoverable relation. Therefore, even if they colluded with each other, they could not uncover a user's identity. Therefore, user privacy is preserved. We believe that our proposal provides users and service providers a more secure, usable, and flexible account management method.

# References

[1] Ryu Watanabe and Yutaka Miyake, "Account Management Method with Blind Signature Scheme" *Engineering and Technology, World of Science, Issue 59*, pp. 2069-2073 (2011).

[2] Security Assertion Markup Language (SAML) V2.0, OASIS (2005), http://www.oasis-open.org/specs/index.php#samlv2.0

[3] OpenID Authentication 2.0 - Final, OpenID Foundation, (2007), http://openid.net/specs/openid-authentication-2_0.txt

[4] Arkajit Dey and Stephen Weis, "PseudoID: Enhancing Privacy in Federated Login," *Proc. of 3rd Hot Topics in Privacy Enhancing Technologies(HotPETs 2010)*, pp.95-107 (2010).

[5] David Chaum, "Blind signatures for untraceable payments," *CRYPTO,* pp.199-203 (1982).

[6] Whitfield Diffie and Martin E. Hellman, "New directions in cryptography," *Trans. on Information Theory, IEEE*, Vol. 22, Issue 6, pp. 644-654 (1976).

[7] Kazutusna Yamaji, Toshiyuki Kataoka, Motonori Nakamura, Tananun Orawiwattanakul, and Noboru Sonehara, "Attribute Aggregation System for Shibboleth based Access Management Federation," *in proc. of 10th Annual International Symposium on Application and the Internet (SAINT)*, *IEEE*, pp.281-284 (2010).

[8] Thorsten Hoellrigl, Jochen Dinger, and Hannes Hartenstein, "A Consistency Model for Identity Information in Distributed Systems," *in proc. of 34th Annual IEEE Computer Software and Applications Conference*, *IEEE*, pp. 252-261 (2010).

[9] Sriram Balasubramaniam, Grace A. Lewis, Ed Morris, Soumya Simanta, and Dennis B. Smish, "Identity Management and its Impact on Federation in a System-of-Systems Context," *in proc. of 3rd Annual IEEE International Systems Conference*, *IEEE*, pp. 179-182 (2009).

# Combating Social Engineering

## A DoD Perspective

Nathaniel D. Amsden
Department of Computer Science
Sam Houston State University
Huntsville, TX

Lei Chen
Department of Computer Science
Sam Houston State University
Huntsville, TX

*Abstract*— **Individuals and organizations are under increasing threats from social engineering attacks. The Department of Defense (DoD) is a lucrative target for malicious attackers, due to the sensitive nature of America's national security assets, tactics, techniques, and procedures. Attackers seek to access this information in whatever manner possible. The rise of social engineering attempts against DoD employees highlights the necessity of defeating social engineering attacks in order to maintain system integrity, thus protecting national security information. Good policies, education and awareness, and common sense defeat social engineering attacks. Formalizing and operationalizing social engineering benefits the DoD both offensively and defensively.**

***Keywords-social engineering; tailgating; social networking; security policy; authentication***

## I. Introduction

The Department of Defense (DoD) places heavy emphasis on the security of its installations, people, and information. Protection of classified information is of utmost importance, with heavy penalties levied against those who, unauthorized, disclose it to others. Social engineering is a threat that must be countered in order to ensure the security of DoD networks and information. Social engineers employ various tactics, techniques, and procedures (TTPs) in order to exploit unsuspecting victims. DoD security officers must understand the TTPs employed by social engineers in order to effectively defeat their attacks. Drafting and implementing policies designed to defeat social engineering attacks are crucial, but are only as effective as they are followed. If the DoD fully recognizes the threat of social engineering, it can operationalize it for offensive purposes. The DoD will gain a better understanding of how to defend against social engineering if it implements it offensively.

Employee education and training must be improved in order to effectively teach employees what social engineering is and how it can be prevented. Current social engineering resistance training is contained in a simple, short online Information Assurance (IA) training module. Social engineering is only briefly covered in the IA training. This training does not meet the needs to fully enable DoD employees to resist social engineering attacks. The difficulty with developing a plan to prevent social engineering lies in the fact that it deals with person-to-person communication. Malware can be blocked at firewalls or caught by anti-virus programs. How does an organization prevent spear-phishing emails, phone calls, or in-person conversations? This is why most social engineering defensive tactics rely heavily on organizational policies and employee education.

## II. Social Engineering Tactics

### A. *Psychological Triggers*

Various psychological triggers and traits of human nature, especially those ingrained into military and DoD culture, increase the likelihood of success when exploited by a skilled social engineer. These include strong affect, overloading, reciprocation, deceptive relationships, diffusion of responsibility and moral duty, authority, and integrity and consistency [1]. The DoD is based on a very rigid command structure. Authority, integrity, and consistency are important to this structure. Commanders, directors, and superiors give instructions and orders to subordinates. Subordinates are expected to execute all tasks given to them. Social engineers exploit this structure through impersonation and name dropping. Impersonating the help desk can trick victims into disclosing a username and password. Social engineers, with minimal knowledge of the military rank structure, can impersonate a higher ranking member. The social engineer convinces the victim the social engineer has authority over them. By pretending to have authority, low ranking members are tricked into giving up information to the social engineer.

Name dropping strengthens the impersonation tactic. Social engineers mention the name of the commander or someone else of importance in the organization. The victim is led to believe the social engineer was asked by the mentioned person, who has authority, to complete whatever the social engineer is asking. The tendency to follow orders of higher ranking people is a military strength, yet a weakness that can be exploited by those with malicious intent. Pretending to be someone else or simply schmoozing are typical examples of how social engineers work to obtain the information they need. They will often contact the help desk and drop the names of other employees. Once they have what they need to gain further access, they will attack a more vulnerable person – someone who has information but not necessarily the clout to challenge anyone of "authority" [2].

### B. *Phishing*

Phishing seeks to trick users into giving up information such as usernames and passwords. Phishers often say an account is about to expire and the victim needs to confirm their account information. Anti-phishing services and toolbars attempt to protect users from phishing attacks. Many users do not understand cues provided by anti-phishing tools or fraudulent websites indicating fake websites. Julie Downs et al recruited 20 people with computer experience, but without any computer security experience. The participants received information regarding a persona they were to portray and to read and react to several emails. Several emails were legitimate whereas the rest contained various forms of phishing attacks [3].

The participants were interviewed regarding their online behaviors and their perception of what made a website trustworthy. The participants reported having seen several cues that alert a user to be suspicious including spoofing "from" addresses (95%), broken images on web page (80%), unexpected or strange URL (55%) and https (35%). Participants identified three main strategies in making decisions about the emails. The strategies include "this email appears to be for me", "it's normal to hear from companies you do business with" and "reputable companies will send emails" [3].

All safety information presented by anti-phishing services and toolbars is relatively useless if the user does not know how to interpret it. Training employees to identify phishing emails is important and can protect the organization. For the DoD, this is often presented in the form of Computer Based Training (CBT). Information Assurance training is required annually, but employees often click through as quickly as possible in order to complete it and move on to "more important" work. Phishing is only briefly covered in the training.

### C. Fradulent Websites

Fake websites seek to lure DoD members into giving out usernames and passwords that can then be used on the real sites. Recently, a fake version of the Air Force Portal was launched. Air Force members seeking the Air Force Portal used Google to search for it. The fake Portal appeared among the top hits. Unsuspecting victims visited the fake Portal, entered their authentication information, and thus had their login information stolen [4]. Users must be careful and should avoid searching for specific websites and instead type in the link directly. An automatic method to detect fraudulent websites, much more than warning of invalid credentials, could be very beneficial to DoD users.

Malicious attackers, recognizing the fact the United States Automobile Association (USAA) banking institution is popular among military members, often create phishing schemes and fraudulent sites to lure military members into giving out login information. Not only do attackers gain access to their bank accounts, but also to military networks if the victims reuse passwords at work.

## III. Social Networking

### A. Social Networking Threats

The use of social media by federal employees is growing tremendously, supported by initiatives from the administration, directives from government leaders and demands from the public. With social media come the threats of spear phishing, social engineering and web application attacks [5]. Spear phishers rely on personal pieces of information about their target. Often, this information is readily available on social media websites. Social media bypasses traditional email security controls and allows attackers alternative methods to send phishing messages and gather information. Federal employees may identify themselves as employees of their department either by using their .gov or .mil email address or by intentionally listing information in their profile.

As their "friends" grow, the network of federal employees expands. Attackers need only to establish a relationship of trust with one person in order to gain a foothold to "friend" other federal employees, harvest info and conduct social engineering attacks. Additionally, enticing victims to install malicious applications on social media websites, such as Facebook, can compromise their account or download unauthorized software to their computer. This is especially risky when victims use social media from their work computers.

Other social engineering websites seek a military audience. They claim to be military only, but have no ties to the military. One in particular is owned by a German company, with a server based in Nova Scotia [6].

### B. The Robin Sage Experiment

The Robin Sage experiment sought to exploit fundamental levels of information leakage stemming from people's haphazard and unquestioned trust. At the end of the month-long experiment, the young, attractive (yet fake) "Robin" accumulated hundreds of connections on social networking sites. These connections included executives at government entities including the National Security Agency (NSA), DoD and Military Intelligence Groups. Much of the revealed information violated Operations Security (OPSEC) procedures [7].

Based on her listed job, many of her "friends" assumed she was trustworthy, having passed trusted government background checks and security clearances. By successfully "friending" renowned security experts, Robin's credibility soared allowing her to create more connections. Close assessments of Robin's profile indicate the false identity. By analyzing profiles and using a little common sense, people can keep themselves safe and not be "friends" with someone who does not exist.

Social engineers build relationships of trust with their targets on social networking sites. The victim trusts the social engineer and opens opportunities for further exploitation when the social engineer begins asking for information. The rise of social networking is a big concern for DoD leaders, as it opens up new attack vectors for social engineers. Social networking adds additional security, OPSEC, and IA concerns. This experiment proves the need for enhanced training regarding the dangers of social networking. It also proves that security is for everyone at all levels of organizations. It is not just for the average employee.

## IV. Combating Social Engineering

### A. A Multi-layered Defense Begins with Policies

Defense against social engineering must be multi-layered. Should one layer be penetrated, other layers are available to halt the attack. Security policies must set the foundation of defense and address social engineering. Combat strategies require action on both the physical and psychological levels. Employee training is essential [8]. Policies such as 100% shredding, no tailgating, and challenging others not wearing identification (IDs) aid security measures and deter social engineers.

100% shred policies greatly decrease all potential printed pieces of information that a social engineer could use to research the organization (and any potential secrets he could find!). Social engineers will dumpster dive, given the opportunity, in order to find any and all information that could be used to exploit others into giving him access to unauthorized systems.

Security policies must address a number of areas in order to be a foundation for social engineering resistance. It should address information access controls, setting up accounts, access approval, and password changes. It should also deal with locks, IDs, paper shredding, and escorting of visitors. The policy must have discipline built in and, above all, it must be enforced [1]. Policies should be reviewed at least every five years, with at least 20% in review each year [9].

### B. Eliminate Tailgating

Badges raise another issue. Everyone, including visitors, should wear access badges indicating status [10]. This helps reduce the threat of people overstating their authority. Some DoD units allow tailgating, that is, following someone through a controlled access door. The first person swipes his/her card and inputs his/her Personal Identification Number (PIN), gaining access. He or she then holds the door for people following, only verifying that they have an appropriate badge. If they are careful, they will also verify the picture looks like the person owning the badge. A social engineer can print a fake ID card to look exactly like the organization's standard ID cards. By following someone entering a building or secure area, it is possible to gain entry after the first person enters the appropriate security measures simply by flashing one's badge. Occasionally, the person checking does not even look at the badge or make sure the picture on the badge looks like the holder. It is humorous to note that security personnel will occasionally wear badges with a Mickey Mouse picture and attempt to tailgate into secure areas. It is a quick way to test employees to ensure they verify the picture on the badge matches the owner.

DoD facilities have the added benefit of multiple entry control points. Individuals must show identification to even enter the perimeter of the installation. This ensures some manner of affiliation with the DoD prior to getting close to restricted areas. Restricted areas then further require additional credentials and access controls in order for an individual to gain access.

By eliminating tailgating, everyone must display valid credentials to the entry control points. This eliminates the possibility of anyone sneaking in without proper authorization. The fourth-factor authentication method allows tailgating, but only by someone who knows and can vouch for the tailgater's access rights. Employees should challenge people walking around without a proper badge, even those people they recognize. They may have had access suspended without other employees knowing.

### C. Employee Training and Education

Security awareness training for all users can also mitigate social engineering attacks [8]. Key personnel should also be resistance trained. Resistance training includes inoculation, forewarning and reality checks. These outline potential social engineering attacks so personnel can recognize and resist them in the future. Inoculation gives employees weak arguments used by social engineers in order to warn them of possible methods of social engineering. Forewarning takes inoculation one step further. Employees are warned of coming attacks and also about the persuasive content of the argument. Reality checks seek to trick the employees, in a controlled manner, into becoming a victim of a mock social engineering attack. This helps them realize they are vulnerable, and puts them at a heightened sense of security for future, real attacks [1]. Understanding the attack vectors and psychological triggers social engineers use can greatly reduce the likelihood of a successful attack.

Many security programs focus on technical security and leave information vulnerable to basic espionage methods. OPSEC addresses processes that could compromise information through non-technical means. "Need to Know" information access helps prevent unnecessary proliferation of information. Other policies restricting the use of open communication lines reduce the potential for the compromise of information. Reporting questionable circumstances and activity can protect information [10].

One of the best methods for educating employees to these risks is to take social engineering stories from current events and post them on an internal web site, or use email for safety tips and informational stories. The security officer can also incorporate these stories into security awareness training sessions held for employees. The stories work like fables of yore, imparting information with a purpose. Telling authentic stories of what happened to the 'other poor guy' increases resistance to these exploits in a non-threatening way, inoculating the employee against a vulnerability to social engineering [9].

### D. Four Authentication Factors

In addition to the three common authentication factors, something you know, something you have and something you are, a fourth authentication factor, someone you know, proves a person's access rights [11]. Social engineers can easily spoof the "something you have" factor by creating a fake ID card or similar. The "something you know" factor is difficult to spoof, since it usually is a password or PIN and thus must be physically given to the person by the organization.

Social engineers seek to circumvent the something you have, something you know, and something you are authentication factors in order to gain access to the desired system or information. When communicating via email or telephone, the something you have and something you are authentication factors cannot be utilized. Something you know, such as a passphrase, can be utilized. If suspicious, ask the caller for a callback number. If they refuse to give one, red warning flags and alarm bells should sound in your mind.

Social engineers often impersonate the help desk or administrators. They call employees claiming they need the potential victim's username and password in order to fix a network issue. By simply calling the organization the caller claims to be from, such as the help desk, the fourth authentication method, someone who knows you, can vouch for the caller and authenticate him. Of course, if the help desk says no one was authorized to call you, then something is obviously wrong. Report the situation immediately.

### *E. Ontological Semantic Technology*

Autonomous systems exist that analyze semantic information from casual and unsolicited verbal and written output of a person of interest. The technology analyzes how the target says things and looks for contradictions in normal patterns of life or from previous statements to detect lies, possible cover ups, setbacks or any other number of possible problems. It is difficult to employ to counteract social engineering due to the brevity of a hit and the relative small amount of conversation pulled during that time. However, since social engineers typically overload conversations with insider terms and name dropping, the system can detect that and alert to a potential social engineering threat [12]. The DoD's culture is full of insider terms and acronyms. A person using these terms and acronyms fits into the culture, whether they are a member of the DoD or not.

This technology is better suited to track potential insider threats rather than social engineering. Yet oftentimes, insiders attempt social engineering against coworkers in order to gain access to additional and/or restricted information. This technology can be deployed in order to track the insider and warn potential victims to be wary while the investigation proceeds. Stringent background checks all but eliminate insider threats.

A way to check if someone is lying is to ask questions about a fake person or situation. By casually asking if they heard about "Bob at the help desk's" accident, the social engineer can be tricked into answering a question about someone who does not exist. If they make up any answer, they are obviously lying. Other ways to detect lying is through contradictions. Do they contradict themselves in conversation by saying different things? Ontological semantic technology can analyze conversations over periods of time and flag potential contradictions for further analysis.

## V. Formalizing and Operationalizing Social Engineering

Formalizing social engineering in cyberspace enables security specialists to not only understand the myriad of different tactics, but also to infer good defenses to prevent social engineering. Lena Laribee et al developed a trust model showing how a social engineer establishes relationships of trust with a victim. The attacker first gathers info, usually freely available, about the victim. The attacker uses this info to exploit three key characteristics of trust: ability, benevolence and integrity. In this way, the attacker seeks to convince the victim that the attacker is a trustworthy person with a need to know or do. Their proposed attack model describes how social engineering attacks are performed. It includes tactics such as friendliness, confidence, persistence, quick-wittedness, impersonation, ingratiation, conformity, diffusion of responsibility and distraction [13].

These models can greatly improve the ability to create countermeasures to social engineering. However, with the rapidly changing nature of cyberspace, the models either need to be generic enough to apply to most situations, or be constantly updated. The DoD has many sub organizations dedicated to modeling and simulation. If cyber operators, information assurance officers, and security officers research, develop, study, and implement trust and attack models, the DoD will be better positioned to understand and combat social engineering attacks.

Operationalizing offensive social engineering will benefit military operations and aid defensive strategies against social engineering. Social engineering is compatible with existing Air Force and Joint military doctrine [6]. In a cyber sense, the DoD does not utilize the offensive capabilities of social engineering to its full potential. "Weaponizing" social engineering provides the benefit of increased US military gain in the cyber realm and a better understanding of ways to defeat social engineering attacks against our organizations. Operational units deployed in theater utilize minor forms of social engineering to build friendly relations with local citizens. Troops build trust with the locals in order to mutually benefit both sides. This aspect of social engineering has no malicious intent, unlike social engineers who lie in order to obtain something through deception.

In order to effectively operationalize social engineering, a framework must be developed for measuring social engineering's effectiveness in the operational realm, training plans must be created, and TTPs must be developed [6].

## VI. Conclusion and Future Work

This paper explained various aspects of social engineering and how they affect the DoD. The military culture of the DoD makes its employees more vulnerable to certain tactics implemented by social engineers. The increase in social networking use among DoD employees creates new attack vectors that must be properly guarded. The DoD must implement a multi-layered defense to protect itself from social engineering. Establishing strong policies, including 100% shred policies and no tailgating, form the foundation of any defense strategy. They are in place in some DoD units, but not all. Strong policies are useless when they are not followed, thus employee education is a key step.

The current social engineering awareness training, only briefly and ineffectively covered in the Information Assurance

CBT, does not adequately train users to identify or combat social engineering attacks. The DoD would do well to develop better education and training methods to protect against social engineering. It should invest in developing social engineering attack models in order to better understand incoming threats and how to counter them. By taking it one step further and operationalizing social engineering, the DoD will be able to exploit enemy systems and defend against similar attacks.

We are developing a multi-phased research plan with the end goal of developing a comprehensive social engineering training program. First, we will develop a detailed questionnaire to gather data on general knowledge of social engineering tactics, techniques, and defenses. We will send this survey to members of various DoD units we have contacts at in order to get a broad variety of responses. Ideally, the organizational responsibility of each unit will be different, such as a communications unit, a test and engineering unit, a surveillance unit, a network security unit, etc. The different unit types will provide a breadth of experience and responsibilities, in order to get a diverse feel of the knowledge of the topic. Gathering enough data may pose a challenge, as the average response rate to surveys, at least within the Air Force, is about 10%. Those surveys are also shorter than we are planning. To help motivate those surveyed, we will inform them that the questionnaire is for a master’s degree project. Also, we will send it to units where we have contacts, who can assist us in distributing and explaining the purpose of the questionnaire. Hopefully this will help increase the response rate.

Second, we will analyze the responses to determine the strengths and weaknesses of the general knowledge of those surveyed with regards to social engineering. This analysis will enable us to determine which social engineering tactics the average user is more susceptible to.

Third, we will develop a social engineering attack model and a social engineering defense model. These models will be specific to nuances of the DoD organization.

Fourth, we will create a comprehensive social engineering defense training program. Information gleaned from the questionnaire will enable us to determine how best to develop the training program.

## References

[1] D. Gragg, (2002, Dec.). *A multi-level defense against social engineering* [Online]. Available: http://www.sans.org/reading_room/whitepapers/engineering/multi-level-defense-social-engineering_920

[2] T. R. Peltier, “Social engineering: concepts and solutions,” *EDPACS*, vol. 33, pp. 1-13, 2006.

[3] J. S. Downs, M. B. Holbrook and L. F. Cranor, “Decision strategies and susceptibility to phishing,” in Proc. *Symp. Usable Privacy and Security (SOUPS)*, Pittsburgh, PA, 2006, pp. 1-12.

[4] R. Boland, (2011, Dec.). *Military website spoofing is no laughing matter* [Online]. Available: http://www.afcea.org/signal/articles/templates/Signal_Article_Template.asp?articleid=2814&zoneid=254

[5] E. Crane, (2009, Sept.). *Guidelines for secure use of social media by federal departments and agencies* [Online]. Available: http://www.cio.gov/Documents/Guidelines_for_Secure_Use_Social_Media_v01-0.pdf

[6] B. E. Skarda, “Operationalizing offensive social engineering for the Air Force,” M.S. thesis, Dept. Elect. and Comp. Eng., AFIT, WPAFB, OH, 2008.

[7] T. Ryan, “Getting in bed with Robin Sage,” in *Blackhat USA 2010*, Las Vegas, NV, 2010.

[8] S. Granger, (2002, Jan 9). *Social engineering fundamentals, part II: combat strategies* [Online]. Available: http://www.securityfocus.com/infocus/1533

[9] W. Arthurs, (2001, Aug. 12). *A proactive defense to social engineering* [Online]. Available: http://www.sans.org/reading_room/whitepapers/engineering/proactive-defence-social-engineering_511

[10] I. S. Winkler, (2009, May 13). *Case study of industrial espionage through social engineering* [Online]. Available: http://csrc.nist.gov/nissc/1996/papers/NISSC96/paper040/WINKLER.PDF

[11] J. Brainard, A. Juels, R. L. Rivest, M. Szydlo and M. Yung., “Fourth-factor authentication: someone you know,” in *Conf. Comput. and Commun. Security*, Alexandria, VA. 2006, pp. 168-178.

[12] V. Raskin, J.M. Taylor and C. F. Hempelmann, “Ontological semantic technology for detecting insider threat and social engineering,” in *Proc. 2010 Workshop New Security Paradigms*, 2010 © ACM. doi: 10.1145/1900546.1900563.

[13] L. Laribee, D. S. Barnes, N. C. Rowe and C. H. Martell, “Analysis and defensive tools for social-engineering attacks on computer systems,” in *Information Assurance Workshop, 2006 IEEE*, West Point, NY, 2006, pp. 388-389.

# ISO 27001 Gap Analysis - Case Study

**Ibrahim Al-Mayahi, Sa'ad P. Mansoor**
School of Computer Science, Bangor University, Bangor, Gwynedd, UK

**Abstract**— *This work describes the initial steps taken toward the development of an Information Security Management System for the UAE e-government. To achieve this goal it was decided to obtain the ISO 27001 certification, which is the leading standard in information security. Gap analysis was performed on four selected organisations within the UAE e-government to determine their compliance against the ISO 27001 standards. This process will help identify the weakness in the existing system and highlight the any associated risks to the UAE e-government. In this paper a Management, Technical and Operational (MTO) model is presented. This model gives greater focus and provides a framework which is more aligned to the organisations structure and responsibilities. The results of benchmarking based on the ISO27001 standard, and the method used to measure the maturity level for each security control domain are presented.*

**Keywords:** Gap Analysis, ISO27001, Compliance, Information Security, Auditing

## 1. Introduction

Information security is critical for todays organisations, global exposure to threats means that they must protect themselves from external and internal threats. Wiander [1] describes the importance of building an information security management system based on ISO 17799 standards. The study concludes that there was internal resistance to change and this was due to lack of information available to the employees within the organisation. Valdevit et al. [2] shows that there is growing interest from SMEs to be ISO27001certified in order to improve their IT security and to achieve that; a suitable GAP analysis tool was developed.

In his paper Dey [3] describes the development of an information security system, and show that there should be proper analysis and design, involving the entire organisation, starting from the senior management to the end users. The conclusion was they all should take appropriate roles in establishing and implementing of an information security system within the organisation. Technology solutions need to be implemented appropriately to fight against threats and risks or to automate certain processes. Policies and procedures need to be established in order to define who will do what, when and how, in order to prevent the threats. A detection mechanism is required and once a thread is detected, it will take corrective measures to fix any damages. There should be a cultural change too within the organisation to deal with information and its security in general.

The concept of an e-government is to provide access to government services to the public sector (citizens and businesses) at anytime over an open network. This leads to issues of security and privacy in the management of the the information systems. To develop a secure e-government system the organisations involved, is required to have an ISMS (Information Security Management System). The reason to develop the ISMS for the UAE e-government was strategic decision agreed by the management board, to meet the following organisational requirements [4]:

- A central government requirement to develop the UAE e-government.
- The desire to meet various regularity requirements, particularly around computer misuse, data protection and personal privacy.
- To manage information more effectively for each organisation within the e-government.

To measure the risks of data misuse, loss, or disclosure organisations normally implement a number of security controls. In the UAE e-government case it was decided to use a suitable security standard as a benchmark. The objectives of the standard itself is to provide a model for establishing, implementing, operating, monitoring, reviewing and maintaining the information system, based on a business risk approach. Most organisations, in order to ensure compliance with the various regulations and corporate governance rules around securing key information, adopt the ISO27001 standard. The compliance assessments evaluate 133 controls of the requirements that are designed to achieve the 39 objectives of the standards within 11 key domains [5].

The objective of this research is to prepare the groundwork for the development of the UAE e-government, and to achieve that it is important to implement Information Security Management System. Four different organisations within the UAE e-government were used as case studies. All these organisations manage and operate their own information security, which implies that they are running an implicit Information Security Management System without a systematic risk assessment according to ISO27001.

There are some of main known risks that have been identified, but the problem is they are not properly classified and no methodology is implemented to deal with these risks. Furthermore, there is no document procedure to control the whole setup, which implies that there is no clear overview that exists about whether actions are taken or not. Thus, a risk management system must be implemented and the gap analysis is the first step toward achieving this.

## 2. GAP Analysis

Compliance is the process of comparing the applied controls of an organisation with those in ISO27001 in this case. The Gap analysis is a tool or a technique that enables an organisation to compare its actual performance with the standards [6]. It is different from risk assessment in the fact that it compares the object against some target (that could be desired performance level or standard), whereas risk assessment is not measured against a target. Both Gap analysis and risk assessment evaluate the answer to *"where are we?"* but in the case of Gap analysis it is measured against *"where we want to be"*.

For this study the four different organisations that were used as case study's were the core electronic service departements under UAE e-government. The gap assessment was initially carried out on the information that has been shared with section managers on a sample basis. Sample cases were taken in each of the areas to check the compliance to the standard. The next step was to assess the compliance of all the sections within the chosen departments. This was achieved by interviewing the relevant managers and their teams to have a clear picture of the business, reproducible results and consistency, together with the review of documentary evidence in order to verify the compliance level. Table 1 shows a list of the 11 key domains and the people responsible for each of them.

Table 1: List of interviewees for each Domain

| | Domain | Interviewee |
|---|---|---|
| A5 | Security Policy | Director & All Teams |
| A6 | Organisation of Information Security | Head of Electronic Audit |
| A7 | Asset Management | Head of Quality Management |
| A8 | Human Resources Security | Head of Network & Operations |
| A9 | Physical and Environmental Security | Head of Cyber Crimes Section |
| A10 | Communications & Operation Management | Local Branch |
| A11 | Access Control | Data Entry |
| A12 | IS Acquisition, Development & Management | Consultant |
| A13 | Information Security Incident Management | Database Specialist |
| A14 | Business Continuity Management | Management Team |
| A15 | Compliance | Legal Departement |

## 3. MTO Model

The ISO27001 eleven security domains do not provide insight into which group in the organisation is responsible for an activity. Thus, as part of this research a model based on the organisations structure was developed. This model provides greater focus and better understanding on where within the organisation the responsibility lies for each domain. The security domains are grouped into three categories based on responsibility:

- Management Controls, which include the following domains: security policy, organisation of information security and compliance.
- Technical Controls, which include the following domains: asset management, physical and environmental security and communications & operations management.
- Operational Controls, which include the following domains: systems acquisition, development & maintenance, access control, IS incident management and business continuity management.

This model provides a common language for all to view and manage information security activities. It could be considered as a framework for measuring and monitoring performance and integrating better management practices, which are more aligned to traditional organisational structure and responsibilities.

## 4. Maturity Model

The concept of maturity models is used regularly in the field of Information Systems as an approach for organisational assessment. Any systematic framework for carrying out benchmarking and performance enhancement that has continuous improvement processes, can be considered a maturity model. Generally, in the constituent literature, maturity implies perfect or explicitly defined, managed, measured, and controlled systems [7]. It is also a progression in the demonstration of a specific ability or in the accomplishment of a target from an initial to a desired end stage.

There are common mature modules available and these are NIST, CITI-ISEM, COBOT, SSE/CM and CERT/CSO and all these have between 5-6 levels of maturity, and for the purpose of this study it was decided to use the COBIT model, because it is focused toward auditing specific procedural awareness and adaptation [8],[9]. The COBIT Maturity Model is an IT governance tool was used to measure how well the management processes are developed with respect to internal controls. Such capability can be exploited by auditors to help management fulfil its IT governance responsibilities.

A fundamental feature of the model is that it allows the organisation to measure its current maturity level against a specific standard, in this case ISO27001. As a result, it can discover practical improvements to the internal controls of the IT system. The maturity levels are not a goal, but rather they are a means to evaluate the adequacy of the internal controls with respect to the e-government business objectives [10]. The model focuses on auditing specific procedures. This definition of maturity has several important characteristics:

- Provides the blueprint for a complete security program.
- Inform management of the order in which to implement security elements.
- Leads toward the use of best practice standards (in our case the ISO 20071).

This approach toward a detailed security maturity model (Security Program Maturity Model) takes a management systems approach. It involves the existence or non-existence of the 11 controls (domains) which comprise the ISO27001. A list of questions was used to capture the compliance of the organisation under different scenarios, and also to establish the maturity level for each of the 11 controls.

The maturity values are determined by the security requirements of the organization. During implementation two issues needed to be addressed the questions and their maturity values. This was resolved by designing the questions using the ISO27001 standard controls and carefully determining and agreeing on their maturity values (weight). Below is the list of agreed maturity values and their description:

- Nonexistence (0): there is no recognition of the need for internal control.
- Ad-hoc (1): there is some recognition of the need for internal control.
- Reputable but initiative (2): controls are in place but are not documented.
- Defined (3): controls are in place and are adequately documented.
- Managable and measurable (4): there is an effective internal control and risk management environment.
- Optimize (5): An organisation wide risk and control program provides continuous and effective control and risk mitigation.

To establish the initial maturity benchmark, the relevant staff where contacted from the four departments. Then, they were interviewed individually, and asked to answer the questions related to their domain.

## 5. Gap Analysis Results

The current levels of compliance against the principle of code of practices have been categorised using the following definitions:

- Compliant: The organisation is fully compliant with the specific are of ISO27001.
- Partially compliant: The organisation has gone some way towards being compliant, but still requires additional work to be undertaken.
- Non-compliant: The organisation does not have the controls in place to satisfy the requirement of ISO27001.

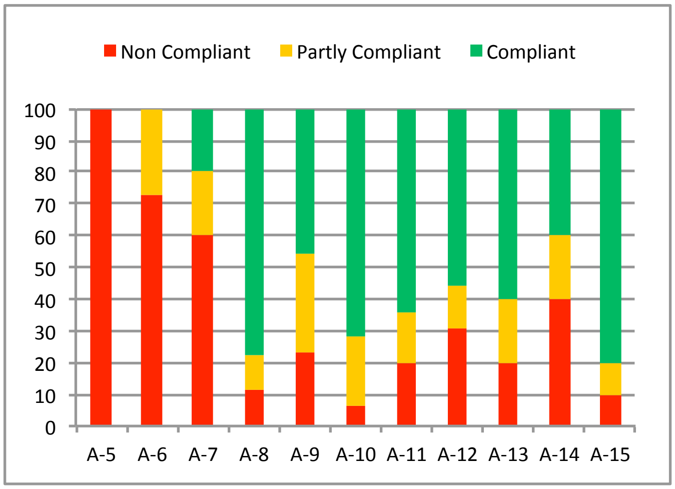


Fig. 1: Gap analysis compliance level

The result shown in Figure 1 indicates that some of the controls are more mature than others; it is evident that control A-5 show 100% non-complaince. This is due to the nonexistence of an approved security policy. It can be seen that the controls A-6 and A-7, exhibit high percentage of non-compliance, and once again this is due to the lack of implementation of effective security policy within these two controls. While the rest of the controls seems to have higher percentage of compliance and this is due to internal security procedure being put in place by the team responsible for each section.

The next analysis carried out was to identify the compliance of each section of the organisation based on the MTO model. The result is shown in Figure 2, and it is clear that the management has more than 60% non-compliance, this is primarily due to lack of information security policy. Meanwhile the technical and operation sections have higher percentage of compliance and this is due to internal security measures put in place.

Table 2 shows the compliance level for all the 133 requirement controls, and it can be seen that:

- 56.4% of the controls that were reviewed found out to be compliant with ISO27001 standards.
- 18.8% of the controls that were reviewed found out to be partly compliant with ISO27001 standards.
- 24.8% of the controls that were reviewed found out to be non-compliant with ISO27001 standards.

This indicates that there are large number of controls that meet the standard required, and bearing in mind that this is the first attempt to test the organisation compliance, it is quite an encouraging outcome.

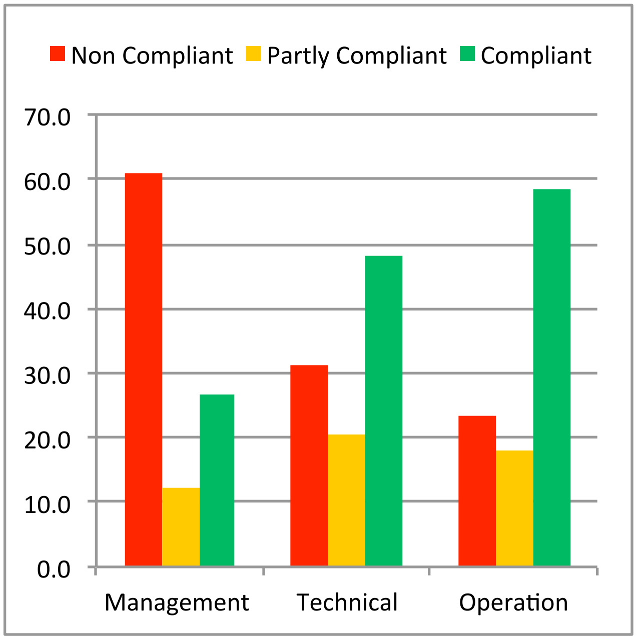


Fig. 2: MTO model results

Table 2: Domain compliant level

| Domain | Req. | Compliant | Partly compliant | Non compliant |
|---|---|---|---|---|
| A-5 | 2 | 0 | 0 | 2 |
| A-6 | 11 | 0 | 3 | 8 |
| A-7 | 5 | 1 | 1 | 3 |
| A-8 | 9 | 7 | 1 | 1 |
| A-9 | 13 | 6 | 4 | 3 |
| A-10 | 32 | 23 | 7 | 2 |
| A-11 | 25 | 16 | 4 | 5 |
| A-12 | 16 | 9 | 2 | 5 |
| A-13 | 5 | 3 | 1 | 1 |
| A-14 | 5 | 2 | 1 | 2 |
| A-15 | 10 | 8 | 1 | 1 |
| **Total** | **133** | **75** | **25** | **33** |

Figure 3 displays the results of the maturity benchmarking against ISO27001, and the scores used for benchmarking are explained below:

- Maturity score below 1.65: The organization should start implementation of overall security measures.
- Maturity score between 1.66 and 3.25: The organization has taken significant steps to enhance security.
- Maturity score above 3.26: The organization fulfils defined measures, thus the probability of high risks is marginal.

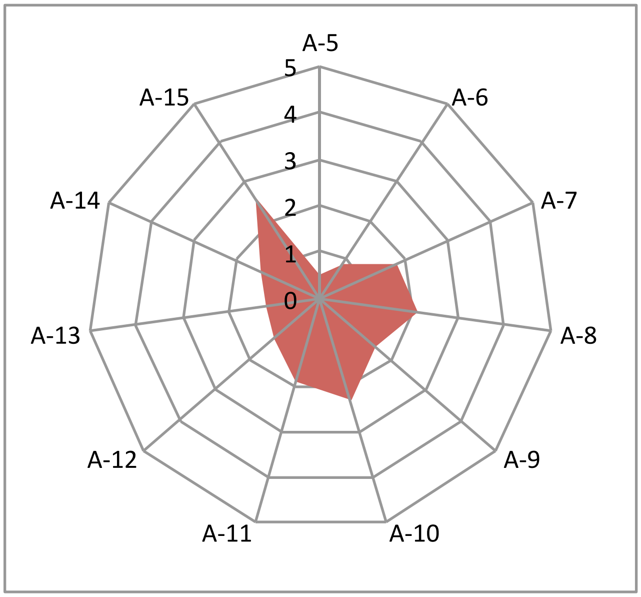


Fig. 3: Maturity benchmarking for each domain

It is obvious that some of the controls are more mature than others, for example the compliance, communication human resources security and asset management score lies between 1.67- 3.25. This implies that work has been done to improve the security of the organisations involved in this research. In the meantime there are some controls that lie in the reign below 1.65, which implies that the operation is dependent on knowledge and motivation of individuals, many control weaknesses exist and are not adequately addressed. Employees may not be aware of their responsibilities, and action is required to improve the security of these controls.

## 6. Conclusions

An information security management system is an integral part of an organisations management it is required to monitor, review and improve the information security of the organisation. It is a continuous process that deals with security policy development, and put procedures in place to deal with security threats. The gap analysis is initially used to identify the weaknesses in the organisations procedures. This should be a continuous process, as the organisation is required to be revisited to update the gap analysis. This is carried out to ensure long term protection against security breaches.

The security that can be achieved through technical means is limited, and should be supported by appropriate policies and procedures. Identifying which controls should be in place requires careful planning and attention to detail. Information security management requires, as a minimum, participation by all stockholders including, employees, suppliers, third parties and other external parties.

## Acknowledgements

The authors would like to thank the UAE Ministry of Interior and the UAE Ministry of Higher Education for supporting this work.

## References

[1] T.Wiander, “Implementing the iso/iec 17799 standards in practice: experience on audit phases,” in *Proc. Australian Information Security Conference (AISC2008)*, 2008.

[2] T.Valdevit, N.Mayer, and B.Barafort, “Tailoring iso/iec 27001 for smes: A guide to implement an information security management system in small settings,” in *Proceeding of the 16th European Systems & Software Process Improvement and Innovation Conference*. Springer Berlin Hiedelberg, 2009.

[3] M.Dey, “Information security management - a practical approach,” in *Proceeding AFRICAN 2007 Conference*, 2007.

[4] A.Calder, *Information Security Based on ISO 27001/ISO 17799: A Management Guid*. Van Haren Publishing, 2006.

[5] *Information security management systems requirements*, International Standards ISO/IEC 27001 Std., 2005.

[6] B.Karabacak and I.Sogukainar, “A quantitative method for iso 17799 gap analysis,” *Computers and Security Journal, Elsevier*, vol. 25(6), pp. 413–419, 2006.

[7] A.Pederiva, “The cobit maturity model in a vendor evaluation case,” *Information systems Control Journal*, vol. 3, 2003.

[8] S. Woodhouse, “An isms (im)-maturity capability model,” in *IEEE 8th International Conference on Computer and Information Technology Workshops*, 2008.

[9] C.S.Leem, S. Kim, and H.J.Lee, “Assessment methodology on maturity level of isms,” *Knowledge-Based Intelligent Information and Engineering Systems,*, vol. Pt 3, Proceedings. vol. 3683 Springer-Verlag Berlin, pp. 609 – 615, 2005.

[10] S. B.Tuttle, “An empirical examination of cobit as an internal control framework for information technology,” *International Journal of Accounting Information Systems*, vol. 8, pp. 240 – 263, 2007.

# The Zachman Framework, the Owner's Perspective & Security

**L. Ertaul[1], V. Rathod[1]**
[1]Mathematics and Computer Science, CSU East Bay, Hayward, CA, USA

**Abstract –** *The Zachman Framework is one of the oldest Enterprise Architecture Frameworks. It is a unique approach to provide a logical understanding of ever increasing size and complexities of information systems. This paper aims to introduce the Zachman framework in general. Also this paper aims to provide details about the Owner's perspective of the Zachman Framework. This paper also address the security requirements for the Owner's Perspective of an enterprise.*

**Keywords:** Enterprise Frameworks, Enterprise Security Planning, the Zachman Framework.

## 1 Introduction

Nature of business has changed in the past half century. Most businesses have grown from region-specific to global. Goals of business, business processes, supplier chains and business economics have changed from simpler to complex. With globalization, these caused rapid changes in organizations and in their structure. Businesses are now also focusing and relying more on Information Systems. It became more obvious and evident to have a more structured system and architecture for information flow and its integration with business.

The Zachman Framework precisely addressed these requirements and provided logical structure for such a flawless integration. This paper is logically divided into 4 sections. In the first section, the history of the Zachman Framework is introduced. In the second section Zachman Framework is explained. The third section talks about the Owner's perspective in more detail and the following section analyzes general security requirements from the Owner's perspective. The next section discusses the criticisms to the Zachman Framework. And finally, conclusions are given.

## 2 History of Zachman Framework

The Zachman Framework has evolved over time. Although the framework structure still remains the same, it has changed on the graphical representations to gain more generalization and logical representation [1]. This section provides noticeable events in the history of the Zachman Framework.

**1984**: John Zachman created first draft of the Framework. It had only 3 columns. This framework was titled as "Information Systems Architecture". This was a composite model. He used Chen, Bachman, and IMS Root-Segment diagram.

**1987**: The Zachman Framework was first published in the 1987 IBM Systems Journal. It still represented Information Systems; hence it contained only 3 columns.

**1992**: It is still referred as the Zachman Framework for Information Systems Architecture. It was published in IBM Systems Journal. It revolutionized the Information Systems concept that existed at this time. John added the words "Planner", "Owner", "Designer", "Builder", "Sub-Contractor" to Rows 1, 2, 3, 4, and 5 for clarification.

**1993**: It has only 3 columns. It has used the adjectives "Contextual", "Conceptual", "Logical", "Physical" and "Out-of Context" defining the Rows.

**2001**: This version was widely distributed.

**2002**: Updated representation of The Framework to make it more attractive aesthetically. It still contained Information Systems terminology, adjectives.

**2003**: This version is probably the most widely distributed version of the Zachman Framework.

**2004**: This version is also known as the Zachman Framework 2. Earlier versions used Information systems terminology, while this one uses Enterprise Architecture terminology. This version further moved Enterprise Architecture out of the I/T domain and shifted it back into the business domain.

**2008**: This version is the current and most accurate version of the Zachman Framework.

## 3 The Zachman Framework

John Zachman published a new approach towards system development. Traditionally business processes are represented as a series of steps. Zachman proposed a new way for representing these processes. He organized them around the points of view taken by the various participants. [1][2].

The Zachman Framework provides a comprehensive approach towards the Enterprise Architecture. It tries to classify various aspects of business with distinct point of views. This makes it a two dimensional matrix to collect facts, to help make and justify decisions.

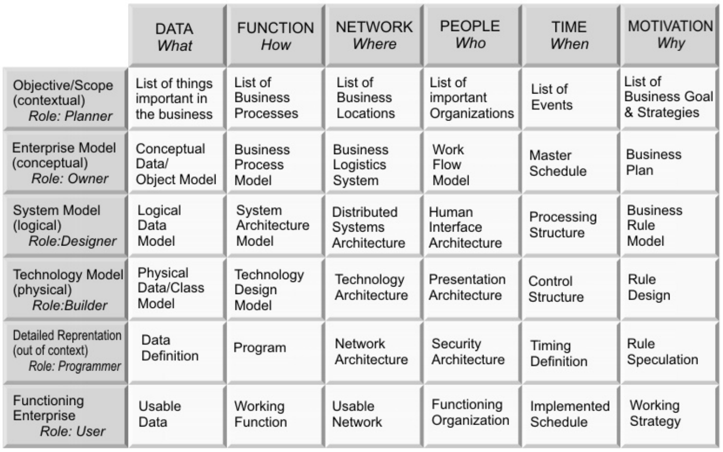
The framework is depicted in figure 1 [1][2].

The framework has 6 different views (Perspectives); each one of the perspectives is depicted as a row in the framework. Each perspective addresses various aspects depicted here as columns.

There are six perspectives.

1. The **Planner's Perspective** represents viewpoint of the group who has undertaken the business in a particular industry. The planners define the scope of the work to be done. It is usually high level abstract information.

2. The **Owner's Perspective** represents the viewpoint of the group who are business owners. Once the planner defines the scope for each of the aspect, the Owner provides some more details about business specific things. This would provide raw data points for the Designer.



| | DATA *What* | FUNCTION *How* | NETWORK *Where* | PEOPLE *Who* | TIME *When* | MOTIVATION *Why* |
|---|---|---|---|---|---|---|
| Objective/Scope (contextual) *Role: Planner* | List of things important in the business | List of Business Processes | List of Business Locations | List of important Organizations | List of Events | List of Business Goal & Strategies |
| Enterprise Model (conceptual) *Role: Owner* | Conceptual Data/ Object Model | Business Process Model | Business Logistics System | Work Flow Model | Master Schedule | Business Plan |
| System Model (logical) *Role:Designer* | Logical Data Model | System Architecture Model | Distributed Systems Architecture | Human Interface Architecture | Processing Structure | Business Rule Model |
| Technology Model (physical) *Role:Builder* | Physical Data/Class Model | Technology Design Model | Technology Architecture | Presentation Architecture | Control Structure | Rule Design |
| Detailed Reprentation (out of context) *Role: Programmer* | Data Definition | Program | Network Architecture | Security Architecture | Timing Definition | Rule Speculation |
| Functioning Enterprise *Role: User* | Usable Data | Working Function | Usable Network | Functioning Organization | Implemented Schedule | Working Strategy |

*Figure 1 The Zachman Framework*

3. The **Designer's Perspective** represents the viewpoint of the group who are systems analysts and wants to represent the business in a disciplined form. The Designer provides logical structure to the raw data points by defining the relevance to these data points. This perspective becomes an architect of the whole enterprise.

4. The **Builder's Perspective** represents viewpoint of the group who implements specific technologies to solve the problems of the business. Once the designers have provided designs / architecture, builder implements the design. In other words, builders are responsible for translating the design into reality.

5. The **Sub-contractor's Perspective**, represents the viewpoint of the group who are hired to do certain specific tasks. Sub-contractor's perspective depends on the builder's perspective. In various cases, domain specific expertise is required for implementation. Sub-contractors are used in those cases.

6. The **Functioning** Enterprise is the system itself. Once all these operations are done, the result is a functioning enterprise.

Each of these perspectives has 6 different aspects, depicted as columns. According to the Zachman, "**What**", "**Why**", "**How**", "**When**", "**Who**" and "**Where**" provides a complete understanding of the subject.

1. **"What" or "Data" column** addresses the understanding of the enterprise data.

2. **"How" or "Function" column** describes various processes involved in dealing with the "Data" columns.

3. **"Where" or "Network" column** describes geographic locations and logistics between the entities.

4. **"Who" or "People" column** describes the people participating in the organizational activities.

5. **"When" or "Time" column** describes when the "function" should be performed.

6. **"Why" or "Motivation" column** describes the end goals, constraints, rules and regulations.

## 3.1 Rules of the Zachman Framework

The Zachman framework defines some rules to alleviate effectiveness of the framework. Following is the list of rules with a brief description. [2]

### 3.1.1 *Do Not Add Rows or Columns to the Framework*

Who, What, When, Where, Why and How are the six primitive interrogatives. According to linguistics, answers to these questions could provide a comprehensive understanding about a subject or an object. Hence all of them are required. In this paper, they are also referred as aspects. Similarly each subject could be explained with 6 different perspectives, depicted as rows. Adding or removing them would either create duplicates or discontinuities. Hence the first rule states that framework will NOT be modified..

### 3.1.2 *Each Column Has a Simple Generic Model*

Each column describes a single and independent aspect of the Enterprise. Therefore the basic model for any of the columns is simple and generic.

### 3.1.3 *Each Cell Model Specializes Its Column's Generic Model*

As each of the columns has a simple and generic model, each cell tends to provide information or perspective that is specific to the row. Therefore each cell model specifies the generic model for each column.

### 3.1.4 *No Meta Concept Can Be Classified Into More than One Cell*

In the Zachman framework, each row is unique and so is each column. Therefore each cell is unique. Each meta-concept will be specific to the cell; therefore it is logical that none of the meta-concepts can be classified into more than one cell.

### 3.1.5 *Do not Create Diagonal Relationships Between Cells*

The Framework is described in plain English. Each perspective defines its own semantics for the aspects or columns. Therefore creating diagonal relationships could lead to semantically in-complete communication. This could lead to big disasters and hence there must not be any diagonal relationships.

### 3.1.6 *Do Not Change the Names of the Rows or Column*

Each name has a semantic meaning, changing names would, in effect, change the meaning for the row or column. In that case, the framework would not be a Zachman Framework anymore.

Naming should be as follows,

For *Generic Frameworks* rows should be named as *Scope*, *Owner*, *Designer*, *Builder*, *Out-of-context*, *Product*. And Columns should be named as *What, How, Where, Who, When, Why*.

For *Enterprise Specific Framework* rows should be named as *Scope*, *Models of the Business*, *Systems Models*, *Technology Models*, *Detailed Representations*, *Functioning Enterprise*. And columns should be named as *Data*, *Function*, *Network*, *People*, *Time*, *Motivation.*

### 3.1.7 *The Logic is Generic and Recursive*

The Framework itself is generic enough to classify descriptive representation of anything and therefore it is enough to analyze anything relative to its architectural composition.

# 4 The Owner's Perspective

Row 2 in the Zachman Framework depicts the Owner's perspective. This perspective is very important because it is defined by the business people who run the organization. This perspective provides a high level design and organization of the enterprise. The Idea here is that the Owner works closely with the planner to provide a high level description of the organization and core guidelines for the business.

The Owner's perspective defines following details for different aspects [3] [4].

In the "WHAT" column the Owner defines the requirements of important data points or documents for the organization. The comprehensive list of these data points provides a Semantic Data Model for the enterprise. Based on this model, a data audit model could also be developed.

In the "How" column the Owner defines the process of the enterprise. Essentially this is the place the Owner defines the Business Process. It is the core part of an organization. It uses the data points defined in "WHAT" column to produce a more relevant output. This may be used to generate dependencies and to trigger processes. All the processes could be classified with the level of criticality.

In the "WHERE" column the Owner defines the location of the business entities. These entities may exist in geographically diverse locations. This leads to the most critical "communication" part of the business. The Owner must consider various scenarios before defining the locations for the entities.

In the "WHO" column the Owner defines roles and responsibilities attached to each person. Here clear distinction between working units are materialized into various departments. This is a generic guideline for defining more granular roles and responsibilities by lower level rows (i.e. Designers & Builders). Typically this aspect addresses the human resources within the enterprise by creating an organization chart. This chart provides information on the desired flow of work-related responsibilities within the enterprise by clearly outlining the characteristics of who does what work.

In the "WHEN" column, the Owner describes the time dimension within the enterprise. Typically this is the place where the Owner defines 'what activities could occur in what sequence?' Typically this is called 'Corporate Calendar'. In all the enterprises there are certain processes which are time critical. The Owner also defines the Master Schedule based on the business processes. This master schedule contains sequences of major events in the business process.

In the "WHY" column, the Owner describes the corporate ethics and business competition, which play a central role in enterprise decisions. Apart from that, this is the place where the Owner defines enterprise-wide standards in order to have a complete control over the quality of the outputs. This also defines the various industry-wide constraints and justifications of various processes. It is also referred as "Constraints" column.

# 5 Security and Owner's Perspective

The Zachman Framework defines as many as six distinct aspects for each of the perspectives; however it fails to clearly define the security requirements for businesses. In today's world, security must not be underestimated. It is as important, if not more, as the other six aspects. So it is imminent to address security requirements of businesses.

If we consider security as an aspect of each perspective, we could actually add an extra column in the framework. This is in direct contradiction with rule 1 "*Do not add rows or columns to the framework*". If we consider security as a new perspective, there are two major problems with that: First it contradicts with rule 1, secondly it is not logical to view security as a new perspective altogether. This could pose a big question about the holistic approach of the framework.

In both cases above, we considered security as supplement to the framework, which is not viable. So the next question is: Could we consider security as an integral part of the framework? The idea here is to consider security parallel to each Meta concept. There are no rules denying that.

So in this section we will add security parallel to the Meta concept of each of the cells in the Owner's perspective.

In the "WHAT" column the Owner has defined the important data points. Security could be added by classifying this data into various security categories or levels. For example, let's say there are three security categories for data: *Highly sensitive*, *Sensitive* and *Public*. *Highly sensitive* data is available to only few people across the enterprise and is stored securely. *Sensitive* data is available to everyone in the enterprise but not outside of the enterprise, and is stored securely. Public data is available to everyone. Based on these categories various security services could be included. For example, *authorization*, *access-control*, *non-repudiation*, *confidentiality* and *availability*.

In the "HOW" column the Owner defines the process security. This cell defines the core business processes. So securing the processes will guarantee the robustness of the process and will provide fail-safe measures. For example, the validation of the inputs of the processes. If the input is faulty, it is hard to validate the output from a process. In business scenarios, one process (say p1) may depend on another process (p0); failure of the process p0 could lead to failure of p1. To handle such scenarios defining access-control and authorization on processes could help modify the processes or the sequence of the processes.

In the “WHERE” column the Owner has defined the location of the business entities, communication channels and logistics. The security here would be to ensure that there is no disconnection among business entities, physical security of the locations, buildings and logistics under any conditions. For example, avoiding building facilities in seismically active zones; providing strong enough infrastructure to withstand highest predictable quakes; providing redundant wired and/or satellite communication channels to ensure connectivity; providing multiple logistics channels. Also there must be backup entities that could be used in case of complete failure of one entity.

In the “WHO” column the Owner has defined the roles, privileges and responsibilities of each person. Each department has specific functions and has various access control mechanisms in place. This whole setup would provide a structure in the organization. The strength of the enterprise is as good as its structure. This affects the access control and the authorization of “WHAT” column, the robustness of the processes in “HOW” column and the logistics between entities in “WHERE” column.

In the “WHEN” column the Owner has defined time dimension with respect to the availability of data points, processes, network setup, people recruitment and various deadlines of business objectives and constraints. In general it’s called a corporate calendar. The security in a corporate calendar is to assess the risks associated with each time line and mitigation of those risks in worst cases.

In the “WHY” column, the Owner has defined a logical reasoning for business decisions. The security here should be the probabilistic validity of the decisions and mitigation in case of failures.

Based on these security requirements the Designer will define a security structure. The precise definitions for security in the Owner’s perspective will increase the robustness of this overall security structure of the enterprise.

## 6 Criticism

Although the Zachman Framework provides a perfect tool for classification of artifacts and delegation of responsibilities, it fails to provide any step-by-step process for building the reference model and Enterprise Architecture [5]. The Zachman Framework is a generic framework and does not add value to the business objectives. For businesses, cost is one of the major decision-making factors, the framework fails to address this [6]. Enterprises in general have become global now. This led to rapid dynamic changes in organization. The Zachman Framework lacks the agility to handle these rapid changes.

Other frameworks may include DoDAF (Department of Defense Architecture Framework), TOGAF (The Open Group Architecture Framework), FEAF (Federal Enterprise Architecture Framework) etc. Each of these architecture frameworks have been designed to be used in specific cases and hence provides value addition to the process of Enterprise Architecture.

DoDAF [7], has a specific use in defense department. DoDAF v2.0 provides as many as 52 models classified into 8 different views and a meta model that helps in choosing the specific model.

TOGAF [8], is process centric framework. It is a detailed method and set of supporting resources for developing an Enterprise Architecture.

FEAF [9], is a conceptual model that defines a documented and coordinated structure for cross-cutting businesses and design developments in the Government.

## 7 Conclusion

The Zachman Framework provides a holistic view of the enterprises. It provides a natural approach towards understanding the Enterprise Architecture. However it does not consider rapid changes on the enterprises. It provides a complete classification for enterprises but fails to provide any value addition to them. As far as security is concerned, it is shown that decisions taken by the Owner will impact the whole security of the enterprise.

# 8 References

[1] The Zachman Framework™ Evolution, John P. Zachman, April 2009. http://old.zachmaninternational.com/index.php/ea-articles/100-the-zachman-framework-evolution

[2] John Zachman, The Zachman Framework For Enterprise Architecture: Primer for Enterprise Engineering and Manufacturing, 2003 http://www.businessrulesgroup.org/BRWG_RFI/ZachmanBookRFIextract.pdf

[3] L. Ertaul, R. Sudarsanam, "*Security Planning Using Zachman Framework for Enterprises*", Proceedings of EURO mGOV 2005

[4] The Framework for Enterprise Architecture Cell Definitions, https://apps.adcom.uci.edu/EnterpriseArch/Zachman/ZIFA03.pdf

[5] Roger Sessions, A Comparison of the Top Four Enterprise-Architecture Methodologies, 2007 http://msdn.microsoft.com/en-us/library/bb466232.aspx

[6] Graeme Simsion, What's Wrong With The Zachman Framework? 2005 http://www.tdan.com/view-articles/5279/

[7] DoDAF, http://cio-nii.defense.gov/sites/dodaf20/

[8] TOGAF, http://www.opengroup.org/togaf/

[9] FEAF, http://www.cio.gov/documents/fedarch1.pdf

# Analytical impact of the Security Techniques on Information System for Decision Making

**Tejpreet Singh[1], Parminder Singh Nandha[2], and Kanwaljit Singh[3]**
[1]Deptt. of Comp. Sc. & Engg. Khalsa College of Engg. & Tech. Amritsar. Punjab. India
[2] CQ University Australia. CQ University Australia. Carnia. Australia
[3] Deptt of Mngt Studies and Comp. App.. Amritsar College of Engg. & Tech. Amritsar. Punjab. India

***Abstract**: Data in Information System (IS) are at risk from human errors, technical errors, and malicious attacks. Illicit information revelation, loss of integrity and data tampering in an IS can be sheltered by technical or manual firewall. The collection of data from internal or external sources is done via network. On the network data are transformed in packets or frame. To protect data from unauthorized users, distressful attacks and harmful viruses, security is a key element that should always be on the network. The priority of the Information Technology profile now in current scenario is to provide valid, accurate and undistracted data to the organization for better decision making and system growth. Various security techniques can be used to protect the data for being pirated. In this paper, the various security techniques and harmful attacks will be discussed and a detailed data analysis will be done. This paper will be beneficial for the organizations to make the data secure for making the decisions more effective and bring benefits.*

***Keywords:** Information System, Management Levels, TCP/IP Network, Security, Vulnerable attacks.*

## 1. Introduction

*IS are used to capture, create, store or distribute classified information.* Confidentiality is the ability to conceal messages from a passive attacker so that any message communicated via the sensor network remains confidential. Authentication ensures the reliability of the message by identifying its origin. Data integrity in sensor networks is needed to ensure the reliability of the data and refers to the ability to confirm that a message has not been tampered with, altered or changed. Availability determines whether a node has the ability to use the resources and whether the network is available for the messages to communicate. All the levels or layers of the management in an organization make the operations on the basis of the data or information that is created, stored and distributed in the Information System. The main goal of data collection and sharing is to make better decisions for the growth of the organization. The decision makers can be at different locations and so as the information, and for the sharing of data, the data flow in either unified network or interconnected network. A unified network is a collection of Information Systems (IS) that are accredited as a single entity. An interconnected network is comprised of separately accredited ISs and/or unified networks, each of which maintain its own intra-IS services and control. The security is applied at internal and external boundaries of the IS

The information system has the internal and external security control system. As the data flows and shared within and outside the information system. The mishandling of data at network environment may make a terrific affect of the decisions either for long run or short run. The security system has to make a check on the data flowing on the network within the information system or outside the IS. While importing or exporting the data between the strategic – tactical-operational layers of the management, or collecting the data from the external environment, the data has to pass through the network model of four layers called TCP/IP model i.e. Transmission Control Protocol Internet Protocol model. The data communication process starts from the Application layer to Link layer of source end, transfer through the network and ends at the Link layer to Application layer of destination end. Fig-2 shows a complete flow of data from source to destination. Each layer of the TCP/IP model performs different functions. The Link layer transmits and receives packets of information reliably across the Information System.

Switches, bridges are used to transmit and receive the data. The Link layer provides error control and flow control of data. A MAC address role comes into play. The Internet layer handles the data routing and forwarding. The Internet layer routes data through various IS networks while travelling to a known host. Routers are machines that decide how to send information from source to destination. Transport layer ensures the reliable arrival of message and provides error checking mechanism and data flow controls. The Application layer defines the standard for interaction at the user or application program level, such as provide GUI to communicate between user and computer or server. To make data communication efficient for the IS, the data flows in the network in form of packets. The packet is divided into two parts: ‘Source address and Destination addresses and ‘encrypted data and identification number.’ When large amount of data are in electronic form, they are vulnerable to many kinds of threats through communication network and information system. The potential of unauthorized access, fraud are not limited to a single location but can occur at any access point in the network of information system. Vulnerabilities

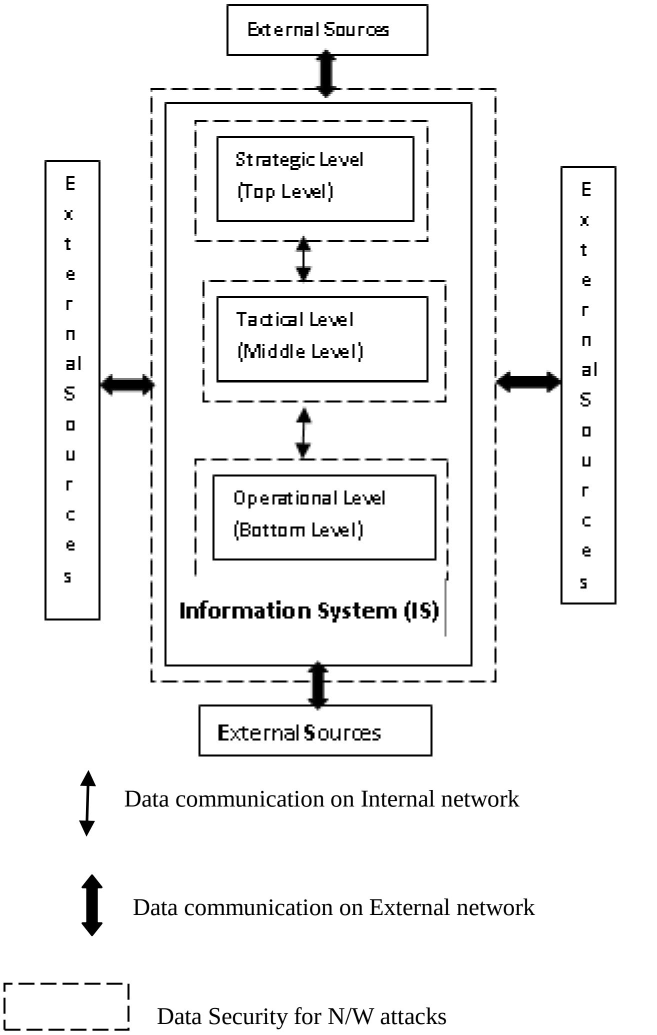


**Fig 1**:- IS with security control from Internal and External attacks

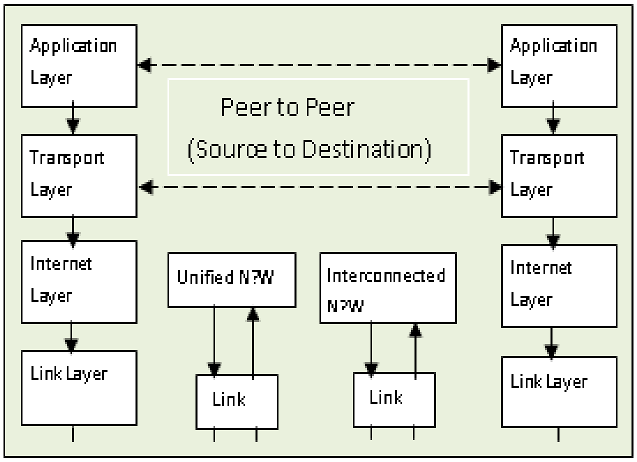
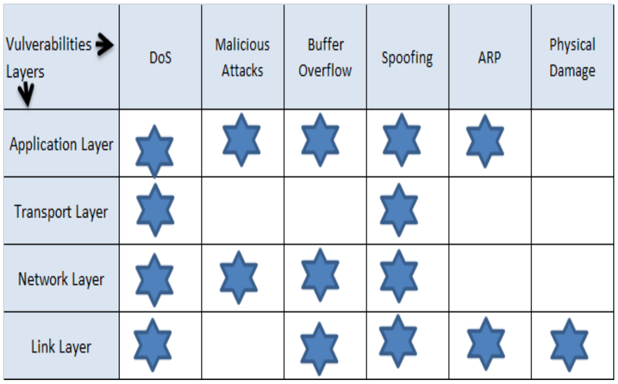


**Fig-2**: Network model for data communication

## 2. Network Vulnerabilities And Management Decisions

Vulnerabilities are weaknesses in victims that allow a threat to become effective. Vulnerabilities exist at each layer of network and level of information system. End user at any level of IS can cause harm by introducing errors or by unauthorized access. Vulnerabilities increase due to the grant permitted at different levels of IS to share the information for decision making. When Internet become a part of corporate network, the IS are even more vulnerable to action from e-mails, instant messages, peer to peer file sharing, etc. Various vulnerabilities are generated at different layers of TCP/IP model. For a successful IS the IS specialist must be concerned about the security of the network for the data sharing. The major vulnerabilities that create the obstacle in the network are Spoofing, Buffer Overflow, Denial of Services, Malicious attacks, ARP and Physical damage. If these vulnerabilities are plug out from the network the communication become more efficient and safe. The Internetwork connection is all it takes to exploit a well known vulnerability and a computer doesn’t check for a user’s intent when logging them on. Majority of the vulnerabilities display themselves as application vulnerabilities, which are closer to user application. Table-1 represents the effect vulnerabilities on the individual network layers. The Application layer vulnerabilities are the Denial of Service (DoS), Buffer Overflow and Malicious Attacks like viruses, worms, Trojan horses, spyware.



| Vulverabilities → Layers ↓ | DoS | Malicious Attacks | Buffer Overflow | Spoofing | ARP | Physical Damage |
|---|---|---|---|---|---|---|
| Application Layer | ★ | ★ | ★ | ★ | ★ | |
| Transport Layer | ★ | | | ★ | | |
| Network Layer | ★ | ★ | ★ | ★ | | |
| Link Layer | ★ | | ★ | ★ | ★ | ★ |

**Table – 1** Tabular representation of Layers with Vulnerabilities.

, Trojan horses, spyware. Virus is rogue software program that attaches itself to other programs or data files in order to execute without users knowledge. Worms are independent computer programs that copy themselves from one computer to another computer over the network to destroy data or halt the computer operations. Trojan horse is a program that appears to be benign but does something other then expected. Spyware program install itself secretly on computer to monitor network surfing activities and serve up advertising. The transport layer manages the end-to-end transmission. The different vulnerabilities occur in transport layer are Denial of Services and spoofing. The next layer is Internet layer; this layer is responsible for routing and forwarding of data. One of the biggest vulnerability is that there is no certain path between the source and the destination that make their data safe. Other vulnerabilities are attacks on routers like password attack, malicious attacks, buffer attack, denial of services and IP spoofing. The next layer is Link layer and deal with physical and logical connections of the packet on the network and also the error control. Link layer work with the help of two sub layers; Media Access Control (MAC) and Logic Link Control (LLC) that are connected via a protocol Address Resolution Protocol (ARP). The various vulnerabilities arises in data-link layer are ARP vulnerabilities i.e. ARP has no means for authentication or validation, ARP spoofing, MAC flooding, loss of power, physical theft of data or hardware, unauthorized change in the environment by removing resources, or by disconnecting physical data-links. The data integrity and confidentiality may be disturbed by the vulnerabilities when the data transfer through different levels of the IS via network media. The management committee has to select the expert members to select the appropriate alternative to make the network risk free from the vulnerabilities. And this problem can be resolved by making a decision to select the various network security techniques and to select the best among them to make the IS free from data risks on the network.

## 3. Anti-Vulnerable Techniques In Information System

The effective and efficient decision makes the rules to fulfill the end user requirements on the basis of the collected data. To reduce the possibility of attacks by vulnerable objects on the required data, various security techniques are being used. The technique is said to be effective if it prevents the network from the maximum network threats. The organization has to choose the most appreciable techniques that are most effective against the intrusions.

| **Vulnerabilities** → **Anti- Vulnerable Techniques** ↓ | DOS | Malicious Attacks | Buffer Overflow | Spoofing | A R P | Physical Damage |
|---|---|---|---|---|---|---|
| Firewall | √ | √ | √ | √ | √ | |
| IDS/IPS | √ | √ | √ | | | |
| IP Hopping | √ | | √ | √ | √ | |
| Physical Security | √ | | | | | √ |

**Table – 2** Tabular representation of Vulnerabilities w.r.t Anti Vulnerable Techniques

The table (Tab-2) represents the 2-dimensional view of the network anti-vulnerable techniques.

### 3.1 Firewall

A firewall intercepts and controls traffic between networks with differing levels of trust. It is part of the network perimeter defense of an organization and should enforce a network security policy The firewall provides the first and last word as to which the data may enter and leave the organization network. Firewall monitors the TCP handshake. If the handshake completes with reasonable period of time, the data packets are passed till the TCP session is closed. Firewall is a combination of software and hardware objects that filter the information flowing through the network to block unwanted flow of data packets. The network attackers and hackers act as a transmit machine to undulate data or become a part of DoS attack. The technologies used by firewall to overcome the DoS are: Network Address Translation (NAT to prevent crackers from seeing our network address). The NAT technology translated the IP address of the network to different IP address for the Internet. Stateful Packet Inspection (SPI provide high degree of security) SPI inspects the data packets flowing on the network to make sure they correspond to an outgoing request and are not an attack from a cracker. Proxy Server: With proxy, all traffic from the Internet is sent to the proxy, the third party watching the transmission only see the proxy not the actual network.

Advanced firewall programs give network extra protection through cookie control, spyware control, adware control and software application control. The firewall that keeps on informing all the conversations the network has on the Internet so that network can be in control. All the above technologies help the firewall to prevent the network and standalone system attached to network from the Denial of Service, Malicious attacks, buffer overflow and spoofing vulnerabilities.

### 3.2 IDS/IPS

Intrusion Detection / Prevention System**:** Intrusion detection System (IDS) is a type of security management system for computers and networks. An IDS gathers and analyzes data from various sources in the network range to identify possible security breaches, which include both intrusions and misuse. In IDS the sensors are located in the network borders. The sensors capture all network traffic and analyse the contents of individual packet for malicious traffic. The recognizing pattern matching attack methods is the oldest method in IDS. It involves identifying an intrusion just by examining a packet and recognizing within a series of bytes, a sequence which corresponds to a specific signature. This method is used as a supplement to filters on IP addresses, destination used by connections and source and destination ports. By filtering the requests the denial of service vulnerability can be overcome and the network will response to the required requests.

An IPS is a network security device that monitors network activities for malicious behavior and can react in real-time to block the activities. Like an IDS, an IPS monitors network traffic. IPS is new generation of IDS. IPS drops the offering packets and then potentially blocks the entire data flow from the suspected hacker. If the traffic that triggers the false positive alert of a customer order, the customer will not wait around for long as his entire session is torn down and all subsequent attempts to reconnect to network are blocked by the well meaning IPS. This will overcome the problem of denial of service. IPS operate online to monitor all network traffic for malicious attacks. The blocking of packets overcome the problem of buffer overflow and malicious attacks.

### 3.3 IP Hopping

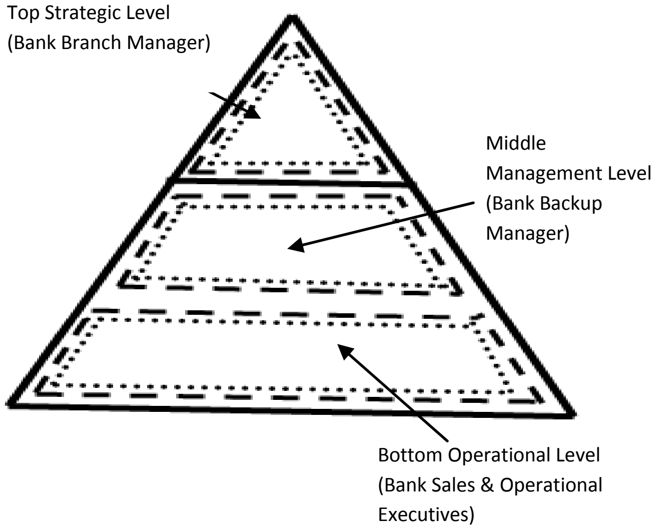
Various attacks can be prevented by changing location or IP address of the active server or desktop proactively within a pool of homogeneous system or with a pre-specified set of IP address ranges. The victim computer’s IP address is invalidated by changing it with a new one. Once the IP addresses change is completed all internet routers will be informed and edge routers will drop the attacking packets. Although this action leaves the computer vulnerable because the attacker can launch the attack at the new IP address, this option is practical for DoS, Buffer Overflow, ARP, Spoofing attacks that are based on IP addresses. On the other hand, attackers can make this technique useless by adding a domain name service tracing function to the attack tools.

### 3.4 Physical Security

Locking the server room, switch box help the network authorities to protect the server and the switch from the network tangling. Latch guards should be installed to prevent prying the door open with a crow bar. Hallway cameras are also good deterrents for un-wanted visitors. Physical access could also allow a hacker to add accounts to the server. Those accounts could be used to access the server remotely and launch a DoS attack. The hard locks with maximum pins and the hidden cameras will prevent the hackers to access the servers to operate with the data and network tools and applications. These physical security tools prevent the network from the denial of service vulnerability

## *4.* Conclusion And Scope

As per the information system all the three layers performs the transaction as per the correct information. Information integrity can be maintained with the handshake of the above mentioned ant-vulnerable techniques. As the information flow through data network and the most effective vulnerabilities that are major threat to the network are DoS, Spoofing and buffer overflow. The Firewall, IPS/IDS and IP Hopping are the strong techniques that are deadly effective on the harmful vulnerabilities. Each IS layer is bordered with the strong Firewall and interiorly scheduled with IPS/IPD and IP Hopping. And the guarding is done by Physical Security. All the discussed techniques make a solid impact on the information flow and maintain the integrity. The bank information system can use the security method to secure the data from the bankers as well as the outsiders. Avoiding the foreign level banker to interfere the information in individual level, the information can easily flow in the bank to provide maximum profit to the bank and make the customer maximum satisfaction regarding its liquidity With proper security and correct information theprovide maximum profit to the bank and make the customer maximum satisfaction regarding its liquidity With proper security and correct information the organization can make the correct decisions proper intercommunication and exchange of information. By implementing as discussed technique the organization can secure the important information from strangers or unauthorized users to make effective decisions for 45% development and competitive growth.



▬▬ (Firewall) ‒ ‒ ‒ (IPS/IDS)
(IP Hopping)
••••••

**Fig-3**: Information System Security Topology

## References

[1] P.Vadivel Murugan, Dr.K.Alagarsamy, "Buffer Overflow Attack – Vulnerability in Stack", International Journal of Computer Applications (0975 – 8887), Volume 13– No.5, January 2011.

[2] Wong, Lai and Cheng, "Value of Information Integration to Supply Chain Management: Roles of Internal and External Contingencies", Journal of Management Information Systems, Vol 28, No. 3, 2011-12. Pg 161-200.

[3] Guo, Yuan, Archer, Connelly, "Understanding Nonmalicious Security Violations in the Workplace: A composite Behaviour Model". Journal of Management Information Systems, Nol. 28, No. 2. 2011. Pg. 203-236.

[4] Prof. Sumir Kumar and Suman Chakarborty, "An Approach for Interanet Data Security", International Journal of Computer Applications (0975-8887), Vol. 9 No 4, November 2010

[5] B. B. Gupta, Student Member, IEEE, R. C. Joshi, and Manoj Misra, Member, IEEE , "Distributed Denial of Service Prevention Techniques ",International Journal of Computer and Electrical Engineering, Vol. 2, No. 2, April, 2010.

[6] Siti Rahayu, Robiah Y, Shahrin S, Faizal MA, Mohd Zaki M, Irda R,"Tracing Technique for Blaster Attack", International Journal of Computer Science and Information Security, Vol.4, No. 1 , 2009.

[7] Dr. G. Padmavathi, Mrs. D Shanmugapriya, “A survey of attacks, security mechanisms and challenges in wireless sensor networks”, International Journal of Computer Science and Information Security, Vol.4, No. 1 , 2009.

[8] Ivan Png, Qiu Wang, “ Information Security: Facilitating User Precautins Vis-à-vis Enforcement Against Attackers”. Journal of Management Information Systems, Vol. 26, No. 2. 2009. Pg. 97-121.

[9] M Eric Johnson, “Information Risk of Inadvertent Disclosure: An Analysis of File-Sharing Risk in the Financial Supply Chain”. Journal of Management Information Systems. Vol. 25, No. 2, 2008. Pg. 97-124.

[10] White Paper, “Host and Network Intrusion Prevention, Competitors or Partners”, Network Assosiates, June, 2004.

[11] Zahir Irani and Peter ED, “The Propagation of Technology Management Taxonomies for Evaluating Investment in Information System”. Journal of Management Information System/2000-2001. Vol. 17 No. 3 pp 161-177.

[12] Frederic Avolio, Avolio Consulting, ” Firewalls and Internet Security, the Second Hundred (Internet) Years “ The Internet Protocol Journal - Volume 2, No. 2, June1999.

[13] Liebowitz J, “Information Systems: Success or Failure?”, The Journal of Computer Information System, Vol. 40, N0.1, 1999, pp.17-26.

[14] The Book “Information System Security Review Methodology, A guide for Reviewing Information System Security in Government Organizations” by EDP Audit Committee International Organization of Supreme Audit Institutions, Vol 1, II, October 1995

[15] Swanson, EG. “Information System Implementation”. Richard Irwin Inc. Homewood IL 1988.

[16] Lucas HC. “Performance and the Use of an Information system”, Management Science, Vol. 21, No.8, 1972, pp 908-919.

# SESSION

# SECURITY APPLICATIONS AND ALGORITHMS

## Chair(s)

### Dr. Nizar Al Holou

# Security Concerns and Disruption Potentials Posed by a Compromised AMI Network: Risks to the Bulk Power System

**M. M. Olama[1], J. J. Nutaro[1], V. Protopopescu[1], and R. A. Coop[2]**
[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
[2]University of Tennessee, Knoxville, Tennessee, USA

***Abstract***— *The advanced metering infrastructure (AMI) of a smart electrical grid is seen as both a network for improving the efficiency of the electrical power system and as a potential target for cyber-attackers bent on disrupting electrical service. In this paper, we examine how a hijacked AMI network might be used to instigate widespread blackouts, and the physical barriers that the electrical system itself poses to such an attack. To this end, we present a simple, but potentially useful, model for gauging the quantity of load that an attacker must control for an attack to be successful. Conversely, the model suggests a scheme for mitigating the attack, but at the cost of decreasing the usefulness of smart meters as devices for the legitimate regulation of electrical load.*

**Keywords:** Advanced metering infrastructure, smart meters, energy management systems, swing equation, cyber security

## 1 Introduction

Smart electric meters, capable of two-way communications and having software and hardware to enable energy management in real-time, are a major part of the advanced metering infrastructure (AMI) of a smart electrical grid. These meters have tremendous potential to improve the efficiency and reliability of the national power system. For example, a washing machine in a household with a smart meter could be set to run only when energy is cheap. This reduces energy costs for the power consumer and reduces the size of demand peaks, which are served with expensive, relatively inefficient sources of energy. Current plans call for nearly 17 million smart meters to be installed in U.S. homes and businesses over the next few years. While proponents of a smart grid and, in particular, an AMI have touted the potential to improve the electricity system, critics have expressed concerns about the susceptibility of AMI meters to cyber-attacks.

Cyber security practitioners [1]-[5] have claimed that smart meters are susceptive to hacking, and thereby are a potential enabler of unauthorized access to command and control processes that could be abused to disrupt electrical service. Such claims are often disputed by engineers that operate large electrical power systems. Significantly, the vulnerabilities found in smart meters have not been put to such a use.

However, if such an attack were to be carried out, it might unfold as follow. The attacker gains control over a substantial quantity of load by hijacking a large number of smart meters. Using these meters, the attacker creates a large imbalance between power used and power supplied by switching off the load that he controls. This causes a large and sudden change in the frequency of the power system, thereby forcing some generators to disconnect from it. By repeating this attack several times in the course of a few minutes, large numbers of generators may be forced to disconnect, thereby instigating a large-scale blackout.

This scenario is often used as evidence of a systemic risk posed by vulnerabilities in smart meters and their attendant infrastructure for communication and control. Opponents of these claims cite the intrinsic robustness of a power system to sudden changes of load. Indeed, the inertia of the power system's generators and the presence of automatic controls designed specifically to deal with imbalances of supply and demand are a physical barrier to successful attack.

In this paper, we estimate the physical requirements that must be met by an attack seeking to disable a large power system by the sudden and simultaneous manipulation of numerous electrical loads. We further propose that a random delay imposed on energy management actions can mitigate the most disruptive effects of such an attack. Towards this end, we examine two key quantities in such an attack: the amount of load controlled by an attacker and the swiftness with which it is switched. Our method of analysis is illustrated by its application to data published for the power grid that serves the western United States.

This manuscript has been authored by UT-Battelle, LLC, under Contract DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

## 2 Motivation and method of attack

Among the motives for instigating a computer-based, rather than physical, attack against electrical infrastructure is the relatively small risk of getting caught. Cyber-attacks are notoriously difficult to attribute to a specific source (see, e.g., the discussion by Libicki [6]), and so the risks faced by a cyber-attacker are typically much less than those faced by the instigators of a physical attack. This aspect of a cyber-attack may make it attractive to malicious actors who want to disrupt electrical service but have a low tolerance for risk; this may be particular true for potential attackers who would never consider carrying out a physical attack.

There are two basic avenues for a cyber-attack on electrical loads: 1) through the meter hardware or 2) through a computer that controls metering or other electrical management, functions. Attacking the meters would perhaps be the most effective. However, such an attack requires both known, exploitable flaws in the meter software and a capability to exploit that flaw via remote operations on a very large number of meters; e.g., by the use of a botnet-style worm (see, e.g., [7], [8]).

The other avenue for an attack is via computers (e.g., home computers) that interact with and control some actions of a smart meter or load. For example, a home energy management system that is operated by a personal computer may be susceptible to attack via the Internet. By hijacking this computer, the attacker may be able to modify the power consumption of electrical appliances under its control (see, e.g., [9]). It is likely in this case that the computer controls only part of the load at a home or business, e.g. the PC may be able to remotely turn the furnace on or off, and so such an attack vector may be relatively less effective than one that targets the meter itself.

In any case, a significant hurdle that a cyber-attacker must overcome is to ensure that the infected devices are physically collocated. For an attack to be effective, it must be able to disconnect (or otherwise change the consumption of power by) loads within a specific electrical power system. Hence, all of the hijacked load must reside within the geo-physical region served by the targeted power system.

## 3 Effect of a sudden change of load

The sudden disconnection of an electrical load is felt by a generator as an acute easing of the torque which opposes its turning rotor. Automatic controls act to bring all forces back into balance by adjusting the production of power at the generator to match demand. A sudden reconnection of the same load will have the opposite effect, causing the generator to feel an acute tug in opposition to its spinning rotor. This causes it to slow and, again, automatic controls act to bring supply and demand into balance. If, however, the rotating speed of the machine drifts too far from normal, then automatic protection devices disconnect it from the electrical network. This can turn an abnormal event into a cascading failure.

It is conceivable that an attacker in control of a sufficiently large number of smart meters will use this physical phenomenon to instigate a large blackout. The number of meters required, and the precision with which their switches must be operated, are determined by the physical properties of the generators and the settings of their controllers. The main method in this attack is a change in load that happens more quickly than the automatic controls can respond.

We use a model derived from the swing equation and a simplified speed governor (see, e.g., the power system model presented in [10]) to approximate the system-wide, average change in frequency of a large electrical power system following a sudden change in the real electrical load. This model has two parameters for characterizing the electrical generation system, both of which can be measured (see, e.g., [11], [12]). These are its inertia $M$ and the rate $\tau$ of its response to frequency excursions. This model also has one parameter for describing the change in load; this is the fraction $\alpha$ of the base electrical load $P_e$ that is shed.

The change $\omega$ of the frequency of the power system during the event is related to the changing power $P_m$ output by the generators and power $P_e$ demanded by the loads by

$$\dot{\omega} = \frac{1}{M}\Big(P_m - P_e\big(1 - \alpha u(t)\big)\Big)$$

$$\dot{P}_m = -\tfrac{1}{\tau}\omega \tag{1}$$

where $u(t)$ is the step function and the model begins in steady state with $\omega(0) = 0$ and $P_m(0) = P_e$.

Though simple, this model captures the three salient features that act unavoidably to oppose an attack relying on the manipulation of load. These features are (i) the inertia of the generators, which opposes sudden changes in frequency; (ii) the speed governors that act to correct an imbalance before dangerous changes in frequency are realized; and (iii) the rate at which those governors must act, which is determined in turn by the generation inertia and the size and rate at which the imbalance forms.

Only the frequency excursion is of interest here; solving Eqn. (1) we get

$$\omega(t) = \alpha P_e \sqrt{\tfrac{\tau}{M}} \sin\!\left(\tfrac{t}{\sqrt{M\tau}}\right) \tag{2}$$

which has a maximum amplitude $\omega_{\max}$ at

$$\omega_{\max} = \alpha P_e \sqrt{\tfrac{\tau}{M}} \tag{3}$$

This derivation shows that the maximum frequency deviation is proportional to the fraction of load that is removed by the step change. Using

$$\epsilon = P_e\sqrt{\frac{\tau}{M}} \tag{4}$$

Equation (3) can be written as

$$\omega_{\max} = \epsilon\,\alpha \tag{5}$$

and given measurements of $\alpha$ and $\omega_{\max}$ for an observed event, $\epsilon$ can be calculated.

If $\omega_{\max}$ is greater than the maximum excursion tolerable by the power system (again, a quantity which is known or can be estimated) then the system is at risk. Such a sudden change in load can be prevented in at least three ways.

First, hardware in the meter itself can impose a random delay and thereby force a ramped response from the population of meters. If the rate of response is sufficiently slow, then it will give automatic controllers sufficient time to safely adjust the power output of their generators. If implemented in trustworthy hardware, this protection mechanism is effective regardless of how the attack is instigated; that is, it cannot be disabled by software faults or computer-based attacks.

Second, business logic, implemented in software, can monitor for unsafe load changes and refuse to execute them, force the change into safe limits, ask the operator for confirmation, or all three. This kind of process control supplements other security measures which seek to ensure that requests come only from authorized operators, that worms and viruses do not infect the smart meters and enable malicious operation of the electrical switch, and to prevent other similar kinds of contingencies. However, the software that implements the protective business logic is itself subject to cyber-attack, and so its effectiveness as a security measure cannot be guaranteed.

Third, AMI installations could be deliberately limited in their scope. Large-scale penetration of AMI would be much more difficult given isolated AMI networks that employ different types of metering hardware, operating systems, and networking protocols. This is probably infeasible as a long term security solution, but could serve as a risk mitigation strategy while pilot deployments are rolled out. A staged deployment will provide opportunity for a utility to discover security problems, gain valuable operating experience in this respect, and to do so before the population of advanced meters reaches a threshold sufficient for causing wide spread disturbances.

In the next section, we address the first approach as a practical means for reducing the risks of an AMI being misused as a tool for the widespread disruption of the electrical service. Specifically, we show with an analytical model how imposing a random delay in the meter mitigates the impact of a cyber-attack on the power system.

# 4 Mitigating strategy

The maximum frequency deviation of the power system, $\omega_{\max}$, due to a sudden change in load can be reduced by enforcing a random time delay between a request that the smart meter opens or closes its switch and the moment at which the action is actually carried out. The switching delay must be determined at random for each request, and the delay implemented by a device not accessible in any way via remote access to the meter. If a hardware delay can prevent a sudden change in electrical load, automatic speed controls in the generators can prevent a dangerous jump in $\omega$.

This security control is modeled in a way similar to the attack in (1), but with the step $u(t)$ replaced with the ramp $\beta(t)$ as

$$\dot{\omega} = \frac{1}{M}\left(P_m - P_e\left(1-\alpha\beta(t)\right)\right)$$
$$\dot{P}_m = -\tfrac{1}{\tau}\omega \tag{6}$$

where $\beta(t)$ is the ramp function

$$\beta(t) = \begin{cases} t/\gamma & t<\gamma \\ 1 & t\ge\gamma \end{cases} \tag{7}$$

and $\gamma > 0$ is the duration over which the load change occurs. This is illustrated in Fig. 1.

Solving Eqns. (6) and (7) for $\omega$ gives

$$\omega(t) = \begin{cases} \frac{P_e\alpha\tau}{\gamma}\left(1-\cos\left(\frac{t}{\sqrt{\tau M}}\right)\right) & t<\gamma \\ \frac{P_e\alpha}{\gamma M}\left(\cos\left(\frac{t-\gamma}{\sqrt{\tau M}}\right)-\cos\left(\frac{t}{\sqrt{\tau M}}\right)\right) & t\ge\gamma \end{cases} \tag{8}$$

If the power system is stable in the sense that it will ultimately damp out any excursion (and this must be the case while the power system is operating) then it is sufficient to consider just the first swing (i.e., half period); subsequent swings will have decreasing amplitudes.

From this perspective, there are two possibilities. If the interval $\gamma$ is sufficiently large, then the first swing occurs while $t < \gamma$. Its magnitude is

$$\omega_{max}(t<\gamma) \le 2\alpha\tau P_e/\gamma \tag{9}$$

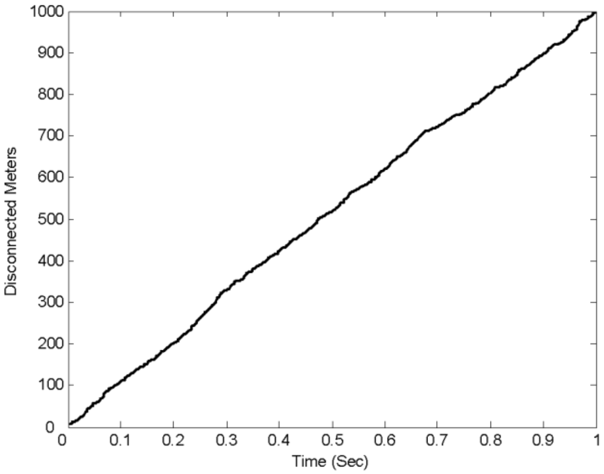


Fig. 1. Simulated ramp response of 1000 meters with enforced delays selected at random from [0,1].

and this limit is reached at

$$t_{max}(t < \gamma) = \pi\sqrt{\tau M} \tag{10}$$

In this case, the rate of the response of the generators to the imbalance counteracts the magnitude and suddenness of the change in load.

More desirable for the attacker is that $\gamma$ be small, in which case inertia carries the system through its first swing. This is the case of practical interest, for which $\omega$ is bounded by

$$\omega_{max}(t \geq \gamma) \leq \frac{P_e\alpha}{\gamma M}\left(\cos\left(\frac{\pi - \frac{\gamma}{\sqrt{\tau M}}}{2}\right) - \cos\left(\frac{\pi + \frac{\gamma}{\sqrt{\tau M}}}{2}\right)\right) \tag{11}$$

and this limit is reached at

$$t_{max}(t \geq \gamma) = \frac{\pi\sqrt{\tau M}}{2} \tag{12}$$

Indeed, if $\gamma$ is very small, then Eqn. (11) can be approximated by

$$\omega_{max}(t \geq \gamma) \approx \frac{P_e\alpha}{M\sqrt{\tau M}} \tag{13}$$

Thus, a conservative limit for $\omega_{max}$ is given by

1. Eqn. (11) while $\gamma$ is less than the $t_{max}$ in Eqn. (12),
2. Eqn. (11) from $\gamma$ equal this $t_{max}$ until Eqn. (11) is less than Eqn. (9), and
3. Eqn. (9) afterwards.

# 5 Illustration

To illustrate this method of analysis, we estimate the combinations of $\gamma$ and $\alpha$ required to cause particular frequency excursions in the western United States. The data used for these calculations is derived from a June 14 event in the western U.S. (as governed by WECC; see [12], especially Figures 1 and 2). In that event, 4.589 GW of generation was lost resulting in a 0.4 Hz frequency excursion. Chassin *et. al.* measured the system inertia during this event as 17.8 GW.sec$^2$. The base load $P_e$ for this specific event is not reported in their paper, but 90 GW is the midpoint for the range of values reported and using this for $P_e$ gives a notional 5% change in load. With this data and Eqn. (13), the control constant $\tau$ is calculated as 0.0224.

Fig. 2 illustrates the conservative limit for $\omega_{max}$ using the above data. When the ramping time is small, the size of the change in load is the dominant factor. This observation is consistent with the approximation in (13). The size of the excursion can be controlled, however, by lengthening the ramping interval; using the data above, the magnitude of the excursion falls quickly for $\gamma > 1$ second.

This observation suggests that the impact of hijacked loads on the power system may be mitigated with an enforced ramping time. One place to enforce this is in the meters themselves; an inexpensive circuit may be introduced into the electronics that selects a delay uniformly in $[0, \gamma]$ for that meter's switch. The aggregate effect of this delay is described by (7), and if $\gamma$ is large enough then the population of smart meters adjusts the total load $P_e$ at a tolerable rate. This scheme prevents an attacker from causing an unacceptably large change in frequency by placing a physically enforced limit on his actions. Moreover, if this delaying circuit is installed when the meter is manufactured then its protective function comes at a very small price.

Though this scheme prevents rapid changes in load at the meters, it does not prevent a similar type of attack at other points in the electrical system; e.g., sub-stations typically have a capability to disconnect large amounts of load for the purposes of emergency load shedding. Moreover, it is not known whether the risk posed by an attack on smart meters justifies the proposed limit on their use: rapid control of load via smart meters is desirable when it is used to regulate (rather than disrupt) frequency.

# 6 Conclusions

Risk that may be posed by smart meters is of particular concern because these meters, unlike a control center or substation, are readily accessible to an attacker. Smart meters are easily purchased, giving a potential attacker the opportunity to study these devices in detail; and smart meters that are installed in a home or business may be linked directly to the home or business computer network. Indeed, this feature is being used in some energy management systems, and it raises the possibility of hijacking loads via the Internet.

The model developed in this paper is a tool for estimating the impact that hijacked loads may have on an electrical power system. Though the model is simple and the calculated estimate necessarily rough, the resources required to construct this estimate are minimal and can be obtained relatively easily (again, see [11], [12]). The model makes concrete the argument that inertia and speed controls are a barrier to causing widespread disruption of electrical service by the use of hijacked loads. The technological sophistication and engineering resources required to overcome this barrier remains a topic for future analysis.

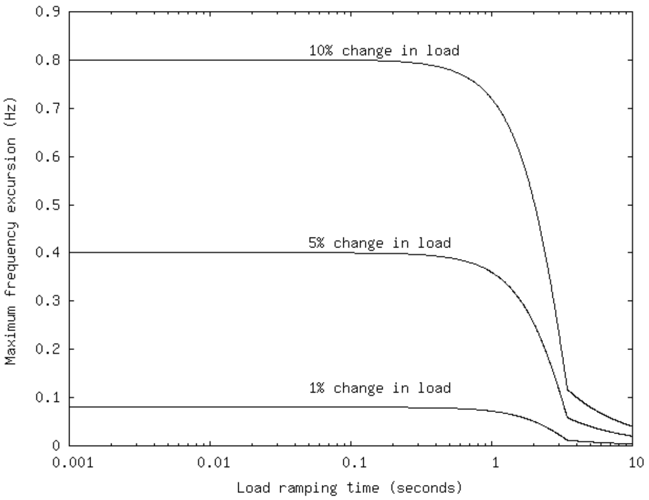


Fig. 2. Estimate of the maximum frequency excursion as a function of the load ramping time.

## 7 References

[1] A Risk-based Approach to Determining Electronic Security Perimeters and Critical Cyber Assets, Technical Report, IOActive Inc., 2009.

[2] Study of Security Attributes of Smart Grid Systems – Current Cyber Security Issues, Technical Report, Office of Electricity Delivery and Energy Reliability, U.S. Department of Energy, April 2009.

[3] S. Harris, “Chinese Hackers Pose a Clear and Present Danger to U.S. Government and Private-Sector Computer Networks and May be Responsible for Two Major U.S. Power Blackouts,” *National Journal Magazine*, May 31, 2008.

[4] M. Carpenter, T. Goodspeed, B. Singletary, E. Skoudis, and J. Wright, “Advanced Metering Infrastructure Attack Methodology,” version 1.0, http://inguardians.com/pubs/AMI_Attack_Methodology.pdf, Jan. 2009.

[5] R. Berthier, W. H. Sanders, and H. Khurana, “Intrusion Detection for Advanced Metering Infrastructures: Requirements and Architectural Directions,” *Proc. of the IEEE International Conference on Smart grid Communications*, pp. 350 – 355, 2010.

[6] Martin C. Libicki, Cyberdeterrence and Cyberwar, RAND Corporation, 2009.

[7] B. Stone-gross, M. Cova, L. Cavallaro, B. Gilbert, M. Szydlowski, R. Kemmerer, C. Kruegel, and G. Vigna, “Your Botnet is My Botnet: Analysis of a Botnet Takeover,” *Proc. of the 16th ACM Conference on Computer and Communications Security (CCS’09)*, pp. 635-647, Nov. 2009.

[8] S. Staniford, D. Moore, V. Paxson, and N. Weaver, “The Top Speed of Flash Worms,” *Proc. of the 2004 ACM Workshop on Rapid Malcode (WORM’04)*, pp. 33–42, New York, NY, USA, 2004.

[9] M. LeMay, G. Gross, C. A. Gunter, and S. Garg, “Unified Architecture for Large-scale Attested Metering,” *Proc. of the 40th Annual Hawaii International Conference on System Sciences (HICSS)*, pp. 115-124, 2007.

[10] J. Arrillaga, C. P. Arnold, and B. J. Harker, *Computer Modeling of Electrical Power Systems*, New York, Wiley 1983.

[11] T. Inoue, H. Taniguchi, Y. Ikeguchi, and K. Yoshida, “Estimation of Power System Inertia Constant and Capacity of Spinning-Reserve Support Generators Using Measured Frequency Transients,” *IEEE Transactions on Power Systems*, vol. 12, no. 1, pp. 136-143, Feb. 1997.

[12] D.P. Chassin, Z. Huang, M.K. Donnelly, C. Hassler, E. Ramirez, and C. Ray, “Estimation of WECC System Inertia Using Observed Frequency Transients,” *IEEE Transactions on Power Systems*, vol. 20, no.2, pp. 1190-1192, May 2005.

# Practical Security in E-Mail Applications

Franziskus Kiefer[1], Alexander Wiesmaier[2], Christian Fritz[1]

[1] Technische Universität Darmstadt
Hochschulstraße 10, 64283 Darmstadt, Germany
kiefer@cdc.informatik.tu-darmstadt.de
c_fritz@rbg.informatik.tu-darmstadt.de
[2] AGT Group (R&D) GmbH
Hilpertstraße 20a, 64295 Darmstadt, Germany
awiesmaier@agtinternational.com

***Abstract*—This paper deals with practicability issues of encrypted e-mails. A quick survey on the status quo indicates that popular e-mail clients lack substantial practicability qualities, for example searching in encrypted e-mails. Other approaches such as De-Mail provide solutions, but offer transport encryption only. We present and discuss a number of improvements to the practicability of e-mail encryption. These enable efficient searching in encrypted e-mails as well as subject encryption and the use of cryptographic functions in calendar applications. We present a prototype, called CryptoBird, providing a proof of concept for the proposed core features.**

***Keywords*—Calendar, CryptoBird, E-Mail, Encryption, PGP, Practicability, Searching, S/MIME**

## I. Introduction

Only few people use encryption to protect their e-mails [11]. While in private communication this might be acceptable, this is a serious issue in case of governmental or business usage as plain e-mails are inherently insecure [21].

Nowadays, all major e-mail clients (and servers) support encrypted e-mail transport using SSL. Once configured correctly, this is transparent to the users and protects their e-mails on their way through the Internet. It provides a reasonable level of privacy protection and is recommendable for day to day private e-mail communication. But in other scenarios such as handling confidential business e-mails or sharing devices with others, transport security is not sufficient, as the e-mails are stored in plaintext.

Possible solutions for this problem are using hard disk encryption or applying e-mail encryption, which both suffer from low popularity [11], [19]. In the work at hand we focus on problems related to e-mail encryption.

Reasons for the unpopularity of e-mail encryption might be unawareness and usability or comfort issues. A lot of usability research in terms of cognitive walkthroughs and user studies has been performed on e-mail encryption [29], [23], particularly on the software Pretty Good Privacy (PGP). We consider the following practicability issues with encrypted e-mails as most serious:[1]

1) Using asymmetric cryptography to encrypt e-mails requires some kind of public key infrastructure (PKI). While the use of X.509 [6] and the Secure/Multipurpose Internet Mail Extensions (S/MIME) [27] is suitable for companies, universities or other organizations, the centralized PKI approach is not practicable for private use. PGP [34] and its infrastructure offer good possibilities, but its correct application is too complicated for the regular user [33].
2) Even if the PKI problems are solved, the handling of encrypted e-mail correspondence comes with some drawbacks. One of the most serious issues is losing the ability to conduct full text searches on e-mail folders when using end to end encryption (which is desired for security reasons). More disturbances appear when using the contents of encrypted e-mails outside the e-mail application, e.g. in calendar applications, which are usually not capable of decrypting the e-mail content on the fly. This leads to either unreadable (encrypted) or insecure (decrypted) calendar entries.
3) Even though the e-mail body is encrypted, the e-mail header is usually not. This leads to a couple of security risks, of which the most serious one probably is that the subject of encrypted e-mails is not encrypted. Even worse is the fact that the user might not be aware of this.

### A. Contribution and Delimitation

The work at hand focuses on practicability and security aspects of encrypted e-mail communication and how handling encrypted e-mails differs from handling plaintext e-mails. We also show how these differences can be minimized while still complying to existing standards, resulting in a substantially better aligned user experience.

This work is not a usability study on (plaintext or encrypted) e-mail communication. Long-term issues of public key cryptography are also out of the scope of this work.

[1] To see other reasons why users decide whether to use encrypted e-mails or not we refer to [15].

### B. Outline

First, we define practicability criteria and use them for a survey on common e-mail clients (Sec. II). Then, we propose and discuss solutions which are in line with existing standard mechanisms and protocols to solve the identified problems (Sec. III). After that, we discuss related work and compare it to our solution (Sec. IV). We demonstrate the feasibility of our solution presenting a prototypical implementation (Sec. V). Finally, we give a conclusion and discussion on remaining open problems (Sec. VI).

## II. Practicability of Secure E-Mails

In this section, we define and discuss criteria for the practicability of secure e-mail clients. With these criteria in mind we analyze existing e-mail clients and identify their shortcomings before proposing our solution to these problems in the subsequent section.

### A. Practicability Criteria

Our main approach for optimizing secure e-mail handling is to provide equal treatment of encrypted and plaintext e-mail communication. We consider the smooth integration of security features with the existing user interfaces and work flows as one of the main concerns for practicability. The user should not have significant additional expenditure when using encrypted e-mails and user interaction should take place in a way the user is familiar with from dealing with plaintext e-mails. This is expressed by the following practicability criteria:

**Automatic encryption and decryption (automation):** To achieve a seamless integration into the e-mail client, the user must be able to write an e-mail in the usual way and toggle encryption so that the e-mail is *sent* and *stored* encrypted. Encrypted text should be automatically decrypted for viewing without requiring user interaction.

**Efficient key management (key management):** The user should not be obliged to perform demanding key management tasks. However, in PGP the user has to decide whether a key is trustworthy or not. But there are some actions that can be done automatically, for example selecting the correct key for encrypting or decrypting an e-mail (this overlaps with the automation criterion). The client might also search for a key on a (selectable) key-server or offer to import keys received via e-mail into the keyring. Due to the central approach of S/MIME, its certificate and key handling is easier. Keys usually can be downloaded from defined servers (which are assumed to be trustworthy) or come along with S/MIME signed e-mails from where they can be imported automatically into the client's key management. There is also the topic of management of expired keys and certificates, which is especially important in case of long-term archiving of encrypted e-mails. Although long-term storage is an important issue, it is not the most pressing one and is not considered in the work at hand.

**Feature Range (feature range):** The features offered for encrypted e-mails should not differ from the ones for plaintext e-mails (for example the possibility to search for keywords). In addition to the e-mail functionality, there are several integrated functionalities, such as converting e-mails into calendar entries that work fine with plaintext e-mail content. They should be usable with encrypted content, too.

**Seamless integration (integration):** Security related functionality should be provided via the existing interfaces, such as tool bars, context menus or filters. If this is not possible, the interfaces should be extended in the style and spirit of existing features and user interfaces. Another important topic is that the security functionality behaves as expected (which is currently not always the case, cf. Sec. III).

**Support of common standards (standard support):** There are different standards which have to be supported. Regarding e-mail security, these standards are S/MIME [27], PGP/MIME [5] and PGP/INLINE [5] supporting e-mail encryption and signing.

### B. Client Analysis

With these criteria in mind we evaluate common e-mail clients regarding their practicability when using encryption features. Our distinction between "encrypted content" and "cleartext" refers to e-mails as well as calendar entries. Table I summarizes the survey. The table is to be read in the following way: 'yes' means that the feature is included in the e-mail client. '+' implies that the functionality in question can be added by installing add-ons or extra programs. 'no' stands for missing functionality. If a feature is only available for a certain standard, this is indicated by the name of the supporting standard, e.g. 'S/MIME'. The detailed results are described in the following paragraphs.

| Client | Automation | Key Mgnt. | Feat. Range | Integration | Std. Sup. |
|---|---|---|---|---|---|
| Apple Mail | yes/+ | yes | no | no | yes |
| Evolution | yes | + | no | yes | yes |
| Gmail | +S/MIME | + | no | no | + |
| MS Outlook | no | + | no | S/MIME | + |
| Thunderbird | + | + | no | yes | + |

TABLE I
FEATURE COMPARISON OF E-MAIL CLIENTS

**Apple Mail** [1] (aka Mail.app) is the standard e-mail client for Mac OS X (and iOS devices). It supports S/MIME encryption and signing. With the help of the GPGMail [18] plug-in, it is possible to perform PGP encryption and signing (using the desktop application). E-mails are decrypted when the decryption button is pressed. GPGMail can find keys of recipients automatically in the keyring and is even able to download missing keys from a given server. It is not possible to search in encrypted e-mails. All in all only the criteria *key management* and *standard support* are sufficiently fulfilled.

**Evolution** [25] is an open-source e-mail client for Linux and is shipped with GNOME. It is able to encrypt, decrypt, sign and verify PGP- and S/MIME e-mails automatically. It uses a keyring provided by programs such as the free PGP implementation GnuPG. There is no possibility to directly add keys received by e-mail. Searching for text in encrypted e-mails is not supported. Evolution complies with the criteria *automation*, *standard support* and *integration*. The *key management* criterion is only partially fulfilled.

**Gmail** [17] is a very popular example for web-mail interfaces. Additional software like GnuPG enables PGP usage with Gmail, but requires additional manual actions to be performed by the user. Recurity[28] recently published a first prototype[2] of a PGP plugin for the Chrome[16] browser and Gmail service. Another add-on that works with various web browsers is Penango [26]. It allows to encrypt and sign e-mails with Gmail using the S/MIME standard. The Penango Firefox add-on takes advantage of the build-in certificate management of Firefox. To our knowledge  it is not possible to search in encrypted e-mails. None of the practicability criteria defined above is adequately fulfilled.

**Microsoft Outlook** [24] can handle S/MIME-encrypted e-mails but is not able to process PGP e-mails on its own. There are some approaches to add PGP compatibility, for example a project called gpg4win [20], which is a collection of different tools including a PGP plug-in for Microsoft Office 2003/2007. Currently, the plug-in does not work with Microsoft Outlook 2010.

The commercial software PGP Desktop [30] is a stand-alone solution that handles PGP-encrypted texts. It can be embedded (as an Add-In) into Outlook 2010 and other e-mail clients such as Thunderbird. The decryption requires some manual action: the encrypted e-mail has to be moved to the PGP Viewer window where it gets decrypted. The decrypted e-mail can be copied to the Outlook 2010 inbox - so the e-mails can be searched afterwards, but they are not stored encrypted any longer which might be unwanted due to confidentiality risks. In summary none of the practicability criteria is met completely.

**Mozilla Thunderbird** is a multi-platform e-mail client which can be easily extended by add-ons. It is able to handle S/MIME encrypted e-mails. By using the add-on Enigmail [31] and GnuPG [13] PGP-encrypted e-mails are supported and processed automatically. Enigmail assists the user by offering an automatic import of keys sent by e-mail and also picks the correct keys for the recipients from the keyring. Thunderbird is lacking a possibility to search in encrypted e-mails. As mentioned above, PGP Desktop can be used with Thunderbird, but regarding our criteria it is not as powerful as the Enigmail add-on. In summary, Thunderbird and its add-ons support *automation*, *key management*, *integration* and *standard support*. The criterion *feature range* is only partly fulfilled.

### C. Practicability Conclusion

None of the introduced e-mail clients meets all our practicability criteria. In particular, no e-mail client is compliant with the *feature range* criterion, not even with third party plug-ins. For most clients encryption features are only available after installing extra software or add-ons. The e-mail clients that provide calendar functionality are not prepared to handle encrypted calendar content.

## III. Enhancing the Practicability of Secure E-Mails

In the following we consider an e-mail client with calendar functionality that handles S/MIME encrypted e-mails as well as PGP/INLINE and PGP/MIME encrypted e-mails.

We have seen that currently no e-mail client exists which provides equal treatment of encrypted and plaintext e-mails, in particular meeting all our criteria. In this section we propose and discuss concrete improvements to overcome the shortcomings discussed in Section II. As all of the investigated clients miss this quality, we focus on the equal treatment of plaintext and encrypted e-mails. Furthermore, we deal with some additional improvements that enhance the security of encrypted e-mails. We consider the ability to search in e-mails as one of the most critical shortcomings when it comes to the handling of encrypted e-mails. As shown in [32], searching in e-mails is one the most efficient ways to organize them.

In the following we investigate possibilities to reach a seamless integration of encrypted e-mails into existing clients. Additionally, we provide some details on our prototypical implementation of the proposed improvements.  We implemented an open source prototype as Thunderbird add-on called CryptoBird as proof of concept[3] (cf. Section V).

In addition to the previously defined criteria we consider the following:

In order to provide reasonable efficiency when working with big amounts of data, some kind of indexing or caching might be necessary. In this case, it is important to work with an encrypted index or cache to avoid compromising the confidentiality. An encrypted index or cache might require an additional password, which should be integrated into the password manager of the e-mail client.

Current implementations of e-mail encryption show a strange and risky behavior: the encryption applies to the body only and does not include the header, especially not the subject line. To meet the user's expectation when encrypting a message, some kind of header encryption has to be applied to encrypted e-mails. This way the user is not lulled into a false feeling of security when encrypting e-mails.

### A. Searching in Encrypted Content

As mentioned before, we consider conducting full text searches in the bodies of encrypted e-mails as one of the most pressing matters. This pays attention to the paradigm shift from ordering to searching big amounts of data, especially when dealing with e-mails [32]. While this is possible in some e-mail clients for the currently displayed e-mail, it is currently impossible to conduct reasonable full text searches over multiple encrypted e-mails in folders or whole mailboxes.

A main problem which has to be solved here is the efficiency. A search in encrypted e-mails would take a long time if each e-mail has to be decrypted on the fly while searching. This lack of efficiency has mainly two reasons. First, the straight forward full text search in stored e-mail bodies is slow

[2] http://gpg4browsers.recurity.com

[3] The prototype can be found at https://code.google.com/p/cryptobird/.

anyway. Second, the cryptographic operations need additional time.

We solve this in CryptoBird by indexing all decrypted e-mails and storing them in a secure database. This database offers encrypted data storage while preserving efficient querying features. This procedure can also be applied for other sensitive content like calendar entries. Section V gives more technical details on this topic.

This solution especially supports the *integration* and *feature range* criteria while preserving efficiency.

### B. Calendar Integration

Another issue is the conversion of encrypted e-mails into calendar entries (events or tasks) using the integrated calendar of the e-mail client. During the conversion a new calendar entry is created from an e-mail and inserted into a selectable calendar (local or remote). The new calendar entry is pre-filled with the body of the e-mail and is displayed for editing. The e-mail's subject is used as title of the calendar entry. Furthermore, the conversion between tasks and events as well as all other calendar functions should work whether the content is encrypted or not.

It should be possible (and the default behavior) to keep the content encrypted for confidentiality reasons (especially when using third party online calendars). Then, the plaintext content should be displayed in cleartext to the user and all invitees in an automated manner. This can be achieved with the same techniques as with e-mails sent to multiple recipients, i.e. encrypting the session key with the public key of each participant. It should also be possible to easily generate a calendar entry containing the plaintext body of an encrypted e-mail. This is useful when the user wants to generate a public calendar entry or not all invitees are able to view encrypted calendar entries. Then, the user should be able to edit the calendar entry before saving it to have the possibility to delete confidential information. The other way round — converting a calendar entry into an e-mail — also has to be supported. In this case the same applies analogously. It should also be possible to search in encrypted calendar entries (cf. previous subsection).

### C. Encryption Toggling

Another useful feature is toggling the encryption status of existing e-mails or calendar entries. This means (bulk) encryption or decryption of stored e-mails or calendar entries. This should be possible for individual or multiple selected items or for entire e-mail folders or calendars. We discuss the usefulness of both directions:

*Encrypting plaintext e-mails / calendar entries:* It is obvious that encrypting cleartext that has been submitted via an open network or has been stored in an untrusted environment does not annihilate the disclosures it may already have suffered. The point in encrypting such data is to decrease future risk of disclosure. An everyday example where this is useful is loosing the device the e-mail client is stored on, for example a laptop or a memory stick. By encrypting all e-mails (although received in cleartext) a possible thief or finder of the device is prevented from reading the e-mails. This could also be achieved by encrypting the entire device, but a recent study [19] shows that only very few users actually use hard disk encryption. An interesting feature in this scenario is the automatic encryption of all stored e-mails or calendar entries by the e-mail client. This would for example protect all server side data of Web based e-mail clients in case the user's account is hacked.

Sometimes cleartext e-mails or calendar entries never leave the trusted environment. For example, by sending intra-organization e-mails from computers within an organization The same applies for calendar entries created and stored in local calendars. When leaving the trusted environment, e.g. when the organization is about to outsource its e-mail and calendar services into the Cloud, it seems appropriate to encrypt these e-mails and calendar entries. This is also applicable to private users.

As we see, encrypting plaintext e-mails or calendar entries has useful applications and should be supported by the e-mail clients.

*Decrypting ciphertext e-mails / calendar entries:* At first glance, the permanent decryption of ciphertext seems unreasonable, in particular if the information is still to be protected. An applicable example is a local storage which is considered secure (for example if hard disk encryption is applied). By decrypting e-mails / calendar entries all cleartext based features of the e-mail client or additional software (such as spam or malware scanners) can be used, while the transport is protected by encryption[4]. If this is to be considered secure, it has to be made sure that the e-mail server / calendar server and the connection to the server is also secure, as local decryption might lead to decryption on the server (e.g. when using IMAP folders). In this scenario, automatic decryption of all stored e-mails by the e-mail client might be applied. It is important that the e-mail is per default encrypted again when it is forwarded or the user replies to it.

Another scenario where permanently decrypted e-mails or calendar entries might be appropriate is when a user leaves the company and has to deliver the information to his successor or superior. He can pass the information on without revealing his private key. A similar situation occurs when the user is forced by court order to reveal his e-mail communications or business appointments.

As we see, permanently decrypting ciphertext e-mails or calendar entries has reasonable scenarios. A further (practicability related) scenario is restoring accidentally encrypted plaintext e-mails.

### D. Header Encryption

One major drawback that might lead to serious information disclosure is the following characteristic of e-mail encryption standards: Even if the body is encrypted, all header information including the subject is sent in plaintext. If a user

[4] This is exactly what happens in the upcoming De-Mail solution (cf. Section IV).

does not consider this, he might unwillingly disclose sensitive information. We discuss two different mechanisms to realize header encryption.

*Enveloping:* [27, 3.1] proposes a way to secure headers, especially the subject, of e-mails. The proposal is from the S/MIME standard but is also applicable to PGP/INLINE and PGP/MIME encrypted e-mails. In order to protect the header fields, the sending client wraps a full MIME message in a message/rfc822 [7] wrapper in order to apply security services to the header fields. This is possible since the entire e-mail (header and body) consists of printable characters and thus, can be handled as a message body itself. As discussed in [22], this approach comes with some drawbacks like duplicated header fields. According to our knowledge, no commonly used e-mail client implements this feature.

*Field Encoding:* The second possibility to secure header fields is to encrypt them separately one by one. The subject header field, for example, can be encrypted using standard S/MIME or PGP mechanisms. As both produce Base64 encoded ciphertexts, these can be stored in the header field instead of the original content. Before sending an e-mail the user can decide to encrypt for example the subject. The decisions may also be linked, so that the subject is encrypted as soon as the e-mail is encrypted, using the same encryption mechanism. The encryption and decryption has to be performed transparently. Obviously, the encrypted subject has to be handled the same way as encrypted bodies (cf. Subsection III-A). Due to the lack of standardization both sender and receiver have to make sure to be interoperable. Otherwise, the receivers e-mail client may display the Base64 encoded ciphertext as subject.

### *E. Future of Encrypted E-Mails*

Contemporary e-mail clients support encryption algorithms that meet the security demands of the near future. But new algorithms – e.g. post-quantum algorithms – are developed. For long-term security it should be possible to react to the progress in cryptographic research by including new algorithms in a flexible way. We created a proof of concept called ThunderCrypt [2] to provide the possibility to use cutting edge cryptographic algorithms in e-mail communication, but we do not cover this topic in the work at hand.

## IV. Related Work and Comparison

To our knowledge there is no previous work that deals with equal treatment of encrypted and plaintext e-mail communication, but there are some other approaches that offer improved handling of encrypted e-mails. We analyze the approaches which try to solve some of the identified problems. While our approach enhances the practicability and security and complies with the given standards and protocols as far as possible, other approaches deviate from existing standards.

### *A. Client Site Approaches*

One possibility to overcome the described shortcomings of deployed encryption mechanisms is to use enhanced cryptographic schemes. We present two client site approaches that cover only certain aspects of our main approach.

**Opportunistic Encryption:** A very convenient way of avoiding practicability problems in e-mail encryption is to use so-called Streams with opportunistic encryption as proposed in [14]. Using opportunistic encryption, the e-mail client tries to find a key matching the receivers e-mail address. If the e-mail client is not able to find any matching key it falls back to the unencrypted mode. This is completely transparent for the user. The proposed Streams additionally create a sanitized e-mail header and encapsulate the original one in the e-mail to protect the subject.

From the practicability point of view, opportunistic encryption is very good in the sense that the encryption process does not interfere with the functionality. As it spares the user to interact with the encryption at all, the user does not need any knowledge of the underlying cryptography which is good for practicability. But regarding the security, opportunistic encryption is disastrous. The user does not have control over the actual security measures applied to his message. Furthermore, even if transferred encrypted, e-mails are stored in plaintext on the receiver's device, so that they are locally vulnerable. Thus, opportunistic encryption meets most of our criteria but lacks the standard support and also has some privacy issues.

**Encryption with Keyword Search:** Another proposal that covers only one aspect of our approach is the possibility to search for predefined keywords (tags) in encrypted e-mails [4], [3]. The focus is to hide all information from a third party (for example the e-mail server), which is nonetheless able to fulfill requests based on the tags. The approach does not allow to perform full text searches and is not standardized.

### *B. Server Site Approaches*

Besides the possibilities to use enhanced cryptographic mechanisms in e-mail clients (on user site), there are also techniques where the security is achieved by server side measures (on provider site). We present De-Mail [10] as an example. It is an e-mail project promoted by the German federal government to make e-mail transfer secure and legally binding. Assuming that Alice wants to send an e-mail to Bob using De-Mail, she logs into her De-Mail account to send the e-mail without performing any additional steps. The connection between Alice and her e-mail provider is secured by standard mechanisms (e.g. mutual authenticated SSL/TLS [8]). The e-mail provider encrypts the e-mail and sends it (using a secure channel) to Bob's e-mail provider. Bob's e-mail provider decrypts the message and checks its integrity before storing it in Bob's inbox. When Bob now checks his e-mails (using a secure connection), he sees Alice's e-mail and can be sure that Alice wrote it and nobody else has been able to read or modify its content (if he trusts the De-Mail provider).

A major concern regarding such techniques is the privacy of the data, since the involved e-mail providers are able to read all messages in plaintext. Thus, the user has to put unconditional trust into all involved De-Mail providers. This trust is to be established by a certification process supervised by the German Federal Office for Information Security (Bundesamt für Sicherheit in der Informationstechnik, BSI) [12], [9]. Furthermore, only the transport is secured. The e-mail is not secured locally and on the servers of the provider. Registration is possible since July 2010, the service is available for corporate customers since march 2012 and some providers state to start the service for private customers in 2012. Apart from a small pilot scheme in 2010, there are no practical experiences or studies on the ease of use yet.

### C. Related Work & Conclusion

There is no implementation of opportunistic encryption available for current mainstream e-mail clients. Even if it was available, both sender and receiver had to use the same plugin as opportunistic encryption is not standardized. The same applies for encryption with keyword search.

Despite the fact that De-Mail service providers most likely will charge the users for sending e-mails, it is not possible to send or receive a De-Mail with a regular e-mail account since the e-mail provider is not able to process the De-Mail.

Therefore, none of these approaches are satisfying. They do not solve the practicability problems depicted in Section II or miss to use standardized protocols.

## V. CryptoBird

As a proof of concept we implemented features described in Section III as a Thunderbird add-on called CryptoBird. In the following we describe the core features and how they implement the mechanisms described in Section III to make a step towards equal treatment of encrypted and plaintext e-mails. Thunderbird with the Enigmail and Lightning add-ons offers the basic functionality to handle plaintext and encrypted e-mails as well as calendar entries, which we extended to fulfill the following criteria:

### A. Searching in encrypted e-mails

One of CryptoBird's core features is to allow the user to conduct full text searches in the bodies of encrypted e-mails over entire folders or mail boxes as introduced in Section III-A. Thereby, it supports all common standard mechanisms, namely S/MIME, PGP/INLINE and PGP/MIME.

To realize the search in encrypted e-mails the secure database software Derby[5] is used. The software allows password based encrypted storage of all database contents while still providing the common database functionalities, especially efficient data querying. Upon arrival of a new encrypted email, it is indexed and stored in the database. CryptoBird traces e-mails throughout their life cycle. Moving e-mails from one folder to another one or moving entire folders, cause the database to be updated so that the search results remain valid. Deleting an e-mail will also delete the according database entry. With CryptoBird, not only plaintext e-mails but also encrypted e-mails are considered when searching for a keyword using Thunderbird's standard search feature. Therefore not only the internal Thunderbird database is queried but also our password protected database.

### B. Encrypted Calendar Entries

Using the e-mail client as personal information manager (PIM), a calendar feature is integrated in the e-mail client (in the case of Thunderbird Lightning undertakes this task). Converting an e-mail into an event or task works fine with plaintext e-mails. But encrypted e-mails are only decrypted on the fly to display. Converting an encrpyted e-mail to an event or task, Lightning inserts the ciphertext as description in the case of PGP/INLINE encrypted e-mails. S/MIME and PGP/MIME encrypted e-mails lead to empty events and tasks. CryptoBird offers the possibility to convert at least PGP/INLINE encrypted e-mails to calendar events (cf. Subsection III-B) for now. The event or task can be stored in plaintext or encrypted. Encrypted events and tasks lead to the same problem as with encrypted e-mails. To overcome this, CryptoBird decrypts events and tasks on the fly, similar to the decryption of e-mails, and stores them in the database for searching.

### C. Header Encryption

CryptoBird does not only enhance the feature range for encrypted e-mails, but also brings additional privacy features. To solve the privacy problems of plaintext e-mail subjects in encrypted e-mails, CryptoBird allows to encrypt the subject field using PGP as explained in Subsection III-D using the encoding technique. To ensure equal treatment, the decrypted subjects are also stored in the encrypted database to consider them in the keyword search.

### D. CryptoBird Conclusion

Having the criteria from Section II in mind, CryptoBird proofs the possibility of fulfilling all five criteria. As CryptoBird complies to the established standards, the regular e-mail encryption features stay interoperable with standard e-mail clients without CryptoBird. However, due to the lack of standardization in the case of header encryption and encrypted calendar entries, all participating parties have to install the CryptoBird extension to read encrypted subjects and calendar entries.

## VI. Conclusion and Open Issues

We saw that the practicability of security in e-mail applications is an important yet not sufficiently addressed issue. The different mainstream e-mail clients provide different degrees of practicability in this respect, but none of them provides a solution comparable to the comfort of using plaintext e-mails. We discussed the four in our eyes most pressing practicability issues namely searching in encrypted content, calendar integration, toggling the encryption, and header encryption. We showed how it is possible to solve them while sticking to the most common standards, thereby guaranteeing interoperability. As a proof of concept we presented CryptoBird, a prototypical Thunderbird add-on implementation.

[5] http://db.apache.org/derby/

An open issue, regarding the subject and calendar encryption, is the lack of standardization. An adaption of the according RFCs may be reasonable here. Further projects should increase the practicability of CryptoBird by implementing the remaining not yet implemented features described in Section III. An indexed database may also be desirable for efficiency reasons. An integration of the PGP related features of CryptoBird into Enigmail seems be reasonable. It is also thinkable to integrate CryptoBird's S/MIME enhancements into Enigmail. In order to evaluate the usability enhancements CryptoBird provides, conducting a user study would be the next step.

## References

[1] Apple Inc. Apple mail. http://www.apple.com/macosx/whats-new/mail.html, 2012. Accessed: 18/02/2012.

[2] S. Arzt. Design and implementation of a cryptographic plugin for e-mail clients, Nov 2009. Whitepaper.

[3] J. Baek, R. Safiavi-Naini, and W. Susilo. Public Key Encryption with keyword Search Revisited. In *International conference on Computational Science and Its Applications*, pages 1249–1259. Springer, 2008.

[4] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano. Public Key Encryption with keyword Search. In *Advances in Cryptology Eurocrypt*, volume 3027, pages 506–522. Springer, 2004.

[5] J. Callas, L. Donnerhacke, H. Finney, D. Shaw, and R. Thayer. OpenPGP Message Format. RFC 4880 (Proposed Standard), November 2007. Updated by RFC 5581.

[6] M. Cooper, Y. Dzambasow, P. Hesse, S. Joseph, and R. Nicholas. Internet X.509 Public Key Infrastructure: Certification Path Building. RFC 4158 (Informational), Sept. 2005.

[7] D. Crocker. Standard for the format of ARPA Internet Text Messages. RFC 822 (Standard), Aug. 1982. Obsoleted by RFC 2822, updated by RFCs 1123, 2156, 1327, 1138, 1148.

[8] T. Dierks and E. Rescorla. The Transport Layer Security (TLS) Protocol Version 1.2. RFC 5246 (Proposed Standard), Aug. 2008. Updated by RFCs 5746, 5878, 6176.

[9] Federal Office for Information Security (Bundesamt für Sicherheit in der Informationstechnik). Akkreditierung von de-mail-diensteanbietern, 2011. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Egovernment/De_Mail/De-Mail-Akkreditierung-Prozessuebersicht.pdf?__blob=publicationFile.

[10] Federal Office for Information Security (Bundesamt für Sicherheit in der Informationstechnik). BSI - Technische Richtlinie DE-Mail. Technical Directive (BSI-TR-01201), Version 1.00, 2011. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/TechnischeRichtlinien/De_Mail/TR_De_Mail_pdf.pdf?__blob=publicationFile.

[11] Federal Office for Information Security (Bundesamt für Sicherheit in der Informationstechnik). De-mail - eine infrastruktur für sichere kommunikation. https://www.bsi.bund.de/DE/Themen/EGovernment/DeMail/DeMail_node.html, 2012. Accessed: 18/02/2012.

[12] F. O. for Information Security (Bundesamt für Sicherheit in der Informationstechnik). Verfahrensbeschreibung zur akkreditierung von de-mail-diensteanbietern, 2011. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Egovernment/De_Mail/Verfahrensbeschreibung-zur-Akkreditierung-De-Mail-Dienstanbieter.pdf?__blob=publicationFile.

[13] Free Software Foundation Inc. The GNU Privacy Guard. http://www.gnupg.org/, 2012. Accessed: 25/02/2012.

[14] S. Garfinkel. Enabling email confidentiality through the use of opportunistic encryption. In *Proceedings of the 2003 annual national conference on Digital government research*, pages 1–4. Digital Government Society of North America, 2003.

[15] S. Gaw, E. W. Felten, and F.-K. Patricia. Secrecy, flagging, and paranoia: adoption criteria in encrypted email. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 591–600. ACM, 2006.

[16] Google. Chrome. http://www.google.com/chrome, 2012. Accessed: 18/02/2012.

[17] Google Inc. Gmail. http://gmail.google.com/mail/, 2012. Accessed: 14/02/2012.

[18] GPGMail. Gpgmail. http://www.gpgtools.org/gpgmail/index.html, 2012. Accessed: 14/02/2012.

[19] O. Inc. ISecurity Industry Market Share Analysis, Sept. 2011.

[20] Intevation GmbH. Gpg4win - a secure solution... http://www.gpg4win.org/index.html, 2012. Accessed: 11/02/2012.

[21] J. Klensin. Simple Mail Transfer Protocol. RFC 5321 (Draft Standard), Oct. 2008.

[22] L. Liao and J. Schwenk. Header protection for s/mime draft-liao-smimeheaderprotect-05. http://tools.ietf.org/html/draft-liao-smimeheaderprotect-05, 2009. Accessed: 18/02/2012.

[23] D. G. T. Markotten. *Benutzbare Sicherheit in informationstechnischen Systemen*, chapter 4. Rhombos Verlag, Berlin, 2004.

[24] Microsoft Corporation. Mircosoft outlook 2010. http://office.microsoft.com/outlook/, 2012. Accessed: 18/02/2012.

[25] Novell Inc. Evolution. http://projects.gnome.org/evolution/, 2012. Accessed: 18/02/2012.

[26] Penango Inc. Penango. http://www.penango.com/index.html, 2012. Accessed: 18/02/2012.

[27] B. Ramsdell and S. Turner. Secure/Multipurpose Internet Mail Extensions (S/MIME) Version 3.2 Message Specification. RFC 5751 (Proposed Standard), Jan. 2010.

[28] recurity.com. Recurity. http://www.recurity.com/, 2012. Accessed: 18/02/2012.

[29] S. Sheng, L. Broderick, C. A. Koranda, and J. J. Hyland. Why Johnny Still Can't Encrypt: Evaluating the Usability of Email Encryption Software. In *Symposium On Usable Privacy and Security*, 2006.

[30] Symantec Corporation. Pgp desktop email - email encryption software for desktop and laptop computers. http://www.symantec.com/business/desktop-email, 2012. Accessed: 02/02/2012.

[31] The Enigmail Project & mozdev.org. Enigmail - openpgp email security for mozilla applications type. http://enigmail.mozdev.org, 2012. Accessed: 26/01/2012.

[32] S. Whittaker, T. Matthews, J. Cerruti, H. Badenes, and J. Tang. Am I wasting my time organizing email?: a study of email refinding. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, pages 3449–3458. ACM, 2011.

[33] A. Whitten and J. Tygar. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. In *8th USENIX Security Symposium*, 1999.

[34] Zimmermann, Philip R. *The official PGP user's guide*. MIT Press, Cambridge, MA, USA, 1995.

# A Method For Engineering Secure Control Systems With Application To Critical Infrastructures

**James Nutaro, Ike Patterson, Glenn Allgood,**
**Teja Kuruganti, and David Fugate**
**Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA**

**Abstract**—*The protection of physical infrastructure from cyber-attacks is addressed only in part by what are typically thought of as cyber-security controls: strong passwords, encrypted data, and other similar security measures. These defensive measures reduce the likelihood that the computing and communication components will be breached, but do not enable a system to operate in spite of that breach. When a breach does occur, it is necessary to react in a time commensurate with the physical dynamics of the system as it responds to the attack. Failure to act swiftly enough may result in undesirable, and possibly irreversible, physical effects. When assessing the security of a cyber-physical system, it is therefore necessary to understand how physical dynamics and cyber-security solutions collectively determine the outcome of a cyber-event. We outline the first steps towards a method for assessing the physical risks posed by a cyber-attack, assessing the impact of cyber-security solutions on those risks, and using this information to inform both design and investment decisions. We illustrate the proposed method with a model of a chiller system based on the super-computer chillers at Oak Ridge National Laboratory.*

**Keywords:** cyber-security, risk, modeling, simulation

## 1. Introduction

The protection of physical infrastructure from attacks on its computing and communication components is addressed only in part by what are typically thought of as cyber-security controls: strong passwords, process white-lists, encrypted data, authentication of a message's source, and similar security controls. These defensive measures reduce the likelihood that the computing and communication components will be breached, but do not enable a system to operate in spite of that breach. When these security controls fail, it is necessary to detect the breach, understand its intent and purpose, formulate a plan for response, and then act on that plan to mitigate the consequences of the attack. This must be done in a time commensurate with the dynamics of the physical system to prevent unwanted physical effects. Fig. 1 illustrates these elements of a cyber-physical security event.

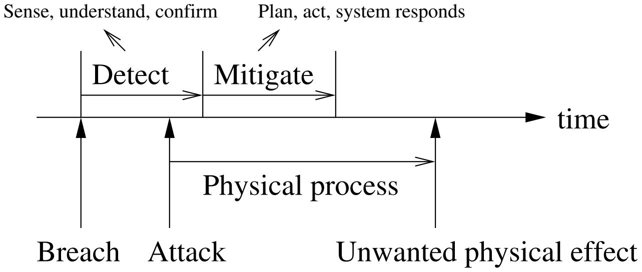


Fig. 1: Time-line of a cyber-attack.

When investing in a system's security, it is necessary to consider the costs and rewards of improving each aspect of defense: prevention, detection, mitigation, and time available for action. Most work on cyber-security emphasizes methods and technologies for prevention, detection, or mitigation. In this paper, we focus on how the effectiveness of each and the dynamics of the physical system together determine the likelihood of an undesired outcome and how this information may be used to set requirements for the performance of security functions. We follow an approach similar to that described by Huang et. al. [1], but here with the purpose of quantifying the relationship between performance (and, therefore, cost) of a security system and the likelihood of an undesired outcome.

To illustrate the proposed method, we model a chiller system based on the super-computer chillers at Oak Ridge National Laboratory. We focus on a particular cyber-attack: the modification of temperature data used by the chiller's control system. A thermodynamic model of the chiller system is combined with a discrete, stochastic model of its cyber-defenses to show how the likelihood of a system failure is collectively determined by the timeliness and effectiveness of its prevention, detection, and mitigation mechanisms given constraints posed by the underlying thermodynamic process.

## 2. Modeling performance and outcomes

Our method is based on an intuitively appealing decomposition of a cyber-security event (see, e.g., [2]). This decomposition has four elements: an attempt to breach security, the detection of a successful breach, the mitigation of that breach, and the consequences of failing to prevent, or to detect and mitigate, the attack. A finer decomposition may be necessary for practical applications. For example, the act of detecting has several parts: sensing of raw data, understanding that data and its significance, and confirming the understanding. Mitigation may likewise be decomposed into planning, acting, and waiting for the end effect of that action.

Nonetheless, in this exposition of the method we will restrict ourselves to the aggregate performance of three stages: prevention, detection, and mitigation. Performance, in this case, is the likelihood that the security system will successfully perform its function at each stage and that this success will occur before damage is done to the system. The chance of a successful defense is therefore critically dependent on the speed with which the security controls act and the rates at which physical effects may be induced.

For any attack to be successful, it must first breach the outer defenses of a system: by cracking a password, installing a rootkit, or by some other means (see, e.g., [3], [4]). In the simplest case, the effectiveness of the outer defenses may be modeled by a random variable $\mathbf{p}$ (for prevent) with outcomes 'breached' and 'not breached'. The greater the probability of a 'not breached' outcome, the more secure the system.

If a breach is successful, it must be detected before mitigating action can be taken. The time to detect a breach is of primary importance. In the simplest case, the effectiveness of the detection system may be modeled by a random variable $\mathbf{d}$ (for detect) whose outcome is a positive number indicating the time until the breach is found and reported. The more quickly the breach is detected, the more secure the system.

After the breach is detected, its physical consequences must be mitigated. As with detection, the time to mitigate is of primary importance. Therefore, the time to mitigate may also be modeled by a random variable $\mathbf{m}$ (for mitigate) whose outcome is a positive number indicating the time until the mitigation strategy blocks the progress of the attack. The more quickly this happens, the more secure the system.

The detect and mitigate phases of a defense must be completed before the attack causes an undesired physical effect. The rate at which physical effects occur is captured by models of the physical system being protected. For the chiller system, this is a thermodynamic model that describes the rates of heat accumulation and dissipation. These rates place an upper limit on the time $t_f$ in which the security system must act: if the security system is too slow, then it will fail to prevent undesired physical outcomes.

With these quantities, the probability $U$ (for undesirable) of a cyber-attack causing an undesirable physical event is given by

$$U = P\{\mathbf{p} = \text{breach}\} \times P\{\mathbf{d} + \mathbf{m} \geq t_f\} \tag{1}$$

Conversely, the probability $S$ of a successful defense is

$$\begin{aligned} S =& 1 - U = P\{\mathbf{p} = \text{no breach}\} + \\ & P\{\mathbf{p} = \text{breach}\} \times P\{\mathbf{d} + \mathbf{m} < t_f\} \end{aligned} \tag{2}$$

With these expressions, an engineer may determine requirements for each element of the security system - prevention, detection, and mitigation - to reach a desired level of security for the physical system as expressed by $S$ or $U$. If the cost of a particular performance profile for the prevention, detection, and mitigation technologies is known, then the performance requirements may be optimized to achieve a desired level of security at the least cost or, conversely, to determine the best security that may be had for a fixed cost. The model may also be used, as it is in our example, to determine the security afforded by a particular set of performance profiles.

## 3. Thermodynamics of the chiller

Our model of the chiller focuses on the temperature regulation of the computer and the water that carries heat from the computer. A bond graph with the thermodynamic elements of this model is shown in Fig. 2 (the model used here is a simpler version of that presented in [5]). Heat enters the system in the form of electrical power dissipated by the computer's electronics. Some of this heat increases the temperature of the electronics and some of it is exchanged with the computer (hot water) side of the chiller system. The hot water transports heat to the evaporator (cold water) side of the chiller system, where heat is extracted from the inflowing water. The resulting cold water flows back into the computer, where it is heated again.

The computer's electronics, water in the computer, and water in the evaporator are modeled by thermal masses with heat capacity $C_{comp}$, $C_{hot}$, and $C_{cold}$ respectively. Their temperatures are $T_{comp}$, $T_{hot}$, and $T_{cold}$. The heat flow into the computer is determined chiefly by its electric power consumption, which is modeled by the heat flow $Q_{power}$. Heat exchange from the computer to the hot water is modeled by the constant thermal resistance $G_{xchg}$.

Transfer of heat from the hot to cold water is modulated by the mass flow rate $m_{flow}$. This flow rate is regulated to keep the computer temperature at $T_1$. Letting $c_v$ be the specific heat capacity of the water and $K_1$ the control rate for the mass flow, the thermal resistance $G_{flow}$ is given by

$$G_{flow} = (c_v m_{flow})^{-1} \tag{3}$$

$$\dot{m}_{flow} = K_1(T_{comp} - T_1) \tag{4}$$

The heat flow $Q_{evap}$ is the heat extracted by the evaporator. This heat flow is modulated at a rate of control $K_2$ to maintain the cold water temperature at $T_2$ according to

$$\dot{Q}_{evap} = K_2(T_{cold} - T_2) \tag{5}$$

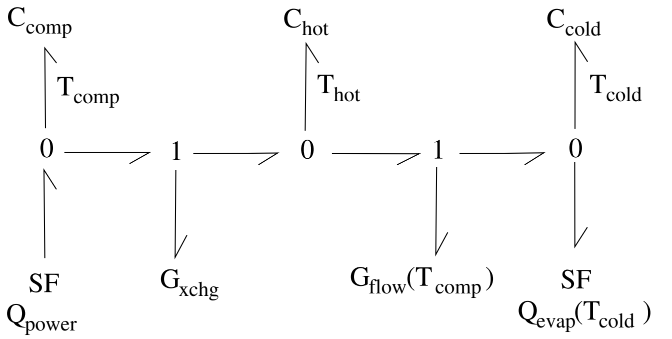


Fig. 2: Bond graph model of the chiller system.

Reading the remaining equations from the bond graph gives the system below:

$$\dot{T}_{comp} = \frac{1}{C_{comp}}\left(Q_{power} - \frac{T_{comp} - T_{hot}}{G_{xchg}}\right) \tag{6}$$

$$\dot{T}_{hot} = \frac{1}{C_{hot}}\left(\frac{T_{comp} - T_{hot}}{G_{xchg}} - (T_{hot} - T_{cold})c_v m_{flow}\right) \tag{7}$$

$$\dot{m}_{flow} = K_1(T_{comp} - T_1) \tag{8}$$

$$\dot{T}_{cold} = \frac{1}{C_{cold}}\left((T_{hot} - T_{cold})c_v m_{flow} - Q_{evap}\right) \tag{9}$$

$$\dot{Q}_{evap} = K_2(T_{cold} - T_2) \tag{10}$$

## 4. Cyber-security of the chiller

The security of the chiller has three aspects: the probability of preventing a breach, the time to detect a breach, and the time to mitigate the breach (i.e., to respond and for the chiller to react to that response). For the sake of illustration, these are modeled in the following way. The probability of preventing a breach is a uniform random variable $\mathbf{p}$ with $\mathbf{p} \leq p$ indicating that an attempted breach has been prevented.

We assume that the attacker immediately exploits a breach to manipulate the temperature sensor for the computer's electronics. This is done by adding $\Delta T$ to the sensor's reports; i.e,. so that in Eqn. 8 is replaced by

$$\dot{m}_{flow} = K_1(T_{comp} - (\Delta T + T_1)) \tag{11}$$

This is equivalent to raising the set point from $T_1$ to $T_1 + \Delta T$, and if this quantity is larger than the shutdown temperature $T_{shutdown}$ then it will lead eventually to the shutdown of the computer.

Given that a breach has occurred, the time to detect the breach is an exponentially distributed random variable $\mathbf{d}$ with mean time to detect $\mu_d$. Similarly, given that the breach has been detected, the time to mitigate the breach is an exponentially distributed random variable $\mathbf{m}$ with mean time to mitigate $\mu_m$. Mitigation corrects the sensor readings by setting $\Delta T = 0$.

Table 1: Parameters values used for the chiller model.

| | | | |
|---|---|---|---|
| $C_{comp}$ | $10^4$ J / K | good $p$ | 0.8 |
| $Q_{power}$ | $10^7$ W | mediocre $p$ | 0.2 |
| $G_{xchg}$ | $10^{-6}$ K / W | mediocre $\mu_m$ | 100 s |
| $C_{hot}$ | 1000 kg $\cdot\, c_v$ | good $\mu_m$ | 50 s |
| $c_v$ | 4186 J / kg K | mediocre $\mu_d$ | 200 s |
| $K_1$ | 0.05 kg / s$^2$ K | good $\mu_d$ | 100 s |
| $C_{cold}$ | 1000 kg $\cdot\, c_v$ | $\Delta T$ | 11 K |
| $K_2$ | 0.005 W / s K | $T_{shutdown}$ | 325 K |
| $T_2$ | 284.1 K | $t_f$ | 350.5 s |
| $T_1$ | 315 K | | |

Assuming a breach has been successful, the thermodynamic model is used to calculate the time $t_f$ to shutdown from the initial thermodynamic equilibrium and using a particular $\Delta T$. Using Eqn. 1 and the models described above, the probability of an undesired outcome is (see, e.g., [6])

$$U = (1-p)F(t_f) \tag{12}$$

$$F(t_f) = \begin{cases} \left(1 + \dfrac{t_f}{\mu_d}\right)e^{-t_f/\mu_d} & \text{if } \mu_d = \mu_m \\ \left(\dfrac{\mu_d}{\mu_d - \mu_m}\right)e^{-t_f/\mu_d} + \left(\dfrac{\mu_m}{\mu_m - \mu_d}\right)e^{-t_f/\mu_m} & \text{if } \mu_d \neq \mu_m \end{cases} \tag{13}$$

This probability may be improved by altering any of the terms that are in the purview of the designer. The term $t_f$ may be altered by changing the thermodynamic properties of the chiller; e.g., by increasing the mass of water. The terms $\mathbf{d}$, $\mathbf{m}$, and $\mathbf{p}$ may be improved by additional spending on security measures; e..g, with intrusion detection tools to improve $\mathbf{d}$, new types of operator training or control systems to improve $\mathbf{m}$, and more sophisticated authentication systems to improve $\mathbf{p}$.

## 5. Numerical example

Table 1 shows the parameter values for the thermodynamic model and the values examined for the variables in the security model. For each security parameter $\mu_m$, $\mu_d$, and $p$ we consider "good" and "mediocre" options; these are selected to ensure some possibility of effectiveness in the context of this example and also to reflect a scenario in which mitigation is generally quicker than detection. The parameter $\Delta T$ is fixed at 11 K, which will shutdown the computer in approximately 350 seconds if the attack is not halted.

Fig. 3 shows the trajectory of the temperature $T_{comp}$ when using mean values for the "good" security parameters. This trajectory is typical of a successful defense. Following the breach, the temperature rises towards its shutdown limit.

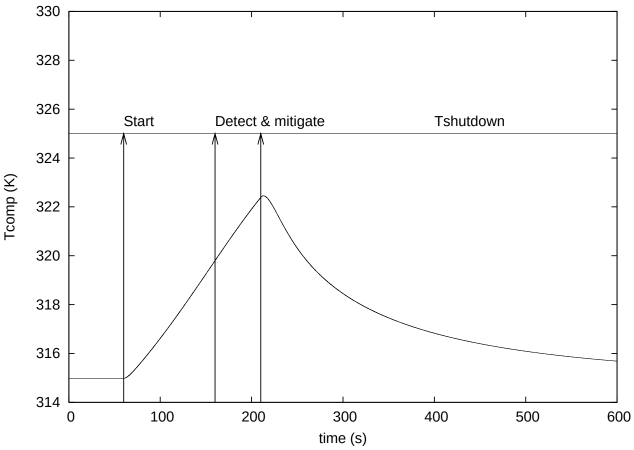


Fig. 3: The trajectory $T_{comp}$ during a successful defense.

Table 2: Probability of a shutdown for each combination of security parameters.

| Case # | Prevent | Detect | Mitigate | U |
|---|---|---|---|---|
| 1 | mediocre | mediocre | mediocre | 0.25 |
| 2 | mediocre | mediocre | good | 0.18 |
| 3 | mediocre | good | mediocre | 0.11 |
| 4 | mediocre | good | good | 0.047 |
| 5 | good | mediocre | mediocre | 0.063 |
| 6 | good | mediocre | good | 0.046 |
| 7 | good | good | mediocre | 0.027 |
| 8 | good | good | good | 0.012 |

Upon mitigation, the temperature gradually returns to its proper set point. If the defense fails, the temperature rises above the shutdown limit and forces the computer to shutdown.

Table 2 shows the probability $U$, calculated using Eqn. 12, of a shutdown for each of the eight combinations of parameter values. A comparison of cases 4 and 5 reveals that in this example a "good" technology for mitigation and detection can compensate for a "mediocre" technology for prevention and vice versa.

To see how $t_f$ and the sum $\mu_d + \mu_m$ interact to affect the security of the chiller, Eqn. 12 may be reformulated as a function of two parameters. Let

$$\mu_{min} = \min(\mu_d, \mu_m) \text{ and } \mu_{max} = \max(\mu_d, \mu_m) \quad (14)$$

and define two new parameters

$$\rho = \mu_{min}/\mu_{max} \quad (15)$$

$$\tau = t_f/\mu_{max} \quad (16)$$

Now $F(t_f)$ can be written in terms of $\rho$ and $\tau$ as

$$F(\rho, \tau) = \left(1 - \rho e^{(1-1/\rho)\tau}\right)\left(\frac{e^{-\tau}}{1-\rho}\right) \quad (17)$$

The parameter $\tau$ is in the range $[0, \infty)$ and $\rho$ is in $[0, 1]$. Fig. 17 shows the surface defined by Eqn. 17. The cases considered in Table 2 are labeled in the plot. Clearly evident in this plot is the importance of reacting quickly to a breach relative to the rate of change in the physical system; i.e., of making $\tau$ large. Indeed, for this purpose it is as effective to design for a large $t_f$ as it is to build a rapid detection or mitigation strategy; e.g., a greater mass of water improves the likelihood of the chiller operating through an attack on its temperature sensor.

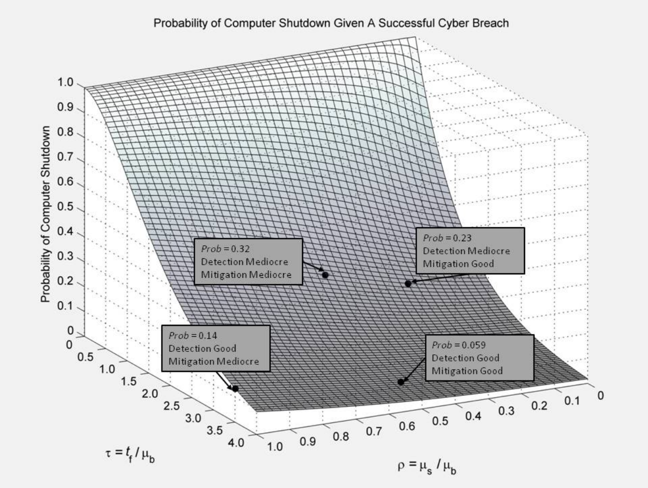


Fig. 4: The surface $F(\rho, \tau)$.

## 6. Conclusions

The proposed method for relating the performance and effectiveness of cyber-security technologies is closely related to simulation methods used for reliability engineering. Significantly, we do not propose to model specific security solutions. Rather, we relate the effectiveness of any solution meeting a particular performance criteria to the risk posed by an attack. In this way, the proposed method may serve as a tool for system engineers when determining requirements for cyber-security in the early stages of a system's life-cycle.

To fully understand the effect that cyber-attacks have on a system's performance requires investigating the impact that humans have as part of the overall, integrated control process. As such, we plan to develop an operator in the loop model and integrate this with our process model to exercise varying tempo, data pedigree, and operator decision costs. Understanding where the human touches the system and the artifacts of such are critical in any controller design, especially when considering system resiliency.

Therefore, our future research will include (i) developing a formal representation of a control action process and identifying the points where the human operator touches the system; (ii) developing a probabilistic model of the operator's decision making process that captures decisions under stress and the time dilation that accompanies it; and (iii) formalizing a procedure and protocol for capturing and analyzing these results that are consistent with industry needs and understanding.

An obstacle to the application of the proposed method is the difficulty of determining the probabilities that describe the security measures and, in future work, the decision making processes of an operator who is responding to a cyber-event. In many cases, experience with particular security measures that are being considered may permit the construction of useful empirical models, but this remains an important area for research. Nonetheless, our work highlights the importance of such models for engineering secure systems, and we hope our work will motivate broad interest in the construction of these models and their applications to cyber-security.

## References

[1] Y.-L. Huang, A. A. CÃąrdenas, S. Amin, Z.-S. Lin, H.-Y. Tsai, and S. Sastry, “Understanding the physical and economic consequences of attacks on control systems,” *International Journal of Critical Infrastructure Protection*, vol. 2, pp. 73–83, 2009.

[2] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, and S. Sastry, “Attacks against process control systems: risk assessment, detection, and response,” in *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, ser. ASIACCS ’11, 2011, pp. 355–366. [Online]. Available: http://doi.acm.org/10.1145/1966913.1966959

[3] A. A. Cárdenas, S. Amin, and S. Sastry, “Secure control: Towards survivable cyber-physical systems,” in *The 28th International Conference on Distributed Computing Systems Workshop*, 2008.

[4] T. T. Kim and H. V. Poor, “Strategic protection against data injection attacks on power grids,” *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 326–333, June 2011.

[5] M. Albieri, A. Beghi, C. Bodo, and L. Cecchinato, “A simulation environment for the design of advanced chiller control systems,” in *IEEE International Conference on Automation Science and Engineering, 2007. (CASE 2007)*, Sept. 2007, pp. 962 –967.

[6] S. Amari and R. Misra, “Closed-form expressions for distribution of sum of exponential random variables,” *IEEE Transactions on Reliability*, vol. 46, no. 4, pp. 519 –522, December 1997.

# SWM: Simplified Wu-Manber for GPU-based Deep Packet Inspection

Lucas Vespa
Department of Computer Science
University of Illinois at Springfield
lvesp2@uis.edu

Ning Weng
Department of Electrical and Computer Engineering
Southern Illinois University Carbondale
nweng@siu.edu

*Abstract*—**Graphics processing units (GPU) have potential to speed up deep packet inspection (DPI) by processing many packets in parallel. However, popular methods of DPI such as deterministic finite automata are limited because they are single stride. Alternatively, the complexity of multiple stride methods is not appropriate for the SIMD operation of a GPU. In this work we present SWM, a simplified, multiple stride, Wu-Manber like algorithm for GPU-based deep packet inspection. SWM uses a novel method to group patterns such that the shift tables are simplified and therefore appropriate for SIMD operation. This novel grouping of patterns has many benefits including eliminating the need for hashing, allowing processing on non-fixed pattern lengths, eliminating sequential pattern comparison and allowing shift tables to fit into the small on-chip memories of GPU stream cores. We show that SWM achieves 2 Gb/s deep packet inspection even on a single GPU with only 32 stream cores. We expect that this will increase proportionally with additional stream cores which number in the hundreds to thousands on higher end GPUs.**

## I. Introduction

Throughput requirements continue to increase for network applications and services. Deep packet inspection (DPI) has the most strenuous throughput requirement and is a key component of network intrusion detection systems [1], [2], [3], [4], [5]. DPI scans payloads for the presence of known attack patterns [6]. Increasing DPI speed can alleviate the bottleneck behavior that current payload search systems impose.

Graphics processing units (GPU) [7] have potential to speed up DPI by processing many packets in parallel. However, the SIMD configuration of GPU processing cores require simplicity for efficient utilization and fast kernel execution. Although deterministic finite automata (DFA) have been implemented in a GPU [8], [9], [10] due to their simplicity of operation, state transition tables require excessive memory and one global memory access for every packet byte processed in the worst case. Therefore, reduced memory algorithms [11], [12] have been used to allow DFA to be efficiently implemented in a GPU [13], however, DFA algorithms are single stride, which limits their speedup capability.

Methods to increase stride [14], [15], [16], [17] have been developed, however, the complexity of these algorithms is not appropriate for the SIMD operation of a GPU. For example, in the Wu-Manber [14] algorithm, when a substring that is a suffix of multiple patterns is encountered in a packet, multiple patterns must be compared to the packet text through hashing. This causes great divergence among processors in an SIMD arrangement, and thus a significant performance degradation.

In this work we present a simplified Wu-Manber like algorithm called SWM, which uses a novel method to group patterns such that the shift tables are simplified and the algorithm becomes appropriate for SIMD operation. Specifically, our grouping method simplifies the algorithm resulting in the following properties. Multiple pattern comparisons never occur, allowing for SIMD operation without path divergence between stream cores. This also allows for the use of shift tables which include the entire length of all patterns, rather than truncating the patterns when creating shift tables. Our pattern groupings also allow the use of 2-byte substrings for the shift tables which creates two other benefits. First, the shift tables are very small and can be stored directly in the local memories of the GPU stream cores which improves performance and determinism. Second, 2-byte substrings can be direct indexed which removes the need to hash the packet text. This also improves performance by removing the hash calculation entirely.

We further optimize SWM for the VLIW arrangement of stream processors and efficient access of the global memory packet buffer. We implement SWM in an ATI Radeon GPU [18], and show that, even on a low end GPU, SWM can achieve 2 Gb/s deep packet inspection.

The remainder of this work is organized as follows. Section II discusses the SWM algorithm. Section III discusses the architecture of the GPU system. Performance analysis and experimental results are presented in Section IV and related work is covered in Section V. The paper is concluded in Section VI.

## II. SWM Algorithm

This section begins by discussing the construction of shift tables and operation of the Wu-Manber algorithm. It continues by discussing our novel method for grouping patterns, shift table construction and SWM algorithm operation. It concludes by presenting some optimizations required for efficient GPU execution.

### A. Multiple Stride Basics (Wu-Manber)

Given a list of patterns P, Wu-Manber constructs a shift table which stores a shift value (in bytes) for all B-byte substrings in the first m bytes of each pattern $p \in P$. Each pattern therefore has one shift value for $m - B + 1$ substrings. The shift value for substring s, is the position of s from the end of the pattern, subtracted from m. Here is an example where m = 5 and b = 2. In the pattern 'HELLO', the substring 'HE' occurs 2 bytes into the pattern so the shift value for 'HE' is $5 - 2 = 3$, meaning that if 'HE' is encountered in a packet, we can shift forward 3 bytes. In the pattern 'HELLO', the substring 'LO' occurs 5 bytes into the pattern so the shift value for 'LO' is $5 - 5 = 0$, meaning that, if 'LO' is encountered in a packet we cannot shift forward because we could potentially pass over the string 'HELLO'. Shift table entries are created for all B-byte substrings, including those that do not exist in any pattern. The stride value for substrings that do not exist in a pattern is $m - B + 1$.

Shift table operation begins by examining B-bytes of a packet starting at offset $m-B+1$. This B-byte value is looked up in the shift table. If the shift value found is non-zero, then we simply stride in the packet by this shift value and examine the B-bytes at this next location. If the shift value is zero then the packet text must be compared to any patterns that share this B-byte substring as a suffix. This could potentially be many patterns that need sequential comparison. This is the most time consuming part of the algorithm. Wu-Manber uses several methods to help with this problem but none are appropriate for a GPU-based multiple stride algorithm. The following are the methods used by Wu-Manber:

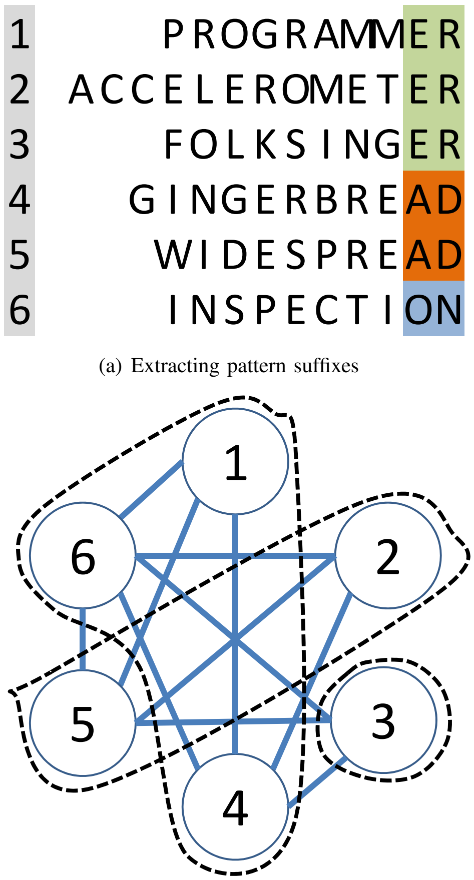
- Using a larger value for B helps reduce the number of patterns that share suffixes, but this also requires more memory for the shift tables
- The shift tables for larger values of B utilize a hash table to reduce memory, but this requires calculating a hash value for each B-bytes examined in the packet. Also, this reduces the average stride because substrings that hash to the same value must store the minimum stride value
- Hashing is used to reduce the time to compare many patterns that share suffixes

The goal of SWM is therefore to avoid these methods and produce a simpler algorithm with the following goals:

- Avoid using larger values for B so that the shift table size is small enough for GPU stream core caches
- Use a direct indexed lookup for each B-bytes to avoid the overhead of hashing and divergence between SIMD processors
- Avoid hashing when comparing patterns to the packet text in order to remove divergence between SIMD processors
- Avoid sequential pattern comparison to the packet text in order to remove divergence between SIMD processors
- Create shift tables with full patterns rather than the first m bytes of each pattern

## B. Overview of SWM

All of the aforementioned properties can be achieved by circumventing the occurrence of patterns that share suffixes. If all patterns have a unique suffix, then any time that a stride of zero occurs, the current B-bytes are known to belong to only one specific pattern. If no two patterns share the same 2-byte suffix, then B can be chosen to be two bytes and the following properties apply:



(a) Extracting pattern suffixes

(b) Compatibility graph

Fig. 1. Example of creating a compatibility graph for patterns 'PROGRAMMER', 'ACCELEROMETER', 'FOLKSINGER', 'GINGERBREAD', 'WIDESPREAD', 'INSPECTION'. Each pattern has a vertex and vertexes are joined if the pattern prefixes are different. Maximal cliques are chosen from the graph to form pattern groups.

- Shift tables for B = 2 are small enough for GPU stream core local memories
- Shift tables for B = 2 can be direct indexed
- Sequential pattern comparison and hashing are not needed to compare patterns to the packet for shift values of zero because there is a one-to-one relationship between patterns and suffixes
- Shift tables can be created for full pattern lengths because the length of the pattern associated with each shift value of zero is known

In order to remove the occurrence of patterns that share suffixes, we must group patterns such that no two patterns in a group share the same two byte suffix. Specifically, we must find the minimal number of groups such that no two patterns in a group share a suffix. This grouping takes advantage of the parallel processing capability of graphics processing units. Each stream core in a GPU can processing one group of patterns. Also, in a modern GPU, even though multiple stream cores must be used to process the full set of patterns, there are still enough stream cores to replicate these groups and process the full pattern set for many packets in parallel.

### C. SWM Pattern Grouping

We create a compatibility graph to group patterns such that no two patterns in the same group share a suffix. In the compatibility graph, each pattern is represented by a vertex. If two patterns have the same 2-byte suffix, their vertices receive no adjoining edge. On the contrary, if two patterns have different suffixes, their vertices are joined by an edge.

We group vertices together by finding maximal cliques ($k_n$ subgraphs with maximal $n$). Each clique becomes a pattern group. This is a minimal graph coloring problem so we use a graph coloring heuristic. Finding a low number of groups to cover all patterns allows greater replication of the entire pattern set in the GPU.

Figure 1 shows an example compatibility graph and corresponding groupings for patterns 'PROGRAMMER', 'ACCELEROMETER', 'FOLKSINGER', 'GINGERBREAD', 'WIDESPREAD' and 'INSPECTION'. There are three resulting pattern groups as shown in Figure 1(b). The groups are ('PROGRAMMER','GINGERBREAD','INSPECTION), ('ACCELEROMETER','WIDESPREAD') and ('FOLKSINGER').

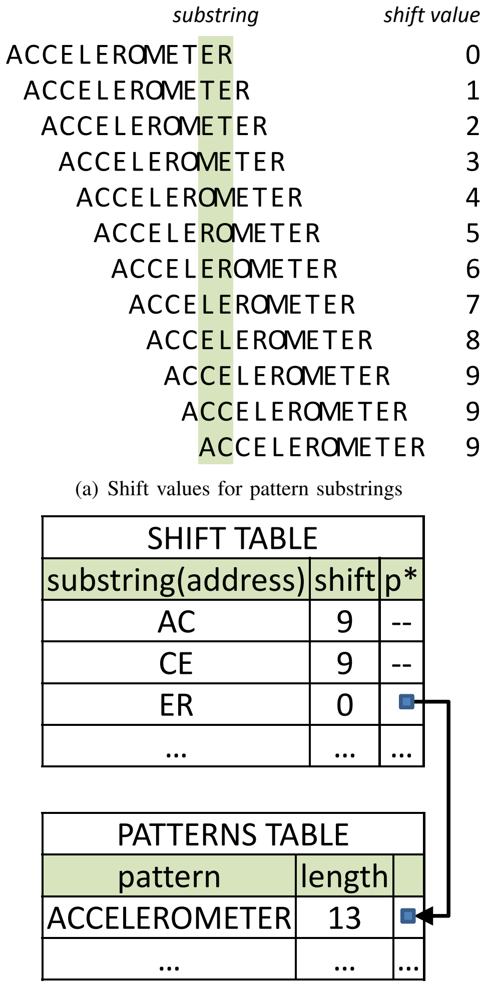
### D. SWM Shift Table Construction

We begin shift table construction by finding the number of characters $m$, in the shortest pattern. We use the value of m to derive a shift value for all 2-byte substrings within the patterns in a set. To find the shift value for any 2-byte substring we use the distance in bytes $v$ from the end of the pattern that the substring occurs. The shift value for any substring is calculated to be MIN(v, $m - B + 1$). Figure 2(a) shows the shift values for the 2-byte substrings in the pattern 'ACCELEROMETER'.

The shift values for each substring are stored in a direct indexed shift table as shown in Figure 2(b). If a substring has a shift value of zero, a pointer is added to the patterns table which contains the length and full pattern from which the substring was derived. This is used for comparing the pattern to the packet when a shift value of zero is found. Because of the pattern grouping, there will be only one pointer in the shift table for each pattern in the patterns table.

### E. SWM Operation

SWM begins by looking up the shift value for 2-bytes ($p$) of a packet starting at offset $m - B + 1$. If the shift value ($v$) is non-zero, this is used as the stride and SWM repeats the process at packet byte (p + v). If the shift value is zero then the pointer in the shift table is used to compare a pattern from the patterns table to the packet text.



(a) Shift values for pattern substrings

SHIFT TABLE

| substring(address) | shift | p* |
|---|---|---|
| AC | 9 | -- |
| CE | 9 | -- |
| ER | 0 | ■ |
| ... | ... | ... |

PATTERNS TABLE

| pattern | length | |
|---|---|---|
| ACCELEROMETER | 13 | ■ |
| ... | ... | ... |

(b) Shift tables

Fig. 2. Extracting shift values and shift table examples for the pattern 'ACCELEROMETER'.

### F. Group Balancing

Because substrings that do not exist in a pattern set receive a default stride value of m, which is the largest possible stride, pattern sets with fewer patterns may have a larger average stride. Since a packet must be processed by all pattern sets before deep packet inspection is complete for a packet, it makes sense that we should try to make the average stride for each pattern set similar, in order to equalize the processing time for each pattern set and minimize the overall latency for processing a packet. We therefore perform further refinement on our pattern groupings. After the minimum number of pattern groups is found, we attempt to equalize the number of patterns in each group without increasing the number of groups.

## III. GPU Architecture

The functionality of SWM is split between the CPU and the GPU, as illustrated by Figure 3. The following sections describe the functionality of SWM in the CPU and the GPU.

### A. CPU

As shown in Figure 3, the CPU host has several responsibilities. These responsibilities include creating the SWM shift

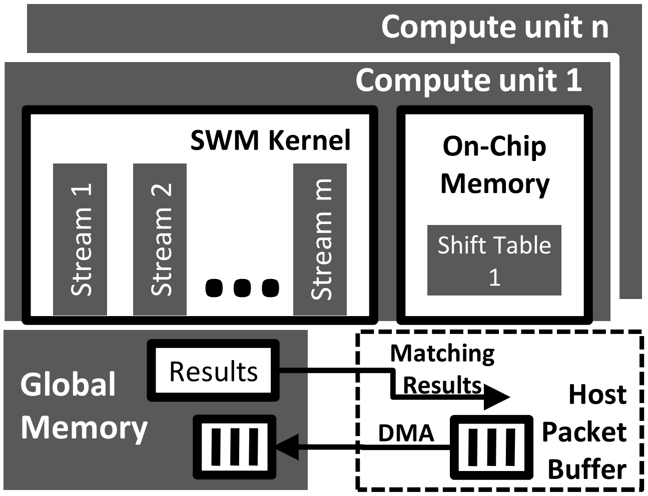


Fig. 3. SWM system architecture. Deep packet inspection is performed by the GPU where each stream core processes a separate packet and any matches are reported back to the CPU.

tables and transferring the tables to the local memory of the GPU compute units. The host maintains a current packet buffer which is mapped to the global memory of the GPU. The host also reads results from the matching buffer on the GPU and reports any potential attack patterns.

### B. GPU

The SWM kernel runs on each stream core in the GPU, as shown in Figure 3. The following are specifics about the functionality of the kernel as well as GPU memory management.

*1) Kernel:* Each stream core on the GPU has a VLIW processor. In order to more efficiently use the VLIW processors we thread multiple, non-adjacent packet sections simultaneously per work item. This increases the utilization of the individual processing elements in each stream core. Most GPUs have the ability to run more work items than available stream cores. The GPU will trade off active work items in order to help hide the latency caused by global memory accesses. The ATI Radeon HD 6450 has 32 stream cores so this is the minimum number of work items that we will run on the GPU.

*2) Memory:* The local data store (LDS) of each compute unit, and the private memory of each stream core, contain the shift tables necessary for SWM kernel operation. Local access to the shift tables allows for faster performance. Packet data is stored in a memory buffer on the CPU host. The map_buffer OpenCL command creates a mapping between this host buffer and a buffer in the GPU global memory. This mapping is used for DMA between the GPU memory and host memory. This method is faster that using a write_buffer command to explicitly write packet data from the host to the GPU global memory. The kernel requests 16 byte vectors from the global packet buffer. Fetching 16 byte vectors most efficiently utilizes the memory fetch unit, which can access 128 bits at a time.

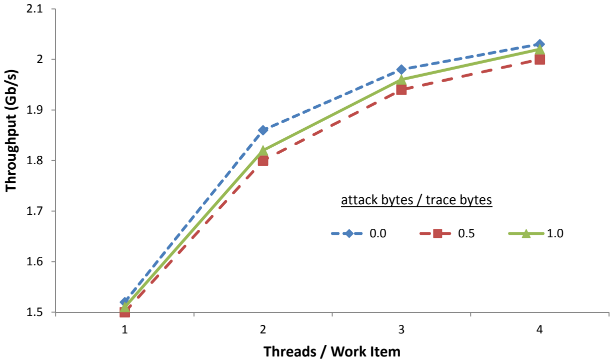


Fig. 4. Kernel performance using 1 to 4 copies of SWM sequentially per work item (kernel instance). Increasing the number of copies increases available instruction parallelism to efficiently utilize the VLIW processors. Increasing the number of payload bytes derived from attack signatures being searched also affects the performance of the SWM kernel.

## IV. Performance Analysis

In this section we evaluate the performance of SWM on our test GPU. We begin by describing the experiment hardware, followed by evaluation of the SWM kernel performance.

### A. Experiment Setup

We implement SWM on an ATI Radeon HD 6450 which has 32 stream cores and 512MB of DDR memory. Our host system contains an Intel I5 processor running at 3.3 GHZ and 8GB memory. The GPU and host interconnect via a PCIe 2.1 x 16 bus. SWM is written using Open Computing Language (OpenCL) [19] which abstracts the programming of various parallel computing devices. Using OpenCL allows SWM to be portable amongst most newer graphics processing units.

### B. Kernel Performance

In this section we evaluate the performance of SWM using different design optimizations. First, we evaluate the performance of the kernel using a varying number of copies of SWM in the kernel. Second we evaluate the performance of SWM using a varying number of work items. We also evaluate SWM by using global memory and local memory to store the state tables.

*1) Kernel Thread Optimization:* Figure 4 demonstrates the performance of SWM using an increasing number of copies of the SWM code in the kernel. We do this by copying the SWM code within the kernel code. This increases the parallel instructions available to the VLIW processors. As shown in Figure 4, the throughput increases when increasing the number of copies of SWM due to the increase in processor utilization.

Figure 4 also shows that SWM achieved a throughput of over 2 Gb/s. This throughput is achieves on a GPU with only 32 stream cores. Other GPUs have many more cores, such as the AMD 6970 which has 640 stream cores. Also observed in Figure 4, changing the payload content to contain a varying percentage of attack strings affects the throughput a minor amount.

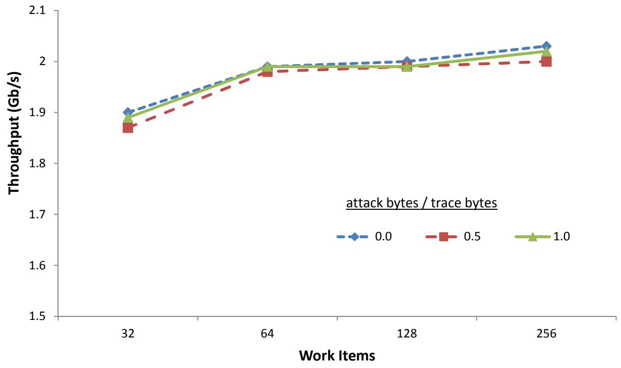


Fig. 5. Kernel performance when the number of work items is increased. Increasing the number of work items allows a stream core to execute one work item while another awaits a memory access.

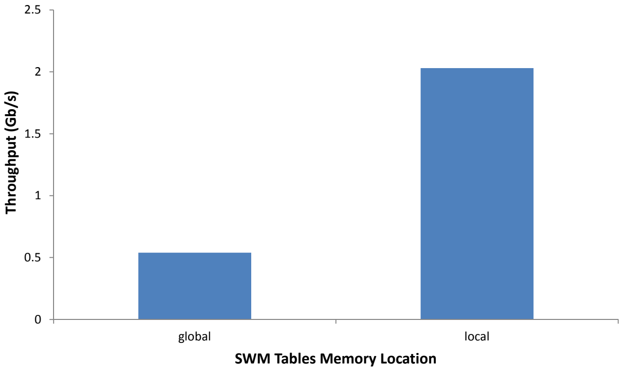


Fig. 6. Performance using global vs local memory to store the SWM shift tables on the GPU.

*2) Work Item Optimization:* Figure 5 demonstrates the effect of using a different number of work items. The lowest number of work items used is 32. With 32 work items, only one kernel instance will run on each stream core. If that stream core is waiting for a global memory access to execute an instruction, then the stream core's ALU will be idle. Increasing the number of work items allows the stream cores to execute a different kernel instance while another instance waits on a memory access. In Figure 5, we increase the number of work items from 32 up to 256. I minor increase in throughput is achieve by increasing the number of work items.

*3) Memory Optimization:* Figure 6 shows the performance of SWM when storing the shift tables in local vs global memory. As expected, storing the shift tables in local memory achieves much higher performance as this method does not suffer the same wait time as when using global memory. Overall, SWM achieves a throughput of about 2 Gb/s.

## V. Related Work

There are two main methods to accelerate deep packet inspection. These are intra-stream parallelism and inter-stream parallelism. In intra-stream parallelism, multiple, contiguous bytes of a packet are scanned simultaneously. In inter-stream parallelism, multiple packets are scanned simultaneously using multiple copies of the pattern matching engine.

Methods have been presented to exploit intra-stream parallelism to increase DPI performance. Wu and Manber [14] and derivatives [15] have produced multiple-pattern, multiple-stride average case algorithms. The complexity of these algorithms is not appropriate for GPU implementation. Brodie et al [20] increases throughput by allowing multiple DFA transitions to be traversed simultaneously. This system uses a specially designed hardware approach and is therefore limited in its implementation possibilities.

Hua et al [16] introduces a variable stride DFA (VS-DFA) which partitions patterns into variable size blocks using a fingerprinting scheme. These blocks are used to construct a multiple byte striding DFA. The same fingerprinting scheme is also used as a preprocessing step on the input source such as incoming packets. This guarantees that the correct size block of characters is fed to the VS-DFA. This preprocessing requires hashing of every byte of the packet before the input is given to the VS-DFA. The VS-DFA operation and the fingerprinting operation must be performed in parallel, again requiring special hardware.

Methods have been presented to exploit inter-stream parallelism to increase DPI performance. Commercial content inspection products use specialized hardware to accelerate pattern matching. Commercial chips such as the LSI Tarari T2000 series [21], the Cavium Networks CN1700 series [22] and the Netlogic NLS2008 [23] are advertised to achieve content inspection speeds of multiple Gb/s. Unfortunately, these are specialized hardware chips and therefore the implementation platforms are very limited.

Graphics processing units (GPU) have been used to exploit inter-stream and intra-stream parallelism. Vasiliadis et al [8], [9] have implemented deterministic finite automata (DFA) in a GPU. Unfortunately, the state transition tables have a large memory requirement. The state transition tables must be stored in global memory and require one global memory access for every byte processed in the worst case. GPEP [13] uses an optimized algorithm called P$^3$FSM which has similar complexity to a state transition table but reduces the memory requirement. However, this algorithm is limited to single stride. A GPU-based Wu-Manber modification [24] has been implemented but does not utilize the Wu-Manber shift table for packet strides and is limited to multiple Mb/s rather than Gb/s. SWM is a simplification rather than a modification of the Wu-Manber, algorithm which is most appropriate for a GPU implementation, allowing for multi-Gb/s speeds.

## VI. Conclusion

Accelerating deep packet inspection is important to the effectiveness of network security due to the continuing need to increase the number and complexity of attack signatures. Given this requirement, graphics processing units can be used to process network traffic in parallel and improve signature scanning speeds. However, GPUs have limitations in terms of the type of algorithm that can be implemented, and multiple stride deep packet inspection algorithms are not congruent with these limitations. SWM solves this congruency problem through a systematic pattern classification system. Our results indicate that SWM can achieve a consistent throughput of 2 Gb/s on a low end GPU.

## References

[1] N. Tuck, T. Sherwood, B. Calder, and G. Varghese, "Deterministic memory-efficient string matching algorithms for intrusion detection." in *Proc. of the IEEE Infocom Conference*, 2004, pp. 333–340.

[2] D. Denning, "An intrusion–detection model," *IEEE Transactions on Software Engineering*, pp. 222–232, Feb. 1987.

[3] M. Roesch, "Snort – lightweight intrusion detection for networks." in *Proc. of the 13th Systems Administration Conference*, 1999.

[4] L. Bu and J. A. Chandy, "Fpga based network intrusion detection using content addressable memories," in *FCCM: Proceedings of the 12th Annual IEEE Symposium on Field-Programmable Custom Computing Machines*, Washington, DC, USA, 2004, pp. 316–317.

[5] V. Paxson, "Bro: a system for detecting network intruders in real-time," *Computer Networks*, pp. 2435–2463, 1999.

[6] *Snort Rule Database*, http://www.snort.org/snort-rules.

[7] S. Han, K. Jang, K. Park, and S. Moon, "Packetshader: a gpu-accelerated software router," in *SIGCOMM '10: Proceedings of the ACM SIGCOMM 2010 conference*, 2010, pp. 195–206.

[8] G. Vasiliadis, S. Antonatos, M. Polychronakis, E. Markatos, and S. Ioannidis, "Gnort: High performance network intrusion detection using graphics processors," in *Proceedings of the 11th international symposium on Recent Advances in Intrusion Detection*, 2008, pp. 116–134.

[9] G. Vasiliadis and S. Ioannidis, "Gravity: a massively parallel antivirus engine," in *Proceedings of the 13th international conference on Recent advances in intrusion detection*, 2010, pp. 79–96.

[10] R. Smith, N. Goyal, J. Ormont, K. Sankaralingam, and C. Estan, "Evaluating gpus for network packet signature matching," in *IEEE International Symposium on Performance Analysis of Systems and Software, 2009*, 2009, pp. 175 –184.

[11] L. Vespa, M. Mathew, and N. Weng, "P3fsm: Portable predictive pattern matching finite state machine," in *20th IEEE International Conference on Application-specific Systems, Architectures and Processors*, Boston, MA, USA, 2009, pp. 219–222.

[12] L. Vespa and N. Weng, "Deterministic finite automata characterization and optimization for scalable pattern matching," *ACM Transactions on Architecture and Code Optimization*, 2011.

[13] ——, "Gpep: Graphics processing enhanced pattern-matching for high-performance deep packet inspection," in *IEEE International Conference on Internet of Things (iThings 2011)*, 2011.

[14] S. Wu and U. Manber, "A fast algorithm for multi-pattern searching," Tech. Rep., 1994.

[15] Y. D. Hong, X. Ke, and C. Yong, "An improved wu-manber multiple patterns matching algorithm," in *25th IEEE International Performance, Computing, and Communications Conference. IPCCC 2006.*, 2006, pp. 680–686.

[16] N. Hua, H. Song, and T. Lakshman, "Variable-stride multi-pattern matching for scalable deep packet inspection," in *IEEE INFOCOM 2009*, April 2009, pp. 415–423.

[17] L. Vespa, N. Weng, and R. Ramaswamy, "Ms-dfa: Multiple-stride pattern matching for scalable deep packet inspection," *The Computer Journal*, pp. 285–303, December 2010.

[18] *ATI Radeon 6450 GPU, note=http://www.amd.com/us/products/desktop/graphics/amd-radeon-hd-6000/hd-6450/pages/amd-radeon-hd-6450-overview.aspx#1 year= 2011*.

[19] *OpenCL (Open Computing Language*, http://www.khronos.org/opencl.

[20] B. C. Brodie, D. E. Taylor, and R. K. Cytron, "A scalable architecture for high-throughput regular-expression pattern matching," *In SIGARCH Computer Architecture News*, pp. 191–202, 2006.

[21] *LSI Tarari T2000*, LSI Corporation, 2011.

[22] *Cavium NITROX CN17XX*, Cavium Networks, 2011.

[23] *Netlogic NLS2008 NETL7*, Netlogic Microsystems, 2011.

[24] N.-F. Huang, H.-W. Hung, S.-H. Lai, Y.-M. Chu, and W.-Y. Tsai, "A gpu-based multiple-pattern matching algorithm for network intrusion detection systems," in *22nd International Conference on Advanced Information Networking and Applications. AINAW 2008.*, March 2008, pp. 62–67.

# Securing Your Containers

## An Exercise in Secure High Performance Virtual Containers

**Adam Miller**

Department of Computer Science
Sam Houston State University
Huntsville, TX 77341
ajm023@shsu.edu

**Lei Chen**

Department of Computer Science
Sam Houston State University
Huntsville, TX 77341
LXC008@shsu.edu

**Abstract** – ***In this research we introduce a new way for container based virtualization to be used in a highly secure fashion. As the industry requires for condensing enterprise infrastructure, the need to use virtualization technologies is becoming a necessity. In the realm of virtualization technologies there are many popular hypervisors (also called Virtual Machine Manager, or VMM), but they all fall short in terms of performance when compared to container based virtualization technologies. However, the virtualization technologies that utilize containers often become victims to security vulnerabilities in ways that hypervisors are abstracted away from. In this paper we introduce a security mechanism that can be used to thwart the shortcomings of popular container based virtualization technologies. We incorporate SELinux along with Linux Containers (LXC) that make use of a recent Linux kernel feature called cgroups. It shows that by incorporating these two technologies we are able to achieve both high level of security and high performance environment where container based virtual servers can run and be utilized for the enterprise.***

**Keywords**: Containers, Virtual Machine Manager, SELinux, virtualization, security, hypervisors

## 1 Introduction

Virtualization is the creation of the virtual version of system and network resources, including hardware, operating system, storage devices, and network resources [11]. The essential benefit of virtualization is that computer processing power is treated as a utility and service that clients can buy and subscribe. Cloud computing is considered a natural evolution based on virtualization. Virtualization aims to centralize administrative tasks while improving scalability and workloads.

In 1967 IBM introduced the System/360 Model 67 which contained the first recorded instance of virtualization. Nowadays companies such as VMware, Microsoft, Citrix (formerly XenSource), Red Hat, and Oracle all offer virtualization products of their own in order to try and win their place in the industry's datacenters. One thing all of these technologies have in common is that they rely on hypervisor technologies.

Hypervisors are an abstraction layer between the virtual machine kernel and the actual hardware allowing for "fake" hardware resources to be presented to each virtual operating system while leaving the management of the actual hardware resource to the hypervisor. This abstraction layer helps in many ways but at the same time adds overhead to the system as resource management is no longer solely dependent on the operating system but upon the hypervisor. If this overhead could be removed and containers within the operating system can be created to allow other "contained" instances of the operating system, it would thwart the speed impact of hypervisors [1].

The rest of the paper is structured as follows. Next in Section 2, we conduct background study of container based virtualization especially focusing its security concerns. Section 3 discusses the details of Secure Linux Containers, including Linux Containers, Research Environment, Implementation, Process Isolation, Test Design, and Research Contribution. Simulation, experiments and results are discussed in Section 4. Conclusion is drawn in Section 5 and references are listed in Section 6.

## 2 Background

While container based virtualization thwarts the speed impact of hypervisors, it introduces a new concern about security allowing different systems to exist within the same file system hierarchy. There will be an added layer of security needed in order to enforce this new virtualization construct. Lowering the number of context switches can be achieved by reducing the number of system calls required for the applications running within these containers [2]. Once the performance gains are introduced there will be a need to enhance the security measures by implementing a custom SELinux context per container.

The security advancements found in SELinux will bring the further isolation needed in order to keep the separate running instances of guest operating systems apart from one another by utilizing the SELinux Mandatory Access Control mechanisms [3][4][5][6][7].

In the industry there are a number of virtualization technologies clouding the landscape. The large commonplace issue is that all of these technologies are reliant on a hypervisor technology to manage the resources on the back end. This consequently introduces a heavy overhead that has proven to impact up to 28.1% higher CPU utilization [1] when compared to a much lighter weight solution such as the container based technologies. The main issue this new paradigm introduces is the concern of keeping compromised processes from exiting their directory structure and entering into a neighboring virtual environment. Our research targets this security concern and provides a solution to it.

# 3 Secure virtual containers

## 3.1 Linux Containers

There are a number of container based virtualization technologies today in the open source world, among which the more prominent ones are OpenVZ and BSD Jails. Our research focuses on an emerging technology, Linux Container (LXC), because of its deep roots in reliance on a Linux kernel level construct that breaks resources into “control groups” referred to cgroups.

LXC is based around cgroups which are implemented and executed via a virtual filesystem. This filesystem is mounted and its files are modified traditionally using shell scripts and GNU coreutils as will be demonstrated throughout the course of the research. The general hierarchy of the filesystem layout for cgroups is as follows in Figure 1.

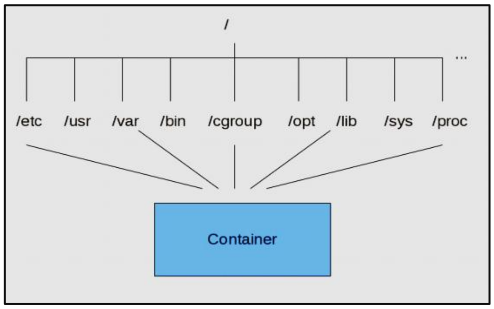


Figure 1. General hierarchy of filesystem layout for cgroups

Beyond the virtualization technology itself we will need to bring in the security aspect. SELinux is the de facto standard of high security in the Linux environment. The SELinux concepts revolve around using context for files, processes, and transactions and using policy to police these transactions by enforcing mandatory access control. The basic overview of this in practice is outlined in Figure 2.

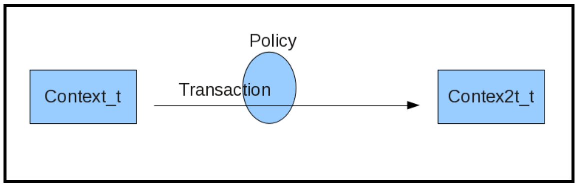


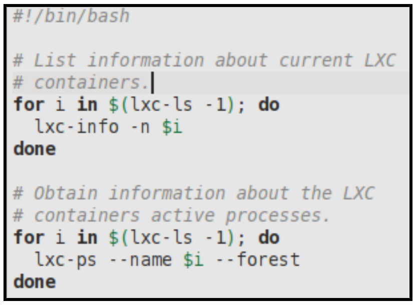
Figure 2. Mandatory Access Control

## 3.2 Research Environment

In this section we discuss the configuration and setup of the environment in which the research was performed. We used a system of modest resources in order to further highlight the performance gains. The system used in this research and development is Fedora GNU/Linux Distribution version 14 (codename Laughlin), and the version of LXC being used is 0.7.2 as per the release currently in the stable distribution repository at the time of this writing. The version of SELinux Utils that was used is 2.0.96 and the version of SELinux Targeted Policy is 3.9.7 also as per the releases available in the stable repositories. The LXC configuration and test setup was implemented using the example configuration files found in the LXC man (7) page for simplicity and ease of standardization.

## 3.3 Implementation

One of the most powerful elements of any UNIX-styled operating system is its shell environment combined with a special set of Core Utilities (conventionally known as coreutils). Code listing 1 shows a small example obtained from the LXC man (7) page that offers a glimpse into the powerful system level utilities that we have at our disposal.



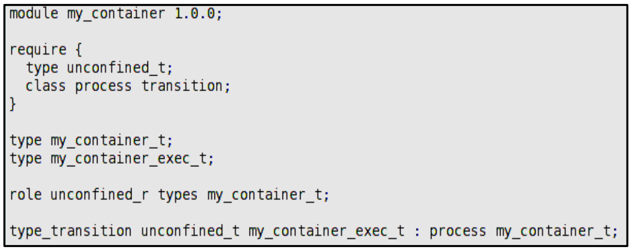
```
#!/bin/bash

# List information about current LXC
# containers.
for i in $(lxc-ls -1); do
  lxc-info -n $i
done

# Obtain information about the LXC
# containers active processes.
for i in $(lxc-ls -1); do
  lxc-ps --name $i --forest
done
```

Code Listing 1. Example from LXC man page

This example offers insight into the abilities of the scripting environment built into the shell of the system that we used to administer the testing environment.

Next we need to create our SELinux context and policy in order to apply to our container. This was done using a combination of constructs, one being the new

SELinux type and the other being the policy as defined by code listing 2.



```
module my_container 1.0.0;

require {
  type unconfined_t;
  class process transition;
}

type my_container_t;
type my_container_exec_t;

role unconfined_r types my_container_t;

type_transition unconfined_t my_container_exec_t : process my_container_t;
```

Code Listing 2. SELinx policy module

This code listing is the SELinux policy module that provides the simplest configuration needed. We have created two new types and allowed unconfined types operating within our container's context to continue to do so but nothing else. All non-allowed actions will be denied so when we create the directory structure that will contain the root filesystem for the container based virtual machine it will be necessary to label the entire directory structure with our my_context_t type as well as set the default context of the highest level parent directory with this type, such that its children and processes executed from within it will inherit these context attributes. It is this premise that keeps our container secure [3][4][5][6][7].

## 3.4 Process Isolation

Now that the core security is in place by isolating an entire container based virtual environment confined within its own SELinux context and the policy has been created to keep the context transactions isolated within itself. The following is a fraction of the upstream Linux kernel documentation listed at the time of this writing for the 2.6.35 kernel from kernel.org:

"A *cgroup* associates a set of tasks with a set of parameters for one or more subsystems.

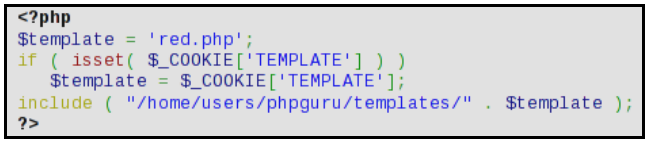
A *subsystem* is a module that makes use of the task grouping facilities provided by cgroups to treat groups of tasks in particular ways. A subsystem is typically a "resource controller" that schedules a resource or applies per-cgroup limits, but it may be anything that wants to act on a group of processes, e.g. a virtualization subsystem.

… " [8]

This is relevant as it describes a "subsystem" such as the LXC system that is being used in the course of this research. Without this ground work in which we can build upon, very little of this would be possible.

## 3.5 Test Design

With the LXC containers setup and configured, the SELinux policy and contexts in place, and the subsystem running we are ready to start to run some services within the environment in order to test the implementation. The easiest way to do this is to run an old outdated version of some server software with known vulnerabilities so that we can identify if our SELinux enforcement will actually protect the environment. The likely candidate is to run an older version of a web service program like Apache but to make it simple we chose to run a version of an interpreted module based language within the apache environment. Here we use php because of its wide popularity. It is target to many exploits and one of which we can use is a simple directory traversal exploit which will afford us the ability to determine if the SELinux containers are truly functional and can be done without causing any heavy amount of damage to our testing environment. Another reason why this is a solid example to show the security measures being enforced is that directory traversal vulnerabilities are often not due to an exploit in the implementation language such as php but of the code itself that was used to develop the web aspplication. The following listing (Code Listing 3) is an example of php code that leaves the server it is run on open to directory traversal. (NOTE: This code was obtained from wikipedia.org on 05/04/2011, author unknown.)



```
<?php
$template = 'red.php';
if ( isset( $_COOKIE['TEMPLATE'] ) )
   $template = $_COOKIE['TEMPLATE'];
include ( "/home/users/phpguru/templates/" . $template );
?>
```

Code Listing 3. Example of php code with traversal vulnerabilities

This could easily be exploited by an attacker using a properly constructed HTTP GET request to the server which would then warrant a response of either a directory listing or the contents of a file pending the permissions and SELinux context in place. The expected result of our research is that the potential attacker would not be able to obtain system level information by exploiting this aspect of our system regardless of the bugs introduced by novice web developers.

## 3.6 Research Contribution

The publication landscape in this field focuses on the architecture of the entire virtual stack with enhancements proposed in the area of hypervisor algorithms, resource management, and performance enhancements. These topics range from publications that cover the top to bottom architectures that are used in virtual environments [9] to those of performance characteristics based on the differences between various virtualization approaches [10].

With the introduction of the new research we can utilize the emerging technology based on highly developmental frameworks that are reliant on constructs which are now formally part of the upstream Operating System kernel. The system that has formed around this kernel by using the GNU userspace has been the basis for the currently most widely used Enterprise UNIX-styled Operating System as produced by Red Hat. By proposing research based upon already existing proven technologies that are heavily used in the industry, it is hoped that the methods could be quickly implemented in such markets. Another aspect of this research is built upon SELinux which is an enhancement originally introduced by the United States of America National Security Agency and is largely advocated as a default inclusion in the Red Hat Enterprise Linux Operating System. We used these two primary elements to combine for a more powerful secured virtualization construct.

The formal outline is simply that the performance gains of container based virtualization combined with the mandatory access control of SELinux will offer both enhanced performance and high level of security for virtual environments.

## 4 Simulation, experiments, and results

In this section we walk through the simulation environment that was utilized, the experiments performed upon the environment and present the results that were found in order to show that the research performed enforces the proposed result.

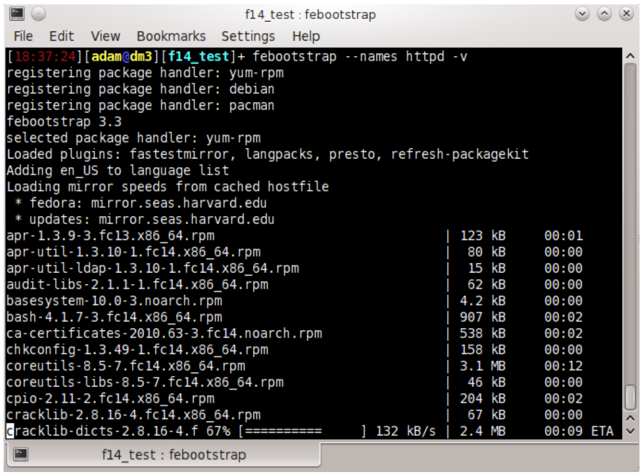
### 4.1 Simulation Preparation

We use the LXC constructs in order to define the container in which to run our virtual GNU/Linux environment. At the time of the writing the current stable release of Fedora GNU/Linux is 14 which is used in this research. We first configure an LXC environment using the example files provided by the LXC utilities manual pages in order to create a base LXC container. Then a utility called febootstrap is made for creating miminal bootstrapped operating systems within our previously constructed container. Figure 3 shows a screenshot of the creation of that contained virtual instance.

From here we initialize the apache instance and run a piece of known vulnerable php code in order to attempt the exploit against files from an adjacent container.

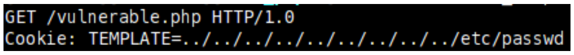
### 4.2 LXC Without SELinux Context

In this scenario we assume that the file permissions in an adjacent container were accidentally left in the state of “chmod 777” which allows all users all access to the files. In a situation where someone was running a virtual private server hosting company or similar business unit this is not an uncommon occurrence. In the following output we see that this directory traversal was easily obtained. Figures 4 and 5 show the HTTP GET request and successful response from the server respectively.



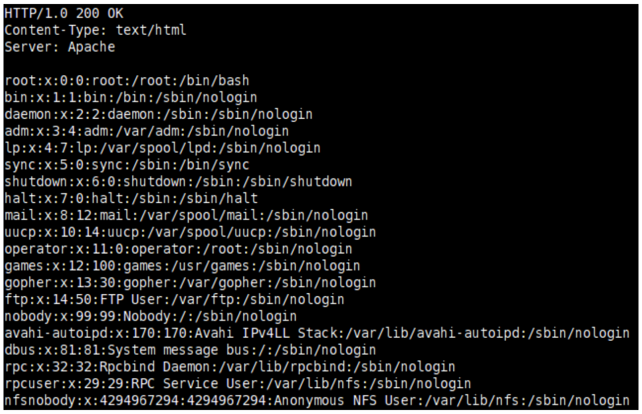
```
[18:37:24][adam@dm3][f14_test]+ febootstrap --names httpd -v
registering package handler: yum-rpm
registering package handler: debian
registering package handler: pacman
febootstrap 3.3
selected package handler: yum-rpm
Loaded plugins: fastestmirror, langpacks, presto, refresh-packagekit
Adding en_US to language list
Loading mirror speeds from cached hostfile
 * fedora: mirror.seas.harvard.edu
 * updates: mirror.seas.harvard.edu
apr-1.3.9-3.fc13.x86_64.rpm                              | 123 kB     00:01
apr-util-1.3.10-1.fc14.x86_64.rpm                        |  80 kB     00:00
apr-util-ldap-1.3.10-1.fc14.x86_64.rpm                   |  15 kB     00:00
audit-libs-2.1.1-1.fc14.x86_64.rpm                       |  62 kB     00:00
basesystem-10.0-3.noarch.rpm                             | 4.2 kB     00:00
bash-4.1.7-3.fc14.x86_64.rpm                             | 907 kB     00:02
ca-certificates-2010.63-3.fc14.noarch.rpm                | 538 kB     00:02
chkconfig-1.3.49-1.fc14.x86_64.rpm                       | 158 kB     00:00
coreutils-8.5-7.fc14.x86_64.rpm                          | 3.1 MB     00:12
coreutils-libs-8.5-7.fc14.x86_64.rpm                     |  46 kB     00:00
cpio-2.11-2.fc14.x86_64.rpm                              | 204 kB     00:02
cracklib-2.8.16-4.fc14.x86_64.rpm                        |  67 kB     00:00
cracklib-dicts-2.8.16-4.f 67% [==========     ] 132 kB/s | 2.4 MB     00:09 ETA
```

Figure 3. Creating contained virtual instance



```
GET /vulnerable.php HTTP/1.0
Cookie: TEMPLATE=../../../../../../../../../etc/passwd
```

Figure 4. HTTP GET request



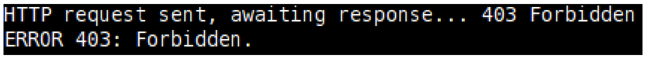
```
HTTP/1.0 200 OK
Content-Type: text/html
Server: Apache

root:x:0:0:root:/root:/bin/bash
bin:x:1:1:bin:/bin:/sbin/nologin
daemon:x:2:2:daemon:/sbin:/sbin/nologin
adm:x:3:4:adm:/var/adm:/sbin/nologin
lp:x:4:7:lp:/var/spool/lpd:/sbin/nologin
sync:x:5:0:sync:/sbin:/bin/sync
shutdown:x:6:0:shutdown:/sbin:/sbin/shutdown
halt:x:7:0:halt:/sbin:/sbin/halt
mail:x:8:12:mail:/var/spool/mail:/sbin/nologin
uucp:x:10:14:uucp:/var/spool/uucp:/sbin/nologin
operator:x:11:0:operator:/root:/sbin/nologin
games:x:12:100:games:/usr/games:/sbin/nologin
gopher:x:13:30:gopher:/var/gopher:/sbin/nologin
ftp:x:14:50:FTP User:/var/ftp:/sbin/nologin
nobody:x:99:99:Nobody:/:/sbin/nologin
avahi-autoipd:x:170:170:Avahi IPv4LL Stack:/var/lib/avahi-autoipd:/sbin/nologin
dbus:x:81:81:System message bus:/:/sbin/nologin
rpc:x:32:32:Rpcbind Daemon:/var/lib/rpcbind:/sbin/nologin
rpcuser:x:29:29:RPC Service User:/var/lib/nfs:/sbin/nologin
nfsnobody:x:4294967294:4294967294:Anonymous NFS User:/var/lib/nfs:/sbin/nologin
```

Figure 5. Successful HTTP response

The output here shows that we are able to obtain the output from the vulnerable containers. This was the expected result and shows how this can be problematic in practice.

### 4.3 SELinux Enforcing

In the second scenario we have a configuration identical to the first in respect to the virtual container, but on the back end we are enforcing SELinux policy and we can see that the attempt to perform a directory traversal is thwarted by our SELinux enforcement. Figure 6 shows the screenshot.



```
HTTP request sent, awaiting response... 403 Forbidden
ERROR 403: Forbidden.
```

Figure 6. Attempt to perform directory traversal forbidden

## 5 Conclusion and future work

In this research we have shown that not only can we obtain highly available, high performance, and highly scalable virtualization infrastructure using container based virtualization [1] but we can also provide a high level of security inside these containers using the new paradigm enforced by SELinux. These concepts combined have proven to alleviate the host server administration needs concerned with the virtual containers from impeding upon one another. Future work will be to explore options of this application and to fine tune the approach for more sophisticated architectures.

## 6 References

[1] Stephen Soltesz, Herbert Potzl, Marc E. Fiuczynski, Andy Bavier, and Larry Peterson. "Container-base operating system virtualization: a scalable, high-performance alternative to hypervisors." *Proceedings of the 2nd ACM SIGOPS/EuroSys European Conference on Computer Systems 2007* (EuroSys '07), ACM, New York, NY, USA, 275-287.

[2] Yih Huang, Angelos Stavrou, Anup K. Ghosh, and Sushil Jajodia. "Efficiently tracking application interactions using lightweight virtualization." *Proceedings of the 1st ACM workshop on Virtual machine security* (VMSec '08). ACM, New York, NY, USA, 19-28.

[3] Giorgio Zanin and Luigi Vincenzo Mancini. "Towards a formal model for security policies specification and validation in the SELinux system." *Proceedings of the ninth ACM symposium on Access control models and technologies* (SACMAT '04). ACM, New York, NY, USA, 136-145.

[4] Gaoshou Zhai, Wenlin Ma, Minli Tian, Na Yang, Chengyu Liu, and Hengsheng Yang. "Design and implementation of a tool for analyzing SELinux secure policy." *Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human* (ICIS '09). ACM, New York, NY, USA, 446-451.

[5] Bjorn Vogel and Bernd Steinke. "Using SELinux security enforcement in Linux-based embedded devices." *Proceedings of the 1st international conference on MOBILe Wireless MiddleWARE, Operating Systems, and Applications* (MOBILWARE '08). ICST, Brussels, Belgium, Belgium, Article 15 , 5 pages.

[6] Fabrizio Baiardi, Daniele Sgandurra. "Securing a Community Cloud." *Distributed Computing Systems Workshops, International Conference on*, pp. 32-41, IEEE 30th International Conference on Distributed Computing Systems Workshops, 2010.

[7] Gaoshou Zhai, Yaodong Li. "Analysis and Study of Security Mechanisms inside Linux Kernel." *Security Technology, International Conference on*, pp. 58-61, International Conference on Security Technology, 2008

[8] Kernel Docmentation on cgroups maintained by kernel developers abroad, the following were listed or original authors/modifiers: Written by Paul Menage <menage@google.com> based on Documentation/cgroups/ cpusets.txt

[9] Jeff Daniels. "Server virtualization architecture and implementation." *Crossroads* 16, 1 (September 2009), 8-12.

[10] Jeanna Neefe Matthews, Wenjin Hu, Madhujith Hapuarachchi, Todd Deshane, Demetrios Dimatos, Gary Hamilton, Michael McCabe, and James Owens. "Quantifying the performance isolation properties of virtualization systems." *Proceedings of the 2007 workshop on Experimental computer science* (ExpCS '07). ACM, New York, NY, USA , Article 6.

[11] Virtualization, retrieved on June 20th from *http://en.wikipedia.org/wiki/Virtualization*

# Incorporating Soft Computing Techniques into Anomaly Intrusion Detection Systems

**Yingbing Yu**[1], **Han Wu**[2]
[1]Department of Computer Science & Information Technology
Austin Peay State University, Clarksville, TN, USA
[2]Department of Mathematics & Statistics
Minnesota State University, Mankato, Minnesota, USA

**Abstract** - *One critical threat of inside attacks facing many organizations is from masqueraders, internal users or external intruders who exploit legitimate user identity and manipulate the system of performing malicious attacks. Intrusion detection systems can be used to build a user profile and a large deviation from the past behavior patterns indicates a possible illegal access from a masquerader. In this paper, we first introduce the typing biometrics of keystroke patterns and apply the probabilistic neural network for the classification. A second behavior profiling approach from command sequences generated by a user is to build the behavior model using the finite automata. New activities of users are compared with the finite automaton for the deviation. A fuzzy inference system is then applied to integrate the two input variables to evaluate the overall threat of the possible masquerader existence. Experiments show promising results with a high detection rate and a low false alarm rate.*

**Keywords:** Anomaly Intrusion Detection, Masquerader Detection, Probabilistic Neural Network, Fuzzy Systems

## 1 Introduction

Intrusion detection systems (IDSs) attempt to perform the process of monitoring computer networks and systems for violations of security policy (a set of laws, rules, and practices that define the system boundaries) [1]. IDSs can be categorized into two classes based on different detection approaches. Misuse (knowledge or signature-based) IDSs look for specific patterns that define a known attack. The information about known attacks and vulnerabilities of a system is encoded into “signatures”. Any actions that trigger the matches will be reported as “attempts” of intrusions. Anomaly (behavior-based) IDSs assume the deviation of normal activities under attacks and perform abnormal detection compared with predefined system or user behavior reference model.

Anomaly IDSs can be used to build system or user behavior profiles to detect inside attacks from masqueraders, internal users or external intruders who exploit legitimate users identification and password that one may obtain illegally and then perform malicious attacks from the inside. Inside abuse of network access has been cited as the second most cited forms of attacks [3]. To prevent a system from attacks due to identity theft, the effective approach is to monitor user behavior and report any suspicious activities. Alarms are alerted when a user behaves out of characters and a large deviation with the behavior profile is detected.

The goal of this paper is to distinguish a masquerader from genuine users which is a challenging task due to the problem of concept drift, where the observed user behavior may change with different tasks, time, general knowledge level and such other uncertain elements [4]. In this paper, we introduce a soft computing based model to detect masqueraders based on two different user behavior profiling approaches - typing biometric as keystroke patterns and user activities as command sequences generated in a UNIX/Linux system.

The probabilistic neural network (PNN) is applied for the classification of keystroke patterns as normal or abnormal after the biometric template is constructed. Command sequences ordered by date and time are used to build a finite automaton which is identified with the behavior model for a user. A command block of new activities from a user is compared with the finite automata and the deviation is evaluated as memberships with fuzzy sets. A fuzzy inference system is further introduced to integrate the information from the typing biometrics and behavior profile from the command sequence analysis to evaluate the overall threat of a case as the possible masquerader existence in a computing system.

The rest of this paper is organized as follows. Section 2 is the literature review that discusses related research in masquerader detection from command sequences and typing biometrics. Section 3 presents the typing biometrics template for a user and the classification of patterns using the probabilistic neural network. In section 4, we introduce the finite automata based behavior profiling from user command sequence and the deviation evaluation. Section 5 proposes the fuzzy reasoning system from the input of typing biometrics and behavior profiling of commands to evaluate the degree of overall threat. Experimental results conducted in a data set are given in section 6. The paper concludes with section 7, which discusses the future research work.

## 2 Literature Review

User behavior profiling can be used for classification, future behavior prediction and masquerader detection. Traditionally user behavior in a system is characterized by parameters such as login frequency, location frequency, last login, session elapsed time, password fails, location fails, amount of network traffic, resources used by user in a session and so on [5]. In this paper, typing biometrics of keystroke patterns and command sequences concatenated by date order are selected to capture a user normal behavior model and detect masqueraders based on the degree of deviations.

De Ruet et al. developed a software methodology that improves security by using typing biometrics to reinforce password authentication mechanisms [2]. Typing biometrics is the analysis of a user's keystroke patterns. Each user has a unique way of using the keyboard to enter a password; for example, each user types the characters that constitute the password at different speeds. The methodology employs fuzzy logic to measure the user's typing biometrics.

Machine learning and statistical methods have been widely used in the literature for the behavior profiling from the analysis of command sequences. Davison and Hirsh developed a model called IPAM (incremental probabilistic action modeling) to predict sequences of user actions [6]. Single-step command transition probability is estimated from training data. Balajinath introduced GBID (Genetic Based Intrusion Detector) to model individual user behavior with a 3-tuple vector which is learnt later by a genetic algorithm [11]. Ryan used a back propagation neural network NNID (Neural Network Intrusion Detector) to identify users simply by what commands and how often they use, called the ‘print’ of a user [7].

Lane and Brodley [8] selected a machine learning algorithm IBL (instance based learning) to measure the similarity between the most recent 10 commands of a user and the profile extracted from the past. The similarity measure is the count of matches of a new sequence with the sequences from a user’s commands history, with a greater weight assigned to adjacent matches. Schonlau [9] selected several statistics-based methods to detect masqueraders, including uniqueness, Bayes one-step Markov, Compression, Multi-step Markov chain etc. Maxion and Townsend [10] applied Na ïve Bayer classification algorithm to user profiling with command-line data, which shows improvement over the best approach of Schonlau [9].

## 3 Typing Biometrics Classification

Typing biometrics is used for the analysis of a user’s keystroke patterns. Each user types the characters that constitute a command at different speeds. Our method is to build a biometrics template to be used for the anomaly detection. When an authorized user is accessing a system and typing commands, the interval time between two successive characters in a unique command is recorded. For example, if the user types a command “dir”, the time intervals between the characters “d” and “i”, “i” and “r”, “r” and the “ENTER” key will be stored as the biometric characteristic.

If the command contains parameters, the time between the space and a regular character is also recorded. For example, if a user types the command “dir –p”, we will also calculate the time interval between the characters of “r” and the “BLANK” key. A user may type capital letters instead of lower case ones, and this will involve a larger time interval since the user needs to press the “CapsLock” key at the same time. We have not considered this scenario and experiments show that most users prefer to always type commands in lower case. In addition, we only consider those frequently typed commands by a user. The reason is that later a neural network is used for the classification and a relatively large training data set is required for that.

The user typing biometrics in the form of time interval of successive characters in commands will be used as the typing template for the user. On subsequence access to the system, each of the user command typing will go through a neural network system for the classification as normal from the genuine user or abnormal from a masquerader. We have chosen the supervised probabilistic neural network (PNN) for the classification purpose. PNN [12] has the learning and generalization ability of back-propagation multi-layer neural networks (BPNN) and is simpler and faster. It can identify the commonalities in the training examples and then perform classification of unseen examples from the predefined classes.

PNN consists of three feed-forward layers: input layer, pattern layer, and summation layer. An input layer contains as many elements as the number of characters in a command. The pattern layer represents a neural implementation of a version of Bayes classifier and can provide an optimum pattern classifier to minimize the expected risk of wrongly classifying an object. It gets closer to the true underlying class density functions as the number of training samples increases if the training set is an adequate representation of the class distinctions. The summation layer (output layer) has two elements of normal and abnormal classes to denote if the typed command is from the genuine user or a masquerader. A compete transfer function on the output of the second layer picks the maximum of these probabilities, and produces a “1” for the normal class and a “0” as an abnormal case.

Since a user session usually involved the execution of multiple commands, we pick up a block size of 50 commands for the masquerader detection. To determine if the whole block of commands is from the genuine user or a masquerader, we use a fuzzy inference system for the deviation evaluation from the biometric template. To denote the degree of deviations, three fuzzy sets “low”', “medium” and “high” and fuzzy membership functions are defined to measure the magnitude of these values. The common triangular fuzzy membership is chosen. The common triangular fuzzy membership is chosen in Figure 1. Given an example, if 8 commands in a block of 50 are classified by the NN as abnormal, the degree of this case from a masquerader is low with a degree 0.6 and medium with a degree 0.4.

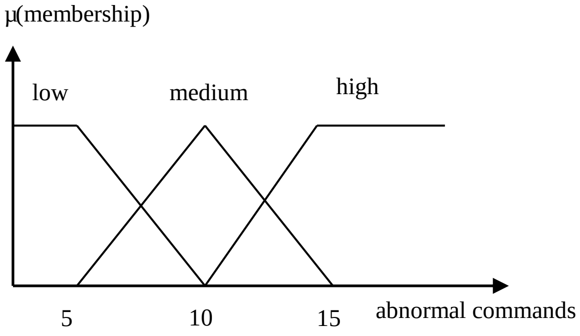


Figure 1. Memberships of Deviation for Typing Biometrics

# 4 Profiling from User Commands

To detect masqueraders, we also analyze user activities in the form of commands sequences in addition to the typing biometrics. This approach is more effective in UNIX/Linux system where most of the activities involve the shell commands execution though it can also extend to the Windows systems. We first present the finite automata model to capture the behavior patterns from the sequential data of commands. Then we describe how to measure the mismatch with the profile and evaluate the degree of deviation using fuzzy logic.

To construct finite automata as the user behavior model, the sequence of user commands concatenated by temporal order is divided into blocks of *n* commands in each, called a pattern. From the very beginning of a sequence, each time a pattern is searched and the position moves forward by 1 using the sliding window technique. For a sequence of *l* commands, there will be (*l* - *n* +1) patterns. If a pattern is not seen before, a new state is created in the finite automaton to hold it and a connection from previous state to the new one is added. The connection is labeled as the last element of the previous pattern and a number indicating the transition frequency.

New activities of user command sequences are divided into patterns of the same size. Each pattern is compared with the finite automaton to determine the degree of mismatch denoted by numerical values. The sum of mismatch values of all the patterns in a case is regarded as overall mismatch for the test case. Not only the exact mismatch of a pattern with the finite automaton, but also if there are correct transitions between states, will be checked to determine the two values. Three rules have been created for different scenarios.

1. Finite automaton doesn’t have a state associated with a pattern, mismatch=1.

2. Finite automaton has a state associated with a pattern but without a transition from previous state, or the existed transition not correctly labeled, then mismatch=0.5.

3. Finite automaton has a state associated with a pattern and a correctly transition. If the frequency of this transition occurred very few in the finite automaton, then mismatch = 0.2. If the transition occurrence is high, mismatch = 0.

A fuzzy inference system and the corresponding fuzzy membership definition in the Figure 2 are given to evaluate the overall deviation of a case (a sequence of commands in a block) based on the command sequence analysis.

**Rule 1.1**: If the mismatch is low, then the deviation with the profile is low.

**Rule 1.2**: If the mismatch is medium, then the deviation with the profile is medium.

**Rule 1.3**: If the mismatch is high, then the deviation with the profile is high.

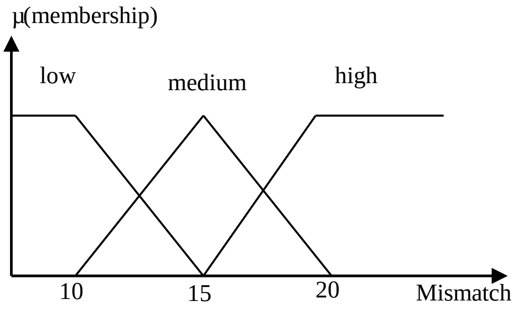


Figure 2. Fuzzy Memberships of Deviation

## 5 Fuzzy System Threat Evaluation

The information from the user typing biometrics and command sequence analysis is integrated for the analysis since this will provide more accurate judgment about new behavior patterns. We create a fuzzy inference system based on the two factors to evaluation the overall threat as the possibility of masquerader existence.

The variable "biometric deviation" is selected as one input to denote the degree of deviation from the biometric template discussed in section 3. The variable "profiling deviation" is used as another input to represent the mismatch with the established profile from command sequences. For the output, we have chosen five different classes "very low", "low", "medium", "high", and "very high" to denote the degree of the threat belonging to each fuzzy set. The membership functions are defined in Figure 3 using the triangular memberships.

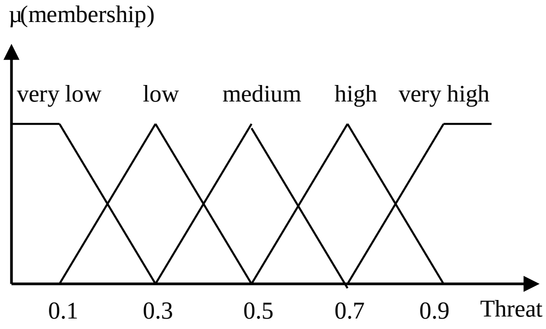


Figure 3. Fuzzy Membership of Overall Threat

Since each of the two input variables can be in one of three fuzzy sets "low", "medium" and "high", the number of possible fuzzy rules is very large and only those mostly effective ones in our experimental trials are selected for the fuzzy inference system

**Rule 2.1**: If the biometric deviation is high, and the profiling deviation is high, then the possibility of masquerader is very high.

**Rule 2.2**: If the biometric deviation is high, and the profiling deviation is medium, then the possibility of masquerader is high.

**Rule 2.3**: If the biometric deviation is medium, and the profiling deviation is high, then the possibility of masquerader is high.

**Rule 2.4**: If the biometric deviation is medium, and the profiling deviation is medium, then the possibility of masquerader is medium.

**Rule 2.5**: If the biometric deviation is medium, and the profiling deviation is low, then the possibility of masquerader is low.

**Rule 2.6**: If the biometric deviation is low, and the profiling deviation is medium, then the possibility of masquerader is low.

**Rule 2.7**: If the biometric deviation is low, and the profiling deviation is low, then the possibility of masquerader is very low.

After the membership values of facts with respect to each antecedent in a rule are determined, the MAX-MIN method is applied to measure the impact of fuzzy rules and the highest membership is selected. For each fuzzy rule, the output membership is obtained using the MIN implication operator to select the minimum membership value of antecedents in the premise. If several fuzzy rules generate different membership values associated with the same fuzzy set, the MAX implication operator is used to select the largest one as the final result.

The center of area (COA) defuzzification method is applied to get a single non-fuzzy crisp output of final threat ($\Delta Threat$) to measure the degree of the deviation with the behavior profile. The COA formula is:

$$\Delta Threat = \frac{\sum_{k=1}^{n} \mu_k * center(k)}{\sum_{k=1}^{n} \mu_k} \qquad (1)$$

where $n$ is the number of fired rules, $\mu_k$ is the degree of membership of rule $k$, and *center(k)* is the peak-value where the fuzzy set for the rule $k$ has the maximum membership values. This final *ΔThreat* value can be compared with a threshold value to determine the degree of the threat to the system. In our experiments, if the overall *ΔThreat* is great than 0.6, a test case will be classified as abnormal from a masquerader. Otherwise the case will be regarded as “normal” that the genuine use is accessing the computing system.

## 6 Experimental Results

The proposed model based on soft computing techniques is applied to detect masqueraders using the data generated in a network laboratory. During that time, we have collected user commands of about 35 students who worked on Linux systems over a period of about three months. Normal data was generated when they worked on several programming and lab assignments. The masquerader data was generated when the students tried to compromise computers in a local area network, as part of lab assignments.

The data generated in the first two months in the laboratory is used to build the typing biometrics and behavior profile for each user. The data generated in the third months when some of the users conducted attacks on the network is served as the masquerader cases. Since some user didn’t generate a large amount of data and only 25 users are selected as intrusion targets and the normal training data set is 2148 command blocks of 50 commands in each. The number of normal test cases is 436 blocks and the number of anomaly cases from masqueraders is 545.

For the 545 cases of masquerader data, 508 are classified correctly as “abnormal” and only 37 cases are mistakenly identified as “normal”. For most users, the model can achieve a high masquerader detection rate between 90% and 95%. The average detection rate is about 93.2% (508/545) and the missing rate is 6.8% (37/545). We also tested the 436 normal cases to check if the model can successfully do the classification. If a normal one is classified incorrectly as abnormal, a false alarm was generated. In total, only 18 normal cases are incorrectly identified as “abnormal” and the false alarm rate is 4.1%.

In addition, we have conducted further experiments with other different block sizes (# of commands in a block) to compare the performance of detection and false alarm rate. For a real intrusion detection system, it is the ultimate goal to detect masqueraders within a short time interval and alert the system earlier to prevent further information loss. Based on the results, we have found that the interval of about 50 commands execution achieves both a high detection rate and lower false alarm rate. When a larger test block size is selected, the model can still achieve a high detection rate but will introduce a higher false alarm rate.

If the block size is above a threshold (e.g., 100 commands), the model has a low detection rate and high false alarms. The possible reason is that the probability of overlapping behavior patterns increases rapidly when target size is above a certain threshold. In practice, an appropriate size can be selected based upon specific security policies for an organization. In general, it is fairly reasonable and effective for an anomaly IDS to detect potential masqueraders after just about 50 commands execution.

## 7 Conclusions

In this paper, we introduce the soft computing techniques in the area of intrusion detection systems to detect masquerades based on two different approaches of user behavior profiling. Typing biometrics of keystroke patterns from users can be collected in the training phase as the biometric templates. Another behavior profiling approach from user activities of commands execution is also applied where the sequential data is to build a finite automaton. To integrate the valuable information from both the approaches, a fuzzy inference system is presented based on the two inputs – deviations by comparing with the biometric template and finite automaton.

We also want to extend the current research of masquerader detection. As we have notices that in the last decade GUI and Internet-based applications have been deployed in both UNIX and Windows systems. A large of part of user activities associated with these applications may not involve individual commands directly entered into the system, but instead consist of mouse clicks on icons. The behavior modes from this kind of activity will differ significantly from those discussed in this paper. Future work would address these questions.

## 8 References

[1] R. Bace, Intrusion Detection. Macmillan Technical Publishing, 2nd Edition, Indiana, 2000.

[2] Willem G. de Ru, Jan H.P. Eloff. Enhanced Password Authentication through Fuzzy Logic. *IEEE Expert*, Volume 12, Issue 6, Nov.-Dec. 1997 Page(s):38 – 45.

[3] Computer Security Institute and Federal Bureau of Investigation, “CSI/FBI Computer Crime and Security Survey,” Computer Security Institute publication.

[4] Terran Lane. Machine Learning Techniques for the Computer Security Domain of Anomaly Detection. *PhD Thesis*, Purdue University, W. Lafayette, IN, August 2000.

[5] Dorothy E. Denning. An Intrusion Detection Model. *IEEE Transactions on Software Engineering*, 13(2): 222–232, February 1987.

[6] B. Davison and H. Hirsh. Predicting Sequences of User Actions. *Predicting the Future: AI Approaches to Time-Series Problems*, 1998 AAAI Workshop, July 1998, Madison, Wisconsin, 5–12.

[7] J. Ryan, M. Lin, R. Miikkulainen. Intrusion Detection with Neural Networks. *Advances in Neural Information Processing Systems*, M. Jordan et al., Eds., Cambridge, MA: MIT Press, 1998, pp. 943-949.

[8] Lane, T. and Brodley C. E. Temporal Sequence Learning and Data Reduction for Anomaly Detection. *ACM Transactions on Information and System Security*, 2(3). Pp. 295—331, 1999.

[9] Schonlau, M., DuMouchel, W., Ju, W., Karr, A., Theus, M., Vardi, Y. Computer Intrusion: Detecting Masquerades. *Statistical Science*, 16(1): 58-74, 2001.

[10] Roy Maxion and Tahlia Townsend. Masquerade Detection Using Truncated Command Lines. *International Conf. on Dependable Systems & Networks*, Washington, DC, 23-26 June 2002: 219-228.

[11] B. Balajinath, S. V. Raghavan. Intrusion Detection Through Learning Behavior Model. *Computer Communication* 24(2001) 1202-1212.

[12] Specht, D.F. Probabilistic Neural Networks. *Neural Networks*, vol. 3, pp.109-118, 1990.

# SESSION

# NETWORK SECURITY

## Chair(s)

**Dr. Gregory Vert**

# Environment Feature Map for Wireless Device Localization

T. Kaiser[1], P. Card[2], and K. Ferens[1]

[1]Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, Manitoba, Canada

[2]Department of Research and Development, Seccuris, Inc., Winnipeg, Manitoba, Canada

**Abstract—***This paper proposes the use of an artificial neural network to learn the environmental impairments of an organization's premises in the application of wireless device localization for data management and security. The purpose is to locate a wireless device in a given protected area (organizations' premises). Sets of contaminated signal strength samples, one for each grid-location of a protected area, train an artificial neural network (ANN) to learn the obstacles (environmental features) of a building or structure and classify any element of a contaminated set to the originating grid location from where the device was transmitting. This work consists of two phases: phase 1 simulates the contaminated subsets by applying random perturbations of each coordinate in the protected area. By controlling the amount of random perturbation, an ANN may be trained to handle a higher degree of errors than normally expected. Results show that the ANN has very good classification performance. Phase 2 will apply real signal strength samples and these results will be reported in a subsequent publication.*

**Keywords-**wireless device localization, tracking, data security, artificial neural network, received signal strength.

## 1 Introduction

The detection, location determination, and tracking of wireless-devices in the proximity of an organization's premises are growing concerns. Many organizations need to know if a rogue intruder device is attempting to gain access to their network and protected data, and to quickly take corrective action to thwart and deny access. At the same time, known clients need to be given access in a secure and controlled way, abiding by the organization's security policies. For example, many authorized clients visiting an organization with smart phone, laptop, or PDA, require limited access to the organization's internal information. These devices cannot be managed easily because people will non-deterministically visit and leave an organization with previous or possibly new devices at any time. Furthermore, an organization may limit access depending on the location of the device within the premises of the organization itself. For example, clients in the reception area would be granted, perhaps, stock quote data, while a client in a laboratory would be granted experimental results. An organization would like to automatically detect a wireless-device in the proximity of their wireless network, and based on the calculated location and subsequent identification of the user of the device, grant access to a limited amount and type of information behind the protected wireless network.

There are many different types of measurements which can be taken and used in node localization, i.e., to determine the physical location of a wireless device in a wireless network, such as time of arrival (TOA), angle of arrival (AOA), and received signal strength (RSS).

These measurements will be contaminated with channel errors, which include time varying impairments and environmental impairments. The time varying impairments result from interference by other deterministic narrowband wireless technologies (e.g., Wi-Fi, Bluetooth, ZigBee, cordless phones, and microwave oven) and stochastic broadband electromagnetic (EM) noise signals. Simple time averaging of the TOA, AOA, and RSS measurements would mitigate the effects of the time varying impairments to the extent that these impairments are purely random, but they would be ineffective for the non-random components. Other more complex forms of time averaging or signal processing techniques would be required to take into account the not-so-quite random properties exhibited by these types of impairments, such as signals which contain constant sloped and exponential decay power spectrum density [1].

The environmental impairments, such as fading, shadowing, multipath, non-line-of-sight propagations, reflections, and attenuation, result from obstacles in the environment (buildings, steel structures, and electronic equipment). A basic approach to mitigating the effects of environmental impairments is to create, a priori, a wireless feature map of the environment, which would consist of electromagnetic contours of the environment that roughly represent and map out the obstacles. Once a map is obtained, a rogue device may be detected within the environment, since its presence would disturb the known waveform patterns in the map.

There are basically two approaches which can be used to learn a feature map of the environment. The invasive approach uses reference, anchor, and agent nodes to periodically broadcast probe signals to one another. These probe signals will experience the obstacles and environmental impairments listed above. A cooperative distributed or centralized algorithm would gather the received waveform patterns from each of the probe receivers and analyse the results to form a map. This method is expensive in terms of the additional nodes required to receive the probes, the complex and custom antennae required for the measurements, and the high level of complexity required for interpreting the received waveforms and correlating and associating this information to obstacles in the environment. This great degree of expense may not be tractable or desired.

The other approach to creating a feature map is non-invasive and only uses the signals emanating from the target device alone to learn the environment. This paper proposes to capture received signal strength meaning for target devices, and to use an artificial neural network (ANN) to learn the characteristic features of a specific environment (client's building). By providing the ANN with many contaminated samples of signal strength, for each known location in the protected area, the ANN will learn to associate (map) the impaired samples to the originating coordinate triplet. The learning phase of the ANN can be applied to each new environment in which the system is deployed, and the learning phase can be done dynamically while the system is running to constantly improve its classification performance.

The localization problem faced by the ANN in this application is analogous to the task which a forward error correcting code (FEC) would perform in a communications system [2]. A FEC protects each information word from bit errors by adding parity bits to each information word (resulting in code word) in such a way as to separate each code word as far as possible in the FEC vector space. Analogously, to remove the contamination, the ANN needs to separate the subsets of contaminated signal strength samples as far as possible in the ANN vector space to form a decision boundary between each subset. Overlap of subsets will cause the ANN to incorrectly map the received signal strength to the originating coordinate triplet, in the same way a FEC will fail to detect or correct errors if a code word experiences more bit errors than the code can handle.

The remaining parts of the paper are organized as follows: section II discusses related work and identifies the extensions and contributions of this work. Section III gives the details of applying an artificial neural network to classify signal strength and perform the trilateration. Section IV discusses the simulation experiments performed to determine the feasibility of the ANN approach. Finally, conclusions and future work are given.

## 2 Related Work

There are generally two main approaches that have been used to detect and localize wireless devices in a network: cooperative and autonomous. The work done by [3] [4] aimed to determine a fundamental limit to accuracy for a given localization algorithm. Their analysis used received waveforms themselves (to exploit all information in the signal) rather than utilizing only the signal metrics extracted from these waveforms, such as time-of-arrival and received signal strength. The method requires localization information from individual anchors and a priori knowledge of the agent's position in the network, which may not be known in a general scenario. Also, this method is based on the cooperative approach, which relies on (possibly rouge) target devices to cooperate in the localization process.

Another work that is based on the cooperative approach is [5]. This work describes how time-of-arrival (TOA), angle-of-arrival (AOA), and received-signal-strength (RSS) measurements can be used cooperatively to detect a device's location in wireless sensor networks; but this work requires a higher infrastructure cost. Along the same lines, several off-the-shelf technologies for wireless device location detection have been proposed, such as GPS, cell phone, and infrared (IR). These cooperative based approaches require the installation of custom hardware and/or software on the target device, or establishment of anchors and agents in the network, and they must work in conjunction with the remote detection system to identify the target's physical location. However, the required subsystems may not be installed, or they can be easily removed and defeated. Or, the data may be altered on the target device before transmission to the detection system. Consequently, even a novice hacker may avoid detection or misrepresent their location, leaving the system ineffective.

To work around these problems this paper uses an autonomous system, which does not need any special hardware or software installed on the target device, or any cooperation from the target device. The proposal is to use trilateration of the received signal strength of the target device to determine its physical location. However, received signal strength, even when sent from a fixed-location, may vary from ideal conditions because of radio signal reflections off walls and other structures, attenuation caused by moving objects, and the co-existence problem. Furthermore, each company's unique indoor/outdoor structural characteristics will have different effects on a radio signal, and so the cost-benefit ratio for each company will be different and will affect their competiveness. Consequently, the variance of signal strength results in a corresponding variance in detected location. Prior art took average and standard deviation measurements to mitigate the variance of measurements, but these were ineffective and the target was incorrectly detected in many cases. To mitigate the effects of the variance in signal strength measurements, [6] took other first order statistical measurements, which may lead to the same problem as averaging and standard deviation methods face. This paper proposes a method that is similar to [6], but it differs in that this paper uses artificial neural network's implementation of trilateration. The ANN will extract higher order statistical properties of a subset of contaminated received signal strength samples and map these samples to the originating coordinates from which the target device was transmitting.

# 3 Algorithm Description

The idea is to capture a subset $S_i$ of contaminated signal strength samples emanating from a device, which was transmitting from known spatial coordinates ($x_i$,$y_i$,$z_i$). A subset of samples is acquired for each known coordinate triplet in a protected area. The known coordinate triplet *i* and the acquired subset of samples represent a set $S_i$ of training vectors for the ANN. A superset S = {$S_j$, where j = 1…N} represent all training vectors for each location in a protected area. The ANN is trained with the set $S_j$. To properly classify each contaminated subset, the ANN will be forced to push away as far as required the contaminated subsets so that they do not overlap in the vector space and so that a decision boundary can separate each of the subsets. After training, the ANN would be tested with contaminated signal strength occurrences, which it has not seen before, to determine its classification performance.

## 3.1 Random Perturbations

To simulate the effects of signal impairments (reflection, attenuation, etc.), each of the coordinates in a (10x10x10) grid were permuted by 5%, as shown in Fig. 2. (In the simulation, instead of permuting signal strength measurements, an equivalent method of permuting coordinates was done.) We generated 500 random perturbations for each of the coordinate points in the grid Fig. 2. The number of grid points is given by (1), and the total number of perturbations was 500,000.

$$Targeted\ Samples = 10^3 = 1000 \qquad (1)$$

## 3.2 Relative Distance Measurement

A 10x10x10 grid was used to position the client and the three receivers. Three receivers were placed at a fixed coordinate in the grid, while the client's coordinates varied in the grid. The location of any wireless client within the grid point was determined by calculating its relative coordinates. The coordinates of the client relative to the receiver is given by (2).

$$\{X_{CR}, Y_{CR}, Z_{CR}\} = \{(x_r - x_c), (y_r - y_c), (z_r - z_c)\} \qquad (2)$$

Here, CR indicates the coordinate of the client relative to the receiver, r indicates wireless receiver's coordinate and c indicates wireless client's coordinate.

We took 500 random perturbations of each relative client position, for each of the 1000 possible client positions in the grid. We fed this data to the ANN as input to train all possible erroneous coordinates for every coordinate point in the grid. It took approximately 48 hours to generate the random data for ANN input.

## 3.3 Artificial Neural Network

Neural networks could best describe for input-output mapping, using relationship between inputs (perturbed relative client coordinates) and targeted outputs (absolute coordinates for every point in the grid). The training data set is feed initially to the ANN with a specific learning rate and it started adjusting the errors. The error information is fed back to the system which formulates all tuning to their parameters in an orderly mode. The training of the network continues many times until the network achieves its targeted output. After the training phase, we setup the ANN for the validation and testing phase with partial input of data sets, which were excluded from the training phase. Simplification occurs for a realistic output. The breakdown of the partitions of the input datasets is given below:

Training Data: 70%
Cross-Validation Data: 15%
Testing Data: 15%

Below algorithms were used for the ANN system:

Data division: Random
Free Parameters: Appropriate initial values of weight
Training: Scaled Conjugate Gradient Algorithm
Performance- Mean Square Error, Regression, Histogram
Samples: 500 samples for each case
Cases: 1000

We use multi-layer perceptron architecture with one output layer, one input layer and one hidden layer. By varying the number of processing neurons from 2 to 10 in the hidden layer, and applying scaled conjugate gradient as a training algorithm for the system; we achieved best performance.

In the conjugate gradient algorithms a search is performed along conjugate directions, which produces generally quicker convergence than steepest descent directions. The scaled conjugate gradient algorithm (SCG) was designed to avoid the time-consuming linear search. This algorithm is too complex to explain in a few lines, but the basic idea is to combine the model-trust region approach, with the conjugate gradient approach.

We represent the obtained error as *E(n)* in (3) [7],

$$E(n) = \frac{1}{2} \sum_{i \in C} ((t_i(n) - o_i(n))^2 \qquad (3)$$

Here N is the number of samples with which network has trained, *C* includes all the neurons in the output layer, and $t_i$ and $o_i$ are the targeted output and actual output values for *i-th* neuron respectively. Therefore average error $E_{av}$ was obtained by summing *E(n)* over all *n* which is given by (3) [7],

$$E_{av} = \frac{1}{N} \sum_{i \in C}^{N} E(n) \tag{4}$$

The purpose of the transfer functions is to calculate a layer's output from its net input. The activation function applied in the hidden layer was non-linear and in the output layer is linear. The nonlinearity of the hidden layer activation functions which were based on the hyperbolic tangent sigmoid function provided the generalization. The adaptability was needed for a proper and accurate location system with adaptive slope in (5),

$$f(v,a) = \frac{1-e^{-kv}}{1+e^{-kv}} \tag{5}$$

Here, the slope parameter 'k' could tune so that error is minimized. Another advantage of sigmoid function is that these are differentiable functions, which is a requirement for the proposed ANN.

For a single epoch $w_k$ is a vector of weights and biases, $\nabla E(w_k)$ is the current gradient and $\alpha$ is the learning rate. Thus, the relationship of these parameters for an epoch can be written as (6),

$$w_{k+1} = w_k - \alpha \nabla E(w_k) \tag{6}$$

$\alpha$ is a significant parameter in artificial neural networks is the learning rate. It affects the learning capability of the ANN and an appropriate value is required to perform a satisfactory training. The correct setting of the learning rate is often dependent on the size and type of input data and is typically chosen through experimental testing. Its value can also be adapted during the training phase, therefore becoming time dependent. The value for the learning rate chosen was 0.01. This low value is related to the nature of the input values that were between -1 and 1. SCG demonstrate the linear convergence which is a sub class of Conjugate Gradient Methods and a faster algorithm as it avoids a time consuming line-search per learning iteration. The error function of all weights can be written as (7),

$$J(w) = \frac{1}{2N} \sum_{i=1}^{N} \sum_{j=1}^{n} (t_{ij} - o_{ij})^2 \tag{7}$$

Where, $w$ represents all the weights in the network, $N$ is a total number of training patterns, $n$ is the number of neurons in output layer, $t_{ij}$ is targeted output of the *i-th* neuron in the output layer to the *j-th* training input, and $o_{ij}$ is the actual output of the *i-th* neuron in the output layer to the *j-th* training input.

It took approximately 24 hours to train the ANN. The training process of a neural network must be tolerable so that problems like poor fitting or over-fitting should not arise. Due to over- fitting, ANN collapses all the generalization aptitude. In contrast, a poor training results on scarce learning.

After the training process the ANN is ready to receive data and calculate the wireless client position. When a primary try did not generate good results, minor improvement could seek by retrain the data. If retraining did not help, then hidden layer neuron size could increase to get better performance. After the training process, the ANN is ready to receive data and determine location by calculating the wireless client coordinate. Then Euclidian distance could be calculated from (8),

$$D = \sqrt{((x_r - x_c)^2 + (y_r - y_c)^2 + (z_r - z_c)^2)} \tag{8}$$

Here, $x_r$ ,$y_r$ ,$z_r$ are the coordinates of each receiver and $x_c$,$y_c$,$z_c$ are coordinates of wireless client. Training automatically stops when generalization stops improving, as indicated by an increase in the mean square error of the validation samples.

### 3.4 Trilateration

Trilateration was performed with the 2-dimensional data (ignoring the z-coordinate). A system of 3 equations based on Pythagorean Theorem was solved to determine the intersection of the three circles. The relative coordinate measurements from each of the wireless receivers were essential for each point. Then Euclidian distance could be calculated from (8).

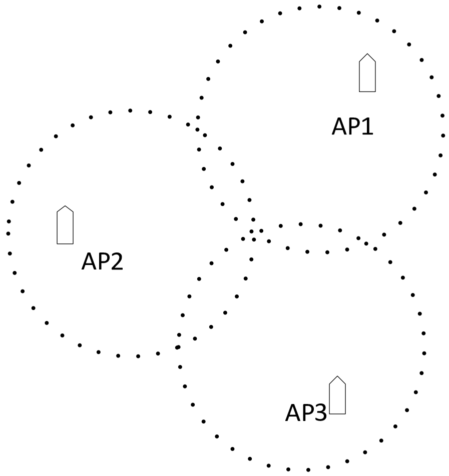


Fig. 1 Trilateration for 2D coordinate system.

## 4 Experiments and Results

A MATLAB simulation was created to determine the location of the wireless device using the artificial neural network.

### 4.1 Generation of Data

A 10m X 10m X 10m grid was created (Fig. 1), whose coordinates ranged from [1,1,1], [1,1,2]……, to [10,10,10]. The relative client coordinates were fed to the ANN input. 500 samples of random perturbations for each of the coordinates with maximum 5% error were generated. Each coordinate was given the same learning parameters and simulation time. In the 3D grid, the fixed coordinates of the receivers were as follows: Rx1=[0, 0, 0]; Rx2=[7, 7, 7], Rx3=[12, 12, 12].

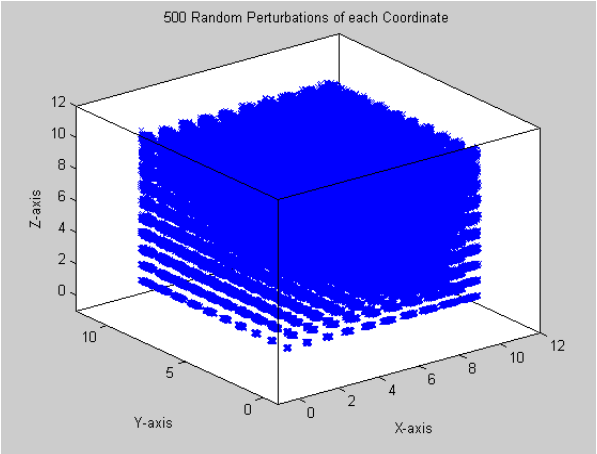


Fig. 2 Random perturbations of each coordinate.

Indoor RSS value goes through attenuation and reflection, which results in erroneous coordinates for a single coordinate. Thus we generated 500 random perturbations for each coordinate position. We fed the ANN with a 9-component training vector as input (Fig. 3) that was obtained from the client coordinates relative to the three receivers.

Every receiver provides a 3-input coordinate for each of the 500 erroneous coordinates. Thus we generated input data for all 1000 target sets. Fig. 1 shows 500 samples of random perturbations for each of the coordinates with maximum random perturbation of 5%.

### 4.2 Artificial Neural Network Setup

The data was collected using three receivers, specifically 500 times for all experimental cases. The calibration was such that the subsequent measurements of relative distance of coordinates between wireless client and receivers were fed to the ANN as input cell array. For every receiver, we had received relative distance for 3 different coordinates which was collectively considered as a single input set. Then training, verification and testing phases were set using the agreeable approach so that the data are randomly divided for effective representation. In all, 1000 experimental cases were recorded as tolerable representation within the model.

We created the ANN with three layers. One input layer, one output layer and another is hidden layer. The number of neurons in the hidden layer is taken as 10. The quality of generalization depends of this value and free parameters. Multi-layer feed-forward network with sigmoid hidden neurons and linear output neurons could fit multi-dimensional mapping problems arbitrarily well. Also it provided dependable data and enough neurons in its hidden layer. An ANN is an adaptive network of interconnected weight consisting of processing elements or neurons that generalize or classify patterns and relationships between inputs and outputs through a process known as training. The number of neurons can be adjusted to define a network size in the fitting network's hidden layer.

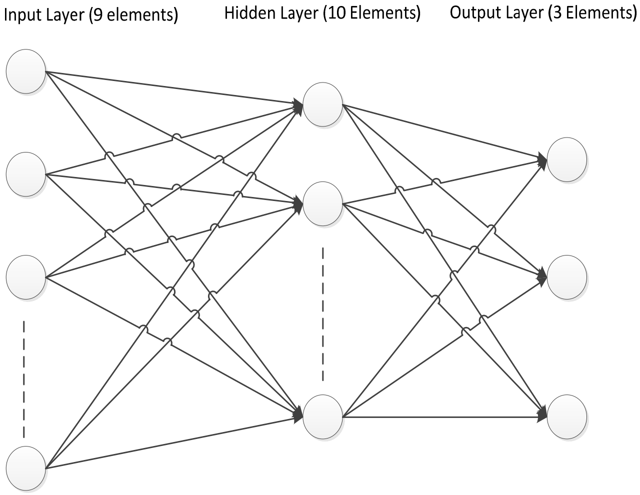


Fig. 3 Multi-layer perceptron architecture.

### 4.3 Training State

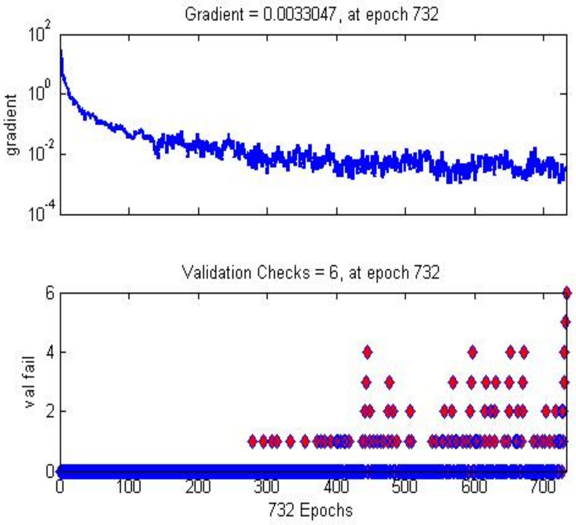


Fig. 4 Various Training States.

In the conjugate gradient algorithms the investigation is performed along conjugate directions, which gives a quicker convergence. Training rate was defined as 0.1. Neural network was mapped between a data set of numeric inputs and a set of numeric targets. Trainings are presented to the network and the network is adjusted according to its error. Validations are used to measure network generalization, and to halt training when generalization stops improving. Testing has no effect on training and so provides an independent measure of network performance during and after training.

From the above training states, it has observed that network achieves its goal in 732 epochs.

## 4.4 Mean Square Error

MSE showed the difference of the real results to the expected ones. Mean Squared Error is the average squared difference between outputs and targets. Lower values are better. Zero means no error.

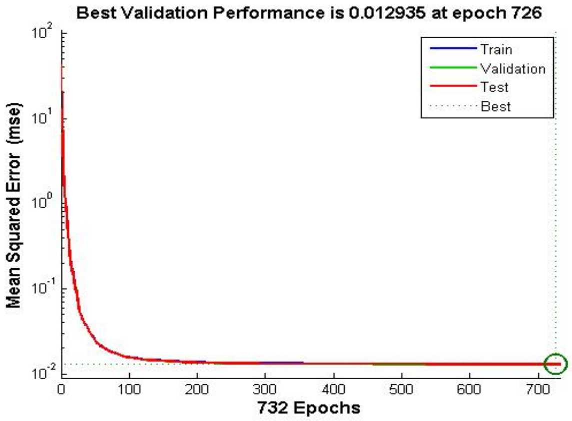


Fig. 5 Mean Square Error (MSE) V.S. Epochs.

Training error decreased, unless it had reached balanced state. From Fig. 5, we found that ANN achieves the best training performance, best validation performance and best testing performance at MSE value of 1.30363 $e^{-2}$, 1.29353 $e^{-2}$, 1.29836 $e^{-2}$ respectively at 732 epochs. It was observed that the training had a sharp fall of MSE value at 50 epoch (approx). Afterwards, there was a insignificant change in the MSE value. Then at 732 epoch, it achieved best classification of the training data. The training stopped at 732 epoch as it had met the target.

## 4.6 Regression

The regression, R, value measures the correlation between outputs and targets. An R value of 1 means a close relationship, 0 means a random relationship among real and expected outcomes. To compare the results of the ANN, we performed a basic trilateration of the 2-dimensional data.

### 4.4.1 Trilateration

Fig. 7 shows a regression analysis of the trilateration method on which the same data the ANN was trained. As shown in the figure, the average regression value for trilateration was R = 0.92.

### 4.4.2 ANN

For the ANN, for each of the training, testing, and validation phases, the average regression value was R=0.99921, which indicates a very close relationship between output and target.

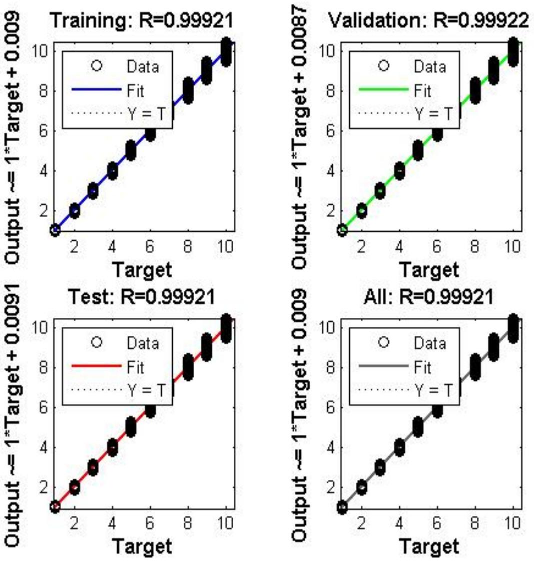


Fig. 6 Regression Analysis at Training Phase, Validation Phase and Testing Phase.

The regression analysis shows that the rate of hits for the ANN for each of the Training, Validation, and Testing Phases is better than that for trilateration. The ANN rate of hits on average is about 8% better than that of trilateration (ANN 99%, Trilateration 92%).

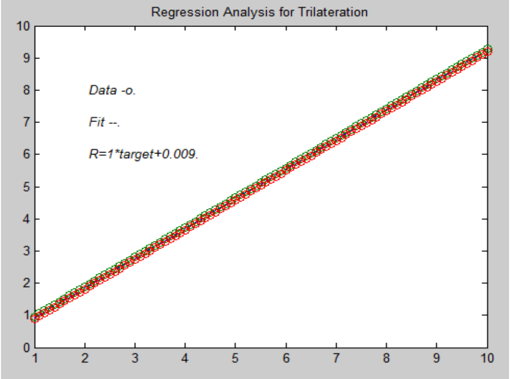


Fig. 7 Regression Analysis for Trilateration.

### 4.6 Network Evaluation

An error histogram was generated for each the Training, Validation, and Testing Phases, as shown in Fig. 8. As the figure shows most of the errors occurred during the training phase with very small error values (0.006118). The number of instances or error becomes more negligible as the value of the error increases. This means the network performed fairly well.

Initially, training was tried with 70% of the input data. If a first try did not generate good results then marginal improvement was needed by retraining the network. If retraining did not help, then network size was increased to get better performance.

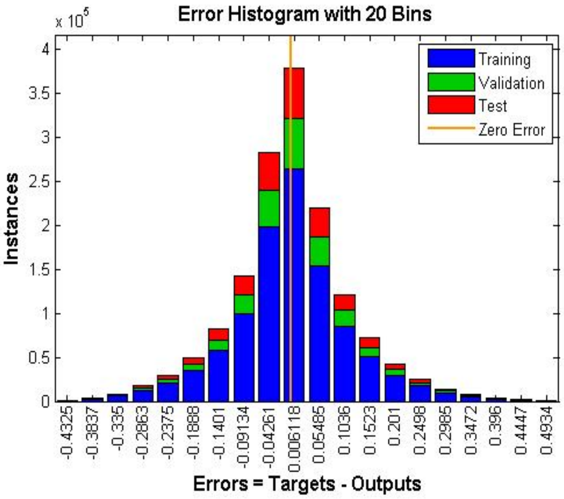


Fig. 8 Error histogram for training, validation and test.

## 5 Conclusions and Future Work

This paper presents a simulation of node localization using an ANN. We used a total of 500,000 input sets. Among them 70% data were used as training sets, 15% data were utilized as validation sets, and remaining 15% data were considered as testing sets. Based on the error histogram, we found the error for validation and testing phases are one fourth than that of the learning phase. Network performs best validation performance 0.012935 at 736 epochs with 10 neurons in hidden layer. The regression analysis shows that the rate of hits for the ANN for each of the Training, Validation, and Testing Phases is better than that for trilateration. The ANN rate of hits on average is about 8% better than that of trilateration (ANN 99%, Trilateration 92%).

Future work will progress to Phase 2, which will capturing RSS measurements of devices and use these data as input to the ANN for training, validation, and testing. Phase 2 will begin with capturing RSS measurements of devices in an open field (to approximate free space). The captured data will be analyzed to determine its random nature; the random nature will be modeled using a correlated fractal model. New data will be generated using the fractal model (instead of the random perturbation method) and fed to the ANN for learning. Once the ANN has learned the new data, an ANN free-space model will have been established. The ANN free-space model will thereafter be used as a basis for comparison with captured RSS data from within the sponsor company's office space. The ANN's representation of the office space will be compared with its free-space representation, and, therefore, the building's electromagnetic map will be defined by the ANN.

## 6 Acknowledgements

This research work was supported by the NSERC Engage Grants (EG) Program (EGP 428305 – 11).

## 7 References

[1] L. Woo, K. Ferens, W. Kinsner and M. Potter, "Analysis of Modulated Monofractal Noise for Noise Modeling in Wireless Networks," *IEEE Transactions on Electromagnetic Compatibility,* vol. 53, no. 2, pp. 524-530, 2011.

[2] K. Ferens, C. Love, A. Indrayanto, A. Langi and W. Kinsner, "A neural network Hamming encoder and decoder," in *IASTED International Conference on Computer, Electronics, Control, and Communication*, Calgary, 1991.

[3] M. Z. Win and Y. Shen, "Fundamental Limits of Wideband Localization — Part I: A General Framework," *IEEE Transactions on Information Theory,* vol. 56, no. 10, pp. 4956 - 4980, 2010.

[4] Y. Shen, H. Wymeersch and M. Z. Win, "Fundamental Limits of Wideband Localization—Part II: Cooperative Networks," *IEEE Transactions on Information Theory,* vol. 56, no. 10, pp. 4981 - 5000, 2010.

[5] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses and N. S. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *Signal Processing Magazine, IEEE,* vol. 22, no. 4, pp. 54 - 69, 2005.

[6] G. L. U. P. S. Kraxberger, "WLAN location determination without active client collaboration," in *Proceedings of IWCMC*, 2010.

[7] I. Vilovii and B. Zovko-Cihlar, "WLAN Location Determinatoin Model Based on the Artificial Neural Network," in *Inernational Symposium ELMAR*, Zadar, Croatia, 2005.

# A Survey of Peer-to-Peer Attacks and Counter Attacks

**Yu Yang and Lan Yang**
Computer Science Department
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768, USA

**Abstract--***Peer-to-Peer (P2P) network is a distributed network architecture that partitions tasks or workloads among peers (nodes). Similar to traditional Internet, P2P networks are open to many attacks. In this research work we survey the defensive measures against general attacks as well as P2P specific attacks. We take BitTorrent (a P2P communications protocol for file sharing) as an example to illustrate defense strategies for Rational attack and Index Poisoning attack, present an algorithm named Self-Registration to defend against Sybil attack, and clarify terminologies for defending Eclipse attack. We summarize and classify the various possible defense mechanisms for both general and P2P specific attacks.*

**Keywords**: Peer-to-Peer (P2P); attack; defense; general attacks; specific attacks

## 1 Introduction

Peer-to-Peer (P2P) technology implements peers (nodes) of equal standing with other peers (nodes) in a P2P network. Each node not only accepts the service, but also provides the service, and nodes can exchange information directly. P2P networks make good use of network resources by utilizing the idle resource of the nodes to develop an efficient information sharing platform. At present, P2P technology is widely used in file sharing protocols such as BitTorrent and Dropbox, as well as in instance message communication systems such as Skype. Similar to traditional Internet, P2P networks are open to many general attacks, such as Denial-of-Service (DoS) attack, Distributed Denial-of-Service (DDoS) attack [9], Man-in-the-middle attack [9], Worm propagation [3], and Pollution attack [4]. To defend these general attacks, technologies and mechanisms for ensuring network safety usually come from security companies (for example, the Verizon Business [13]) and the common network knowledge, such as encryption mechanisms and authentication technologies. Also, some well-known safety measures, such as firewall, anti-virus software, and security operating systems, provide the relative defensive strategies. P2P networks can also be the victim of some P2P specific attacks. Rational attack [7], Index Poisoning attack [4], Sybil attack [14], and Eclipse attack [14] are P2P specific attacks. The secure mechanisms for defending these P2P specific attacks are from a variety of sources. In this research work, we survey general attacks as well as P2P specific attacks and analyze defense strategies for each attack surveyed. In particular, we use BitTorrent as an example to illustrate the defensive measures against Rational attack and Index Poisoning attack. We present an algorithm called Self-Registration [2] to defend against Sybil attack, and clarify terminologies that are used to defend Eclipse attack. The rest of the paper is organized as follows. In section 2, five types of general network attacks and their defense mechanisms are presented. In section 3, P2P specific attacks and their corresponding defensive strategies are described. Finally, summary and classification of attacks and defenses including analysis of attack behaviors, defense strategies, risk analysis and level of defense are presented.

## 2 General Attacks and Defenses

### 2.1 Denial-of-Service (DoS) Attack

DoS attack is an attack on a computer or a network, attempting to make a computer resource unavailable to its intended users. In P2P networks, the most common form of DoS attack is an attempt to flood the network with bogus packets, thereby preventing legitimate network traffic. Another method is to drown the victim node with fastidious computation so that the node becomes too busy to answer any other queries [9].

Defenses:

A widely used technique to hinder DoS attacks is “pricing” [9]. In this technique, the host will submit puzzles to its clients before continuing the requested computation. When an attacker attempts to flood his victim, he has to solve a puzzle first, thus it becomes more difficult for the attacker to launch a successful DoS attack.

### 2.2 Distributed Denial-of-Service (DDoS) Attack

DDoS attack is an attacking technique based on the DoS attack [9]. The system of DDoS attack includes four parts as Figure 1 shows.

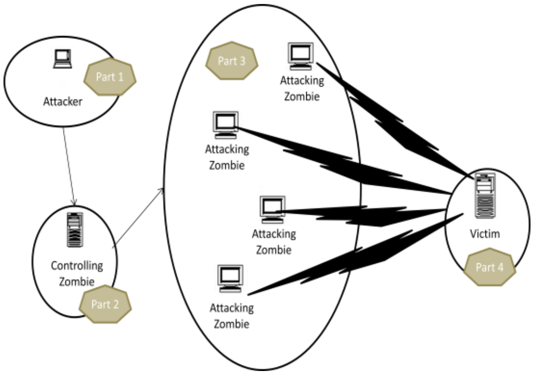


Figure 1. DDoS Attack

The first part is the actual attacker, who controls the part 2 and part 3. Part 2 and part 3 are often personal computers with broadband connections that have been compromised by a virus or Trojan. The difference between part 2 and part 3 is: from the point of view of part 4, the victim, the attacking comes from part 3, the attacking zombies while part 2 only issues an attacking order from the actual attacker without actually attending an attack. The detailed parts of DDoS attack can also be developed as shown in Figure 2.

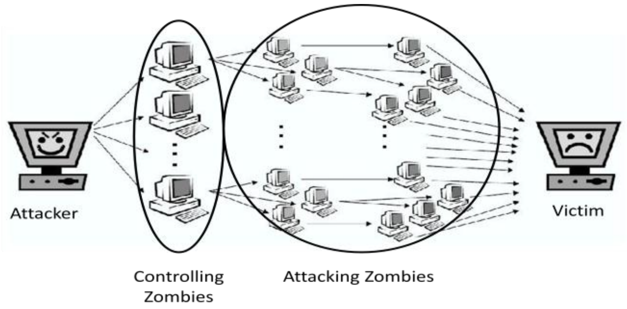


Figure 2. Developed DDoS Attack

In the developed DDoS attack, the hacker controls more than one controlling zombies, and each controlling zombie also controls a lot of attacking zombies and so on [9]. So, in DDoS attack it is hard to trace the actual attacker, because the attacker is often indirectly involved.

Defenses:

DDoS attacks are extremely hard to block due to the enormous numbers and diversity of machines involved in the attack. However, there are still many companies proposing countermeasures to defend against DDoS attack. Take Verizon business security team for example [13]. In the online broker's business, when hackers use DDoS attack to launch some attacks, the companies will lose revenue, productivity and reputation. The attacks will cause the broker's clients to experience timed-out pages, slow loading times, and overall non-responsiveness to user inquiries. And the company will receive the notice to demand an extortion in order to stop the crippling attacks or prevent the coming attacks.

There are three steps to prevent DDoS attacks:

First, let the broker company's Internet traffic through Verizon business, which will help the clients to filter a series of malicious information. Second, security team offers a monitoring and detection capability that constantly searches incoming DDoS attack. This warning system also gives the broker the ability to determine the extent of an attack and respond with the proper level of mitigation that could help protect against losses. Finally, the brokers can have their own blacklist or whitelist, which allow the brokers to terminate blacklisted traffic before it reached the brokers' Internet site while allowing whitelisted traffic to always be permitted [13].

## 2.3 Man-in-the-middle Attack

Man-in-the-middle attack is an indirect intrusion, and the attacker inserts his computer undetected between two nodes [9].

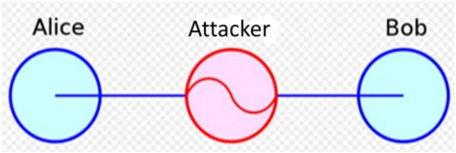


Figure 3. Man-in-the-middle Attack

In Figure 3, Alice and Bob are normal users. The attacker in the middle can intercept data, modify data and send data without being detected by Alice and Bob.

Defenses:

So far from our literature survey we haven't yet found any effective defense strategies for this type of attack. However, deriving from the common network knowledge we propose the following defense strategies. First, encryption mechanism should be used to protect the information to be transmitted. The information is encrypted with some encryption methods before being transmitted. Even though the intruder intercepts the information, he is unable to decrypt the message without knowing how to decrypt the message [12]. Also, authentication technologies should be used to detect Man-in-the-middle attack. The authenticator includes redundant information about the message contents, such as who created the authenticator, who is the sender of the messages. In other words, authentication is used to verify and distinguish the authenticity and validity of a user [8] [10]. The purpose of this technique is to distinguish legal users from illegal users.

## 2.4 Worm Propagation

Worm transits the copies of itself from one node to others through the network communication, and starts by itself. Worm can be propagated through file, email, web server, and so on [3].

Defenses:

The defense strategies we recommend here are to use some common network measures that have already been widely used in many computer systems. The first one is using firewall. Most of the time, worm scans a certain port in the computer to infect, and firewalls can block the port that worm needs. Also, we can use some anti-virus software to protect our computers. The anti-virus software includes the virus signature, if some attributes of the file correspond to the attributes in virus signature, the anti-virus software can delete or isolate that file [11]. The last defense has been offered by security measures from operating system developers. For example, OpenBSD operating system concentrates on the aspect of security and possesses many security features such as protecting the operating system from buffer overflows or integer overflows, which makes an attacker without any ideas of what data segment he should overwrite [9].

## 2.5 Pollution Attack

The practice of this attack is to replace a file in the network by a false one, and this polluted file is of no use to the clients [4]. The attacker makes the target content unusable by changing the contents or part of it into another irrespective content, and then makes this polluted content available for sharing. In order to attract people to download the polluted content, the polluted content needs to disguise itself as the target content, such as having the same format and similar size. It also needs to keep high-bandwidth connections.

Defenses:

From the user's side, the downloaded file that has been polluted is not harmful to our computers, but it is just of no use. Therefore, in our opinion, once a user finds out that the downloaded files are polluted files, the user should remove the files from the P2P system.

# 3 P2P Specific Attacks and Defenses

## 3.1 Rational Attack

In most P2P systems, self-interested behavior at the expense of the system can be classified as a Rational attack [7]. For instance, Figure 4 shows a possible scenario of Rational attack.

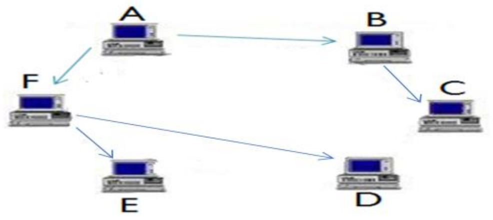


Figure 4. Rational Attack

In the P2P system shown in Figure 4, node A wants to distribute content. To decrease the upload bandwidth burden on the node A, only a small number of nodes such as node B and node F are directly connected to it. The content is then propagated from node B and node F to additional peers such as node C, D and E. Because of the self-interested behavior in most P2P systems, a self-interested node may realize that it can save expensive upload bandwidth if it chooses not to share. If a large number of nodes are self-interested and refuse to contribute, the system may destabilize [7]. In this case, if enough nodes such as B and F become self-interested, the system cannot guarantee a reasonable level of uploads and downloads.

Defenses:

Here we take BitTorrent as an example to illustrate the countermeasure of Rational attack. BitTorrent is popularly used for file distribution. In BitTorrent, there is an algorithm called Choking algorithm [1] [5], which can guarantee a reasonable level of upload and download reciprocation. If peers just download, and never upload, they should be penalized.

Terminology in Choking algorithm:

*Pieces and Blocks*: transmission unit on the network.

*Interested and Choked*: peer A is interested in peer B when peer B has pieces that peer A does not have. Otherwise, peer A is not interested in peer B. Peer A chokes peer B when peer A decides not to send data to peer B. Otherwise, peer A unchokes peer B.

*Planned optimistic unchoked peer*: a random peer that is choked and interested.

*Active peer:* a peer has sent at least one block in the last 30 seconds.

The flowchart is shown in Figure 5 describes details of the Choking algorithm.

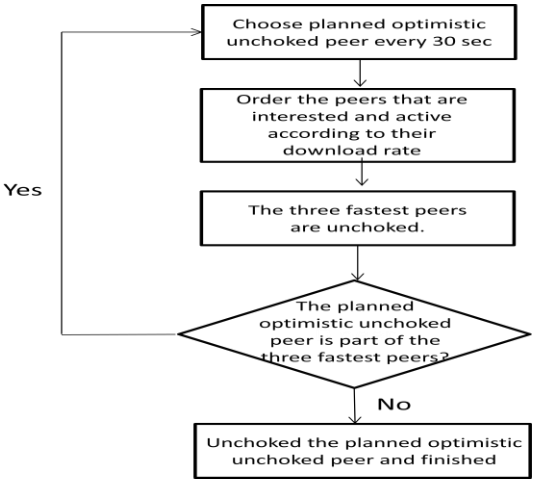

Choose planned optimistic unchoked peer every 30 sec
Order the peers that are interested and active according to their download rate
The three fastest peers are unchoked.
The planned optimistic unchoked peer is part of the three fastest peers?
Yes
No
Unchoked the planned optimistic unchoked peer and finished

Figure 5. Flowchart of Choking Algorithm

## 3.2 Index Poisoning Attack

Most P2P file sharing systems have indexes, allowing users to discover locations of desired content. Index poisoning aims at the index querying process of users and makes it hard to find correct content in P2P network. The attackers simply insert large numbers of invalid peer information into the index to hinder the users from finding correct resource [4]. For example, BitTorrent is easy to be attacked by Index poisoning. In BitTorrent, first, we need to download a complete file known as a seed with the extension .torrent. The .torrent contains information about the file, such as its length, name, and a tracker. The tracker acts as an information exchange center from which peers obtain necessary information about other peers, which are downloading the same file. When a peer starts a BitTorrent task, it first advertises its information into the tracker, and then the peer contacts the tracker and gets a list of other peers' information. When a tracker receives an advertisement for a task from a peer, it does not authenticate the advertisement and does not verify whether the content is truly available with the advertised information or not. The attacker deliberately advertises large quantity of invalid peer information of the targeted content. So, when a user attempts to download the content corresponding to the task, his BitTorrent client always fails to establish connection with the other peers, due to the high probability of connecting to invalid peers [4].

Defenses:

There are two measures to defend against the Index poisoning attack. The first one is to authenticate versions and advertisements [6]. Like some rating websites and forums, the content has been initiated with a moderator to manage disputes. The second method is rating sources [6]. If these are good sources, which advertise and upload files they actually have, the corresponding peers will get high rating scores. If these are bad sources, whose index poison and pollute the system, the corresponding peers will be blacklisted.

## 3.3 Sybil Attack

Many P2P systems introduce a redundant backup mechanism to protect integrity and privacy. A P2P system must ensure that each network entity ID indicates only one entity. If an entity acts as a number of multiple identities, this entity can control a significant part of networks. Such attack is defined as Sybil attack. Sybil attack will destroy the redundancy in P2P network [14].

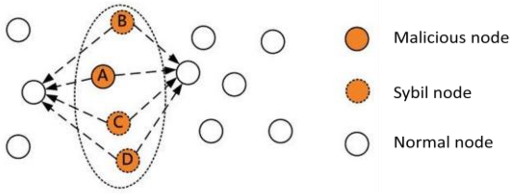


Figure 6. Sybil Attack

In Figure 6, when a normal node makes redundant backup, it selects a group of entities such as node A, B, C and D that have different IDs. But in fact, node B, C and D actually do not exist, as they are the malicious nodes created by the attacker, so the backup cannot finish.

Defenses:

The countermeasure is an identity registration procedure called "Self-Registration" [2], which is shown in Figure 7 and explained below:

A new node hashes its IP address and port to calculate its identifier, and then register its identifier at already registered nodes, which are the registration process of the new node. After that, the new node requests to join P2P network. Other registered nodes have the ability and responsibility to identify whether the new node is real or not. If the new node is not fake, it will be accepted by the P2P network.

*Registration nodes*: in this procedure nodes are verified that they are not fake nodes.

*New nodes*: In this procedure, a node is checked that its ID and Registration ID are one-to-one mapping.

The Self-Registration algorithm consists of two parts, the "Registration node" and the "New node". The functionality of both parts is described in Figure 7.

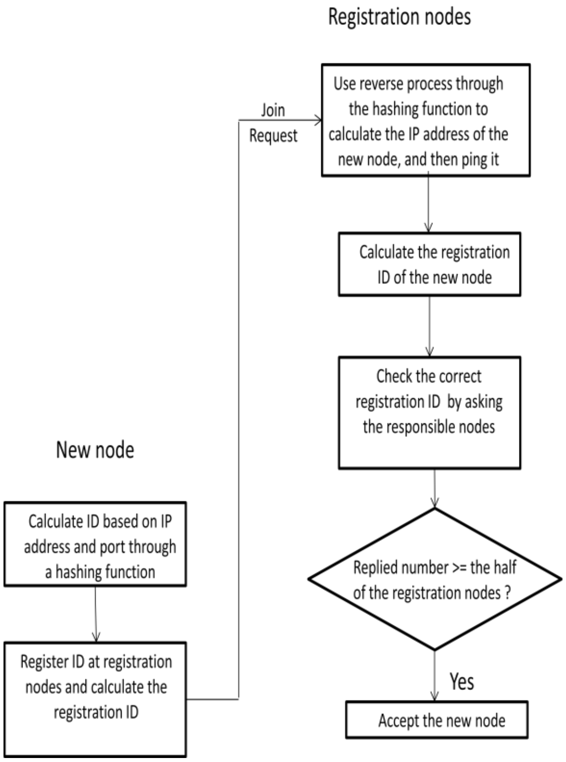

Registration nodes
New node
Calculate ID based on IP address and port through a hashing function
Register ID at registration nodes and calculate the registration ID
Join Request
Use reverse process through the hashing function to calculate the IP address of the new node, and then ping it
Calculate the registration ID of the new node
Check the correct registration ID by asking the responsible nodes
Replied number >= the half of the registration nodes ?
Yes
Accept the new node

Figure 7. Self-Registration Algorithm

### 3.4 Eclipse Attack

In an Eclipse attack, an attacker controls a large part of a good node's neighbors. In this situation, the union of malicious nodes works together to fool a good node by writing their addresses into the neighbor list of a good node. By using Eclipse attack, an attacker can control the significant part of a network, even the entire network. Thus, nodes cannot forward message correctly and the whole network cannot be managed. A Sybil attack can be considered as a specific Eclipse attack, if the attacker generates great amount of identifications to act as neighbors of a good node [14]. For instance, a scenario of an Eclipse attack is shown in Figure 8.

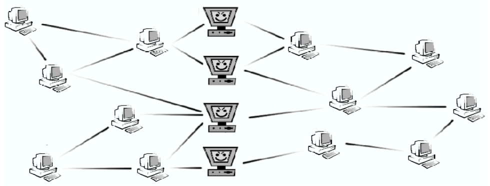


Figure 8. Eclipse Attack

In Figure 8, the malicious nodes separate the network into two subnetworks. No matter what methods are used to communicate within two subnetworks, the normal nodes cannot avoid connecting with one of the malicious nodes. So, the entire network has been controlled by malicious nodes.

Defenses:

Before introducing the countermeasure to against an Eclipse attack, we need to clarify two terminologies, which are indegree and outdegree. Indegree means the number of direct routes coming into a node and outdegree means the number of direct routes going out of a node. The idea to defend against Eclipse attack is to bound both indegree and outdegree of the attacker nodes. This method can be described as follows. First, we apply the countermeasure to the Sybil attack. This process assures there is no possibility of Eclipse attack based on a Sybil attack. Then we concentrate on how to deal with the indegree and outdegree of the attacker nodes. Each node in P2P networks maintains a list of its neighbors. We make a node periodically query the neighbor lists of its neighbor peers. If the items on the replied neighbor list are greater than the indegree bound, or that node is not on its neighbor's list or the size of returned neighbors is greater than the outdegree bound, it means an Eclipse attack happened [14].

## 4 Conclusions and Future Improvement

In this paper, we describe a list of network attacks that are common in current P2P networks. Some of these attacks are general attacks occurring over the traditional Internet that also applies to P2P networks, while others are specific attacks against P2P networks. General attacks described in this paper include DoS attack, DDoS attack, Man-in-the-middle attack, Worm propagation, and Pollution attack. P2P specific attacks include Rational attack, Index Poisoning attack, Sybil attack, and Eclipse attack. Countermeasures to defend each of the general and specific attacks in P2P networks are discussed and analyzed. BitTorrent is used to illustrate the defensive measures against Rational attack and Index Poisoning attack. Examples are used to illustrate various attacks in P2P network. In the following Table 1, we clarify the defense measures and the behaviors of the attacks. Table 1 also summarizes the risk analysis and the level of defense. The summary is derived from the information we collected and analyzed from the above described attacks and defense strategies on P2P networks.

Future will includes more in-depth study of effective defense strategies for various attacks on P2P networks, and survey multiple attacks on one Peer-to-Peer network.

Table 1: Summary of Attacks and Defense Strategies

| **Name of Attack** | **Behavior** | **Defense strategy** | **Extent of Danger** | **Level of Defense** |
|---|---|---|---|---|
| Denial-of-Service (DoS) | 1. Flood the network with bogus packets.<br>2. Drown the victim in fastidious computation. | Pricing | Medium | Easy |
| Distributed Denial-of-Service (DDoS) | Hacker controls the controlling zombies, through the controlling zombies to control attacking zombies to launch the attack. | Through the trusted server, provide warning system, and created blacklist and whitelist for trusted visits. | High | Hard |
| Man-in-the-middle | An attacker inserts himself undetected between two nodes, and intercept, modify and send data between those two nodes. | Encryption mechanism and authentication technology | Medium | Medium |
| Worm Propagation | Transits the copies of itself from one node to others automatically. | Firewall, anti-virus and some safety operating system | Medium | Medium |
| Pollution | Share a file, which is unused. | Remove it | Low | Easy |
| Rational | Download the resource and refuse to upload. | Choking algorithm | Medium | Medium |
| Index Poison | Poison the index information to make the node hard to find correct content. | Authenticate versions and advertisements, rating sources | High | Medium |
| Sybil | An attack controls a number of identities | Self-Registration algorithm | High | Hard |
| Eclipse | The malicious nodes work together to fool the good nodes. | Indegree and Outdegree method | High | Hard |

## 5 References

[1] B, Cohen, Incentives Build Robustness in BitTorrent. In 1st International Workshop on Economics of P2P Systems, pp. 1-5, June 2003.

[2] J. Dinger, and H. Hartenstein, Defending the Sybil Attack in P2P Networks: Taxonomy, Challenges, and a Proposal for Self-Registration. In Proceedings of the First International Conference on Availability, Reliability and Security. Institut fur Telematik, Universitat Karlsruhe (TH), Germany, 2006.

[3]X. Fan, and Y. Xiang, Propagation Modeling of Peer-to-Peer Worms. In 2010 24th IEEE International Conference on Advanced Information Networking and Applications. Central Queensland University, Rockhampton, Australia,2010, pp. 1128-1135.

[4] J. Kong, W. Cai, and L.Wang, The Evaluation of Index Poisoning in BitTorrent. In 2010 Second International Conference on Communication Sofware and Networks. Northewestern Polytechnical University, Xi'an, China, 2010, pp. 382-386.

[5] A. Legout, U. Guillaume, and M. Pietro, Understanding BitTorrent: An Experimental Perspective. In IEEE/INFOCOM’05, 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Institut Eurecom, Sophia Antipolis, France, 2005, pp. 2235-2245.

[6] J. Liang, N. Naoumov, and K.W. Ross, The Index Poisoning Attack in P2P File Sharing Systems. In 25th IEEE International Conference on Computer Communications. Polytechnic Univerisy, Brooklyn, NY, 2006, pp. 1-12.

[7] S. J. Nielson, S. A. Crosby, and D. S. Wallach, A Taxonomy of Rational Attacks. Department of Computer Science, Rice University, Houston, Texas, 2005.

[8] L. L. Peterson, and B.S. Davie, Computer Networks: A Systems Approach. Elsevier, Inc. San Francisco, CA 2007.

[9] B. Pretre, Attacks on Peer-to-Peer Networks. Department of Computer Science, Swiss Federal Institute of Technology (ETH) Zurich, Swiss, 2005, pp. 6-15.

[10] W. Stallings, Cryptography and Network Security: Principles and Practices. Prentice Hall, Upper Saddle River, NJ, 2005.

[11] F. Su, Z. Lin, and Y. Ma, Effects of Firewall on Worm Propagation. Proceedings of ICCTA 2009. Research Institute of Networking Technology, Beijing University of Posts and Telecommunications, Beijing, China, 2009, pp.880-884.

[12] A.S. Tanenbaum, Computer Networks. Prentice Hall PTR, Upper Saddle River, NJ, 2003.

[13] Verizon business, Major Online Stock Broker Turns to Verizon Business to Help Stop a Potentially Devastating DDoS Attack. Verizon business, 2008.

[14] L. Wang, Attacks Against Peer-to-Peer Networks and Countermeasures. TKK T-110.5290 Seminar on Network Security. Helsinki University of Technology, Finland, 2006.

# A System for Detecting a Port Scanner in 3G WCDMA Mobile Networks

**K. Sekwon**[1], **O. Joohyung**[1], **I. Chaetae**[1], **and K. Inho**[2]
[1]Korea Internet & Security Agency, IT Venture Tower, Jungdaero 135, Songpa, Seoul 138-950, Korea
[2]SKtelecom, SK T-Tower, 11, Euljiro 2-ga, Jung-gu, Seoul 100-999, Korea

**Abstract** - *Currently, there has been a 3G mobile networks data traffic explosion due to the large increase in the number of smartphone users. Unlike the traditional wired infrastructure, 3G mobile networks have limited wireless resources and signaling procedures for complex wireless resource management. However, mobile network security for various abnormal and malicious traffic technologies was not ready. So malicious or potentially malicious traffic originating from mobile malware infected smart devices can cause serious problems to the 3G mobile networks, such as DoS and scanning attacks in wired networks. In this paper, we describe the port scanning attacks and propose a port scanning detection system based on the Threshold Random Walk (TRW) algorithm in 3G mobile networks. In 3G WCDMA mobile networks, the proposed system detects a variety of port scanning attack in real time. The results of applying the 3G WCDMA mobile network show that the proposed systems are practical and effective.*

**Keywords:** 3G, WCDMA, Port Scanner, TRW

## 1 Introduction

Currently, 3G mobile networks such as WCDMA and CDMA 2000 have been built. As of December 2005, there were over 300 million CDMA subscribers worldwide. Emerging 3G mobile network standards such as EV-DO and HSDPA promise to deliver broadband mobile internet services with peak rates of 2.4 Mbps and 14.4 Mbps, and HSPA+ will allow uplink speeds of 11Mbps and downlink speeds of 42Mbps, respectively. Also 3G mobile networks with a higher mobility than a Wi-Fi environment was provided.

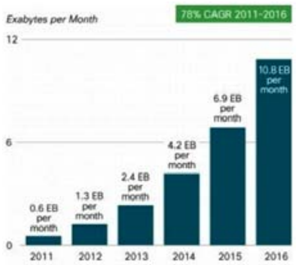


Fig. 1. Cisco Forecasts 10.8 exabytes per month of mobile data traffic by 2016.

However, there has been a 3G mobile network data traffic explosion due to the large increase in the number of smartphone users. Also, new mobile services to satisfy the various needs of smartphone users are being developed day by day. In other words, this means an increase in data traffic over the mobile networks. Fig. 1 shows the mobile data traffic growth forecast[1][2].

Unlike a traditional wired infrastructure, 3G mobile networks have limited wireless resources and signaling procedures for complex wireless resource management. So this data traffic is not a problem in wired networks, but in mobile networks this can be a threat.

Especially, not all of the data traffic that flows into the mobile network is normal. There are segments of unnecessary or abnormal traffic. In the existing wired network, this was not a serious problem, but it can be a threat for the mobile networks, which has a narrow bandwidth and limited wireless resources. In practice, the speed for the data service of 3G mobile networks within crowded locations of smartphone users apparently decreases, or the data service cannot be properly provided[3].

Malicious or potentially malicious traffic originating from mobile malware infected smart devices can cause serious problems to the 3G mobile networks, such as a DoS and scanning attack in wired networks. Unlike the traditional wired network, mobile network security for various abnormal and malicious traffic technologies was not ready. Mobile networks, such as a communication facility, can be viewed as a national infrastructure. If mobile networks are not supported by appropriate security technologies, they can be a target of cyber terrorism by hackers, which can cause serious economic and social losses to mobile communication service providers.

In this paper, we describe port scanning attacks and propose a port scanning detection system based on the Threshold Random Walk (TRW) algorithm in 3G mobile networks[4][5]. And the test results of proposed system is presented in the 3G WCDMA mobile network. This paper is structured as follows: first, in Section 2, we overview the 3G WCDMA mobile networks and TRW algorithm for detecting the port scanner. In Section 3, the port scanning attacks on

3G WCDMA mobile networks are described. In Section 4, we propose the port scanner detection system based on the TRW algorithm. In Section 5, the test results of the proposed systems is presented in the 3G WCDMA mobile network. Finally, conclusions and future works are given in Section 6 and 7.

# 2 Background Information

In this section, we overview the 3G mobile networks and TRW algorithm that can detect a port scanner.

## 2.1 3G WCDMA Wireless Networks

The 3G mobile network technology described in this paper is the WCDMA, which has been adapted as the 3G mobile network technology in many countries. The network structure of the WCDMA is mainly separated by UE, UTRAN, and the Core Network (CN), as Fig. 2.

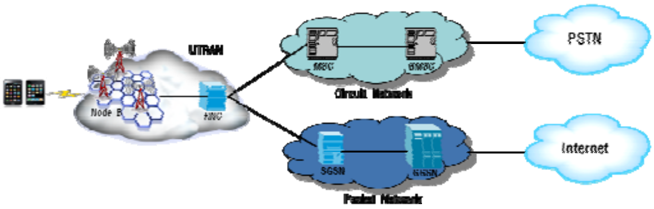


Fig. 2. 3G WCDMA mobile networks.

User Equipment (UE) means a terminal of users that are connected to the 3G mobile network, and UTRAN is a network that controls the wireless resources of terminals. The Core Network can be sub-divided into the Circuit Switched Network for call service and the Packet Switched Network for data service.

The main devices of data transmission over the packet network are the Serving GRPS Support Node (SGSN) and the Gateway GPRS Support Node (GGSN). The SGSN is in charge of the service management over the packet network, which is for data service. The GGSN manages IP allocation to terminals and converts the data from the packet network to IP packets to support communications with other Internet networks. Typically, there are multiple SGSNs, each of which serves the GPRS users that are physically located in its serving area.

Another key component of a 3G mobile network is the Radio Network Controller (RNC), which is the point where the wireless link layer protocols terminate. The RNC manages the radio resources for radio access. The RNC provides the interface between a mobile that is communicating through a NodeB and the network edge. This includes the management of radio transceivers in NodeB, admission control, channel allocation, and management tasks such as handoffs between NodeBs, and deciding on the power control parameters. The functionalities of a NodeB include wireless link transmission/reception, modulation/demodulation, physical channel coding, error handling, and power control.

In this hierarchical architecture, multiple mobiles communicate with a NodeB, multiple NodeBs communicate with an RNC, and multiple RNCs talk to the SGSN/GGSN. Each device uses a different protocol and tunneling. Between the RNC and SGSN is the "Iu-PS" interface, which is usually used in an ATM (Asynchronous Transfer Mode) network. The SGSN and GGSN use a protocol called the GPRS Tunneling Protocol (GTP) to transmit user data and the interface is known as the Gn interface[6]. GTP can be categorized as GTP-U for the packet data, GTP-C for signaling, and GTP (prime) for billing[7]. Fig. 3 shows a 3G WCDMA protocol stack.

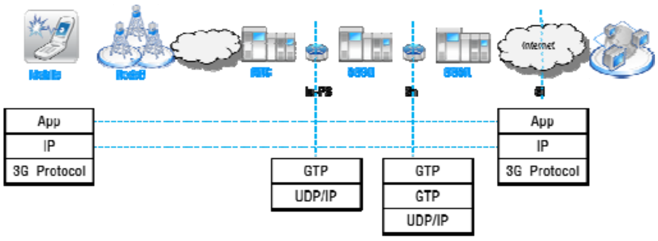


Fig. 3. 3G WCDMA protocol stack.

The UE is representative of smartphones but recently it also includes a variety of devices like laptops and tablets. In particular, a variety of UEs that do not have the function for 3G mobile communication but that have the function for Wi-Fi communication can communicate via 3G mobile networks by using the tethering feature of a smartphone.

The incoming traffic through 3G mobile networks are not only from smartphones but are also from multiple devices such as notebooks and netbooks. Thus, the traffic (in mobile network) will now be observed in various forms.

## 2.2 Related Works

Recently, there is increasing concern about the management and security of 3G mobile networks due to an increase in data traffic that flows into it. Here, not all of the data traffic that flows into the mobile networks is normal[3]. Many detection and corresponding technologies against anomaly traffic in 3G WCDMA mobile networks have been proposed.

F. Ricciato defined anomaly traffic, which might occur in 3G mobile networks, such as scanning or flooding traffic[8]. And V. Falletta tested the syncount and TRW algorithm that can detect port scanners in 3G networks[4].

The TRW algorithm considers two hypothesis, $H_0$ and $H_1$, where $H_0$ is the hypothesis that the given remote source is benign and $H_1$ is the hypothesis that the remote source is a scanner. It also assumes that, conditional on the hypothesis

$H_j$, the random variables $Y_i \mid H_j$, $i = 1, 2, \ldots$ are independent and identically distributed. The distribution of the Bernoulli random variable $Y_i$ can then be expressed as:

$$\begin{aligned} &\Pr[Y_i = 0 \mid H_0] = 0.8, \quad \Pr[Y_i = 1 \mid H_0] = 0.2 \\ &\Pr[Y_i = 0 \mid H_1] = 0.2, \quad \Pr[Y_i = 1 \mid H_1] = 0.8 \end{aligned} \tag{1}$$

where $Y_i$ is a random variable that represents the outcome of the first connection attempt by remote source to the *i-th* distinct local host.

$$Y_i = \begin{cases} 0 & \text{Connection Success} \\ 1 & \text{Connection Failure} \end{cases} \tag{2}$$

And, it calculates the likelihood ratio $\Lambda(Y)$ as follows:

$$\Lambda(Y) = \frac{\Pr[Y \mid H_1]}{\Pr[Y \mid H_0]} = \prod_{i=1}^{n} \frac{\Pr[Y_i \mid H_1]}{\Pr[Y_i \mid H_0]} \tag{3}$$

Finally, the likelihood ratio is compared to an upper threshold, $\eta_1$, and a lower threshold, $\eta_0$. If $Y_i$ then a remote source is a scanner. If $\Lambda(Y) \geq \eta_1$ then a remote source is benign. And if $\eta_0 < \Lambda(Y) < \eta_1$ then it waits for the next observation and updates $\Lambda(Y)$. Fig. 4 shows the flow diagram of the TRW algorithm that can detect a port scanner.

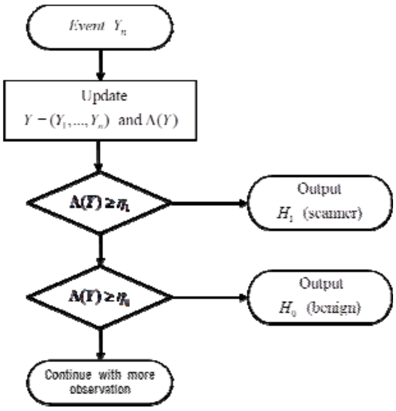


Fig. 4. The flow diagram of TRW algorithm.

# 3 Port Scanning Attack

Scanning attacks are performed to find out the network architecture or the network vulnerability of any systems. It causes a high volume of traffic because of sending traffic to multiple systems at a remote site. Scanning traffic over a general wired-network cannot be a serious problem. In the 3G WCDMA mobile network, however, the problem shows a different aspect. Most of the scanning traffic causes paging traffic and it makes traffic volume weighted in the 3G WCDMA mobile network[3]. In addition, the critical information of 3G WCDMA mobile network configuration equipment, such as the IP address and port number, are exposed by scanning attacks. Therefore, scanning attacks are more fragile than a wired network at the 3G WCDMA wireless network, because it is a closed type of service structure and has a narrow bandwidth and limited wireless resource.

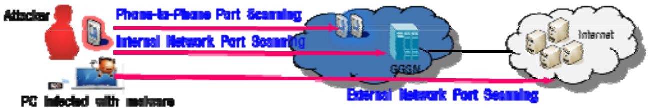
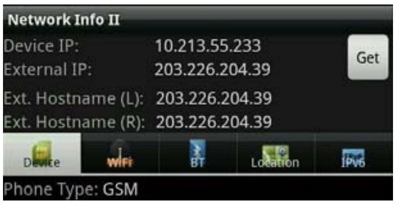

Attacker
Phone-to-Phone Port Scanning
Internal Network Port Scanning
GGSN
Internet
PC infected with malware
External Network Port Scanning

Fig. 5. Port scanning attacks over 3G WCDMA mobile networks.

Fig. 5 shows port scanning attacks in 3G WCDMA mobile networks. An attacker can do scanning attacks using port scanner applications (Port Scanner, TCP Port Scanner, Net Scan, etc.) or tools (Nmap, Superscan, etc.) as follows:

## 3.1 Phone-to-Phone Port Scanning

Check the IP address assigned to the smartphone using applications such as Network Info II



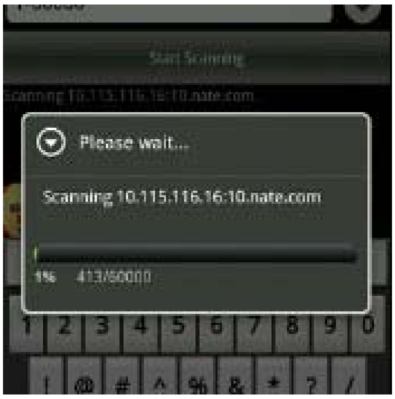
Network Info II

Device IP: 10.213.55.233
External IP: 203.226.204.39
Ext. Hostname (L): 203.226.204.39
Ext. Hostname (R): 203.226.204.39

Phone Type: GSM

Fig. 6. Check IP address assigned to the smart-phone via Network Info II.

Perform a port scanning attack on the target smartphone of the same IP range using the Port Scanner application



Scanning 10.115.116.16:10.nate.com

Please wait...

Scanning 10.115.116.16:10.nate.com

1% 413/60000

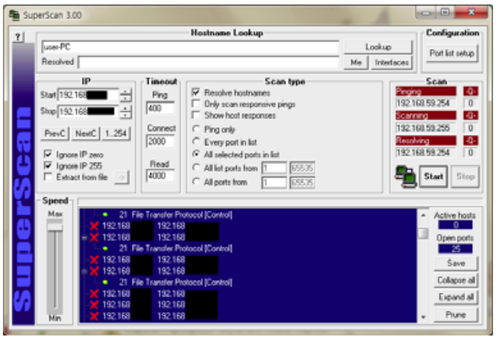
Fig. 7. Port scanning attack on target smart-phone using the Port Scanner application.

## 3.2 Internal Network Port Scanning

Connect a PC to the 3G WCDMA mobile network via the tethering function of the smartphone

Obtain the internal network IP information of the mobile communication service provider via tracert

Perform port scanning attacks on the internal network using the port scanning tool, Superscan



SuperScan 3.00

Hostname Lookup: user-PC
Resolved:

IP
Start 192.168[redacted]
Stop 192.168[redacted]
☑ Ignore IP zero
☑ Ignore IP 255
☐ Extract from file

Timeout
Ping 400
Connect 2000
Read 4000

Scan type
☑ Resolve hostnames
☐ Only scan responsive pings
☐ Show host responses
○ Ping only
○ Every port in list
● All selected ports in list
○ All list ports from 1 65535
○ All ports from 1 65535

Scan
Pinging 0
192.168.59.254 0
Scanning 0
192.168.59.255 0
Resolving 0
192.168.59.254 0

```
    21 File Transfer Protocol [Control]
 192.168[redacted] 192.168[redacted]
 192.168[redacted] 192.168[redacted]
    21 File Transfer Protocol [Control]
 192.168[redacted] 192.168[redacted]
 192.168[redacted] 192.168[redacted]
    21 File Transfer Protocol [Control]
 192.168[redacted] 192.168[redacted]
 192.168[redacted] 192.168[redacted]
 192.168[redacted] 192.168[redacted]
```

Active hosts 0
Open ports 25

Fig. 8. Port scanning attack using Superscan.

## 3.3 External Network Port Scanning

Connect a PC to the 3G WCDMA mobile network via the tethering function of the smartphone

Perform port scanning attacks on the target server in the external network

# 4 System Architecture

In this section, the architecture of the proposed system for port scanner detection in 3G WCDMA networks will be described in detail.

The overall architecture is shown in Fig. 9. It consists of two separate systems, which are the GTP Packet Capture and Parser and the Traffic Flow Management and Port Scanner Detector. The first system captures in/outbound GTP-C packets and outbound GTP-U packet in a Gn interface with an average 6.5Gbps, and extracts the necessary information. The second system consists of the Traffic Flow Management Module that manages the user's session/data flow information and the Port Scanner Detection Module that detects port scanners.

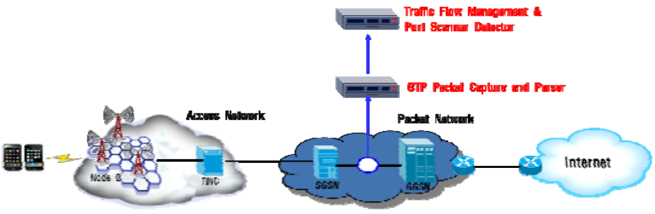


Fig. 9. The structure of port scanner detection systems.

## 4.1 GTP Packet Capture and Parser

Fig. 10 shows the structure of a GTP Packet Capture and Parser. In this system, the DAG Card based on the data stream capturing technique is applied for capturing GTP packets in real time. It captures mirrored GTP packets in the Gn interface and stores them in the buffer. Then it checks the type of GTP packet messages and extracts the fields of GTP messages depending on the type as shown in Table I. If the GTP packet is outbound GTP-U, the GTP Packet Capture and Parser discards the packet.

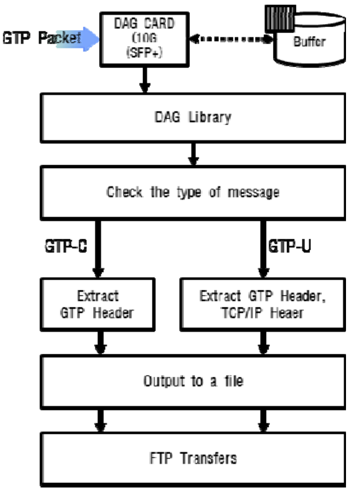
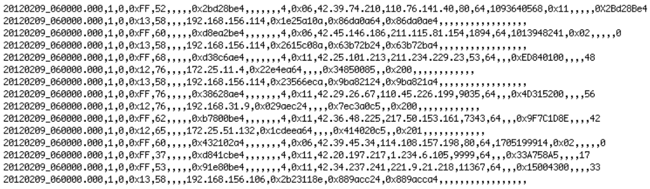


Fig. 10. The structure of GTP Packet Capture and Parser.

TABLE I. THE EXTRACTION FIELD OF GTP MESSAGE IN ACCORDING TO THE TYPE OF MESSAGE

| Type of Message | | GTP Header | GTP Ext Header | TCP(UDP)/IP Header |
|---|---|---|---|---|
| GTP-C | Create request (0x10) | Protocol Type, Reserved, Message Type, Total Length, | IMSI, APN, MSISDN, GSN, TEID, Downlink TEID(Ctl), Downlink TEID(Data) | |
| | Create response (0x11) | | GSN, TEID, Upnlink TEID(Ctl), Uplink TEID(Data), End user Address | |
| | Update request (0x12) | | GSN, TEID, Downlink TEID(Data) | |
| | Update response (0x13) | | GSN, TEID, Upnlink TEID(Ctl), Uplink TEID(Data) | |
| | Delete request (0x14) | | TEID | |
| | Delete response (0x15) | | | |
| GTP-U (0xFF) | | | TEID | Version, Protocol, Source IP, Dest IP, Dest Port, TTL, Sequence Num, Control Bit |

The output is saved as a file, as shown in Fig. 11. Then, the output file is transmitted to the Traffic Flow Management & Port Scanner Detection via FTP every 1 minute.

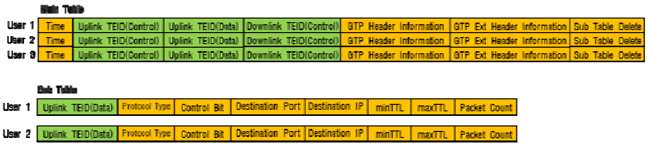


```
20120209_060000.000,1,0,0xFF,52,,,,,0x2bd28be4,,,,,,,4,0x06,42.39.74.210,110.76.141.40,80,64,1093640568,0x11,,,,,0X2Bd28Be4
20120209_060000.000,1,0,0x13,58,,,,192.168.156.114,0x1e25a10a,0x86da0a64,0x86da0ae4,,,,,,,,,,,,,,,,
20120209_060000.000,1,0,0xFF,60,,,,,0xd8ea2be4,,,,,,,4,0x06,42.45.146.186,211.115.81.154,1894,64,1013948241,0x02,,,,,0
20120209_060000.000,1,0,0x13,58,,,,192.168.156.114,0x2615c08a,0x63b72b24,0x63b72ba4,,,,,,,,,,,,,,,,
20120209_060000.000,1,0,0xFF,68,,,,,0xd38c6ae4,,,,,,,4,0x11,42.25.101.213,211.234.229.23,53,64,,,0xED840100,,,,48
20120209_060000.000,1,0,0x12,76,,,,172.25.11.4,0x22e4ea64,,,,0x34850085,,0x200,,,,,,,,,,,
20120209_060000.000,1,0,0x13,58,,,,192.168.156.114,0x23566eca,0x9ba82124,0x9ba821a4,,,,,,,,,,,,,,,,
20120209_060000.000,1,0,0xFF,76,,,,,0x38628ae4,,,,,,,4,0x11,42.29.26.67,110.45.226.199,9035,64,,,0x4D315200,,,,56
20120209_060000.000,1,0,0x12,76,,,,192.168.31.9,0x029aec24,,,,0x7ec3a0c5,,0x200,,,,,,,,,,,
20120209_060000.000,1,0,0xFF,62,,,,,0xb7800be4,,,,,,,4,0x11,42.36.48.225,217.50.153.161,7343,64,,,0x9F7C1D8E,,,,42
20120209_060000.000,1,0,0x12,65,,,,172.25.51.132,0x1cdeea64,,,,0x414020c5,,0x201,,,,,,,,,,,
20120209_060000.000,1,0,0xFF,60,,,,,0x432102a4,,,,,,,4,0x06,42.39.45.34,114.108.157.198,80,64,1705199914,0x02,,,,,0
20120209_060000.000,1,0,0xFF,37,,,,,0xd841cbe4,,,,,,,4,0x11,42.20.197.217,1.234.6.105,9999,64,,,0x33A758A5,,,,17
20120209_060000.000,1,0,0xFF,53,,,,,0x91e80be4,,,,,,,4,0x11,42.34.237.241,221.9.21.218,11367,64,,,0x15004300,,,,33
20120209_060000.000,1,0,0x13,58,,,,192.168.156.106,0x2b23118e,0x889acc24,0x889acca4,,,,,,,,,,,,,,,,
```

Fig. 11. The output of GTP Packet Capture and Parser.

## 4.2 Traffic Flow Management

In WCDMA mobile networks, all of the users are assigned unique control and data tunnel endpoint identifier (TEID), respectively, through which the control and data messages are sent and received. Information, such as MSISDN, IMSI, and IP, which could identify users, does not exist in all of the GTP messages. Therefore, it is necessary for user-specific traffic information extraction to manage the TEID contained in a GTP message. In this paper, control and data traffic flow are managed based on the TEID through the analysis of GTP messages.



Main Table

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| User 1 | Time | Uplink TEID(Control) | Uplink TEID(Data) | Downlink TEID(Control) | GTP Header Information | GTP Ext Header Information | Sub Table Delete |
| User 2 | Time | Uplink TEID(Control) | Uplink TEID(Data) | Downlink TEID(Control) | GTP Header Information | GTP Ext Header Information | Sub Table Delete |
| User 3 | Time | Uplink TEID(Control) | Uplink TEID(Data) | Downlink TEID(Control) | GTP Header Information | GTP Ext Header Information | Sub Table Delete |

Sub Table

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| User 1 | Uplink TEID(Data) | Protocol Type | Control Bit | Destination Port | Destination IP | minTTL | maxTTL | Packet Count |
| User 2 | Uplink TEID(Data) | Protocol Type | Control Bit | Destination Port | Destination IP | minTTL | maxTTL | Packet Count |

Fig. 12. The main table and sub table for traffic flow management.

The Traffic Follow Management module consists of a main table and a large sub-table as shown in Fig. 12. Each user's session information, such as TEID, MSISDN, IMSI, and IP address, is managed in the main table, and the transmission information of each user's data traffic, such as Protocol Type, Control Bit, Destination IP, and TTL, is managed in the sub-table. Here, the Uplink Data TEID is used to map the Main Table and Sub-Table.

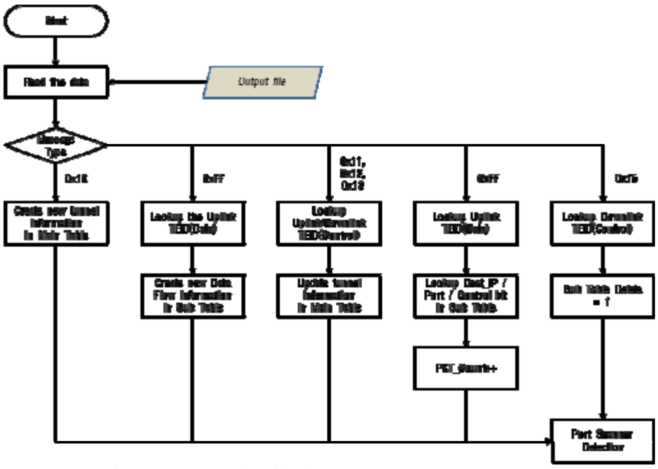

Start
Read the data
Output file
Message Type
0x10
0xFF
0x11, 0x12, 0x13
0xFF
0x15
Create new tunnel information In Main Table
Lookup the Uplink TEID(Data)
Lookup Uplink/Downlink TEID(Control)
Lookup Uplink TEID(Data)
Lookup Downlink TEID(Control)
Create new Data Flow information In Sub Table
Update tunnel information In Main Table
Lookup Dest_IP / Port / Control bit In Sub Table
Sub Table Delete = 1
PKT_Count++
Port Scanner Detection

Fig. 13. The flow diagram of traffic flow management.

Fig. 13 shows the flow diagram of traffic flow management. In the output file received from the GTP Packet Capture and Parser, a new row is added in the main table if the type of message is GTP-C Create Request (0x10) and then the new user's session information is written. If the type of message is GTP-C Creates Response (0x11) or Update Request/Response (0x12/0x13), the user's session information of row with same uplink Control TEID is updated. If the type of message is GTP-C Delete Response (0x15), the Sub Table Delete variable of row with same Download Control TEID is set to 1. Here, Sub Table Delete is a variable used to delete the sub table, and it is changed by Port Scanner Detection that is described in Section 4.3.

When the type of message is same as GTP-U (0xFF), sub table is created if there isn't the sub table of user with same uplink Data TEID or one of destination IP, destination port, Control bit is different, then the user's data transmission information is written. If all of destination IP, destination port, Control bit is same, Packet Count is increased. Once all the packets in the file are analyzed, the Scanner Detection Module is performed. The Port Scanner Detection Module is described below.

## 4.3 Port Scanner Detection

It is limited that a smartphone connects to a large remote source at the same time, and the frequency of repetitive connection attempts is low. Thus, we detected port scanning attacks in case the smartphone fails to connect to multiple IP addresses or to multiple ports of a specific IP address, and the user that caused the port scanning attack is identified based on the traffic flow. In this paper, the TRW algorithm that is described in Section 2 is adopted for detection of the port scanner. In the TRW algorithm, the response to a connection attempt is essential. However, there isn't a response to a connection attempt because the GTP Packet Capture and Parser don't capture inbound GTP-U packets in a Gn interface. So we can't know whether the connection attempt is successful or not. Therefore, we assumed that the connection attempt by a remote source to the distinct local host is successful if the Packet Count is 1, and that it is unsuccessful if the Packet Count isn't 1.

Fig. 14 shows the flow diagram for detecting these port scanning attacks. The marked parts in Fig. 14 are calculated repeatedly until all of the user's data transmission information is analyzed. And, the repeat count is different depending on the Packet Count. Here, $S_n$ is the probability of success for the connection attempt and $F_n$ is the probability of failure regarding the connection attempt. We assume $S_n$ and $F_n$ to be as follows:

$$S_n = 0.8, \quad F_n = 0.2 \tag{4}$$

The likelihood ratio is compared to an upper threshold of $\eta_1$, and a lower threshold of $\eta_0$. We assume $\eta_1$ and $\eta_0$ respectively as follows:

$$\eta_1 = 99, \quad \eta_0 = 0.01 \qquad (5)$$

If $\Lambda_n \geq \eta_1$, a remote source is a scanner. If $\Lambda_n \leq \eta_0$, a remote source is benign. Then, the Sub-Table Delete variable is set to 0 and $\Lambda_n$ is initialized to 1. And, if $\eta_0 < \Lambda_n < \eta_1$, the Sub-Table Delete variable is set to 1. It then waits for the next observation and updates $\Lambda_n$. If the Sub-Table Delete variable is 0, all of the user's data transmission information is deleted. If not, it is not deleted for the next observation

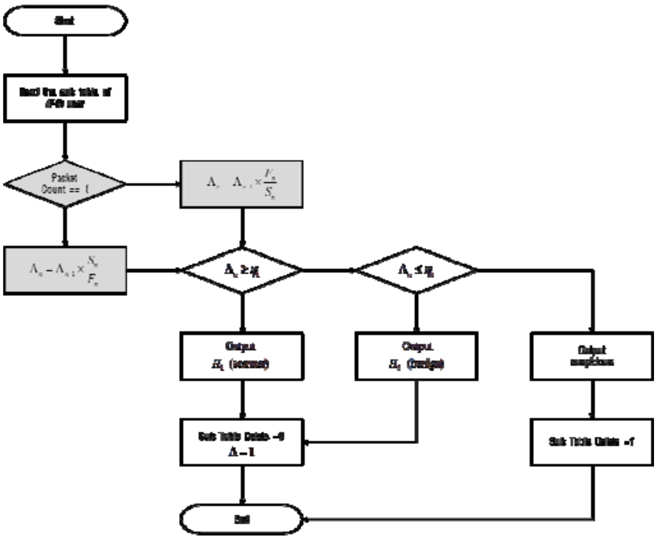


Fig. 14. The flow diagram for detecting port scanning attacks.

## 5 Test Results in 3G WCDMA Networks

The proposed system was tested in the 3G WCDMA mobile network that operates in Korea. Fig. 15 shows the test environment in the 3G WCDMA mobile network. The GTP Packet Capture and Parser system and the Traffic Flow Management and Port Scanner Detection system are installed in the Gn interface of the mobile communication service provider. The input of the GTP Packet Capture and Parser system is the traffic tapping the in/ outbound GTP traffic from one of the GGSNs And the input is the traffic of approximately 2.5 million subscribers with an average of 6.5Gbps. In addition, we performed scanning attacks to evaluate the performance of the proposed systems via tethering. The traffic was incoming to the core of the 3G WCDMA mobile network.

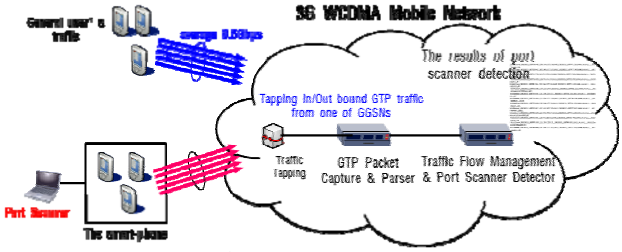

3G WCDMA Mobile Network
General user's traffic
average 6.5Gbps
The results of port scanner detection
Tapping In/Out bound GTP traffic from one of GGSNs
Traffic Tapping
GTP Packet Capture & Parser
Traffic Flow Management & Port Scanner Detector
Port Scanner
The smart-phone

Fig. 15. The test environment for port scanner detection in the 3G WCDMA network.

In test, the GTP Packet Capture and Parser system captured GTP traffic without loss and that the Traffic Flow Management Module managed the user's session/data flow information without delay. And, the Port Scanner Detection Module normally detected all of the scanning traffic caused by us, as shown in Table II.

In Table II, false positive means that the Port Scanner Detection Module misjudges normal user's data traffic with scanning traffic. An example of false positive is traffic to access a particular service like an apple push server (IP address : 17.149.36.71~337, Port : 5223). To prevent a false positive we registered the IP address and port used to access the particular service on a whitelist. The traffic with a registered IP address and port are excluded from port scanning detection.

TABLE II. THE TEST RESULTS OF PORT SCANNER DETECTION MODULE IN 3G WCDMA MOBILE NETWORK

| (A) The number of occurrence of port scanning attack | (B) The number of detection | (C) The number of False positive | (D) The number of additional detection | Results | |
|---|---|---|---|---|---|
| | | | | Detection rate | False positive rate |
| 2,000 | 2,027 | 8 | 19 | 100% | 0.39% |

Here, the detection rate and false positive rate are calculated as follow:

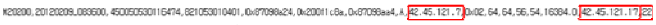
$$\textit{Detection Rate} = \frac{B-(C+D)}{A} \times 100 \qquad (6)$$

$$\textit{False Negative Rate} = \frac{C}{B} \times 100 \qquad (7)$$

We may need to pay attention to the additional detection as (D). It represents the number of scanning attacks among general user's traffic, and the results of the analysis is as follows:

### 5.1 Phone-to-Phone Port Scanning

Through the results of the detection log analysis we found that a smartphone scans over port 22 (ssh) of other smartphones in the same IP range. The IP address of the target smartphone is 42.45.121.7~42.45.121.17. And, the scanning attack is performed in smartphones from the fact that TTL is 64.

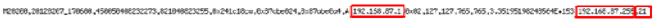


```
K20200,20120209,083600,450050530116474,821053010401,0x87098a24,0x20011c8a,0x87098aa4,4,42.45.121.7,0x02,64,64,56,54,16384,0,42.45.121.17,22
```

Fig. 16. The detection log of phone-to-phone port scanning.

### 5.2 Internal Network Port Scanning

Through the results of the detection log analysis we found that a smartphone scans over port 21 (ftp) of the internal network. And, the scanning attack is performed via tethering from the fact that TTL is 127.

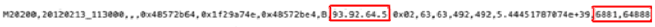


```
M20200,20120207_170600,4580504082232273,8210400823255,0x241c18ce,0x37cbe024,3x87abe6a4,#,192.158.87.1,0x02,127,127,765,765,3.3519519824056 4E+153,192.168.87.255,21
```

Fig. 17. The detection log of phone-to-phone port scanning.

### 5.3 External Network Port Scanning

Through the results of the detection log analysis we found that a smartphone scans over ports 6,881-64,888 of the system in the external network. The IP address of the target system is 93.92.64.5. And the scanning attack is performed via tethering from the fact that TTL is 127.



```
M20200,20120213_113000,,,0x48572b64,0x1f29a74e,0x48572be4,B,93.92.64.5,0x02,63,63,492,492,5.44451787074e+39,6881,64888
```

Fig. 18. The detection log of phone-to-phone port scanning.

## 6 Conclusion

Unlike a traditional wired infrastructure, mobile networks have limited radio resources and signaling procedures for complex radio resource management. So, unwanted traffic is not a problem in wired networks, but for mobile networks it can be a threat. Also, the previous mobile networks seemed to have been relatively safe from external threats because of their close characteristics. However, security threats, which were also in the wired networks, appeared after the switch to the 3G mobile network, and there are practical cases of data service disorders that have been caused by these security threats. Especially, the propagation of smartphones has rapidly expanded and various mobile services appear to be attracting more users, so that security threats to the mobile network have a greater ripple effect.

In this paper, we have proposed systems that are based on the TRW algorithm to effectively detect a port scanner. The proposed systems capture the GTP packet in the Gn interface and manage the user's session/data flow information. The proposed system detects scanning attacks over the multi-port of a target system and one port of IP range. Test results in the 3G WCDMA Network of mobile communication service providers show that the proposed system accurately detect port scanners. Moreover, the proposed system are running on the 3G WCDMA Network, which is operating Korea, without delay.

## 7 Future Work

We captured in/outbound GTP-C and outbound GTP-U packets in the Gn interface and detected port scanners. We will analyze various services that are using specific ports, and improve the accuracy of port scanner detection by capturing inbound GTP-U packets in the Gn interface. Additionally, we will apply the proposed system to the 4G LTE Network of mobile communication service providers that are operating Korea.

## Acknowledgment

This research was supported by the KCC(Korea Communications Commission), Korea, under the R&D program supervised by the KCA(Korea Communications Agency)" (KCA-2012-11912-06001).

## 8 References

[1] Mobile Traffic Data(2011~2016), CISCO VNI Mobile, 2012.

[2] Global Mobile Data Traffic. By Type, Morgan Stanley, 2010.

[3] F. Ricciato, "Unwanted Traffic in 3G Networks," *ACM SIGCOMM Computer Communication Review*, Vol.36, Issue 2, pp. 53-56, April, 2006.

[4] V. Falletta, F. Ricciato, "Detecting Scanners: Empirical Assessment on a 3G Network," *International Journal of Network Security*, Vol.9, No.2, 2009, 143-155.

[5] J. Jung, V. Paxson, A. W. Berger, and H. Balakrishnan, "Fast Portscan Detection Using Sequential Hypothesis Testing," *Proceedings of the IEEE symposium on Security and Privacy*, pp. 211-225, May, 2004.

[6] H. Holma, A. Toskala, WCDMA for UMTS – Radio Access for third Generation Mobile Communications 3rd, Willey, 2004.

[7] 3GPP, GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface (Release 10), TS 29.060 V10.2.0, 2011.

[8] V. Falletta, F Ricciato, P. Romirer-Maierthofer "Traffic Analysis at Short Time-Scales: An Empirical Case Study from a 3G Cellular Network," *IEEE Transactions on Networks and Service Management*, Vol.5, No.1, pp.11-21, March, 2009.

# Peer to Peer File Sharing

## Security Concerns, Unwanted Traffic Detection and Filtering

Leonardo Carvajal
Department of Computer Science
Sam Houston State University
Huntsville TX, USA

Lei Chen
Department of Computer Science
Sam Houston State University
Huntsville TX, USA

***Abstract***—Peer to Peer (P2P) file sharing systems have been introduced in business networks yet they are causing certain security problems in the network infrastructure. Securing P2P networks can be challenging as data traffic in such networks is difficult to manage and peers may have very different security settings and configurations. For example, P2P applications can use different port numbers as a configuration parameter, or simply use a random port number. Furthermore, not every P2P application supports encryption and decryption. This paper summarizes the existing solutions for detecting and stopping unwanted data traffic in P2P networks.

**Keyword:** Peer to Peer, P2P, sharing, security, threats, attacks

## I. Introduction

The use of P2P file sharing applications has greatly increased in recent years. Peer to Peer networks support hundreds of millions of users and generate the majority of Internet traffic [1]. However, P2P applications are also causing problems, including security threats, vulnerabilities, excessive network usage, and legal problems associated with copyrights. Peer to Peer file sharing applications establish multiple TCP connections using different ports between peers to transfer data making it difficult to control network saturation [11]. Moreover, malicious code can take advantage of the regular use of the P2P networks to propagate messages, introduce executable files into a system and trick users into downloading and executing infected files. Therefore, network administrators and Internet service providers are required to monitor unwanted network traffic and create policies on the usage of these applications in order to enforce data integrity, confidentiality and availability, as well as the illicit trade of copyrighted material. In the past few years, there has been a vast amount of research towards enforcing network security in peer to peer networks by combining existing mechanisms to detect unwanted network traffic and reinforce organizational policies.

There are different solutions for different companies. The type of organization can vary in size, e.g. small or large, and type, e.g. public or private. This paper expands on existing solutions to identify and stop unwanted peer to peer network traffic, as well as other tools to protect peer to peer applications against attacks. From the network administrator point of view, these solutions would protect the network infrastructure from infringing activities due to vulnerabilities in P2P applications [9].

This paper is structured as follows: section II presents the security issues in peer to peer file sharing networks; section III describes solutions proposed by other researchers about how to detect malicious peer to peer activity; section IV presents other tools to protect peer to peer networks against unwanted traffic; and the paper finishes with conclusions.

## II. Issues in Peer to Peer File Sharing Networks

Unlike client/server architecture, P2P network services are provided by many nodes simultaneously functioning as both clients and servers. In a P2P networks, nodes play an important role: they control the exchange of data, allow users to share resources, support communication, and provide directory services as well as real time collaboration tools [4]. Decentralized P2P networks spread services among all nodes. An effective attack to peer to peer networks may shut down the nodes offering specific file resources, and attacks to a single node may or may not have an effect on the entire network. There are several issues that are found in P2P file sharing systems.

### *A. Unpredictable Network Usage*

Peer to Peer applications normally take as much bandwith as available [12]. Files available in P2P networks are generally larger. Typical peers serve multi-megabytes of files overloading the network. For example, an audio file is usually from 3 to 5 megabytes, and a video file can be hundreds of megabytes.

### *B. Exposure of Sensitive Data or Personal Information*

P2P users have been observed unintentionally or intentionally sharing private files, including sensitive corporate information [1]. Some users of peer to peer networks do not know about basic computer security. Therefore, these users can share their entire hard drive, allowing attackers to obtain sensitive data, such as operating system files, applications files, and registries. In addition, P2P networks are well known for the distribution of malicious code. Many of the shared files are infected with malware and are spread to peers.

### C. DDOS Attacks

Many attackers are looking for controllable peer to peer networks users, or zombies [3]. These zombies send packages to selected targets in order to get the victim's resources. As a result, the victims will not be able to provide its services. In addition, downloading files can consume bandwidth and may decrease the availability of other network services or systems.

### D. Danger of Legal Action

In many cases, P2P file sharing networks are used to support illegal activities because many available files in P2P networks are copyrighted [1]. These illegal activities may not be limited to the end user and may be extended to the network sponsor. While many countries do not enforce penalty and punishment on copyright infringement offenses, other countries do. P2P networks provide sharing infrastructure that is harder to track and difficult to block, providing cover for espionage and criminal activity [2].

### E. Content Verification Susceptible to Attack

Attackers can introduce files without content, modify files, or share files with malicious code [6]. Therefore, integrity verification of the requested content should be verified. P2P file sharing applications use different ports. Moreover, opening these ports may give access to attackers to the computer network, or attackers can take advantage of the P2P applications vulnerabilities [5].

### F. Malware

Malicious code also exists in peer to peer networks [14]. Malware has the ability to spread across P2P infrastructures by replicating themselves. Malware is placed in shared folders and has names of popular movies, music, or applications in order to catch the attention of the users. Moreover, malicious code also uses other attack vectors including denial of service and has the ability to open backdoors making users' confidential files available to other peers [14].

## III. DECTECTION OF P2P ACTIVITIES

This section presents two mechanisms which are used to detect P2P activities.

### A. Intrusion Detection Systems

Based on the detection of encrypted traffic generated by one of the most popular P2P applications, GoalBit, authors in [8] propose a method to detect peer to peer traffic using intrusion detection systems, specifically using a set of rules. Due to the nature of the analysis, the proposed rules are signature based, focusing on identifying patterns. This method relies on the findings of repetitive string series on the data field of the IP packets, during the link phase or other critical connection points when encryption is not used. This approach is implemented in Snort (the most popular intrusion detection system) to detect signatures and block traffic matching from the protocol signatures. The following rules have been taken from [8] to demonstrate how this method works:

Rule 1000506/TCP traffic: alert tcp any any → any any (msg: "LocalRule: GoalBit tracker Cookie"; content:" | 47 67 61 6c 42 69 74 20 70 72 6f 74 6f 74 6f 63 6f 6c |"; dsize: 77; threshold: type both, track by_src, count 1, seconds 10, sid: 1000506; rev:1;)

In this rule, all TCP traffic coming into the network is scanned to find a GoalBit signature on IP packets. Once the signature is detected, the IDS track the source's IP address and if at least 1 event of the SID is fired, this rule alerts once every 10 seconds.

Rule 1000509/TCP traffic: alert tcp any any → any any (msg: "LocalRule: GoalBit Pattern | 00 0d 06 00 00 |"; flow: established; content:" | 00 0d 06 00 00 | " ; threshold: type both, track by_src, count 3, seconds 10, sid: 1000509; rev:2;)

Rule 1000565/TCP traffic: alert tcp any any → any any (msg: "LocalRule: GoalBit Pattern TCP payload size (1460 bytes)"; content:" | 62 72 6f 61 64 63 61 74 65 72 7b 70 69 65 | " ; depth:90; dsize: 1460; threshold: type both, track by_src, count 3, seconds 10, sid: 1000565; rev:1;)

Rule 1000566/TCP traffic: alert tcp any any → any any (msg: "LocalRule: GoalBit Pattern TCP payload size (1456 bytes)"; content:" | 67 6f 61 6c 62 69 74 5f 74 72 61 63 | " ; dsize: 1460; threshold: type both, track by_src, count 3, seconds 10, sid: 1000566; rev:1;)

The above rules were created during periods when encryption was not being used, such as at the start and in the middle of the transmissions [8]. Rule 1000506 was triggered when it receives the first bytes of the TCP session. This rule checks any encrypted or non encrypted communication to find any GoalBit signatures. Rules 1000565 and 1000566 were created for large payload sizes. Packets with this payload signature, for the most part, were not encrypted. It is very important to mention that the total accuracy of this detection method rate is 96% [8].

### B. Multi-Phased P2P Flow Model

Authors in [10] proposed a method that consists of three steps based on detecting malicious traffic. First, the flow grouping step involves clustering of TCP/UDP connections. In this step, authors track packets to determine if they are normal transmissions or flooding attacks. The segmented connection is the unit of the grouping. If there is a TCP session, ACK packets are discarded with a payload size of zero. Flows are processed to determine the similarity of each flow. Flows are considered different if their time gap is longer than 240 seconds and the threshold is greater than 0.5. Clustering of a flow occurs if a flow is link to at least one other flow, even though it is not link to all flows in a cluster. Second, the flow compression computes the state value of each flow of group and extracts the transition information. In order to define a state, the authors in [10] use seven features of the clustered flow. Each value of the features can only be 0 or 1. The features and values are as follows: protocol (TCP (0), UDP (1)), port (inside port (random port (0), reserved port (1)), outside port (random port (0), reserved port (1)), connection count (connections ≥ minimum count (0), connections < minimum count (1)), connection interaction (round trip (0), one way(1)), packet count comparison (inbound ≤ outbound (0), inbound > outbound (1)), and traffic volume comparison (inbound ≤ outbound (0), inbound > outbound (1)). The F values are the following: protocol (PT (64)), port (IP (32),OP(16), CC(8), traffic (CI (4), PC(2), TV(1)). In the last step, the algorithm constructs a matrix based on transitions of flow modeling. The detection engine uses the ratio computed from the probability-based models [10]. The detection rate of this model is 97% [10].

## IV. Approaches against worms and unwanted P2P traffic

This section presents several ways to detect passive and active worms, application management tools to monitor and control unwanted P2P traffic, and network security policies to prevent the use of P2P applications for illegal file sharing. In addition, P2P topologies are also considered to obtain accurate information from other peers.

### A. *Passive Worm Detector Based on Hash Values*

Based on hash values, authors in [13] proposed a method of detecting passive worm and malware. Files acquire an identifier for the content of each file that is hashed. Therefore, different versions of the same file have distinct hash values. Passive worms have the ability to propagate to other peers as files are copied to other hosts. Although worms are replicated with different file names, their code will be the same. This means that the hash value of the infected files will be the same [13]. Extracting the hash values and looking to see if those values represent multiple files indicates that malicious code might be present in a P2P infrastructure. However, having multiple files and same hash values detecting worms will not be easy. So, the way to detect worms is based on the popularity of the hash which increases in a short period of time [13]. This detection system has the following elements: data collector, which acquire the IP addresses of peers, shared folders, hash values, and obtain file names and other information; finding hash values that increase over time is a task of the popularity analyzer; worm detectors track hash values which increase over time to determine if malware is present on the P2P infrastructure.

### B. *Active Worm Attacks*

In [16] authors define a propagation P2P attack model based on three worm attack strategies: random based, attack, offline based attack, and online based attack. The random-based attack happens when the worm peer chooses IP addresses randomly of victim peers in order to launch the attack. In the offline P2P-based attack, infected peers obtain the IP addresses of offline peers. This information is maintained in a list called the hit-list [16]. The attack is launched based on the hit-list and the infected peers can continue launching the attack using the random-based attack. In the online P2P-based attack, after adhering to the P2P at the system's initial time, infected peers launch the attack to their neighbors. At the same time infected peers can infect other peers using the random-based attack [16]. In order to evaluate how the active worms attack affects and propagates on P2P systems, authors take into account the P2P characteristics or parameters and attacker parameters. The P2P system characteristics are P2P size, P2P vulnerability, P2P topology degree, and structured and unstructured P2P. The attacker parameters are the attack scan rate and the system's initial infected worm instances [16]. The following are the results obtained by the authors in [16] based on the topology degree in structured P2P systems, this P2P parameter only has impact on the online P2P –based attack. Based on the topology degree in unstructured P2P systems, the power-law distribution was used to determine the degree distribution to other P2P network hosts. Therefore, this method only determines the topology degree. Based on P2P size, this characteristic will have impact on both offline and online based attacks. Based o P2P vulnerabilities, this parameter will depend on how well protected the peers are in home environments as well as in organization environments.

Authors in [17] analyze the impact of how a new worm propagation threat is spread in BitTorrent due to its vulnerable topology to active worms. In contrast with [16], where authors do not take into account the cooperation of worm infected peers to share the attack information, authors in [17] consider the level of cooperation on the infected peers. Based on the same parameters or characteristics of the P2P systems and attacker parameters mentioned in [16], in [17] authors include the Internet parameters. These parameters are the connection speed, patch rate when an infected machine becomes impenetrable, and death rate when an infection is detected on a peer and removed without patching [17]. These are the results obtained in [17]: based on the impact of the attack strategy, the BitTorrent Worm (BTW) attack can reach its speed of propagation up to 300% compared with traditional scanning method. Based on the impact of P2P system size, the results show that BTW performance can differ. If the network size is large, the attack performance is higher. Based on the impact of P2P topology degree, the results shows that if the topology degree increases, peers are open to the BTW and the speed of propagation also increases.

### C. *P2P Aplication Management Tools*

In [12] Lai mentions that almost 2.5 billon downloads occur every month using P2P applications. Organizations are making request to ISPs and network administrators to eliminate potential threats and illegal P2P file sharing. However, most companies have insufficient budgets to employ enough staff members for their network operation and even less resources to manage P2P usage [12].

An unsuccessful method that many companies use to block peer to peer traffic is blocking P2P traffic ports using hardware or software firewalls. P2P applications can use different ports to overcome port blocking [12].

Monitoring tools such as Network Instruments Observer can identify the top users of the network, break down web traffic and generate Internet traffic activity reports per user or by department. These tools give a real time picture of actual protocols running across the network, and help network administrators collect information and troubleshoot network issues.

Bandwidth management tools allow network administrators to detect and stop P2P traffic [12]. NetEnforcer can limit the use of bandwidth, prioritize network traffic per application and per user, control the bandwidth utilization and costs associated, while protecting and enhancing service quality for all network users. Using NetEnforcer or similar tools, companies can prioritize business-critical applications. This tool also includes application layer protocol monitoring and application signature detection to control P2P applications [12].

Another bandwidth management tool, Packet Shaper allows administrators to set policies that provide a limit on the bandwidth usage on application type identifying peer to peer application traffic. It permits bandwidth management according to the priority of the application. This tool can prevent denial of service attack. It detects and stops SYN floods and ICMP packets [12].

A new P2P detection tool called Watchdog can detect encrypted peer to peer traffic. SSL encrypted peer to peer file transfer sessions on any port as well as sessions that are hidden behind HTTP proxies can be detected and tracked by this tool [12]. This tool is capable of blocking file transfers.

Audible Magic's CopySense appliance handles illegal peer to peer file sharing of copyrighted works [12]. This tool filters illicit traffic of copyrighted content, allowing network administrators to manage and control network traffic. Audible Magic's CopySense utilizes a database of file signatures for copyrighted media. It can identify over 3.5 million recorded songs [12]. Infractions are tracked and addressed in real time, reducing the use of the network for unwanted traffic. The Integrated Computer Application for Recognizing User Services (ICARUS) tool can block the infringing and non-infringing P2P traffic. In contrast, Audible Magic's CopySense Network Appliance only blocks the infringing use of P2P file sharing applications.

### *D. P2P Network Security Policies*

Network security standards, policies and procedures must be followed and enforced to prevent the use of P2P applications for illegal file sharing. Policies should focus on the prevalent use of this technology that is only for distribution of copyrighted content. Other concerns of peer to peer file sharing applications include network utilization, network security, malicious code and inappropriate content. Policies will support the primary usage of the network for operations of organizations' daily business. In most cases, violations of security policies can result in firing employess and criminal prosecution under state and federal statutes.

### *E. P2P Topology Path Length and Hierachy*

Authors in [15] present aspects that will affect the DoS resilience of P2P systems based on: hierarchy and k - regular topology. These models are based on the probability of obtaining accurate information. In a P2P hierarchy, supernodes are target for DoS attacks because they store the directory of the files that are shared, as well as the information about the connectivity with other supernodes in order to replay clients' file petitions. The following are the results obtained in [15] with the hierarchy model: having a P2P infrastructure with corrupted nodes, if client's petition needs just one supernode hops to reach the solicited file, the probability of obtaining the correct replies is 81%. However, if a client's petition needs 5 supernode hops to reach the solicited file, the probability of obtaining correct replies is 53.1%. Therefore, while paths are longer, the possibility of obtaining correct replies is reduced. In K-regular Topologies, topologies are adjacent nodes that have the same number of neighbors (k) [15]. However, the number of neighbors may not be the same if the peers consider an anonymous connection. Therefore if a peer's file petition is requested, it may require a higher number of hops. This means that the probability of obtaining correct replies is lower. Therefore, attacks to peer to peer file sharing systems are higher [15].

## V. Conclusion

The popularity of P2P file sharing applications has increased security risks for organizations. Most organizations are concerned about how these kinds of applications saturate the network infrastructure with music, videos and other organizations' resources not related to the goals of the organization. Another problem is that files downloaded to organizations' computers might be illegal copies of copyrighted material. Information Technology departments use a variety of mechanisms to prevent the unauthorized use of P2P applications within the organization. This paper has presented the most relevant approaches and tools to detect and prevent unwanted P2P activities, including strong policies and other mechanisms such as scanning and blocking network traffic of suspicious activities.

## References

[1] Danny Hughes, Kevin Lee, and James Walkerdine , "On the Penetration of Business Networks by P2P File Sharing," in *Second International Conference on Internet Monitoring and Protection*, 2007.

[2] Rosslin John Robles, Min-Kyu Choi, and Eun-suk Cho, "A Paradigm Solution to P2P Security Issues," in *Int. e-Commerce Advance Science and Technology*, 2009, pp. 3-7.

[3] Jiri Schafer and Kamil Malinka,"Security in Peer to Peer Networks Empiric Model of File Diffusion in BitTorrent," in *4th Int. Conf. Internet Monitoring and Protecting*, 2009, pp. 39-44.

[4] Huu Tran, Michael Hitchens, Vijay Varadharajan, and Paul Watters, "A trusted Access Control Framework for P2P File-Sharing Systems," in *Proceedings of the 38th Hawaii International Conference on Systems Science*, 2005, pp. 1-10.

[5] Amuthan A, Marimuthu.G, and Kaliaperumal.G, "Secure Trust Management Model for Peer to Peer File Sharing System," in *Int. J. Recent Trends Engineering and Technology*, 2010, pp. 90-96.

[6] M. Eric Johnson, Dan McGuire, and Nicholas D. Willey,"The Evolution of the Peer to Peer file Sharing Industry and the Security Risks for Users," in *Proceedings of the 41st Hawaii International Conference on System Science*, 2008, pp. 1-10.

[7] Alexis Ulliac and Bogdan V. Ghita, " Non Intrusive Identification of Peer to Peer Traffic," in *2010 Third International Conference on Communication Theory, Reliability, and Quality of Service*, pp. 116-121.

[8] Andre F. Esteves, Pedro R. M. Inacio, Manuela Pereira, and Mario M. Freire, "On-Line Detection of Encrypted Traffic Generated by Mesh-Based Peer-to-Peer Live Streaming Applications: The Case of GoalBit," in *2011 IEEE International Symposium on Network Computing and Applications*, pp. 223-228.

[9] Kevin W. Hamlen and Bhavani Thuraisingham, "Secure Peer-to-peer Networks for Trusted Collaboration," in *2007 Collaborative Computing; Networking, Applications and Worksharing*.

[10] Sang-Kyun Noh, Joo-Hyung Oh, Jae-Seo Lee, Bong-Nam Noh, and Hyun-Cheol Jeong, "Detecting P2P Botnets using a Multi-Phased Flow Model," in *2009 Third International Conference on Digital Society*, pp. 247-253.

[11] Wei Li, Shanzhi Chen, Yaning Liu, and Xin Li, "Aggregate Congestion Control for Peer-to-Peer File Sharing Applications," in *Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, 2008, pp. 700-705.

[12] Wayne Lai. (2004, January 5). *Managing Peer-to-Peer Applications in Dormitory* [Online]. Available: http://www.sans.org/reading _ room/whitepapers/tools/managing-peer-to-peer-applications-dormitory-networks_1348

[13] Sahar Fahimian, Amirvala Movahed, and Medhi Khrrazi, " Passive Worm and Malware Detection in Peer to Peer Netwroks," in *2010 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing*, pp. 561-565.

[14] M. Eric Johnson, Dan McGuire, and Nicholas D. Willey, "The Evolution of the Peer to Peer File Sharing Industry and the Security Risks for Users" in 2008 *Proceedings of the 41st Hawaii International Conference on System Sciences,* pp.1-10.

[15] D. Dumitriu, E. Knightly, A. Kuzmanovic, I. Stoica, and W. Zwaenepoel, "Denial of Service Resilience in Peer to Peer File Sharing Systems," in *Proceedings of the International Conference on Measurement and Modeling of Computer Systems,* 2005 pp. 38-49.

[16] Wei Yu, Corey Boyer, Sriram Chellappan, and Dong Xuan, "Peer to Peer System-Based Active Worm Attacks: Modeling and Analysis," in *Proceedings of IEEE International Conference on Communication ,* 2005, pp. 1-7.

[17] Sinan Hatahet, Abdelmadjid Bouabdallah, and Yacine Challal, "A New Worm Propagation Threat in BitTorrent: Modeling and Analysis," in *Proccedings of the International Multiconference on Computer Science and Information Technology,* 2008, pp.791-798.

# SESSION

# CYBERSECURITY + SECURITY EDUCATION

## Chair(s)

**Prof. George Markowsky**
**Dr. Linda Markowsky**

# Implementation Progress, Student Perceptions, and Refinement of a Virtual Information Security Laboratory

**C. Cavanagh**[1] **and R. Albert**[2]
[1]University of Maine at Fort Kent, Fort Kent, ME, USA
[2]Professional Management Division, University of Maine at Fort Kent, Fort Kent, ME, USA

**Abstract -** *Effective information security laboratory exercises are increasingly being viewed as an essential element of information security educational offerings. Such offerings are, in turn, considered by most to be the primary means of raising student awareness and interest in information security and ultimately preparing a much needed, knowledgeable, and skilled information security workforce. Virtual laboratory exercises exist in a variety of formats, each better suited to a particular educational purpose, but all compatible with a distance education delivery modality. The design of a remotely accessible virtual laboratory and the exercises themselves must also provision for a secure environment that encourages experimentation while minimizing the risk of introducing a security incident. The purpose of this paper is to report on progress made in the implementation of a virtual information security laboratory, student experience with and perceptions of the laboratory, and refinement of the laboratories design based on reported advancements in the design and implementation of virtual information security labs and corresponding curricular elements in higher education settings.*

**Keywords:** Cybersecurity, information security, online education, virtual lab

## 1 Introduction

Information security remains a critical topic in society, government, and education today. Educational institutions have been called upon to help raise information security awareness through, among several approaches, expanding university curricula [6]. Professional organizations, in response to this and other related needs expressed by various constituencies, have responded in part, by being more explicit in their curricular recommendations to educational organizations. The Association of Computing Machinery (ACM) for example, has evolved its curriculum recommendations for Computer Science to include "information security" as a core topic in several of the fundamental knowledge areas [1].

Effective information security laboratory exercises are increasingly being viewed as an essential element of information security educational offerings [5, 7, 8]. Practical skills oriented laboratory exercises are an attractive means of raising student awareness and interest in information security and ultimately preparing a much needed, knowledgeable, and skilled information security workforce.

The need to expand appropriate information security learning activities as essential elements of information security instruction, especially hands-on laboratory exercises, should become even more pronounced as university curricula continue to evolve. Similarly, the need to extend access to such laboratory exercises to distance situated students should increase in response to increased utilization of distance education delivery modalities.

The context in which the Maine Information Security Lab (MEISLab) operates requires provisioning for secure access to virtual lab resources by students enrolled in our distance education programs (e.g., online Associate of Science in Information Security). This context also requires making available a variety of exercise formats in addition to virtual labs (e.g., Web labs) that are better suited to a particular educational purpose.

Above all, the laboratory environment and exercises must encourage students to engage in active experimentation that enables attainment of learning objectives while simultaneously minimizing the risk of a student introduced security incident. Hence, the ability to isolate the virtual computer lab was one of the key goals targeted in the design of the MEISLab [3].

## 2 MEISLab Goals

The goals for MEISLab were established following a review of the goals, outcomes, benefits, and recommendations stemming from similar efforts that have been reported [3]. The predominant aim of establishing the MEISLab was to ensure realization of the instructional advantages identified as being associated with delivering a rich learning experience made possible through virtualization.

The specific goals that have driven the design of the MEISLab reflect the following desirable characteristics:

- Accessibility (remote access)
- Observability of host and network events
- Ability to simulate realistic scenarios and various client/server configurations

- Separability of virtual networks including the ability to isolate the virtual lab systems from the campus network
- Remote configurability
- Ability to share resources efficiently
- Provisioning of an appropriate platform in support of different areas of information security instruction
- Support for rapid prototyping of computer and network configurations
- Provisioning of a uniform experience across students
- Simple and cost effective course administration

All of these characteristics support constructivist approaches to instruction [3]. Using virtualization to support lab activities that reinforce problem-solving skills in authentic environments should be considered an essential component of a well-designed virtual information security lab.

To these ends, the lab was designed and implemented to:

- Harness the instructional benefits of virtualization;
- Provide remote access in support of online education modalities including the option for synchronous instruction; and
- Support learning activities that:
    - Ensure students have sufficient background knowledge to maximize comprehension
    - Promote students' appreciation of the ethical dimensions associated with engaging in information security activities
    - Promote students' compliance with all information security and technology use policies
    - Ensure students meet the student learning outcomes and related curricular requirements defined for the information security degree programs.

## 3 Implementation Progress

The virtualization solution that was selected is a "bare metal" VMware vSphere Hypervisor (ESXi) and VMware vCenter Lab Manager 4.0. The Lab Manager product provides access control over local and distance situated students. Local and distance situated students involved in this study participated in an information security course designed to support a hybrid delivery approach and offered during the spring semester of 2012. Online synchronous meeting software/services (e.g., GotoMeeting) was utilized in support of distance situated students on as needed basis. Some have argued the use of synchronous meeting software/services as an essential element of online instruction will become the standard for conducting information assurance instruction going into the future [7].

As noted earlier, one of the key objectives for the MEISLab design was the ability to separate virtual machine traffic from the campus network. This separation is intended to enable students to freely modify the network and virtual machine configurations without risk of introducing potentially malicious traffic or actions on the campus network. In order to achieve this objective we used a creative technical workaround in order to accommodate the VMware Lab Manager installation requirements and ensure proper isolation of network traffic.

When performing the initial installation of Lab Manager the setup wizard requires specification of a physical network that will be designated the default network for Lab Manager [9, 10]. In many common networking configurations, Lab Manager will use this default network to obtain external IP addresses for the virtual machines in order to allow users to remotely access the virtual machines even if the virtual machines are configured to use private IP addresses [9]. As we did not want to make the campus network the default physical network for Lab Manager we utilized a residential-class router to simulate a physical network. This router was configured as a DHCP server that would assign addresses in a private address range. This router was then plugged directly into the physical ESXi server in order to simulate a physical network in Lab Manager. With this router in place, we were successfully able to complete the Lab Manager installation process.

With this configuration implemented, we were able to isolate traffic between the campus and virtual machines in one of two methods. The first method involved attaching the virtual machine's network adapters to a virtual network that had no connection to the Internet [10]. The second method involves connecting the virtual machine's network adapters to a physical network but enabling the block in/out network fencing setting [10]. By specifying the block in/out option the virtual machines are essentially partitioned off from connecting to the physical network and Lab Manager ensures that traffic does not escape [10].

Efforts to build a library of "stock" course-oriented virtual machine images in support of specific instructional lab activities are continuing. Additional effort has been made to ensure the proper training necessary for the synchronous instruction support system and to ensure the availability to students of "Web labs" for the purpose of ensuring students will be better able to fully comprehend and benefit from lab activities by having the prerequisite conceptual knowledge.

## 4 Student Experience and Perceptions

In order to establish how effective the virtual lab implementation was in meeting the goals we had outlined, student feedback was solicited and analyzed. As eight out of 16 of our stated goals dealt with administrative and infrastructure concerns (e.g., centralized logging within the lab and isolation of virtual machine traffic), the students were asked to provide responses to questions focusing on the remaining eight goals.

A web survey consisting of 17 questions was created to gauge how well our implementation met our goals concerning virtual lab accessibility, the ability to simulate realistic scenarios and devices, the separability of virtual networks, remote configurability of virtual machines, the ability to share lab resources efficiently, rapid prototyping of computer/network configurations, problem solving in authentic environments, and quick access to course-oriented virtual machine images. The first section of the survey focused on the accessibility of the virtual lab to students and so we were interested in determining the types of Internet connections used to access the lab. The survey found almost half of the students were accessing the lab from on-campus and the other half from off-campus. The majority of on-campus students were utilizing a wired Internet connection with all off-campus students utilizing a high-speed Internet connection. The most popular type of connection by off-campus students was digital subscriber line (DSL) followed by cable and then satellite. Students found the performance of the lab to be suitable with the majority rating the performance as “acceptable”.

The next section of the survey was dedicated to determining if the students believed that the labs they had completed were models of real-world scenarios. Configuration of virtual private network (VPN) services and a password cracking activity are two examples of lab exercises the students completed. An overwhelming majority of students agreed that the labs did indeed represent real-world scenarios and the majority also agreeing that the configurations of the virtual machines represented real client/server configurations. Furthermore, the majority of students felt that the particular labs they had completed were realistic.

Separation of the virtual networks from one another was an important concern of ours and the next section in the survey was written with that topic in mind. Students were asked if they experienced any odd network behavior while completing their labs and an even spread of responses was recorded. However, the majority of students stated that they did not encounter any odd network behavior while completing their labs. The survey then asked those who did experience odd network behavior if the behavior prevented them from successfully completing the labs. An equal amount of students stated that they were prevented from completing their labs, were not prevented from completing the labs, and were unsure if the behavior prevented them from completing the labs.

Remote configurability of the virtual machines and virtual networks was a very important goal to us, as we want to offer students the freedom to explore new machine configurations and network designs to facilitate their own personal research interests and curiosities. Students were asked if they felt they could configure the virtual machines as they desired and the majority of students answered “yes” for this question. The survey then asked students if they encountered issues with the virtual machines or their configuration and the majority of students responded that they did not encounter issues with the virtual machines. Those users who did encounter issues with the virtual machines stated that their issues were stemming from attempting to use the VMware web browser plug-in. The final question in this section asked students if they felt that they could reconfigure their virtual machines, as they desired with the majority stating “yes” that they could.

We were interested in determining if there were specific times of the day when heavy lab usage could affect performance such as in the late afternoon or on the weekends. Students were asked if they experienced performance issues when accessing the lab during specific parts of the day and the majority of students responded that they did not encounter poor performance during specific times.

The last section of our survey concentrated on the students’ ability to quickly access course-oriented virtual machine images. We asked students if they felt they were able to easily access virtual machine configurations (the virtual labs) stored in the MEISLab library and the majority responded that they could easily access the configurations. Those who were not able to easily access the configurations stated that their issues were with the VMware web browser plug-in and being able to locate the virtual labs assigned to them within the MEISLab.

These student perspectives are similar to those reported by others [2]. Similarly, the authors’ concerns over the practical use of such facilities are similar to those expressed by others [7], especially the need for training of instructional and IT support staff to be more comfortable with the additional levels of abstraction involved with such technology. Further refinement of the design and implementation of the MEISLab is necessary to support even greater levels of effectiveness and student achievement.

## 5 Future Design Refinement

Recent studies have suggested that the choice of laboratory configuration should be based on the complexity of the concepts being taught and the student’s background

[4]. Web labs that consist of a Java-based applet requiring student access to only a web browser and typically focused on a single concept are one configuration option that was initially explored [3]. Web labs will clearly continue to play a significant role in provisioning laboratory exercises accessible to local and distance situated students.

Part of the future refinement of the laboratory exercises will therefore include categorizing the lab exercises based on complexity of the underlying concepts and student background, and then refining the MEISLab configuration to support those exercises for which student knowledge requirements are relatively high (i.e., students who more readily conceptualize a multiple computer environment).

Open source software virtualization platforms (e.g., Oracle VirtualBox) are also important to include when considering implementation options. Such options are particularly appealing for their low/zero licensing cost and for providing opportunities for end-user modification to meet local customization demands and/or environmental conditions. These will be explored further in the continuing evolution of the MEISLab.

# 6 Conclusion

As the number of information security educational offerings continue to grow as significant components of evolving program curricula targeted to address the growing demand for increased information security awareness and preparation of a knowledgeable and skilled information security workforce, so too the need for practical security laboratory exercises accessible in a variety of formats and delivery modalities.

The survey of student perceptions revealed support for MEISLab meeting its goals. In particular, a majority of student respondents rated MEISLab performance as “acceptable”, lab activities as being “realistic”, virtual machines as being “easily accessible” and “reconfigurable”, with no unexplained erroneous behavior.

The design and implementation of the MEISLab will continue to evolve based on students perceptions, complexity of the laboratory exercises relative to achievement of learning objectives, and availability of more affordable and customizable open source virtualization platforms.

# 7 References

[1] Association of Computing Machinery (ACM) (2008). “CS2008 Curriculum Recommendations for Computer Science”. Retrieved May 11, 2012 from http://www.acm.org//education/curricula/ComputerScience2008.pdf

[2] Burd, S. D., Gaillard, G., Rooney, E., & Seazzu, A. F. (2011). “Virtual Computing Laboratories Using VMware Lab Manager”, *Proceedings of the 44th Hawaii International Conference on System Sciences*.

[3] Cavanagh, C. & Albert, R. (2011). “Goals, Models, and Progress towards Establishing a Virtual Information Security Laboratory in Maine”. *Proceedings of the SAM ‘11 Conference*, pp. 496-500. Retrieved May 11, 2012 from http://cerc.wvu.edu/download/WORLDCOMP%2711/2011%20CD%20papers/SAM5057.pdf

[4] Fulton, S. & Schwietzer, D. (2011). “A Concept Focused Security Lab Environment”. *Proceedings of the 15th Colloquium for Information Systems Security Education*, pp. 126-131. Retrieved May 11, 2012 from http://cisse.info/archives/15th-colloquium/papers

[5] Irvine, C. E. (1999). “Amplifying security education in the laboratory”, *First World Conference in Information Security Education, pp. 139-199*. Retrieved May 11, 2012 from http://www.usafa.edu/df/dfe/dfer/centers/accr/docs/schweitzer2009c.pdf

[6] National Security Council (2009). “60-day Cyberspace Policy Review: Assuring a Trusted and Resilient Information and Communications Infrastructure”. Retrieved May 11, 2012 from http://www.whitehouse.gov/assets/documents/Cyberspace_Policy_Review_final.pdf

[7] Nestler, V. & Bose, D. (2011). “Leveraging Advances in Remote Virtualization to Improve Online Instruction of Information Assurance”, *Proceedings of the 44th Hawaii International Conference on System Sciences*.

[8] Schweitzer, D., Gibson, D. & Collins, M. (2009). “Active Learning in the Security Classroom”, *Proceedings of the 42nd Hawaii International Conference on System Sciences, pp. 1-8*. Retrieved May 11, 2012 from http://www.usafa.edu/df/dfe/dfer/centers/accr/docs/schweitzer2009c.pdf

[9] VMware (2011). *Lab Manager Installation and Upgrade Guide*. Retrieved May 11, 2012 from http://www.vmware.com/pdf/labmanager25_Installation_Guide.pdf

[10] VMware (2011). *Lab Manager User's Guide*. Retrieved May 11, 2012 from http://www.vmware.com/pdf/lm40_users_guide.pdf

# Effectiveness of Cybersecurity Competitions

Ronald S. Cheung, Joseph Paul Cohen, Henry Z. Lo, Fabio Elia, Veronica Carrillo-Marquez
Department of Computer Science
University of Massachusetts Boston
Boston, Massachusetts 02125-3393
Email: {cheungr,joecohen,henryzlo,fabioel,veiko}@cs.umb.edu

*Abstract*—**There has been a heightened interest among U.S. government agencies to fund cybersecurity workforce development. These efforts include offering universities funding for student scholarships, funding for building capacity in cybersecurity education, as well as sponsoring cybersecurity competitions, games, and outreach programs. This paper examines the effectiveness of cybersecurity competitions in educating students. Our study shows that though competitions do pique students' interest, the effectiveness of this approach in producing more high quality professionals can be limited. One reason is that the knowledge barrier to compete in these competitions is high. To be successful, students have to be proficient in operating systems, application services, software engineering, system administration and networking. Many Computer Science and Information Technology students do not feel qualified, and consequently this reduces participation from a wider student audience. Our approach takes aims at lowering this barrier to entry. We employ a hands-on learning methodology where students attend lectures on background knowledge on weekdays and practice what they learn in weekend workshops. A virtual networking environment is provided for students to practice network defense in the workshops and on their own time.**

## I. Introduction

It has been known for some time that there is a severe shortage of computer security specialists in the U.S., yet universities are slow to react to this need of educating more cybersecurity professionals. In fact, most universities currently do not offer degrees or concentrations in Information Assurance (IA) or Information Security (IS). A survey of 260 universities in the Northeast (Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, Massachusetts, and New York) shows that less than 8% of the schools offer concentrations or degrees in IA/IS. Over 60% of the schools surveyed do not even offer a single course on network or information security. So, it is quite common for CS/IT majors to graduate from universities without knowing anything about security. This problem has not improved in the past few years when funding for higher education was flat lined or decreased due to the economic downturn.

In light of this shortcoming, President Obama has requested $57 million R&D fund in the FY2013 federal budget for a coordinated cybersecurity research initiative [1]. Together with other efforts, this will fund the NSF Federal Cyber Service: Scholarship for Service (SFS) program that awards scholarships to qualifying students entering the IA and cybersecurity field, and provides funding to higher education enterprises to build up the capacity to educate cybersecurity professionals [2]. The Scholarships for Service track will grant scholarships to students attending schools that have an established IA/IS program. According to our survey, less than 8% of universities in the Northeast can apply for this. The capacity building track is highly competitive and it will take winning schools several years to develop all the necessary IA/IS courses. It probably will take additional years to establish the program in order to graduate students in this area. These programs may help alleviate the cybersecurity workforce shortage in the future, but the impact may not be felt for some time to come.

Several DOD government agencies and public companies are sponsoring cyber defense competitions with the hope of training more cybersecurity professionals in the near term. Schools participate in these competitions or games in order to promote student interest, even though they have no formal cybersecurity programs or courses. At the University of Massachusetts Boston (UMB), the CS Department has no degree program or concentration in IA/IS, but it offers two security courses (IT 428: Information Security and IT 443: Network Security Administration) in the undergraduate IT curriculum. The CS Department formed a cyber defense team and students in the team competed in the 2011 and 2012 Northeast Collegiate Cyber Defense Competitions (NECCDC) [3] and the 2011 MIT Lincoln Lab Capture the Flag (CTF) contest [4]. This research attempts to study the effectiveness of these competitions in increasing the ability of universities to produce more IA and cybersecurity professionals.

## II. Statement of the Problem

Past research has shown that these competitions are very effective in elevating student interest in cybersecurity [5], [6]. One reason is that they provide simulated real-world cyber attacks for students to practice network defense. Students find that interesting because they get a lot of hands-on experience that they cannot get in a classroom. During the competitions, students learn how to work as a team. They are forced to work in an intense atmosphere where they have to band together to solve a common problem, viz., defending their network against outside attackers. Also, students are excited to network with security professionals from industry and learn from them.

Though these cybersecurity competitions do pique student interest, the effectiveness of this approach in producing high quality cybersecurity professionals is limited. One reason is that the knowledge barrier needed to compete in these competitions is extremely high. To be successful at these contests, students have to be proficient in numerous topics, namely, operating systems, application services, software engineering, system administration and networking. A majority of CS majors feel that they do not possess the right skill set. Most universities offer a traditional curriculum in CS which teaches theory of operating systems, compilers and databases. However, the critical skills needed for cybersecurity are hands-on knowledge on script programming, system administration and network configuration. Learning these skills from scratch and be able to use them proficiently during competition take a lot time and effort. Consequently, this discourages participation from a wider student audience.

Another shortcoming for the competition approach is that most universities consider cybersecurity competitions as extracurricular activities. Students spend time on their own preparing for the competitions in addition to carrying their regular course load. When the demand for school work increases, students tend to reduce their involvement in cybersecurity training. For the NECCDC competition, it is not uncommon to see a 50% drop out rate between the Fall semester when students are recruited and the Spring semester when students have to spend a lot of time on preparations. Also, these activities depend mostly on self studies and peer instruction efforts. Those who are not sufficiently motivated to learn new concepts or technologies on their own tend not to show up as often. Since these activities are mostly student organized, there is no penalty for not showing up. This further reduces the number of students from learning cybersecurity.

Furthermore, for schools that do not have a formal program in IA/IS, it is difficult to sustain the interest generated by participating in one cybersecurity competition. Students from these schools work extra hard preparing for the competition and they learn a lot at the event. However, they may not retain any of the knowledge if they do not apply or refresh it after the competition. There are fewer opportunities if the school does not offer any courses or a formal curriculum in cybersecurity. As a result, students will lose interest in this area.

## III. Hands-On methodology

To lower this knowledge barrier, schools participating in cybersecurity competitions are trying to spend extra effort. This includes teaching students on installation of operating systems and applications, configuration of services, setting up a network and its firewall. They also provide a networking environment for students to practice what they learn. The UMB Cyber Defense Team employed a hands-on learning methodology to prepare students for the competitions. Instead of using a standalone network consisting of hardware switches, routers and servers, the team constructed a virtual network based on Virtual Machines (VM) for students to practice network configuration and defense.

### *A. Lectures and Presentations*

Lectures were offered twice a week and they were presented by students that had participated in previous cybersecurity competitions. These students selected the topics based on what they had learned in previous years and they paced the presentations for the new students. Focus was placed on making the lectures interactive and hands-on. Students were encouraged to follow the lessons and do exercises on their laptops. At the end of each lecture, feedback was solicited from these students and this provided guidance for the content of the next lecture.

As an example, the first few lectures offered were on learning and/or brushing up on their basic Linux skills. They were consisted of topics on basic Linux administration, networking and fundamental concepts of cybersecurity. The feedback we got from the students was that the talks could be more advanced. After discussing these basic topics, focus was shifted towards specific tools. In particular, the talk was on what tools were available and how one would use them. At the end of the week, the group held a workshop where they practiced the tools that they had learned during the week.

Students had been spending a lot time each week on the workshops and lectures. As the weeks progressed, the students became overwhelmed with regular classwork. Student participation began to drop off. This was the shortcoming of participating in cybersecurity competition as an extracurricular activity. There were no real incentives for students to attend lectures and workshops aside from students' pure interest in learning the subject. If the preparation for the competition was treated as a part of a regular course, the result would be different. In our case, as the numbers of student dwindled, we refocused the remaining group on competition-specific activities. For example, we invited our IT staff to give detailed lectures on mail server configuration and the DNS.

### *B. Workshops*

A workshop was held at the end of each week. Sometimes students engaged in mock competitions. Students put to use what they had learned from the lectures, presentations and their own practice. Initially the less experienced students were presented with VMs filled with back-doors, some as obvious as open ports that connected directly to root shells, all the way down to cleverly concealed root kits that reopened ports and allowed the attackers (in this case, the more experienced students) to maintain persistence on these machines. Step by step, students taught one another how to detect each of these different vulnerabilities and how to rid the machines of any trace of an attack. Students were able to participate both as attackers and defenders, allowing them to see both sides. The workshop was a hands-on review of the topics that had been covered during the week, and it advanced in difficulty as the weeks progressed. Most participants found these workshops more useful than the lectures themselves. They made students aware of their shortcomings and forced them to apply what they had learned.

The VMs presented to the students also came with the ability to log all activities performed. These logs were on an individual student basis and they were analyzed for the purpose of improving teaching methods and improving skills of each individual student. During the last few workshops, students were encouraged to build their own VMs with vulnerabilities and challenged the more experienced students to find and fix them.

## IV. Virtual Networking Environment

To enable interactive lecture sessions and workshops, we needed a network to demonstrate vulnerabilities and host mock competitions. The virtual networking solution had to be very versatile and unrestricted. The following are requirements that were taken into consideration:

- Allow full control of machines on school network by participants without taking on liability
- Allow full network access to machines from anywhere
- Allow only participants to host and access services
- Allow machines Internet access
- Allow fast provisioning of machines
- Allow auditing of usage of machines by participants

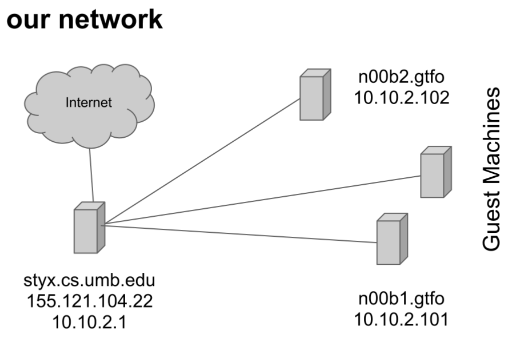
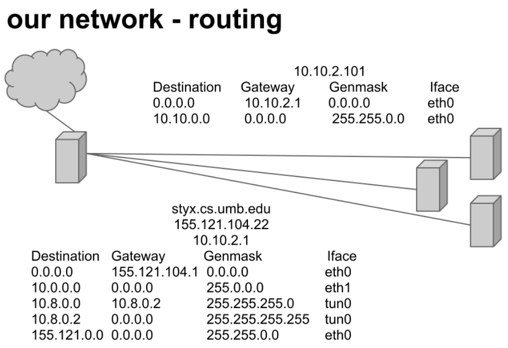


Fig. 1. Our network layout with respect to the Internet



| Destination | Gateway | Genmask | Iface |
|---|---|---|---|
| 0.0.0.0 | 10.10.2.1 | 0.0.0.0 | eth0 |
| 10.10.0.0 | 0.0.0.0 | 255.255.0.0 | eth0 |

| Destination | Gateway | Genmask | Iface |
|---|---|---|---|
| 0.0.0.0 | 155.121.104.1 | 0.0.0.0 | eth0 |
| 10.0.0.0 | 0.0.0.0 | 255.0.0.0 | eth1 |
| 10.8.0.0 | 10.8.0.2 | 255.255.255.0 | tun0 |
| 10.8.0.2 | 0.0.0.0 | 255.255.255.255 | tun0 |
| 155.121.0.0 | 0.0.0.0 | 255.255.0.0 | eth0 |

Fig. 3. The IP routing configuration

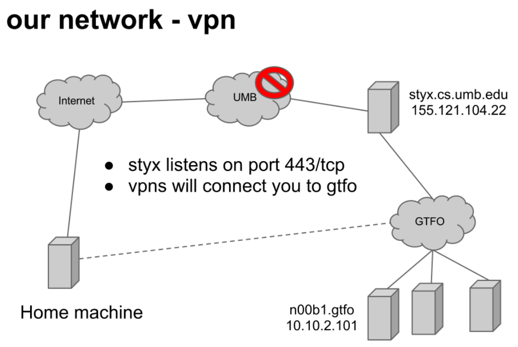


Fig. 2. The network layout with respect to the Internet

### *A. Network*

For both logistical and convenience reasons, a virtual private network (VPN) was used. We needed a powerful VPN that offered tunneling via pseudo adaptors on client machines as well as a connection mechanism that would not be be blocked by most IT departments. OpenVPN [7] was used to accept connections on port 443/tcp. The connections were secured with the 2048-bit TLS encryption using public key cryptography (PKC). OpenVPN supports almost every platform and creates a pseudo interface to allow routing level access to the sandbox network. This method could allow any university to create a sandbox network even with heavy IT restrictions. VPN access to a sandbox network is of primary importance for participants' to do remotely a variety of functions:

- Reverse shells
- Full port range scanning
- Service hosting

The VPN system is displayed in Figure 2. We are able to simulate each participants home or personal machine being plugged into our sandbox network wherever they are. With the VPN system, participants could be given full access to a NAT network without the risk of them launching Internet facing services. Understanding the VPN system was another important aspect of the workshops. We developed VPN configuration tools for Mac and Linux which involved installing a tunneling driver that allowed transparent access to our sandbox network. This allowed students a truly versatile platform for exploit exploration and development without having to worry about university IT policy restrictions.

The networks routing configuration is shown in Figure 3. Here $eth0$ is connected to a public IP address, $eth1$ is connected to the internal switch and guests, and $tun0$ is a pseudo device provided by OpenVPN which handles VPN clients. In our configuration VPN clients are allowed to communicate with other VPN clients which allows for man in the middle (MITM) attacks and other educational vulnerabilities.

### *B. Virtualization*

A virtualization host was used to create both a virtual switch and virtual guests. Virtualization is able to scale in a way that a dedicated standalone network cannot. Virtual machine templates were used to rapidly provision new virtual machines. A predefined list of MAC address to IP was configured in a DHCP server. When a machine was provisioned, its MAC address was the only change that needed to be made. When the machine was powered on, it received the correct IP address right away. This allowed support staff to avoid console access to these machines to configure IP addresses. It also allowed ARP entries to be clearly explained because where was a known mapping between MAC and IP. We set up a virtual Linux machine called Styx which was a dual-homed routable machine that served as a VPN server, NAT translator, Authoritative internal DNS server and resolver, and DHCP server.

### *C. Logging*

A log server was set up along with an in-house keylogger to track participant sessions during workshops. This allowed us to generate statistics about login times and command usage. We used this tracking to see what commands students were still having trouble with after the lectures. This helped us determine what to spend more time on. We can also plot usage of the virtual environment. Figure 4 and Figure 5 show sample statistics of login sessions. It is not surprising that the environment was most often used on Saturdays and between 3:30 and 8pm.

## V. Student Learning Results

After the competition, we conducted a survey to assess the effectiveness of our lectures and workshops. We asked questions

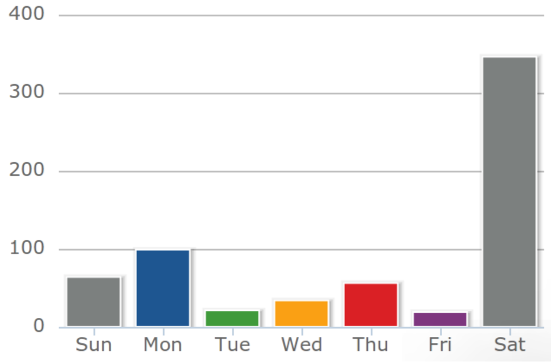


Fig. 4. Total number of logins over two months per weekday

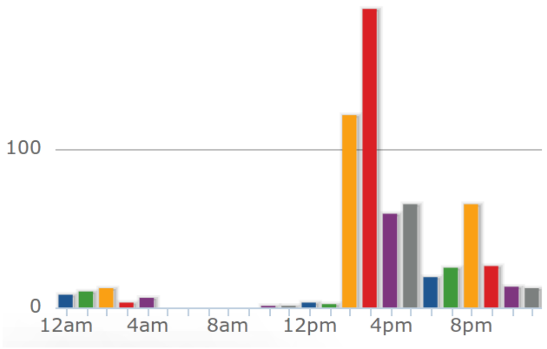


Fig. 5. Total number of logins over two months per hour

about changes in student interest, ability, and participation before and after the NECCDC competition. A total of nine students responded to our survey. Our results are shown as follows:

Figure 6 shows that our lectures and workshops increased student interest in cybersecurity. Many students reported further interest in taking more university courses in security and knowing more about the cybersecurity profession. We see in particular that many students began inquiring about a career in cybersecurity after participating. Indeed, some students suggested that we help find internships, or offer course credit for the competition as a part of our CS/IT curriculum. Less experienced students reported a larger change in interest than those who had competed last year.

Figure 7 shows that students found both our lectures and workshops helpful for learning about cybersecurity and for improving their security and computer skills. Especially for less experienced students, the improvement is more visible.

Our lectures encompassed basic system administration, network troubleshooting, and IT skills, and our workshops provided an environment for which students could practice these skills. Figure 7 shows that students found the workshop to be more beneficial than the lectures. Inexperienced students particularly found the workshops beneficial for learning new subjects. Senior level students and those who have had previous experience found them helpful to practice and apply skills they have learned previously. However, senior students reported that the lectures were less useful for them since they were the ones who were organizing and teaching them.

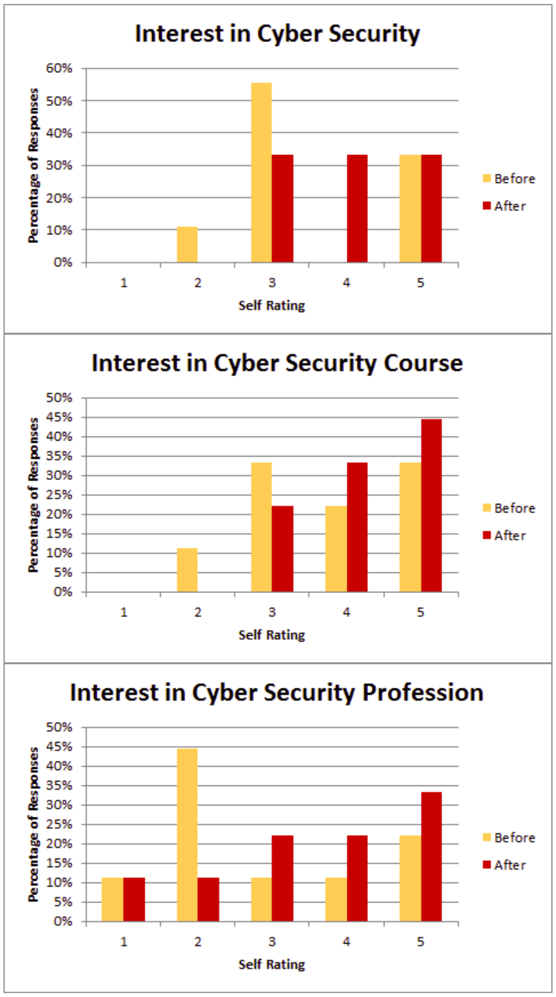


Fig. 6. Student's ratings of their own interest in various aspects of cybersecurity, before and after participation. The scale goes from 1 (low) to 5 (high).

This type of scenario-based learning employed in our workshops seemed to interest students significantly, and may prove to be a key to maintaining student interest and participation. Indeed, many students after the study suggested incorporating more scenarios in the workshops. Topics proposed included simulations of actual cyber attack / defense, identifying weak points in infrastructure, and larger scale attacks. One student even reported monitoring his or her own network more closely in order to learn cyber attack and defense.

Though the lectures and workshops initially attracted about 20 students, as the semester progressed, many students stopped attending. Towards the end, only a handful of students were present, and when the team was finally chosen, only those students who made it to the team continued participating. We conducted a survey of factors affecting student participation to find out the reasons. It reveals that neither boring lecture materials nor subject difficulty were the major factors that discouraged student participation. Surprisingly, even guaranteed inclusion of the individual in the cybersecurity team was not a strong motivator; it seems that most students were interested in the subject itself. The primary reason

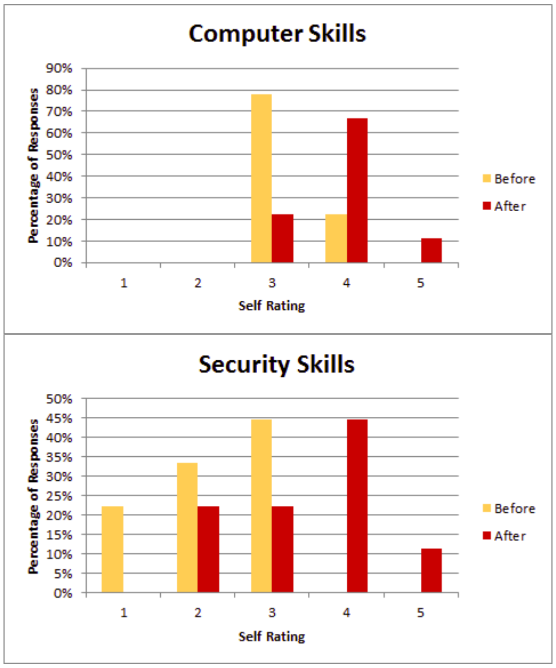


Fig. 7. Student's self reported security and computer skills, before and after participation.

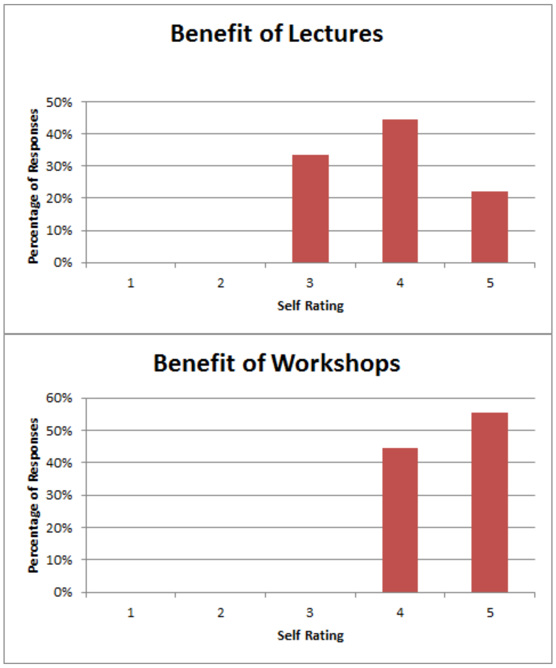


Fig. 8. Student's self reported benefit of the lectures and workshops.

was due to scheduling conflicts and the time commitments students had to made to participate in the competition. The frequency of our lectures and workshops was comparable to those of an university course, and the preparation required was no less. To address this issue, some students suggested incorporating competitions or mock competitions in cybersecurity courses. In this way, they can combine their interest in cybersecurity with their obligations to take classes. Indeed, many were interested after attending our lectures and workshops in taking more courses in security and learning more advanced topics.

The finding that inclusion in the cybersecurity team was not a strong motivator was particularly interesting. In light of the fact that students found real life applications of knowledge extremely helpful, the benefits derived from the competition may just be an instance of hands-on learning. Students were noticeably very excited to be able to practice what they had learned over the week in the workshops. Aside from solidifying their knowledge of concepts and giving them real life experience, students also derived more general benefits from the workshops, such as working in teams, communication, and leadership skills. We believe that the hands-on experience, both in workshops and actual competitions, is beneficial in invoking interest and teaching the cybersecurity subject to students,

## VI. Conclusion

This paper has examined the effectiveness of cybersecurity competitions in educating students. Though competitions do pique student interest, using this approach to increase the ability of universities to produce more IA and cybersecurity professionals is limited. This is due to the high knowledge barrier, competition as a lower priority extracurricular activity, and the difficulty to maintain interest generated by the competitions. As a result, student participation is low. Our study has shown that by combining frequent hands-on workshops with lectures, we can lower the knowledge barrier for students to learn cybersecurity. The effectiveness of cybersecurity competitions can be further improved if they are incorporated into regular courses so that student can have less scheduling conflicts in attending them.

## References

[1] G. C. New(GCN), "Budget reflects shift in scientific R&D, education," Feb. 2012. [Online]. Available: http://gcn.com/Articles/2012/02/13/2013-budget-science-technology-research-funding.aspx?Page=1

[2] "Federal cyber service: Scholarship for service (SFS) - NSF 12-531," http://www.nsf.gov/pubs/2012/nsf12531/nsf12531.htm, Apr. 2012. [Online]. Available: http://www.nsf.gov/pubs/2012/nsf12531/nsf12531.htm

[3] "Northeast collegiate cyber defense competition (NECCDC)," EMC Corporation, Franklin, MA, Mar. 2012. [Online]. Available: http://www.ccs.neu.edu/neccdc2012/index.html

[4] "MIT lincoln Laboratory/CSAIL capture the flag competition," Apr. 2011. [Online]. Available: http://mitctf2011.wikispaces.com

[5] R. S. Cheung, J. P. Cohen, H. Z. Lo, and F. Elia, "Challenge based learning in cybersecurity education," in *Proceedings of the 2011 International Conference on Security & Management*, vol. 1. Las Vegas, Nevada, USA: SAM 2011, Jul. 2011.

[6] J. Werther, M. Zhivich, T. Leek, and N. Zeldovich, "Experiences in cyber security education: The mit lincoln laboratory capture-the-flag exercise," *Cyber Security Experimentation And Test*, vol. 8, 2011.

[7] "OpenVPN," http://openvpn.net/index.php/open-source/downloads.html. [Online]. Available: http://openvpn.net/index.php/open-source/downloads.html

# Who's Knocking at Your Cybercastle's Gate?

**George Markowsky and Linda Markowsky**
School of Computing and Information Science, University of Maine, Orono, ME, USA

**Abstract –** *This paper explores how average users can understand what is involved in securing their computers from attack. We focus on the question of how to determine what is running on your computer. We propose a tutorial that we believe will help the average user understand what information utilities such as netstat can produce and how to interpret that information. We explore customizing this information using Python. We also introduce some graphical utilities that would be of use to most users.*

**Keywords:** netstat, nmap, cybercastle, cybersecurity, TCPview, Process Explorer

## 1 Introduction

Last year at SAM'11, we presented a talk entitled *Using the Castle Metaphor to Communicate Basic Concepts in Cybersecurity Education* [1]. In this paper we develop the metaphor further by looking at who is knocking at the various castle gates. This is part of our ultimate goal which is to create a series of presentations that can motivate average users to take cybersecurity more seriously and which would also teach the average users some steps that they can take. We strongly believe that it is in our best interest to reduce the amount of cybersecurity ignorance among the general public.

**Figure 1. Part of the Outer Wall of Malbork Castle**

Castles inspire many people from an early age and embody the idea of security to many people. Of course, one of the most distinctive features of castles is their walls as seen in Figure 1. It is important for people who are not experts to have some idea of how to defend themselves and their systems.

As noted, the dominant feature, and indeed, the defining feature of castles is their walls. To be useful, castle walls must have openings of various types: doors, gates, windows, drains, arrow slots, etc. A number of these are visible in Figure 1. To defend the castle walls requires understanding who might try to penetrate them and where.

This is not a trivial problem. For that matter, it was not a trivial problem in the past since it was not always be clear to the castle owner which of the people passing into the castle were friends and which were foes in disguise.

## 2 Rootkits

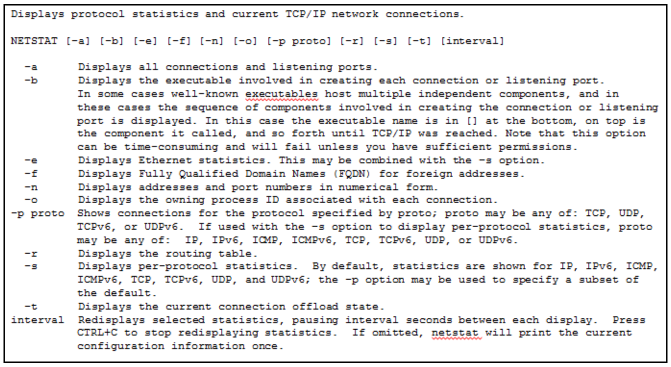
In this paper we assume that the computer in question has not been heavily compromised. In particular, we assume that it does not have a rootkit installed. Rootkits substitute their own versions for the standard utilities so that they can hide their presence on computers. Near the end of the paper we will discuss the question of how the techniques in this paper might help you determine whether you have a rootkit installed on your computer. For now, we will proceed on the assumption that your computer does not have a rootkit installed and that you are working with the native and correct utilities. At a minimum, you should understand how things are supposed to work in their "natural" state.

## 3 The Command Window

Many average users are not familiar with command windows. This is certainly a topic that they should be familiar with if they want to learn more about the security of their computers. We shall start with command window utilities because they give us the most flexibility, but we will also show that there are powerful graphical (GUI) versions of these utilities available for the three major platforms that we consider in this paper: Windows™, Linux and Macintosh™. We will simply assume that the user is aware of the command window and how to use it to issue commands.

## 4 Netstat

Netstat is the core utility that we will base our discussion on. Netstat is found in all three operating systems under discussion. There are variations in the various options that go with netstat and exactly what information it will display, but there is a core of information and options that is the same across platforms.



```
Displays protocol statistics and current TCP/IP network connections.

NETSTAT [-a] [-b] [-e] [-f] [-n] [-o] [-p proto] [-r] [-s] [-t] [interval]

  -a         Displays all connections and listening ports.
  -b         Displays the executable involved in creating each connection or listening port.
             In some cases well-known executables host multiple independent components, and in
             these cases the sequence of components involved in creating the connection or listening
             port is displayed. In this case the executable name is in [] at the bottom, on top is
             the component it called, and so forth until TCP/IP was reached. Note that this option
             can be time-consuming and will fail unless you have sufficient permissions.
  -e         Displays Ethernet statistics. This may be combined with the -s option.
  -f         Displays Fully Qualified Domain Names (FQDN) for foreign addresses.
  -n         Displays addresses and port numbers in numerical form.
  -o         Displays the owning process ID associated with each connection.
-p proto     Shows connections for the protocol specified by proto; proto may be any of: TCP, UDP,
             TCPv6, or UDPv6.  If used with the -s option to display per-protocol statistics, proto
             may be any of:  IP, IPv6, ICMP, ICMPv6, TCP, TCPv6, UDP, or UDPv6.
  -r         Displays the routing table.
  -s         Displays per-protocol statistics.  By default, statistics are shown for IP, IPv6, ICMP,
             ICMPv6, TCP, TCPv6, UDP, and UDPv6; the -p option may be used to specify a subset of
             the default.
  -t         Displays the current connection offload state.
interval     Redisplays selected statistics, pausing interval seconds between each display.  Press
             CTRL+C to stop redisplaying statistics.  If omitted, netstat will print the current
             configuration information once.
```

**Figure 2. The Complete Documentation for Windows 7 Netstat.**

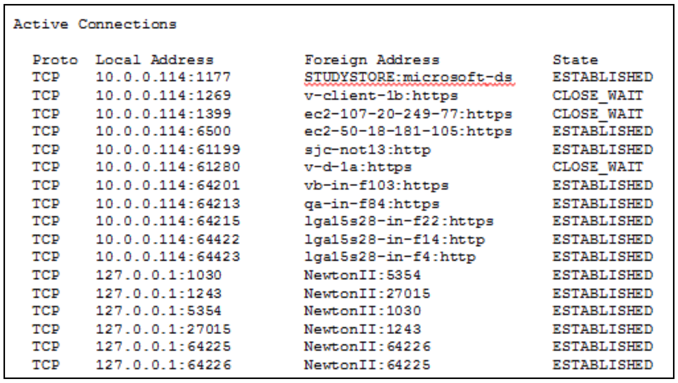
Netstat is a command line utility. Some options require administrator (root) privileges. We will focus on the Windows™ version of netstat since the majority of users are Windows™ users. Netstat has evolved along with the different versions of Windows™. In earlier versions it was quite limited or non-existent, but in Windows 7™ it has become much more powerful. If you are concerned with cybersecurity and are using earlier versions of Windows™ one of the best things you can do is to upgrade to Windows 7™. Windows 7™ netstat does not have all the features of the Linux netstat, but it is a good place to start. It is the only netstat with documentation that can fit on one slide (Figure 2).

By default netstat runs just once and displays the results at the time that you run it. If you put an integer, n, after the command, it will repeat every n seconds until you hit the Ctrl-C key.

Even though it is fashionable for many computer gurus to sneer at Windows™, it has come a long way in the security arena. It has many good GUI utilities, some of which are not available in the *nix world.

It is clear that the instructions in Figure 2 do not convey much to the average user. Often, the way that the average user deals with this problem is to just run the command. In this case just giving the command netstat produces the output shown in Figure 3.

For most people in this audience this is a perfectly easy screen to read, but the average user would most likely be confused by some of the information presented here. In particular, the State column is hard for the average user to understand, and perhaps even experienced users might have a difficult time to accurately explain the meaning of the values in the State column.



```
Active Connections

  Proto  Local Address          Foreign Address            State
  TCP    10.0.0.114:1177        STUDYSTORE:microsoft-ds    ESTABLISHED
  TCP    10.0.0.114:1269        v-client-1b:https          CLOSE_WAIT
  TCP    10.0.0.114:1399        ec2-107-20-249-77:https    CLOSE_WAIT
  TCP    10.0.0.114:6500        ec2-50-18-181-105:https    ESTABLISHED
  TCP    10.0.0.114:61199       sjc-not13:http             ESTABLISHED
  TCP    10.0.0.114:61280       v-d-1a:https               CLOSE_WAIT
  TCP    10.0.0.114:64201       vb-in-f103:https           ESTABLISHED
  TCP    10.0.0.114:64213       qa-in-f84:https            ESTABLISHED
  TCP    10.0.0.114:64215       lga15s28-in-f22:https      ESTABLISHED
  TCP    10.0.0.114:64422       lga15s28-in-f14:http       ESTABLISHED
  TCP    10.0.0.114:64423       lga15s28-in-f4:http        ESTABLISHED
  TCP    127.0.0.1:1030         NewtonII:5354              ESTABLISHED
  TCP    127.0.0.1:1243         NewtonII:27015             ESTABLISHED
  TCP    127.0.0.1:5354         NewtonII:1030              ESTABLISHED
  TCP    127.0.0.1:27015        NewtonII:1243              ESTABLISHED
  TCP    127.0.0.1:64225        NewtonII:64226             ESTABLISHED
  TCP    127.0.0.1:64226        NewtonII:64225             ESTABLISHED
```

**Figure 3. Output from the Netstat Command**

It is important for people to understand that connections are dynamic and change all the time. Using netstat on different occasions will produce different results depending on when the utility was run. We will begin our explanation of the results by focusing on the Protocol column, the first column.

## 5 Protocols

Protocols are sets of rules for how computers communicate – there are many protocols in use today, with new ones being created. The two main protocols that we will discuss are TCP and UDP. You can see TCP listed in the first column in Figure 3. TCP (Transmission Control Protocol) has rules to ensure quality communication and is designed not to lose data. UDP (User Datagram Protocol) is designed for speed and can lose data. Both TCP and TCP use the protocol IP (Internet Protocol) which comes in two versions IPv4 and IPv6. At this time, IPv4 is often just called IP. Variants of TCP that use IPv6 are sometimes referred to as TCPv6. Similarly, UDPv6 is a variant of UDP that uses IPv6. The –p parameter of netstat must be followed by one of TCP, UDP, TCPv6 or UDPv6, and it restricts its output to those values that are limited to the selected protocol.

## 6 IPv4 and IPv6 Addresses

The way computers are found on the Internet is via an address called an IP (IPv4) address. Usually, these addresses are written using standard decimal and consist of four integers in the range 0...255 separated by periods. When IP addresses were first invented they were limited to $2^{32}$ addresses which gives a range from 0 to 4,294,967,295. While this seems like a lot of addresses, it is significantly smaller than the number of people on this planet. In addition, the addresses were distributed in an inefficient manner, so that we have effectively run out of new IP addresses.

The new IPv6 addressing scheme permits $2^{128}$ addresses which gives a range from 0 to approximately $3.4\times10^{38}$ which provides a huge number of addresses for each person who has ever lived or is likely to live. IPv6 addresses are typically given in hexadecimal digits: 8 groups of 4 hex digits each with the groups being separated by colons instead of periods.

There are some complications in reading IPv6 addresses. Because these addresses are so long people use some abbreviations. In particular, leading zeroes are omitted. Thus :0: replaces :0000: and :24: replaces :0024:. Furthermore, if you see :: in an IPv6 address it stands for a number of blocks of zeroes that have been omitted. The number of such blocks is exactly the number necessary to bring the total number of blocks to 8. There is also a % with a number or some text called a *zone index* that gives additional routing information that we will ignore.

In short, if you see something that looks like 192.156.45.2:76 it is an IPv4 address followed by a port number (76) after the colon – we will explain ports in the next section. On the other hand, if you see something like [fe80::dc3:d544:ad26:ef9a%24]:546 it is an IPv6 address with 3 missing blocks of zeroes. In particular, [::] is the address [0:0:0:0:0:0:0:0] and [::1] is the address [0:0:0:0:0:0:0:1]. You may, of course, if you prefer, add all the leading zeroes. Digital locations often have names as well as numerical addresses and netstat can display location information using names instead of addresses. Figure 3 is an example of a display that uses names for locations on other computers.

To make sense of a netstat report there are certain addresses that need to be recognized. Any IPv4 address that begins with the number 127 refers to a loopback which is a connection from a computer to itself. Loopbacks are often used for testing. Typically, you do not need to worry about these addresses. IPv6 has only one loopback address: [::1] or ::1. When names are used, the loopback address is referred to as *localhost.*

One potentially confusing fact is that a computer running netstat has a variety of IP addresses that refer to its self. At this time we will not discuss the reasons for the different self-references. Just bear in mind that the different addresses that appear in the Local Address column are just different references to the computer itself.

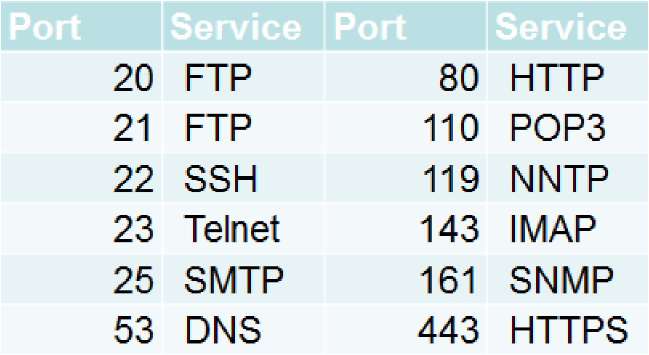
One institution that users should know about is the Internet Assigned Numbers Authority (IANA). Their website www.iana.org is an interesting place to explore and is the authoritative source of information about IP addresses and port numbers. It is good for people to realize that there is some centralized control of the "decentralized" Internet.

## 7 Openings in Walls: Ports

**Figure 4. A Fortified Gate in Malbork Castle**

Ports on a computer correspond to openings in a castle wall – they are essential, but also the weak spot in the defense. Ports are specified by numbers ranging from 0 to 65,535. Many have special meaning as is shown in Figure 5.

The information in Figure 5 comes from IANA and may be found in much greater detail at their website: http://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.xml. Users should know that services can also be run on other ports as well as the associated ports. The important thing is that parties who seek to communicate must agree on the port numbers that they will be using.

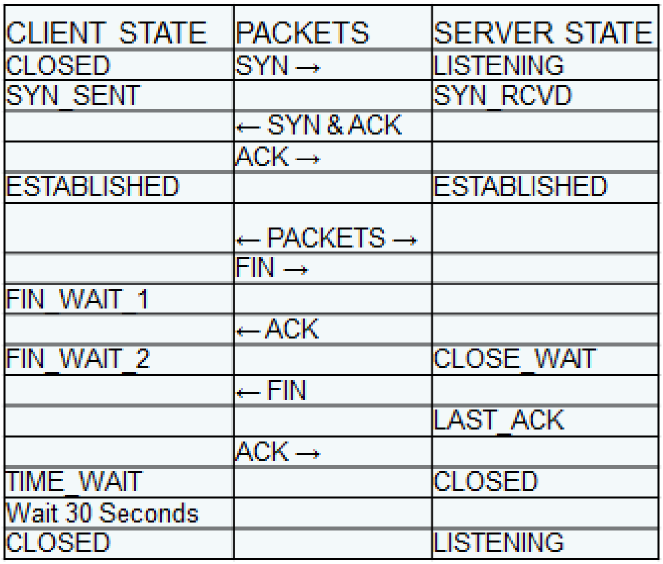


| Port | Service | Port | Service |
|---|---|---|---|
| 20 | FTP | 80 | HTTP |
| 21 | FTP | 110 | POP3 |
| 22 | SSH | 119 | NNTP |
| 23 | Telnet | 143 | IMAP |
| 25 | SMTP | 161 | SNMP |
| 53 | DNS | 443 | HTTPS |

**Figure 5. Some Common Ports and Associated Services**

## 8 TCP States

We have covered the essential details necessary to explain the first three columns of Figure 3. In this section we will explain the meaning of Column 4. Among the rules that are found in a computer protocol are rules that deal with starting a communication and ending a communication. We are familiar with this from our own lives. For example, if you call another person even on a matter of business you do not begin with the matter at hand. Usually there is some preliminary chit-chat that runs something like "Hello, is Joe there?", "I will get him for you.", "Hello, this is Joe.", "Hi, this is John. How are you?", "Fine, how are you", … This chit-chat serves to properly identify the two parties and gauge whether they are ready to communicate. Something similar happens when one computer, called the client, reaches out to make a TCP connection to another computer, called the server.



| CLIENT STATE | PACKETS | SERVER STATE |
|---|---|---|
| CLOSED | SYN → | LISTENING |
| SYN_SENT | | SYN_RCVD |
| | ← SYN & ACK | |
| | ACK → | |
| ESTABLISHED | | ESTABLISHED |
| | ← PACKETS → | |
| | FIN → | |
| FIN_WAIT_1 | | |
| | ← ACK | |
| FIN_WAIT_2 | | CLOSE_WAIT |
| | ← FIN | |
| | | LAST_ACK |
| | ACK → | |
| TIME_WAIT | | CLOSED |
| Wait 30 Seconds | | |
| CLOSED | | LISTENING |

**Figure 6. TCP States**

As you might expect, the details of this reaching out are very clearly defined in a document called RFC 793 (http://tools.ietf.org/html/rfc793). RFC stands for Request for Comments. The RFC reports are maintained by a group called the Internet Engineering Task Force (IETF, http://www.ietf.org/). This group is responsible for defining all of the standards for the Internet and maintains a repository of documents all named RFC #### where #### represents a number. RFC 793 describes many of the details of how TCP is supposed to work. We have reworked this information into the table shown in Figure 6.

By way of explanation, we should add that data between computers is organized into packets. These packets can have different properties. Some common types of packets are SYN, ACK and FIN. We will not discuss these types further, but just state that computers can tell what type of packet is being sent.

It is clear from Figure 6 that getting two computers to talk to each other is not a trivial matter. There are many details to consider. In particular, Figure 6 has no discussion of any error states. For example, what happens if the server sends just an ACK instead of a SYN? There are additional details having to do with numbering of packets, but we do not need to be concerned with this in our paper. The important thing is to understand the various states that the client and server go through depending on where they are in the communication sequence.

## 9 Netstat and Python

A useful exercise for understanding the output of netstat is to write Python programs for analyzing the output and reorganizing it into reports that might be more useful for certain purposes. We selected Python because we are familiar with it and because it is a very easy-to-read computer language that is useful for conveying information even to non-programmers. For the sake of demonstration, we have written a program that we call Pyview.py to process the output of the *netstat –an* command and reorganize the output so that it is grouped by port numbers. Pyview.py is listed in Figure 7.

```
"""
Pyview.py was written by George Markowsky
in May 2012 to illustrate how to
manipulate netstat results in Python. It
produces a listing organized by port
number
"""
import os

def Netstat(): # Transfers netstat
               # results to Python
    fin = os.popen('netstat -an','r')
   # Runs netstat and sets up a stream
    Lin = fin.readlines()
    #Loads results into a list
    fin.close() # Closes stream
    return Lin  # Returns list of lines

def IPv6(addr):
# Returns True iff addr is IPv6 address
    if ':' in addr:
        return True
    else:
        return False # Could be IPv4


def splitPortIP(st):
#Splits st into IP address & Port
    if st == '*.*':
        return '*','*'
    elif st.count(':') > 1: #IPv6 address
        ip,port = st.split(']')
        return (ip+']'),port[1:]
        #adjusts for split at ].
    else: #IPv4 address
        return st.split(':')
        # returns ip,port
```

```
def Ports():
# Groups info by local port number
    Lin = Netstat()
    ports = { }
# Creates a dictionary (assoc. array)
    for i in range(4,len(Lin)):
        lin = Lin[i][:-1].split()
        proto = lin[0]
        ip,port = splitPortIP(lin[1])
        fip,fport = splitPortIP(lin[2])
        if len(lin) > 3:
        # TCP port with a state
            state = lin[3]
        else:
            state = ''
        #UDP ports do not have states
        if IPv6(ip):
            proto += 'v6'
        if port in ports:
               ports[port].append \
             ((proto,ip,fip,fport,state))
        else:
            ports[port] = \
            [(proto,ip,fip,fport,state)]
    return ports

def pcomp(c1,c2):
# Used for sorting by port numbers
    if int(c1[0]) < int(c2[0]):
# If don't convert to integers get
# alphabetical order so that 4 comes
# after 1000
        return -1
    elif int(c1[0]) == int(c2[0]):
        return 0
    else:
        return 1

ports = Ports()
sports = [ ]
# Collects info as pairs of
# (port, list of info)
for p in ports:
    sports.append((p,ports[p]))
sports.sort(pcomp) # Sort using the pcomp
function
for c in sports:
    print "PORT:",c[0]
    for li in c[1]:
        print 6*" "+str(li)
```

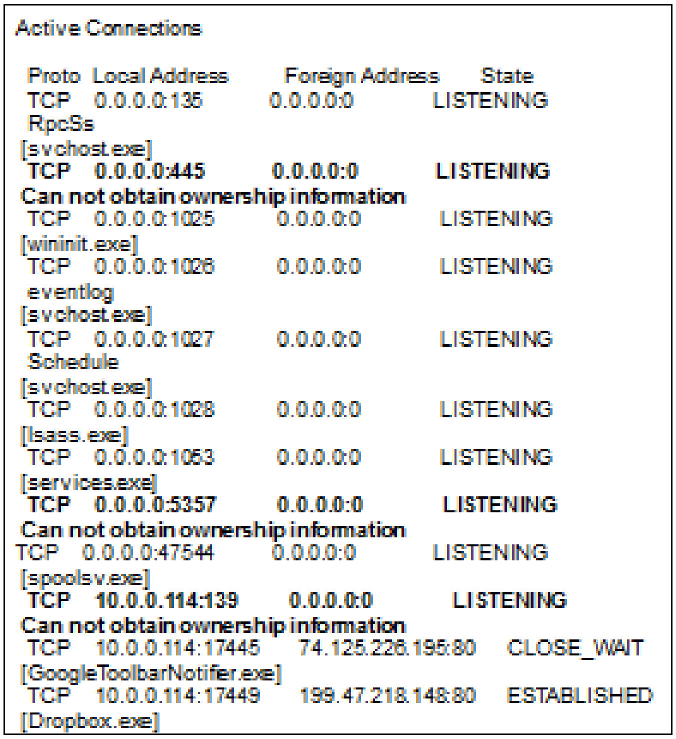
**Figure 7. Pyview.py**

Figure 8 shows part of the output that is produced by running Pyview. Pyview illustrates how to use netstat and Python together to produce custom utilities.

Note that Figure 7 shows that ports can share both IPv4 and IPv6 designations. In a more complete listing you would see that often ports with established connections are also listed as LISTENING. Using Pyview as a starting point, you can begin to manipulate the output of netstat is a variety of ways that might be of more use to you. A good target would be to extend Pyview to handle the output of the *netstat -anb* command which produces information about the programs and services that use a particular port. Figure 9 shows a sample *netstat -anb* output.

```
PORT: 135
     ('TCP', '0.0.0.0', '0.0.0.0', '0', 'LISTENING')
     ('TCPv6', '[::]', '[::]', '0', 'LISTENING')
PORT: 137
     ('UDP', '10.0.0.114', '*', '*', '')
     ('UDP', '192.168.56.1', '*', '*', '')
PORT: 138
     ('UDP', '10.0.0.114', '*', '*', '')
     ('UDP', '192.168.56.1', '*', '*', '')
PORT: 139
     ('TCP', '10.0.0.114', '0.0.0.0', '0', 'LISTENING')
     ('TCP', '192.168.56.1', '0.0.0.0', '0', 'LISTENING')
PORT: 445
     ('TCP', '0.0.0.0', '0.0.0.0', '0', 'LISTENING')
     ('TCPv6', '[::]', '[::]', '0', 'LISTENING')
PORT: 500
     ('UDP', '0.0.0.0', '*', '*', '')
     ('UDPv6', '[::]', '*', '*', '')
PORT: 546
     ('UDPv6', '[fe80::dc3:d544:ad26:ef9a%24]', '*', '*', '')
     ('UDPv6', '[fe80::58ef:5bcf:ebd6:a6db%10]', '*', '*', '')
```

**Figure 8. Some Sample Output from Pyview**



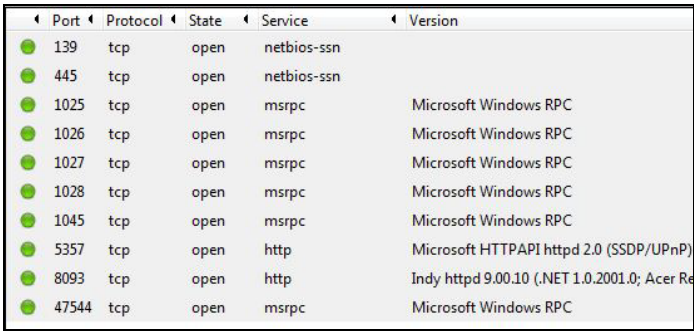
```
Active Connections

  Proto  Local Address          Foreign Address        State
  TCP    0.0.0.0:135            0.0.0.0:0              LISTENING
  RpcSs
 [svchost.exe]
  TCP    0.0.0.0:445            0.0.0.0:0              LISTENING
 Can not obtain ownership information
  TCP    0.0.0.0:1025           0.0.0.0:0              LISTENING
 [wininit.exe]
  TCP    0.0.0.0:1026           0.0.0.0:0              LISTENING
  eventlog
 [svchost.exe]
  TCP    0.0.0.0:1027           0.0.0.0:0              LISTENING
  Schedule
 [svchost.exe]
  TCP    0.0.0.0:1028           0.0.0.0:0              LISTENING
 [lsass.exe]
  TCP    0.0.0.0:1053           0.0.0.0:0              LISTENING
 [services.exe]
  TCP    0.0.0.0:5357           0.0.0.0:0              LISTENING
 Can not obtain ownership information
TCP    0.0.0.0:47544          0.0.0.0:0              LISTENING
 [spoolsv.exe]
  TCP    10.0.0.114:139         0.0.0.0:0              LISTENING
 Can not obtain ownership information
  TCP    10.0.0.114:17445       74.125.226.195:80      CLOSE_WAIT
 [GoogleToolbarNotifier.exe]
  TCP    10.0.0.114:17449       199.47.218.148:80      ESTABLISHED
 [Dropbox.exe]
```

**Figure 9. A Sample *netstat -anb* Output**

Of special interest are the ports for which netstat cannot find ownership information. You can get this information from a variety of places including Google. Some of the additional utilities that we will discuss can also supply some of the information about these unidentified owners.

# 10 Rootkits and External Checking

We mentioned in Section 2 that a rootkit would seek to hide its presence on a computer by replacing standard utilities with its own versions. One of the techniques for spotting rootkits is to investigate a computer from an external source and comparing the internal and external results for consistency. In particular, the very popular open source scanner nmap (nmap.org) [4] can be used for this purpose. Fortunately, for beginners there is a graphical version of nmap called Zenmap that can be used for free to scan your own network. Do not use these programs to scan other people's networks since this can create lots of legal problems for you. Figure 10 shows a slightly truncated version of a Zenmap report. The ports reported as OPEN by Zenmap should be included in the list of ports referred to as LISTENING by netstat.



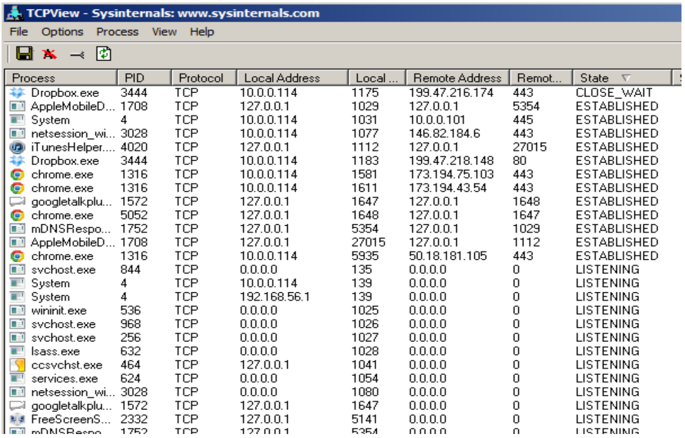
| | Port | Protocol | State | Service | Version |
|---|---|---|---|---|---|
| ● | 139 | tcp | open | netbios-ssn | |
| ● | 445 | tcp | open | netbios-ssn | |
| ● | 1025 | tcp | open | msrpc | Microsoft Windows RPC |
| ● | 1026 | tcp | open | msrpc | Microsoft Windows RPC |
| ● | 1027 | tcp | open | msrpc | Microsoft Windows RPC |
| ● | 1028 | tcp | open | msrpc | Microsoft Windows RPC |
| ● | 1045 | tcp | open | msrpc | Microsoft Windows RPC |
| ● | 5357 | tcp | open | http | Microsoft HTTPAPI httpd 2.0 (SSDP/UPnP) |
| ● | 8093 | tcp | open | http | Indy httpd 9.00.10 (.NET 1.0.2001.0; Acer Re |
| ● | 47544 | tcp | open | msrpc | Microsoft Windows RPC |

Figure 10. A Zenmap Report

A good exercise is for each person to thoroughly investigate which ports are open on each computer of interest. Closing an open port is often a bit involved since you need to understand what program is running on that port and what purpose it serves. A good rule is to close all ports that you are not actively using. Sometimes ports are opened by programs when you start your computer. In some cases these programs are no longer being used and they should be removed from the startup process. Not only does removing these programs make your computer safer, it will often improve its performance. There are programs such as Task Manager and the programs we will describe shortly that can kill a particular process when you direct them to. Unfortunately, if the process is started up when you start your computer it will reappear the next time you reboot your computer. Do not close ports you do not understand because you can disrupt your computer's operations.

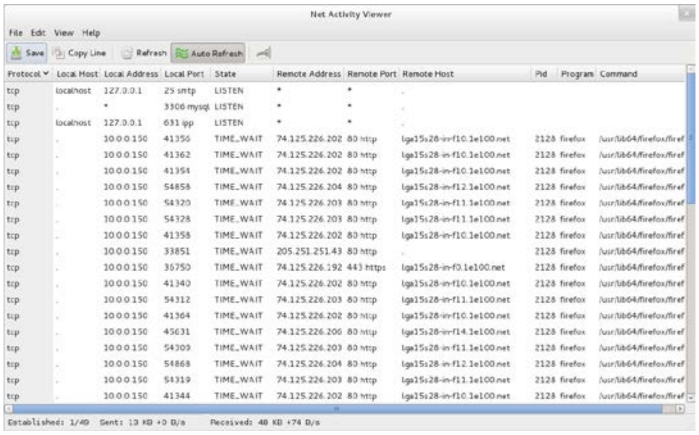
# 11 GUI Utilities

In this section we want to present some GUI utilities that we feel would be useful to most computer users. Figure 11 shows the screen for TCPview. This is a graphical version of netstat and is available for Windows™ systems at http://technet.microsoft.com/en-us/sysinternals/bb897437. You should be able to understand this output based on the earlier discussion in this paper.


TCPView - Sysinternals: www.sysinternals.com
File Options Process View Help

| Process | PID | Protocol | Local Address | Local ... | Remote Address | Remot... | State |
|---|---|---|---|---|---|---|---|
| Dropbox.exe | 3444 | TCP | 10.0.0.114 | 1175 | 199.47.216.174 | 443 | CLOSE_WAIT |
| AppleMobileD... | 1708 | TCP | 127.0.0.1 | 1029 | 127.0.0.1 | 5354 | ESTABLISHED |
| System | 4 | TCP | 10.0.0.114 | 1031 | 10.0.0.101 | 445 | ESTABLISHED |
| netsession_wi... | 3028 | TCP | 10.0.0.114 | 1077 | 146.82.184.6 | 443 | ESTABLISHED |
| iTunesHelper.... | 4020 | TCP | 127.0.0.1 | 1112 | 127.0.0.1 | 27015 | ESTABLISHED |
| Dropbox.exe | 3444 | TCP | 10.0.0.114 | 1183 | 199.47.218.148 | 80 | ESTABLISHED |
| chrome.exe | 1316 | TCP | 10.0.0.114 | 1581 | 173.194.75.103 | 443 | ESTABLISHED |
| chrome.exe | 1316 | TCP | 10.0.0.114 | 1611 | 173.194.43.54 | 443 | ESTABLISHED |
| googletalkplu... | 1572 | TCP | 127.0.0.1 | 1647 | 127.0.0.1 | 1648 | ESTABLISHED |
| chrome.exe | 5052 | TCP | 127.0.0.1 | 1648 | 127.0.0.1 | 1647 | ESTABLISHED |
| mDNSRespo... | 1752 | TCP | 127.0.0.1 | 5354 | 127.0.0.1 | 1029 | ESTABLISHED |
| AppleMobileD... | 1708 | TCP | 127.0.0.1 | 27015 | 127.0.0.1 | 1112 | ESTABLISHED |
| chrome.exe | 1316 | TCP | 10.0.0.114 | 5935 | 50.18.181.105 | 443 | ESTABLISHED |
| svchost.exe | 844 | TCP | 0.0.0.0 | 135 | 0.0.0.0 | 0 | LISTENING |
| System | 4 | TCP | 10.0.0.114 | 139 | 0.0.0.0 | 0 | LISTENING |
| System | 4 | TCP | 192.168.56.1 | 139 | 0.0.0.0 | 0 | LISTENING |
| wininit.exe | 536 | TCP | 0.0.0.0 | 1025 | 0.0.0.0 | 0 | LISTENING |
| svchost.exe | 968 | TCP | 0.0.0.0 | 1026 | 0.0.0.0 | 0 | LISTENING |
| svchost.exe | 256 | TCP | 0.0.0.0 | 1027 | 0.0.0.0 | 0 | LISTENING |
| lsass.exe | 632 | TCP | 0.0.0.0 | 1028 | 0.0.0.0 | 0 | LISTENING |
| ccsvchst.exe | 464 | TCP | 127.0.0.1 | 1041 | 0.0.0.0 | 0 | LISTENING |
| services.exe | 624 | TCP | 0.0.0.0 | 1054 | 0.0.0.0 | 0 | LISTENING |
| netsession_wi... | 3028 | TCP | 0.0.0.0 | 1080 | 0.0.0.0 | 0 | LISTENING |
| googletalkplu... | 1572 | TCP | 127.0.0.1 | 1647 | 0.0.0.0 | 0 | LISTENING |
| FreeScreenS... | 2332 | TCP | 127.0.0.1 | 5141 | 0.0.0.0 | 0 | LISTENING |
| mDNSRespo... | 1752 | TCP | 127.0.0.1 | 5354 | 0.0.0.0 | 0 | LISTENING |

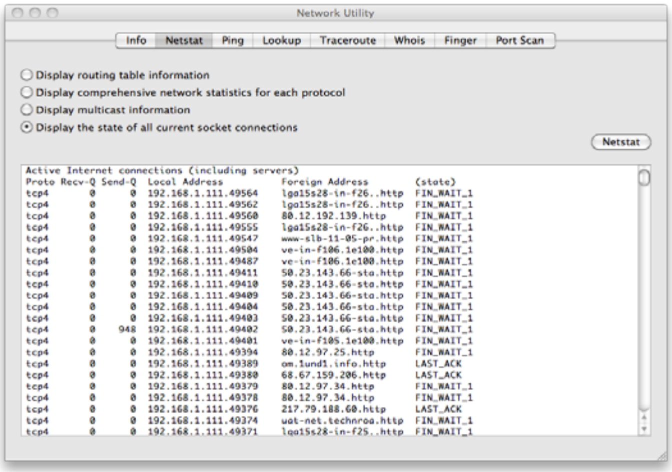
Figure 11. TCPview (Windows™)



| Protocol | Local Host | Local Address | Local Port | State | Remote Address | Remote Port | Remote Host | Pid | Program | Command |
|---|---|---|---|---|---|---|---|---|---|---|
| tcp | localhost | 127.0.0.1 | 25 smtp | LISTEN | * | * | . | | | |
| tcp | . | * | 3306 mysql | LISTEN | * | * | . | | | |
| tcp | localhost | 127.0.0.1 | 631 ipp | LISTEN | * | * | . | | | |
| tcp | . | 10.0.0.150 | 41356 | TIME_WAIT | 74.125.226.202 | 80 http | lga15s28-in-f10.1e100.net | 2128 | firefox | /usr/lib64/firefox/firef |
| tcp | . | 10.0.0.150 | 41362 | TIME_WAIT | 74.125.226.202 | 80 http | lga15s28-in-f10.1e100.net | 2128 | firefox | /usr/lib64/firefox/firef |
| tcp | . | 10.0.0.150 | 41354 | TIME_WAIT | 74.125.226.202 | 80 http | lga15s28-in-f10.1e100.net | 2128 | firefox | /usr/lib64/firefox/firef |
| tcp | . | 10.0.0.150 | 54858 | TIME_WAIT | 74.125.226.204 | 80 http | lga15s28-in-f12.1e100.net | 2128 | firefox | /usr/lib64/firefox/firef |
| tcp | . | 10.0.0.150 | 54320 | TIME_WAIT | 74.125.226.203 | 80 http | lga15s28-in-f11.1e100.net | 2128 | firefox | /usr/lib64/firefox/firef |
| tcp | . | 10.0.0.150 | 54328 | TIME_WAIT | 74.125.226.203 | 80 http | lga15s28-in-f11.1e100.net | 2128 | firefox | /usr/lib64/firefox/firef |
| tcp | . | 10.0.0.150 | 41358 | TIME_WAIT | 74.125.226.202 | 80 http | lga15s28-in-f10.1e100.net | 2128 | firefox | /usr/lib64/firefox/firef |
| tcp | . | 10.0.0.150 | 33851 | TIME_WAIT | 205.251.251.43 | 80 http | . | 2128 | firefox | /usr/lib64/firefox/firef |
| tcp | . | 10.0.0.150 | 36750 | TIME_WAIT | 74.125.226.192 | 443 https | lga15s28-in-f0.1e100.net | 2128 | firefox | /usr/lib64/firefox/firef |
| tcp | . | 10.0.0.150 | 41340 | TIME_WAIT | 74.125.226.202 | 80 http | lga15s28-in-f10.1e100.net | 2128 | firefox | /usr/lib64/firefox/firef |
| tcp | . | 10.0.0.150 | 54312 | TIME_WAIT | 74.125.226.203 | 80 http | lga15s28-in-f11.1e100.net | 2128 | firefox | /usr/lib64/firefox/firef |
| tcp | . | 10.0.0.150 | 41364 | TIME_WAIT | 74.125.226.202 | 80 http | lga15s28-in-f10.1e100.net | 2128 | firefox | /usr/lib64/firefox/firef |
| tcp | . | 10.0.0.150 | 45631 | TIME_WAIT | 74.125.226.206 | 80 http | lga15s28-in-f14.1e100.net | 2128 | firefox | /usr/lib64/firefox/firef |
| tcp | . | 10.0.0.150 | 54309 | TIME_WAIT | 74.125.226.203 | 80 http | lga15s28-in-f11.1e100.net | 2128 | firefox | /usr/lib64/firefox/firef |
| tcp | . | 10.0.0.150 | 54868 | TIME_WAIT | 74.125.226.204 | 80 http | lga15s28-in-f12.1e100.net | 2128 | firefox | /usr/lib64/firefox/firef |
| tcp | . | 10.0.0.150 | 54319 | TIME_WAIT | 74.125.226.203 | 80 http | lga15s28-in-f11.1e100.net | 2128 | firefox | /usr/lib64/firefox/firef |
| tcp | . | 10.0.0.150 | 41344 | TIME_WAIT | 74.125.226.202 | 80 http | lga15s28-in-f10.1e100.net | 2128 | firefox | /usr/lib64/firefox/firef |

Established: 1/49 Sent: 13 KB +0 B/s Received: 48 KB +74 B/s

Figure 12. Net Activity Viewer (Linux)



○ Display routing table information
○ Display comprehensive network statistics for each protocol
○ Display multicast information
⦿ Display the state of all current socket connections

```
Active Internet connections (including servers)
Proto Recv-Q Send-Q  Local Address          Foreign Address        (state)
tcp4       0      0  192.168.1.111.49564    lga15s28-in-f26..http  FIN_WAIT_1
tcp4       0      0  192.168.1.111.49562    lga15s28-in-f26..http  FIN_WAIT_1
tcp4       0      0  192.168.1.111.49560    80.12.192.139.http     FIN_WAIT_1
tcp4       0      0  192.168.1.111.49555    lga15s28-in-f26..http  FIN_WAIT_1
tcp4       0      0  192.168.1.111.49547    www-slb-11-05-pr.http  FIN_WAIT_1
tcp4       0      0  192.168.1.111.49504    ve-in-f106.1e100.http  FIN_WAIT_1
tcp4       0      0  192.168.1.111.49487    ve-in-f106.1e100.http  FIN_WAIT_1
tcp4       0      0  192.168.1.111.49411    50.23.143.66-sta.http  FIN_WAIT_1
tcp4       0      0  192.168.1.111.49410    50.23.143.66-sta.http  FIN_WAIT_1
tcp4       0      0  192.168.1.111.49409    50.23.143.66-sta.http  FIN_WAIT_1
tcp4       0      0  192.168.1.111.49404    50.23.143.66-sta.http  FIN_WAIT_1
tcp4       0      0  192.168.1.111.49403    50.23.143.66-sta.http  FIN_WAIT_1
tcp4       0    948  192.168.1.111.49402    50.23.143.66-sta.http  FIN_WAIT_1
tcp4       0      0  192.168.1.111.49401    ve-in-f105.1e100.http  FIN_WAIT_1
tcp4       0      0  192.168.1.111.49394    80.12.97.25.http       FIN_WAIT_1
tcp4       0      0  192.168.1.111.49389    om.1und1.info.http     LAST_ACK
tcp4       0      0  192.168.1.111.49380    68.67.159.206.http     LAST_ACK
tcp4       0      0  192.168.1.111.49379    80.12.97.34.http       FIN_WAIT_1
tcp4       0      0  192.168.1.111.49378    80.12.97.34.http       FIN_WAIT_1
tcp4       0      0  192.168.1.111.49376    217.79.188.60.http     LAST_ACK
tcp4       0      0  192.168.1.111.49374    uat-net.technroa.http  FIN_WAIT_1
tcp4       0      0  192.168.1.111.49371    lga15s28-in-f25..http  FIN_WAIT_1
```

Figure 13. Network Utility (Macintosh™)

We should note that Linux and Macintosh™ have similar utilities. These are Net Activity Viewer (Linux, Figure 12) and Network Utility (Macintosh™, Figure 13).

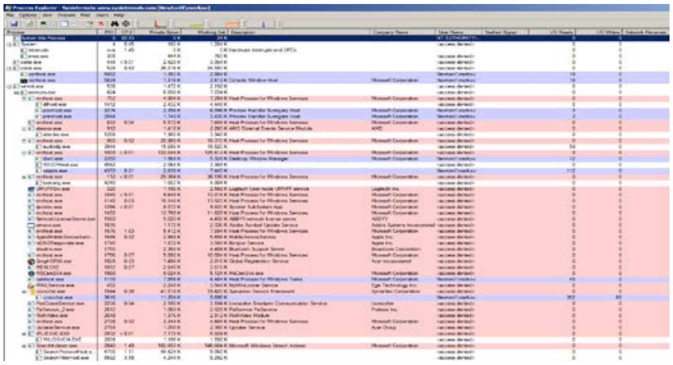
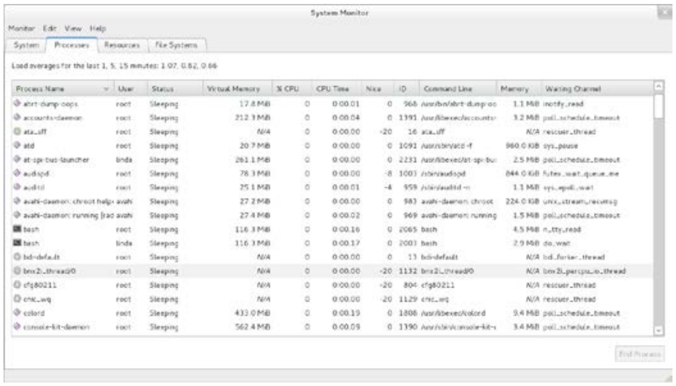


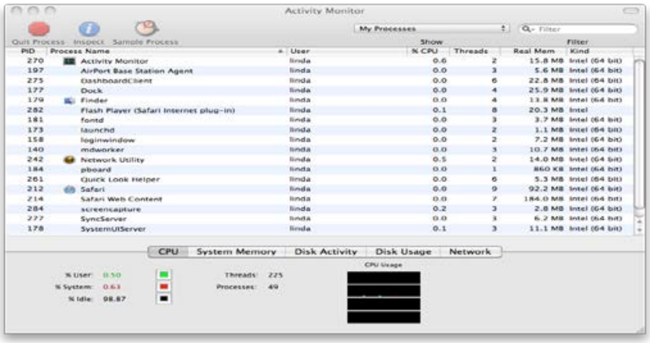
**Figure 14. Process Explorer (Windows™)**



Load averages for the last 1, 5, 15 minutes: 1.07, 0.82, 0.66

| Process Name | User | Status | Virtual Memory | % CPU | CPU Time | Nice | ID | Command Line | Memory | Waiting Channel |
|---|---|---|---|---|---|---|---|---|---|---|
| abrt-dump-oops | root | Sleeping | 17.8 MiB | 0 | 0:00.01 | 0 | 968 | /usr/bin/abrt-dump-oo | 1.1 MiB | inotify_read |
| accounts-daemon | root | Sleeping | 212.1 MiB | 0 | 0:00.04 | 0 | 1391 | /usr/libexec/accounts- | 3.2 MiB | poll_schedule_timeout |
| ata_sff | root | Sleeping | N/A | 0 | 0:00.00 | -20 | 16 | ata_sff | N/A | rescuer_thread |
| atd | root | Sleeping | 20.7 MiB | 0 | 0:00.00 | 0 | 1091 | /usr/sbin/atd -f | 960.0 KiB | sys_pause |
| at-spi-bus-launcher | linda | Sleeping | 261.1 MiB | 0 | 0:00.00 | 0 | 2231 | /usr/libexec/at-spi-bu | 2.5 MiB | poll_schedule_timeout |
| audispd | root | Sleeping | 78.3 MiB | 0 | 0:00.00 | -8 | 1003 | /sbin/audispd | 844.0 KiB | futex_wait_queue_me |
| auditd | root | Sleeping | 25.1 MiB | 0 | 0:00.01 | -4 | 959 | /sbin/auditd -n | 1.1 MiB | sys_epoll_wait |
| avahi-daemon: chroot help | avahi | Sleeping | 27.2 MiB | 0 | 0:00.00 | 0 | 983 | avahi-daemon: chroot | 224.0 KiB | unix_stream_recvmsg |
| avahi-daemon: running [rad | avahi | Sleeping | 27.4 MiB | 0 | 0:00.02 | 0 | 969 | avahi-daemon: running | 1.5 MiB | poll_schedule_timeout |
| bash | root | Sleeping | 116.3 MiB | 0 | 0:00.16 | 0 | 2065 | bash | 4.5 MiB | n_tty_read |
| bash | linda | Sleeping | 116.3 MiB | 0 | 0:00.17 | 0 | 2003 | bash | 2.9 MiB | do_wait |
| bdi-default | root | Sleeping | N/A | 0 | 0:00.00 | 0 | 13 | bdi-default | N/A | bdi_forker_thread |
| bnx2i_thread/0 | root | Sleeping | N/A | 0 | 0:00.00 | -20 | 1132 | bnx2i_thread/0 | N/A | bnx2i_percpu_io_thread |
| cfg80211 | root | Sleeping | N/A | 0 | 0:00.00 | -20 | 804 | cfg80211 | N/A | rescuer_thread |
| cnic_wq | root | Sleeping | N/A | 0 | 0:00.00 | -20 | 1129 | cnic_wq | N/A | rescuer_thread |
| colord | root | Sleeping | 433.0 MiB | 0 | 0:00.19 | 0 | 1808 | /usr/libexec/colord | 9.4 MiB | poll_schedule_timeout |
| console-kit-daemon | root | Sleeping | 562.4 MiB | 0 | 0:00.09 | 0 | 1390 | /usr/sbin/console-kit- | 3.4 MiB | poll_schedule_timeout |

**Figure 15. System Monitor (Linux, Gnome Desktop)**



| PID | Process Name | User | % CPU | Threads | Real Mem | Kind |
|---|---|---|---|---|---|---|
| 270 | Activity Monitor | linda | 0.6 | 2 | 15.8 MB | Intel (64 bit) |
| 197 | AirPort Base Station Agent | linda | 0.0 | 3 | 5.6 MB | Intel (64 bit) |
| 275 | DashboardClient | linda | 0.0 | 6 | 22.8 MB | Intel (64 bit) |
| 177 | Dock | linda | 0.0 | 4 | 25.9 MB | Intel (64 bit) |
| 179 | Finder | linda | 0.0 | 4 | 13.8 MB | Intel (64 bit) |
| 282 | Flash Player (Safari Internet plug-in) | linda | 0.1 | 8 | 20.3 MB | Intel |
| 181 | fontd | linda | 0.0 | 3 | 3.7 MB | Intel (64 bit) |
| 173 | launchd | linda | 0.0 | 2 | 1.1 MB | Intel (64 bit) |
| 158 | loginwindow | linda | 0.0 | 2 | 7.2 MB | Intel (64 bit) |
| 140 | mdworker | linda | 0.0 | 3 | 10.7 MB | Intel (64 bit) |
| 242 | Network Utility | linda | 0.5 | 2 | 14.0 MB | Intel (64 bit) |
| 184 | pboard | linda | 0.0 | 1 | 860 KB | Intel (64 bit) |
| 261 | Quick Look Helper | linda | 0.0 | 6 | 5.3 MB | Intel (64 bit) |
| 212 | Safari | linda | 0.0 | 9 | 92.2 MB | Intel (64 bit) |
| 214 | Safari Web Content | linda | 0.0 | 7 | 184.0 MB | Intel (64 bit) |
| 284 | screencapture | linda | 0.2 | 3 | 2.8 MB | Intel (64 bit) |
| 277 | SyncServer | linda | 0.0 | 3 | 6.2 MB | Intel (64 bit) |
| 178 | SystemUIServer | linda | 0.1 | 3 | 11.1 MB | Intel (64 bit) |

% User: 0.50 🟩
% System: 0.63 🟥
% Idle: 98.87 ■

Threads: 225
Processes: 49

**Figure 16. Activity Monitor (Macintosh™)**

All three operating systems also have handy utilities that give additional information about programs running on the computer. These are Process Explorer for Windows™ which is available at http://technet.microsoft.com/en-us/sysinternals/bb896653 which is shown in Figure 14, System Monitor for Linux (Gnome Desktop, Figure 15) and Activity Monitor for Macintosh (Figure 16).

## 12 Conclusions

The GUI utilities shown in Figures 11 through 16 have many features and capabilities that we do not have time to explore in this paper. We hope that the discussion in this paper has laid the foundation for understanding these utilities and in securing the gates to your cyber castle. Understanding who is knocking at those gates is a key component of securing them.

## 13 References

[1] George Markowsky and Linda Markowsky, *Using the Castle Metaphor to Communicate Basic Concepts in Cybersecurity Education*, Proceedings of the 2011 International Conference on Security & Management, pp. 507-511.

[2] Special ports can be found listed at http://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.xml

[3] Details about the TCP protocol can be found in RFC 793 (http://tools.ietf.org/html/rfc793).

*[4]* Gordon Fyodor Lyon, *Nmap Network Scanning: The Official Nmap Project Guide to Network Discovery and Security Scanning*, Nmap Project, 2009. This book is available through Amazon.com and through nmap.org.

# Developing Small Team-based Cyber Security Exercises

Brandon Mauer, William Stackpole, Daryl Johnson
Networking, Security, and Systems Administration
Rochester Institute of Technology
Rochester, New York
{bem5304,wrsics,dgjics}@rit.edu

*Abstract*—**The growth of the security industry is sparking a significant interest in well-rounded security professionals. Regional and national competitions in the academic community have been developed to help identify qualified candidates to support this industry. A course has been built to allow students to improve their skills in this area. This paper describes the process used to administer events in the support of such a competitive environment, and the process by which appropriate infrastructures are developed.**

***Keywords-cyber security education, security competition; information security***

## I. Introduction

In September 2011, the authors led a seminar course entitled "Cyber Defense Techniques." This course places students into small teams for attack-defend events on small enterprise networks, mimicking the style of the National Collegiate Cyber Defense Competition [2]. A major benefit of providing such an environment is the ability to expose students to the operations of malicious users that commit digital crime, improving their skills at defending systems and networks from attack. In this paper, we discuss the specific roles and responsibilities of our positions in the course, elaborate on pertinent details concerning the seminar and its operation, provide interpretation of the results of the activity performed by the students, and discuss the lessons taken away from this unique experience.

## II. Abbreviations and Acronyms

### A. Abbreviations and Acronyms

Below is a list of terms used throughout this document, along with their meanings:

- CCDC: Collegiate Cyber Defense Competition. One of the United States' premier collegiate cyber security competitions, held annually, culminating in a national competition featuring the winning team from each of the ten U.S. regions in San Antonio, Texas. This event is the basis for the course layout and structure.
- Blue team: the team of students chosen to secure and defend a small enterprise infrastructure and to complete business tasks [1,5].
- Grey team: the team of students responsible for the development of blue team infrastructure and the creation, delegation, and assessment of business tasks. Normally, these two roles would be split into two teams, white for business tasks and black for development of infrastructure [1,4].
- Red team: the team of students who will be acting as penetration testers for a given event. The team's responsibilities include reconnaissance, vulnerability identification, infiltration, data theft, and sabotage, as directed by the grey team [1].
- Inject: a business task for the blue team to complete. Injects are not "mandatory," although failing to perform an inject resulted in no points being awarded. Each inject was given an independent maximum point value and was scored by the grey team after a predetermined time.
- Service check: an operation performed by the event scoring engine to assess the functional and correct operation of a network service based on specific grey team criteria, run at a predetermined interval. A check was only successful if it fully met all criteria. A separate check was used for each network service.
- Service uptime: The amount of checks assessed by the scoring engine to be successful, typically expressed as a percentage.
- NSSA: The Rochester Institute of Technology department of Networking, Security and Systems Administration [3].
- Cyber Defense Techniques: The name of the seminar course described in this paper. It may also be referred to as "the course" or "the seminar."

## III. Roles and Responsibilities

### A. Roles and Responsibilities

A course instructor for Cyber Defense Techniques (Johnson, Stackpole) is given five primary responsibilities:

- To provide overall direction for the course
- To lecture students on technologies and techniques relevant to each team role in the events

- To provide counsel, insight, or assistance to student teams as requested
- To ensure fairness of competition by removing bias, while providing a gradual increase in event difficulty
- To assess student performance to provide an academic grade for the course

A teaching assistant for Cyber Defense Techniques (Mauer) is given three primary responsibilities:

- To provide counsel, insight, or assistance to student teams as requested
- To ensure fairness of competition by removing bias, while providing a gradual increase in event difficulty
- To operate the scoring engine during each event

### B. *Execution of Duties*

The authors have acquired experience at the CCDC on both blue and red teams. This experience proved to be invaluable in providing counsel to students when needed. Understanding the format of CCDC, its goals, and the type of challenges presented during the competitions helped eliminate confusion amongst grey teams as to what types of challenges were appropriate to build into the infrastructure, and streamlined the communication of ideas between the grey team and us. Although any team could freely ask questions, this was most often used by the grey teams to overcome their lack of experience in doing work in this area. The seminar is an advanced course; as such, assistance was only given to students to achieve their desired goals by providing direction and opinion.

The students were placed into specific teams to try to achieve three teams of equal measure. To maintain a high-quality event free of bias, each event infrastructure was analyzed to ensure no steps were taken to gain an advantage from a particular team setup. After the completion of each event, the next grey team was expected to take into consideration the lessons learned from the previous event to provide a more challenging infrastructure for this new event. Each event infrastructure was subject to approval before being put into use to verify the appropriate difficulty level was met.

The grey team was also asked to provide all of the necessary information to properly score a network service so that the scoring engine could be properly configured. A configuration file was then written containing this information, which would be processed by the scoring engine during its operation. The scoring engine used was commissioned for the 2006 National CCDC, written in C and Perl. Each service would be checked by the engine every three minutes. A successful check was recorded if the service provided the expected response as designated by the grey team. Any other response was considered a failure, as the checks were designed by the original programmers to simulate how an end user of a system would attempt to use the service being checked.

## IV. Course Structure and Layout

### A. *General Course Layout*

The course was divided into two major components: lectures and events. The first four weeks of the course were lectures that provided students an understanding of the basic components of each aspect of the event. Such topics included a primer on the use of the nmap network scanner and the Metasploit Framework for penetration testing, as well as illustrating techniques for securing core network services such as the Domain Name System, Internet webservers, and File Transfer Protocol. Students were also briefed on the roles and responsibilities of the grey team, with which many students did not have previous experience. The students were then placed into teams of four, one for each of the three roles needed in each event. The student teams would take a different role for each of the three events to experience red, blue, and grey team operations. Starting in the fifth week, the remaining six weeks were used for the completion of three event rounds.

### B. *Event Layout*

Each event lasted for two consecutive course meetings for a total of three to four hours of activity time, depending on the setup and teardown time needed to return the lab environment to its original state. The events were conducted on an isolated network comprised of eight VMware ESX hypervisor computers, hosting all of the virtual machines the students would be using to attack, defend, and monitor the event. This virtual infrastructure was only in use during course hours to enforce a supervised and controlled competition environment. A preparatory meeting was held before the scheduled start of a given event. At this meeting, the grey team introduced the specifics of each role to its respective team for that event, including the type of infrastructure to be defended, the priorities the blue team should consider for defending the given infrastructure and network services, and priority of targets for the red team. The grey team was also given permission to provide unclear, misleading, or false information if they chose to do so.

The meeting immediately following the completion of the event was reserved for debriefing from the grey team; typical components discussed included any observations they recorded, score analysis of blue team performance on network service uptime and inject completion, and any recommendations they wished to provide. Red team and blue team members were also invited to share their observations and opinions.

In between each event was a week-long period that the grey team would use to consult with us to build the infrastructure for the next event. Some suggestions to the grey team included what services were appropriate in the type of scenario they envisioned to develop, as well as techniques or ideas on how to introduce hidden vulnerabilities into systems the blue team would be defending. Modeling the course after the style of CCDC allows the grey team to provide an infrastructure that is "broken"; systems may not have been fully patched and up to date, services could have been left misconfigured, and hosts may have had backdoors or other malware already installed on them. As previously stated, each event was conducted entirely in a location that had no Internet access, significantly hampering the capability to patch systems, which provided an additional challenge for both red and blue teams.

## V. COMPARISON TO SIMILAR EVENTS

In the United States, cyber security exercises have existed since 2001, with the establishment of the U.S. Military Cyber Defense Exercise (CDX) [4, 6, 7, 8]. The success of such an exercise has accelerated the growth of additional exercises and projects in this area. Such documented exercises include a continuous, live, internal cyber defense activity at the University of Texas at Austin [9], an international capture-the-flag event created by the University of California at Santa Barbara [10], and a semester length graduate course at Texas A&M [8]. Each of these events represents a slightly varied approach to education-oriented cyber security. As the success of these three events continued, a steering committee was established to develop a new cyber defense competition, containing members from the Center for Infrastructure Assurance and Security (CIAS) at the University of Texas at San Antonio, the University of Texas at Austin, and Texas A&M [6]. The resulting competition, the CCDC, is the primary inspiration for the course described in this paper.

The seminar provides many of the same components offered at the CCDC and the exercises referenced above, but no two events are identical. The primary objective shared between each exercise is to provide the students interactive learning opportunities in realistic scenarios. As the course described in this paper is of finite duration, the objectives are more objective and quantifiable over the prescribed period than the event developed by the University of Texas at Austin. Similarly, a capture-the-flag component is not present at the CCDC or in this course. Therefore, the most direct comparison can be made to the CCDC and the semester course found at Texas A&M.

A full-time project by the CIAS at the University of Texas at San Antonio, the CCDC performs scoring, judging, infrastructure development and Red Team assessment using neutral or third party entities and sponsors. The course described in this paper is smaller in scale than the CCDC, and temporal and financial constraints have limited the capability for third party sponsorship and participation in this exercise. However, by allowing the students to take on those additional roles, the students gain a well-rounded and more thorough appreciation for the specific operation of each team as well as an understanding of the considerable effort needed to bring such an exercise to fruition. The professional Red Team, White Team, and Black Team present at the CCDC create an environment that provides student Blue Teams considerable opportunity to demonstrate their cyber defense skill set, and the amount of feedback the professionals can provide is significant. As proponents of a comprehensive security education, the authors propose that structuring the course as described above allows the students to obtain an equivalent amount of knowledge over the duration of the course in a broader scope, by participating as a member of each team present in a given event. The authors recognize the potential for unfair treatment between student teams in an event, as the students will eventually design an event as the Grey Team and participate as Red Team members at some point during the course. As stated above, grey team infrastructure and inject inspection is required before an event begins, and the Grey Team as well as the course instructors actively monitor the competitors throughout the duration of each event round. The authors hope that continued experience in this fashion will allow for a more robust solution to maximize fairness and the opportunity for students to learn in such an environment.

## VI. OBSERVATIONS ON TEAM PERFORMANCE

The interpretation of the results is centric to the interactions with each grey team for several reasons. It was necessary to interface with them directly to ensure the event was running as they intended, as they would be monitoring both teams for the entirety of the event. Any team-specific problems requiring attention would be channeled through the grey team to provide the competing teams more time and to streamline the event as a whole. To ensure flawless operation of the scoring engine, it was necessary to monitor its behavior at all times to correct any issues before they could impact scoring mechanisms.

### *A. Event One*

The experiences observed with each grey team were more varied than anticipated. Many hours were spent with the first grey team discussing their plans to develop a fictitious online medical facility, complete with a website and generated patient records hosted in an SQL database. This event would showcase three major components: government compliance and standards enforcement, misconfigured services, and the OpenSolaris operating system. With a medical company based in the United States, patient records are subject to HIPAA regulations. As a result, the grey team later asked the blue team to provide a compliance report in this regard. The second most prevalent feature of this event was that of misconfigured services; the primary DNS server in this infrastructure allowed all zone transfers and supported dynamic updates from any machine. An attacker who is familiar with manipulating DNS could find a wealth of information about systems on the network and could easily change DNS information to suit their needs.

The first grey team worked tirelessly toward developing a complete, eight node infrastructure with well thought out business injects, centering on system auditing, service improvements, policy, and HIPAA regulation. This first event would set a high bar for the remaining two, as the blue team visibly struggled with the foreign operating system and had a low inject completion rate.

### *B. Event Two*

The second grey team, recognizing the high quality displayed by the previous team, set out to complete an even more ambitious task--to develop a twelve node server farm with separate administration machines. This idea was especially unique for two reasons: the departure from a standard small enterprise scenario and the inclusion of systems already “pre-hardened”. Competitions of this style typically feature a variety of systems, such as client workstations (that would be used by an employee) and servers running corporate services and storing company data. In contrast, the server farm implemented by the grey team was similar to a hosting company. In this scenario, nearly all of the operating systems and data residing on blue team servers would be owned by a customer, a significant departure from the small enterprise convention. The blue team was also given a guarantee that some of the systems provided to them by the "customer" have already been "hardened," although to what extent these systems were hardened was not divulged. Therefore, the new grey team was urged to perform security hardening techniques on some of the systems of their choosing so as not to overwhelm the second blue team with an increased amount of systems to protect. This course, while designed to be challenging to the students, is still only an academic exercise and preparatory course, and with an approximate event time of three to four hours, there simply would not be enough time for either team to fully explore the infrastructure and realize the maximum benefit from this particular exercise.

The second event saw the return of misconfigured services and strongly emphasized the adherence to established service level agreements and contracts the company had entered into with the customer. The grey team explained to the blue team that one "site" had an uptime requirement of 75%, the second at 50%, and the third at 25%. While these levels were simplified, it was effective at forcing the blue team to prioritize the handling of issues as they arose.

The full infrastructure was not deployed as planned. The grey team was only able to deliver nine systems by the start of the event, plus two administration machines and a company-wide pfSense software firewall. The red team was able to capitalize on this and explored further into the network more quickly due to poor configuration of the firewall and one of the Active Directory domain controllers. Nonetheless, there was clear indication that the blue team was aware of the ramifications of violating their service level agreement with the customer (a large point deduction). As a result, only one SLA was not met at the end of the event (50% SLA).

### C. *Event Three*

For the final event, the last grey team chose to implement a small online casino. This scenario was established early on, but they were unsure of how to complete their environment. Two suggestions were offered: a Pluggable Authentication Module (PAM) configuration that would let anyone log in to the Linux systems with any password, or a sendmail email server that would execute arbitrary commands sent to it in email messages. The grey team eventually chose to implement a misconfigured PAM to complement a poorly secured web server.

During the event, the last grey team was very observant and quick to respond to both the red and the blue teams. In particular, when they noticed the blue team was not working cohesively, the grey team initiated a mock fire drill to help the blue team regroup and renew their efforts. This proved to be beneficial; at the end of event three, the blue team had the highest inject completion rate of all three events.

## VII. Event Decision Making Process

Even though the students were responsible for developing the infrastructure for each event and writing injects, there was one component of each event that was featured as a result of the authors' collaboration. As a result, each event had one or two components that not only tied the event together, but represented a realistic component of a small enterprise to which the students may have needed more exposure.

### A. *Event One - OpenSolaris and HIPAA Regulation*

Many of the courses taught in the NSSA department at RIT are taught using Red Hat Linux systems. While Red Hat and its derivatives are enterprise-friendly and capable operating systems, they do not constitute such a large portion of enterprise operating systems that exposing students to other operating systems would seem unnecessary. Solaris, then, seemed the most appropriate choice, due to its orientation to enterprise use, proliferation in technical environments, and its stability when virtualized on the ESX cluster used to host each event. Solaris is a complex and sophisticated operating system that include features offered by few others, and is built on a tested UNIX core platform. Although there was little opportunity to showcase the more advanced features, it was our hope that the exposure to the open-source (and free) OpenSolaris would be an eye-opening experience. Feedback from RIT alumni from the NSSA programs indicate it is evident that Solaris is still used frequently in many types of enterprises that hire administrators, network engineers, or security professionals.

It was also important that students understand that they may be in possession of, or responsible for maintaining the confidentiality of information and documents. Medical records are one such type of information. As medical records are continuing to be made available digitally, regulations such as HIPAA will be more significant than ever. Companies who do business in this industry are bound by law to uphold these requirements concerning digital patient records, while ensuring doctors and other medical employees, as well as patients, can access the appropriate medical records in accordance with their rights. It behooves the persons responsible for the storage, safety, and security of this information to be vigilant in their duties to safeguard this information.

### B. *Event Two - Provider/Customer Model and Service Level Agreements*

We strongly advocated for the development of a server farm/hosting company scenario for the second event. Server hosting companies and cloud computing vendors have changed the landscape in which customers and enterprises do business. This scenario would also provide practical demonstration of the issues that hosting companies and cloud vendors face when providing network services for customers. Combining the aspects of honoring contractual obligations and service level agreements together would help students who plan on working for a service provider to understand the need to develop clearly defined, enforceable, and effective security policies and service level agreements. These documents apply to internal business as well as business between provider and customer. Due to the evident prioritization of the blue team response, meeting two out of three SLA requirements is a step in the right direction.

### C. Event Three - Dysfunctional Authentication and Compliance, Revisited

In the final event, the grey team deployed the infrastructure to the blue team with deliberately misconfigured PAM for all of the Linux systems. As an online casino, this company could potentially be responsible for upholding Payment Card Industry (PCI) compliance for credit card transactions; a database breach could potentially result in the theft of a large volume of credit card information. An attacker can do all manner of malicious things to a system remotely, but the danger is even greater if the authentication mechanisms that provide most of the access control to a system do not function properly. This behavior was particularly difficult to find, as the system would often accept password changes and other control modifications, but they were not enforced. Although many students found it humorous that they could manipulate the system freely, the realization that this system was very poorly secured as a result of a simple misconfiguration resounded clearly.

## VIII. Lessons Learned

We have taken away several important lessons from these exercises. The students expressed that they have noticeably improved their skills, an opinion also shared by the authors. The students more firmly understand the value of team skills, and the observations presented in this paper help us to continue to refine our technique and expand our knowledge of operating this type of environment in which the students compete and learn.

### A. Student Improvement

Based on our observations, students were not only more attuned to finding vulnerabilities in systems (both from an offensive and defensive perspective), but also more easily able to engage the thought processes of incident response and system auditing to improve system security. This experience as a seminar course provides additional learning in areas not typically covered by core curriculum, and is a superb addition to academic credentials and provides a broader foundation for continued study in this area.

### B. Teamwork and Interpersonal Skills

The students have noticeably demonstrated a deeper appreciation of the importance of teamwork in an environment such as this. CCDC is a team competition, and many enterprises have teams of people or departments who work together frequently; consequently, an employee who can work successfully in a team setting can be very valuable. It was evident during the events that teams had initial internal friction over leadership and coordination; fortunately, much of it had been addressed by the end of the course, but teams who could not overcome that challenge admitted that they encountered continued difficulties.

This effect was most profoundly demonstrated by the grey teams. While the students are capable of engaging in both red and blue team exercises in other controlled environments, to our knowledge, the grey team is a unique experience offered in an environment such as this. Understanding the differences in approach from competing in an event versus designing, building, administering, and scoring an event offers a deeper insight into the true goals of the event. The effects of this insight go full-circle; an effective grey team can build an infrastructure that sufficiently challenges the blue team and gives the red team opportunities to hone their skills. Effective scoring of injects requires the capability to quantify that the blue team has achieved the understanding necessary to properly complete assigned tasks. This, in turn, helps discourage teams from completing objectives simply to obtain the points each objective is worth.

### C. Knowledge for Future Endeavors

Our involvement in CCDC in the past was beneficial in helping the students get the maximum benefit from the course, but this knowledge is only part of the solution. Understanding the spirit of CCDC, its goals, organization, and structure ensures the course follows the same path laid out by those responsible for CCDC; however, the competition is not run by specific guidelines that mandate certain systems, devices, or components to be present (or not present). Understanding how systems work when functioning properly, the way individual software components work together to produce a functioning system, and how changes to the system affect its operation (both seen and unseen) are many of the remaining pieces of this elaborate puzzle. Engaging in the development of such an exercise is a large undertaking, which can be improved upon by steadfast practice and learning from the experiences of others. As a result, we are confident that we can use this experience as a foundation for future endeavors in this area and improve the quality of the course as it matures.

[1] ReferencesNational Collegiate Cyber Defense Competition, CIAS http://nationalccdc.org

[2] G. Vigna, "Teaching hands-on network security: testbeds and live exercises," Journal of Information Warfare, vol 3, no 2. 8-25 2003

[3] Networking, Security, and Systems Administration, Rochester Institute of Technology http://nssa.rit.edu

[4] Wayne J. Schepens, John R. James, "Architecture of a cyber defense competition," IEEE International Conference on Systems, Man and Cybernetics, vol 5. 2003, pp 4300-4305.

[5] Gregory B. White, Dwayne Williams, "Collegiate cyber defense competitions," ISSA Journal, October 2005.

[6] Art Conklin, "Cyber defense competitions and information security education: an active learning solution for a capstone course," in *Proceedings of the 39th Hawaii International Conference on System Sciences*. 2006. Honolulu, HI.

[7] Wayne J. Schepens, Daniel J. Ragsdale, John R. Surdu, "The cyber defense exercise: an evaluation of the effectiveness of infromation assurance education" The Journal of Information Security, 2002. **1**(2).

[8] Lance J. Hoffman, Daniel J. Ragsdale, "Exploring a national cyber security exercise for colleges and universities" Report No. CSPRI-2004-08, The George Washington University, Report No. ITOC-TR-04001, *United States Military Academy*. 2004.

[9] G. Chamalese and A. Pridgen, "The success of the UT IEEE communications society," *Proc. 8th Colloquium for Information Systems Security Education*, 10 June 2004, pp. 9-12

[10] G. Vigna, "Teaching network security through live exercises," *Proc. 3rd Ann. World Conf. Information Security Education* (WISE 3), Kluwer Academic Publishers, 2003, pp. 3-18

# Employing Entropy in the Detection and Monitoring of Network Covert Channels

Chaim Sanders, Jacob Valletta, Bo Yuan, Daryl Johnson, Peter Lutz
Department of Network Security and Systems Administration
Rochester Institute of Technology
Rochester, NY 14623, USA
{ces1509, jrv1197, bxyics, dgjics, phlics}@rit.edu

*Abstract*— ***The detection of covert channels has quickly become a vital need due to their pervasive nature and the increasing popularity of the Internet. In recent years, new and innovative methods have been proposed to aid in the detection of covert channels. Existing detection schemes are often too specific and are ineffective against new covert channels. In this paper, we expound upon previous work done with timing channels and apply it to detecting covert storage channels. Our approach is based on the assumption that the entropy of covert channels will vary from that of previously observed, legitimate, communications. This change in the entropy of a process provides us with a method for identifying storage channels. Using this assumption we created proof of concept code capable of detecting various covert storage channels. The results of our experiments demonstrate that we can successfully detect existing and unpublished covert storage channels accurately.***

***Keywords*— covert channel; security; detection; entropy**

## 1. Introduction

Since Lampson [1] originally introduced the idea of covert channels on trusted systems, network based approaches have become more prevalent [2]. These network based approaches make it extremely challenging to secure traditional networks. Covert channels also prove problematic for individuals, companies, and countries in securing their information or data. In general there have been a number of different approaches, each with varying levels of success for detecting and dealing with network covert channels [2]. Most of these proposed schemes are designed to detect a specific covert channel and rely solely on signatures.

Recently, there has been research into using statistical methods in order to better detect covert channels, with mixed results [31]. Typically the major hindrance of these experiments is due to the large number of possibilities at different levels of network communications where a covert channel might occur. While there is recent research into a general detection scheme for network based timing channels [3], the same type of general solution has not been implemented for storage channels.

In this paper we propose an entropy based detection mechanism for storage based network covert channels. Entropy, as defined by Shannon, is a measure of the uncertainty associated with a random variable [4]. While we observe that a network has many different types of data, we postulate that much of it remains constant with the exception of application payloads and checksum fields. From this we concluded that the creation of a storage channel should behave as a statistical outlier when compared to the normal operation of a network.

In order to detect this we use a large baseline estimate of the network traffic and analyze new entries into the network based on the calculated entropy levels that have been seen. The effect of a given packet will then modify the previous entropy estimation allowing, in time, the system to learn what is acceptable for the network and adapt to new technologies, as well as the detection of storage channels.

## 2. Background and Related Work

Covert channels have historically been broken up into two separate categories: storage channels and timing channels [7]. More recently a non-ambiguous definition has been devised that defines a storage covert channel as a communication channel in which “the output alphabet consists of different responses all taking the same time to be transmitted” and a timing covert channel as a channel where “the output alphabet is made up of different time values corresponding to the same response” [10, 12]. Additionally, the idea of behavioral based covert channels has been introduced in [11]. These categories are further broken down by the system they use (network or single system), their stealthiness, and bandwidth. The main characteristic of a covert channel is the aim to hide the fact that a transmission is taking place [5]. Early research has shown it is almost impossible to completely eliminate covert channels. Accordingly, the US government has stated that a covert channel with a bandwidth under 1 bit per second is acceptable [7]. Thus the main goal of research in this field is to develop defenses against covert channels such as removing them, disrupting them, managing their threat and alerting to their presence. Since there is always such a large amount of data, the techniques of alerting, managing and disrupting are more realistic and prevalent in research [2].

There has been extensive research devoted to detecting timing channels. Research started with [6] and the identification of covert channel capacity for auditability. Additional auditability techniques were introduced in [15] which allowed for audited applications that depended on relative timing of operations. The first techniques to prevent timing channels were presented in [9, 14] in the forms of pumping and artificial noise injection, these techniques introduce artificial timing delays to communications in order to eliminate timing channels. In [8] the idea of pumping is evolved into that of jammers which can modulate noise and timing in order to reduce the overall capacity of a given channel. Recent research has established a basis for statistically determining a timing channel from amongst legitimate data and alerting on it [3, 5]. Throughout this whole time frame, individuals who developed covert channels also presented signature based defenses for detecting their own specific covert channels [10, 22, 25].

While many effective techniques have been developed to defend against timing based channels less published work is available to detect covert storage channels. We see early work done on calculating bandwidth and capacity of the channels by [17]. Additionally, [11, 16] discuss the possibility of detecting storage channels at the development stage of production. Recently, [17, 19] used network normalization to remove avenues that could potentially carry a covert channel. As with timing channels, individuals authors also presented signature based methods to detect their own storage channel [20, 21, 23, 24]. In our paper we expound upon previous detection techniques available for timing channels [3], applying them to storage channels to provide a viable solution to detect network based storage channels.

# 3. ENTROPY CALCULATION

## 3.1 Entropy

Our hypothesis for detection of storage based covert channels focuses on the mathematical formula known as information entropy. As previously stated, entropy is a measure of the uncertainty associated with a random variable. Putting this simply, entropy tells us how much randomness exists in a set of values for a given variable. In this case, the random variable is a field in a network protocol header. By this logic, we are able to gather the values for fields and, after collecting a minimum amount, calculate the randomness of the field value. At some point this randomness value should reach a plateau, if the field itself isn't based on a random function. At this point we assume that this is the normal entropy for that field on the network, and that a significant change in entropy would be a sign of the existence of a covert storage channel. As a quick example, the RFC for ICMP denotes that the code field for an ICMP Echo Request should always be zero. If we assume that all devices on the network are obeying the RFC we should be able to say with certainly that the randomness of this field, over time, should equal zero. The presence of a covert storage channel utilizing the code field would thus create a spike in entropy, triggering an alert.

# 4. TEST CHANNELS

We selected a number of covert storage channels in order to test how effective our algorithm was at detecting variations on a network. We developed or implemented the following covert channels which had varying levels of bandwidth and stealthiness along with network impact. Each test channel was chosen because it uses a different network protocol, giving us a wider array of covert channels to test.

## 4.1 ICMPv4 & ICMPv6 Based Covert Channels

In order to first identify a basic covert channel we developed a very simple technique to send data via the ICMP code field. According to the RFC this field should always be zero [29]. This technique was based loosely on the ideas presented in [26] and enabled us to have a moderately stealthy channel from which to develop our test methods around. This concept was then expanded, to include both the identification and sequence numbers in the ICMP Echo header.

To add another protocol to our tests, we created a ICMPv6 covert channel based loosely on [27, 20]. This allowed us to expand our abilities to detect different protocols beyond those based on IPv4. This covert channel uses techniques similar to that of the ICMPv4 variant discussed earlier..

## 4.2 UDP Based Covert Channel

After demonstrating success with an ICMP covert channel, we examined a storage based covert channel presented in [28]. This covert channel modulates the length fields found in network and transport layer protocols to send packets with even or odd lengths, which in turn translate to a one or zero. This gives the potential to transmit a bit of data per packet. Using standardized protocols such as UDP and IP this model proposes a minimum bandwidth ratio of 1:11680, but can easily be increased depending on the maximum transmission unit for a network.

## 4.3 TCP Based Covert Channel

Our last test channel was a very simple covert channel found in the TCP protocol. This channel is one of the classic examples of a storage based covert channel and is presented in [14]. This covert channel uses an unused field in the TCP header to send up to 4 bits of data per packet. It should be noted that this is a dated and deprecated covert channel.

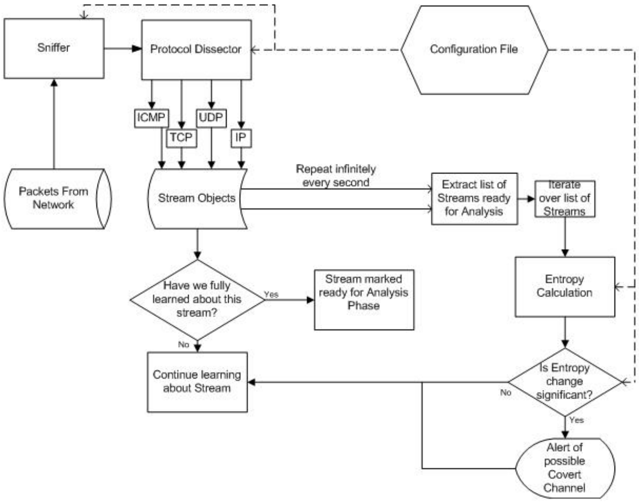
# 5. EXPERIMENTATION

To reinforce our hypothesis that storage channels can be detected using entropy, we developed a proof of concept (PoC) program capable of capturing, dissecting and calculating entropy associated with network traffic. This program uses the network protocol dissectors built into Scapy, a packet manipulation module for Python. After creating our proof of concept code and the test channels described above, we conducted a few experiments to test the overall effectiveness of the code. Our experiment consisted of using random packet captures from an active private network; each packet capture was 50,000 packets in size and was captured at different points during the day. In each packet capture one of the covert channels that we implemented was executed. We parsed these packets into our software and evaluated our success based on the outcome.

## 5.1 Proof of Concept Design

The overall flow of the PoC is summarized in the figure below [*Figure 1*]. The program has two phases: a learning phase, and a analysis phase. Even though the program is designed to be a proof of concept, it is still designed to be robust and expandable. This would allow us to move from just a proof of concept to a usable, implementable piece of code.

*Figure1PoCFlow*


Sniffer
Protocol Dissector
Configuration File
ICMP
TCP
UDP
IP
Packets From Network
Stream Objects
Repeat infinitely every second
Extract list of Streams ready for Analysis
Iterate over list of Streams
Have we fully learned about this stream?
Yes
No
Stream marked ready for Analysis Phase
Entropy Calculation
Continue learning about Stream
Is Entropy change significant?
No
Yes
Alert of possible Covert Channel

### 5.1.1 Learning Phase

In the learning phase, the program establishes a network entropy profile by capturing and storing network traffic. The PoC allows a customizable window size of traffic to be specified. This is designed to provide flexibility for various network implementations. The window size is a requirement of the program that specifies how much traffic will be taken in before the actual entropy comparisons are calculated. This is done to establish a base entropy calculation for the network. By definition the larger the window size the more accurate comparisons will be, leading to less false positives. The window size also serves as the buffer for incoming packets during the analysis phase, therefore, a larger window size, although allowing for more comparisons, also requires exponentially more computation. Another feature we have added during this phase is the ability to manually specify which protocols should be analyzed, thereby decreasing the amount of computation required. Once the specified threshold of network traffic has been reached, the program transitions to the analysis phase.

### 5.1.2 Analysis Phase

The analysis phase is responsible for monitoring and alerting deviations in entropy based on the profile established in the learning phase, as well as to continuously learn and maintain the entropy profile. This process creates a network dependent behavioral based approach to detecting storage based covert channels. We provided a configurable sensitivity level which allows the code to be customizable to a given network. The various setting of the sensitivity parameter control, to some degree, the amount of false positives produced by the code. The sensitivity parameter works hand in hand with the window size parameter specified in the previous section to create a scalable sensitivity that allows the window size to range from a value as small as 64 packets to as large as 500,000 packets and remain accurate.

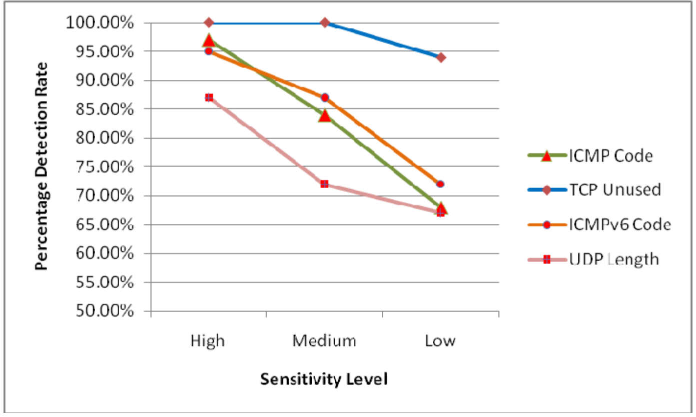
## 5.2 Experimental Setup

We had a total of four experiments that we conducted in which we attempted to detect our test covert channels. Although each tests was performed independently the same application and principles were used for each. For all four of our experiments we used the same default settings. For our experiments we choose the relatively low training window size of 256, that is, the program took its base assessment of the network from the first 256 packets and started evaluating entropy after that point. While with a higher window size we would see initially lower false positive results the total detection rate should arrive at the same level of effectiveness do to its ability to learn the network. The reason why we selected a low window size is so that we didn't discard vital packet information where our covert channels might be occurring in the experiment data. In real world usage training could be done on many thousands of packets prior to full scale usage. The experiments were all designed to transmit the same message for each channel to maintain consistency for detection purposes. We choose the words "this is a test hello world". Additionally, for each experiment we used 3 different levels of sensitivity, what these levels indicate is the amount of variation in entropy on a given field that was allowed without reporting the event as suspicious. The three levels: low, medium and high had 10.04%, 8.007%, and 3.45% variation allowed respectively.

## 5.3 Experimental Results

Our experiments indicated that our original notion that covert channels will behave as statistical entropy outliers was correct. With high sensitivity we achieved, on average, a 94.75% detection rate with a false positive rate less than 1%. While that is an accomplishment, having several hundred false positives could present an issue without proper screening. Low sensitivity achieved nearly as impressive detection rates with an average of 75.25% of covert packets detected with a far lower false positive rate of 124.75 or 0.25%. Ultimately, we believe that the experimental results show that given proper training and configuration our method would result in lower false positive rates while maintaining the expected high detection rates. It is also important to note that false positives that were detected are activity that is not normally seen on the network. That being the case, although it is not a covert channel, the packet has a high probability of being other unwanted traffic. Additionally, although we can see detection rates drop off at low sensitivity in a monitored environment, only one detection of a covert channel should be enough to alert the maintainer of the network to its presence and allow for it to be disabled accordingly.

*Figure 2 – Detection Rates*



### 5.3.1 ICMP Code Field Results

When testing for the ICMP Code channel we originally used high sensitivity, although this approach generated many more false positives, 408 precisely, it alerted us to 97% of the packets that contained the covert channel. When we moved to medium sensitivity where an 8.007% change in entropy per second would alert us, we received similarly high results, achieving an 84% detection rate with 263 false positives. Moving to low sensitivity we only achieved a 68% detection rate but we also only had 126 false detections over our entire capture of 50,000 packets.

### 5.3.2 TCP Unused Field Channel Results

The TCP unused field channel can only transmit 4 bits per packet, as a result very many packet have to be sent in order to get a full message across. This turned out to be a double edged sword in terms of detection for our algorithm. In the short term such a large amount of random traffic was immediately detected on all sensitivity settings. However, we found that if the message was very large the extreme amount of traffic would eventually become expected by the algorithm causing detection rate to drop. Using our test message this channel generated 26 packets, on both high and medium we had a 100% detection rate for this particular channel. When on low sensitivity we dropped to 94%. While the detection rate was high we continued to maintain a comparative steady level of false positives getting 367 false positives on high sensitivity, 197 false positives on medium sensitivity, and 103 false positives on low sensitivity.

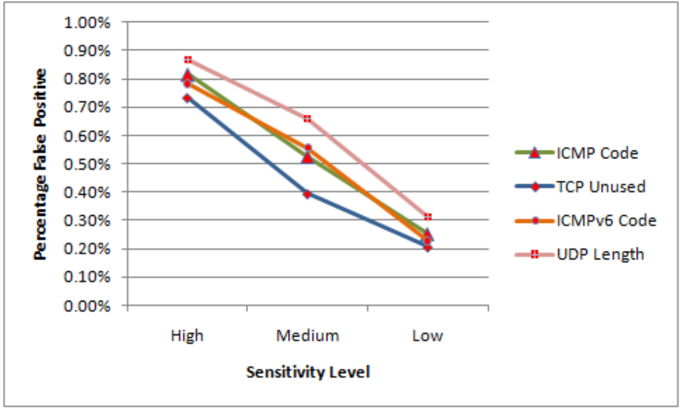
### 5.3.3 ICMPv6 Code Channel Results

Being very similar to the original ICMP channel we expected similar results in terms of detection rates and false positives. We were indeed not disappointed, the variation that we had is most likely a result of the different traffic between the two captures. Our results had 392, 278, and 114 false positives on high, medium, and low sensitivity levels respectively. We achieved slightly better detection rates compared to the regular ICMP channel as well - detecting 72% on low sensitivity, 87% on medium sensitivity, and 95% on high sensitivity. A reason for these slightly better rates could be due to the fact that although the network captures contain IPv6 traffic it is relatively low due to the amount of IPv4 packet. This discrepancy, although true in real life, also makes it harder for our covert channel to blend in thereby making it easier to detect.

### 5.3.4 UDP Length Channel Results

Our final channel dealt with UDP length and had very interesting properties where it didn't try to modify an actual field but rather modified the packet itself in order to add length. While we expected to be able to detect this channel we expected the results to be lower detection rates then the other channels. We had very similar false positives to the other channels having 433, 329, 156 false positives with regards to high, medium and low sensitivity respectively. The detection rates, while existent, were lower here being 87, 72, and 67 percent

*Figure 3 – False Positives Rates*



## 6. LIMITATIONS

In this section we discuss possible limitations of our entropy based system to detect covert channels. A large requirement imposed by this system is the need for intense packet analysis. This scanning can be done on existing packet captures but is most effective on a live network. This type of packet analysis is most analogous to deep packet inspection and requires an in-line component that will slow down the network.

Additionally, at this time, our approach is only possible on protocols for which the specification is known and using unencrypted transmissions. The reason for this is due to the entropy calculation on a per field basis. As a result of the requirements there are several countermeasures to this type of detection, the most obvious being encryption. Other countermeasures include: using unknown protocols and using non-network based covert channels. As with most entropy based systems adjusting the system correctly to minimize false positives, while still reporting legitimate covert channels is pivotal.

As a final limitation, our tool is still not one-hundred percent error free. Because of its fully adaptive capabilities, our proof of concept is prone to false-positive results, especially in high entropy fields. This means that the tool, while effective, might require additional human examination of network traffic.

## 7. Conclusions and Future Work

We expounded on previous work for using entropy to detect timing channels and applied it successfully to the detection of storage based covert channels. We designed and created a system which is implemented for some of the more prevalent packet types and allows easy expansion in the future for other types. We utilized a standard entropy calculation in order to identify a number of different covert storage channels.

In the future we would like to explore the connections between other suspicious activity and the use of covert channels. Additionally, we would like to see if the current detection system could be adapted to not only detect covert channels, but also to block and filter them. This type of system could take advantage of normalization on only suspect packets thereby making the previous work on normalization more effective, turning our passive warden system into an active warden system. Furthermore, we want to investigate ways to further reduce the false positive rate while keeping the detection rate stable.

## References

[1] B. Lampson. “A Note on the Confinement Problem.” Commun. ACM, vol. 16, no. 10, Oct. 1973, pp. 613–615.

[2] S. Zander. “A Survey Of Covert Channels and Countermeasures In Computer Network Protocols.” IEEE Communications Surveys, vol.9, no. 3, 3rd Quarter 2007.

[3] S. Gianvecchio and H. Wang. “Detecting Covert Timing Channels: An Entropy-Based Approach.” Proceedings of the 14th ACM conference on Computer and communications security (CCS '07). ACM, New York, NY, USA, pp. 307-316.

[4] C. Shannon. “A mathematical theory of communication.” Bell System Technical Journal Vol. 27, July and Oct. 1948).

[5] V. Berk et al. “Detection of Covert Channel Encoding in Network Packet Delays.” Technical Report TR536, Revision 1, revised November 2005.

[6] J. K. Millen. “Finite-state Noiseless Covert Channels.” Proc. Computer Security Foundations Workshop 11, Franconia, NH, June 1989, pp. 81-86.

[7] U.S. Department of Defense. Trusted computer system evaluation "The Orange Book". DoD 5200.28-STD Washington: GPO:1985, 1985

[8] J. Giles and B. Hajek. “An information-theoretic and game-theoretic study of timing channels.” IEEE Transaction on Information Theory, volume 48, pages 2455-2477, September 2003.

[9] M. Kang, I. Moskowitz, and D. Lee. A network version of the pump. In Proceedings of the IEEE Symposium in Security and Privacy, pages 144-154, May 1995.

[10] CR. Tsai. "A Formal Method for the Identification of Covert Storage Channels in Source Code." Proceedings of the 1987 IEEE Symposium on Security and Privacy, pp. 74-87.

[11] D. Johnson et al. “Behavior-Based Covert Channel in Cyberspace.” Intelligent Decision Making Systems, pp. 311-318.

[12] I. S. Moskowitz and M. H. Kang. “Covert channels - Here to stay?” Proceedings of the 9th Annual Conference on Computer Assurance (COMPASS’94). National Institute of Standards and Technology, pp. 235–244.

[13] H. Wei-Ming, “Reducing Timing Channels with Fuzzy Time,” Proc. IEEE Computer Society Symp. Research in Security and Privacy, May 1991, pp. 8–20.

[14] T. Handel and M. Sandford. “Hiding data in the OSI network model.” First International Workshop on Information Hiding(Cambridge, U.K.), May-June 1996.

[15] P. M. Melliar-Smith and L. E. Moser. “Protection against covert storage and timing channels.” Proc. Computer Security Foundations Workshop IV, June 1991, pp. 209–214.

[16] L. E. Moser, “Data dependency graphs for Ada programs,” IEEE Transactions on Software Engineering, vol. 16, no. 5, pp. 498-509.

[17] CR. Tsai and V. D. Gligor. “A Bandwidth Computation Model for Covert Storage Channels and its Applications.” Proc. IEEE Conf. on Security and Privacy, 1988.

[18] M. Handley and V. Paxson. “Network Intrusion Detection: Evasion, Traffic Normalization, and End-to-End Protocol Semantics.” Proceedings of the 10th conference on USENIX Security Symposium - Volume 10, 2001.

[19] G. R. Malan et al. “Transport and Application Protocol Scrubbing.” Proc. IEEE Conf. Computer Communications (INFOCOM), Mar. 2000, pp. 1381–1390.

[20] N. B. Lucena et al. “Covert Channels in IPv6.” Proc. Privacy Enhancing Technologies (PET), May 2005, pp. 147–166.

[21] A. Singh et al. “Malicious ICMP Tunneling: Defense against the Vulnerability.” Proc. 8th Australasian Conf. Information Security and Privacy (ACISP), July 2003, pp. 226–235.

[22] N. Schear et al. “Glavlit: Preventing Exfiltration at Wire Speed.” Proc. 5th Wksp. Hot Topics in Networks (HotNets), Nov. 2006.

[23] C. H. Rowland. “Covert Channels in the TCP/IP Protocol Suite.” First Monday, Peer Reviewed Journal on the Internet, July 1997.

[24] daemon9. “LOKI2: The Implementation.” Phrack Magazine, vol. 7, no. 51, Sept. 1997.

[25] S. Cabuk et al. “IP Covert Timing Channels: Design and Detection.” Proc. 11th ACM Conf. Computer and Communications Security (CCS), Oct. 25–29 2004, pp. 178–187.

[26] T. Sohn et al. “Covert Channel Detection in the ICMP Payload Using Support Vector Machine.” Lecture Notes in Computer Science, Volume 2776, 2003, pp. 461-464.

[27] R. P. Murphy. “V00d00n3t – Ipv6 / ICMPv6 Covert Channel.” Slides from DEFCON, 2006.

[28] B. Yuan et al. “A Covert Channel in Packet Switching Data Network.” The Second Upstate New York Workshop on Communications and Networking, Rochester, NY, November 2005.

[29] J. Postel. “Internet Control Message Protocol.” Network Working Group, RFC 792, Sept. 1981, pp. 14.

[30] E. Tumoian and M. Anikeev. "Detecting NUSHU Covert Channels Using Neural Networks." Technical report, Taganrog State University of Radio Engineering, 2005.

# NAT Denial of Service: An Analysis of Translation Table Behavior on Multiple Platforms

**Nathan Winemiller, Bruce Hartpence, Daryl Johnson, and Sumita Mishra**
Networking and Systems Administration Department
Golisano College of Computing and Information Sciences
Rochester Institute of Technology
1 Lomb Memorial Drive
Rochester, NY, USA

**Abstract -** *Network Address Translation or NAT, is a technology that is used to translate internal addresses to globally routable addresses on the internet. NAT continues to be used extensively in almost every network due to the current lack of IPv4 addresses. Despite being exceptionally commonplace, this networking technique is not without its weaknesses, and can be disabled with a fairly straightforward attack. By overpopulating the translation table, the primary mechanism used to translate the internal to external addresses, an attacker can effectively deny all internal users access to the external network. This paper takes an in-depth look at how five different vendors: Cisco, Extreme, Linksys, VMWare, and Vyatta, implement the translation table during active NAT sessions and how they are affected by TCP, UDP, and ICMP variations of the DOS attack.*

**Keywords:** Computer Network Security, Denial of Service, Network Address Translation, NAT

**Track:** Network Security Engineering

## 1 Introduction

Network Address Translation is a technology that is so widely deployed, it can be found in almost every home and in every company that has an enterprise network. With the scarcity of public IPv4 addresses, NAT with port translation has become a necessity when organizations need to provide Internet access to multiple users on the inside of their network. Devices that do the translations keep track of these connections in the NAT translation table. This table maps inside and outside ports and IP addresses so that internal hosts can have multiple concurrent conversations with the outside world. However, this creates a single point of failure in the network. If the translation table becomes too full or non-functional, the internal network could suffer connectivity issues when trying to reach the global network. This makes translation tables a prime candidate for denial of service attacks by a malicious user.

Therefore, it is important to understand how different vendors handle translation table behavior, especially in an enterprise environment which contains devices from multiple vendors. This paper will analyze how several different vendors handle NAT translation table size and how the devices react once the tables have been filled to capacity. Additionally, this paper will determine whether or not a small number of devices on the inside network can easily and effectively deny service to other users on the network by targeting the translation table on these NAT devices.

## 2 Background and Problem Statement

NAT is a networking function that is widely deployed in networks today. The IETF defines NAT as a "method by which IP address are mapped from one realm to another, in an attempt to provide transparent routing to hosts" [3]. NAT with Port Translation or NAPT is the most common and widely deployed version of NAT. The main function of NAPT is the conservation of global IP addresses by mapping a large number of private internal host addresses to a single external host address [1]. Essentially, NAPT provides a way for an intermediate device to map and translate internal IP addresses and ports to an external IP address on a different port.

One of the most important and complicated portions of NAPT has to deal with keeping track of the sessions that are in use. RFC 2663 identifies TCP and UDP sessions by keeping track of the:

- Source IP address
- Source TCP/UDP port
- Target IP address
- Target TCP/UDP port

ICMP tracking is similar except that the NAPT device keeps track of the ICMP query ID instead of TCP/UDP ports [3]. These properties create unique entries in the translation table and can be used to create a virtually limitless number of permutations for use in translation.

### 2.1 The Problem

In order to segment networks from the internet and preserve IP addresses, businesses and households around the world use NAT with Port Translation to map multiple internal hosts to a small number of globally routable public IPs [3]. As far as internal users are concerned, this process is supposed to be transparent, however if this process were to be disrupted, multiple if not all users on the internal network would be affected. The translation table which maps internal ports to external ports for active connections serves as a single point of failure and could be targeted in order to deny service to a large number of users.

Because every conversation that is going from the internal network to the external network has to be mapped and tracked, NAPT is an incredibly processor intensive task when compared to many other services a router could provide. In addition, every packet that goes through the NAT translation has to be rewritten; checksums need to be recalculated, along with many other changes [1]. Consequently, the more translations that occur, the more processor time is consumed to perform the NAT operations.

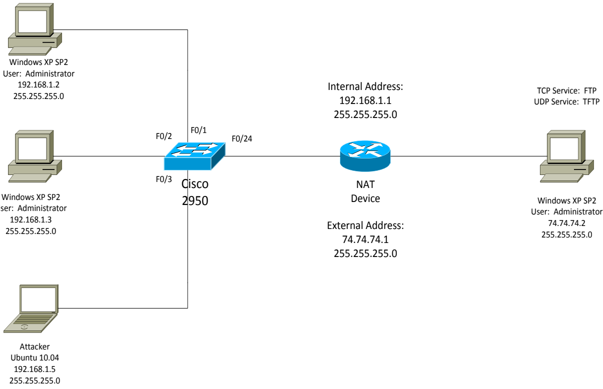
This can lead to performance degradation for all functions on the router if the NAT process consumes the available processing resources. As a result, if an attacker could create a situation where the NAT has so much work to do that it consumes all of the resources, they could cause failures not only for the NAT process, but also other functions that the router is supposed to handle (access lists, routing protocols, DHCP leases, etc.).

While there are many specifications on how NAT with Port Translation should operate, many of the implementation choices are not explicitly defined. Important details such as table entry timeouts for the different protocols, maximum translation table size, and entry tracking are left up to the vendors [3]. If the implementation discrepancies between vendors are significant enough, they could result in significant impact to network performance and application functionality. The next sections will provide scenarios that will demonstrate these issues and will provide data we can use to determine the differences in how each vendor implements NAPT.

# 3 Experimental Design

For all tests involving physical devices, the topology in figure 1 was constructed.

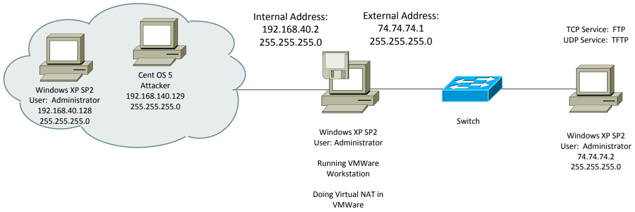
Figure 1: Physical Device Topology



In this diagram there are two networks, one internal and one external. There are two normal hosts on the internal network and an attacker that will launch an attack against the NAT device in the middle. The outside host is there to receive traffic and to provide an end point to determine if a loss of connectivity occurs. The main questions here are: can a single attacker device overwhelm the NAT device with translations thereby blocking legitimate traffic from exiting? What happens when the translation table on the NAT device becomes full?

For all tests involving the VMWare Virtual NAT Process, the topology in figure 2 was constructed.

Figure 2: Virtual Topology



Like the physical topology there is an internal and an external network that the attacker will attempt to disrupt.
Our hypothesis is that the attacker will be able to deny access to the external network by targeting the translation device. By creating a large number of translations, the attacker would be able to deny access to the outside by either filling the translation table or by consuming all of the available resources on the translating device.

## 3.1 Default Device Timers

Since the RFC documents regarding NAT only give recommendations on what the expiration timers should be for TCP, UDP, and ICMP entries, the individual vendors have their own default settings for each of these entries. These values are important because they determine the amount of time an attacker has to fill the table before the entries start expiring.

| | TCP | Syn | Fin/RST | UDP | ICMP |
|---|---|---|---|---|---|
| Cisco | 24 | 60 | 60 secs | 5 mins | 60 |

| | | | | | |
|---|---|---|---|---|---|
| | hours | secs | | | secs |
| Extreme | 2 mins | 60 secs | 60 secs | 2 mins | 3 secs |
| Linksys | N/A | N/A | N/A | N/A | N/A |
| Vyatta | 12 hours | N/A | N/A | 30 secs | 30 secs |
| VMWare | N/A | N/A | N/A | 30 secs | N/A |

**Notes:**

1. TCP Default refers to all translated TCP ports that are not specified in the Cisco IOS (DNS and a few other services have their own special timeouts).

2. Timers on the Linksys device are not configurable by the user.

3. The vmnetnat.conf configuration file for the VMWare NAT process does not have a configurable field for TCP timeout. Additionally, there is no way to view the translation table for diagnostic testing to find out the default timeout.

4. Timers on the Vyatta device are not configurable by the user.

### 3.2 Attack Types

In order to determine the effects of the different protocols on multiple platforms, we had to create a number of scenarios. We tested attacks using TCP, UDP, and ICMP entries into the table. Each of these three tests were done against the five different vendors: Cisco, Extreme, Vyatta, Linksys, and VMWare.

If the attacker proves unsuccessful in filling the translation table for a device using the default timers, the experiments will be repeated for that device using unlimited timeout periods for each protocol. This will allow us to determine the effects of the attack on the device if it were to be successful.

### 3.3 Attack Scripts

The scripts used to execute the attacks were simple bash scripts that called several concurrent sessions of NMAP. These NMAP scans would attempt to reach various addresses on the outside on a variety of different ports. This allows the attacker to generate numerous translation table entries in a very short period of time. The scripts were very similar, with the main difference being in the –s options that changed the protocol being used by NMAP.

Table 1: TCP Attack Script

```
#!/bin/bash
for ((i=0; i<300; i++))
do
        sudo nmap -sS 74.74.74.3-254 1>/dev/null
2/dev/null &
done
```

Table 2: UDP Attack Script

```
#!/bin/bash
for ((i=0; i<300; i++))
do
        sudo nmap -sU 74.74.74.3-254 1>/dev/null
2/dev/null &
done
```

There was one alteration that had to be made during the UDP experiment. The Cisco device would not register the UDP entries in the translation table unless it received some sort of reply from the outside on those ports. Therefore, the UDP Attack Script in the Cisco scenario pointed to the outside host instead of a range of non-existent hosts.

Table 3: Cisco UDP Attack Script

```
#!/bin/bash
for ((i=0; i<300; i++))
do
        sudo nmap -sU 74.74.74.2 1>/dev/null
2/dev/null &
done
```

Table 4: ICMP Attack Script

```
#!/bin/bash
for ((i=0; i<300; i++))
do
        sudo nmap -sP 74.74.74.3-254 1>/dev/null
2/dev/null &
done
```

## 4 Methodology

The topology for each test was set up as shown in the figures above with three PCs connected to a switch behind a NAT device which then connected to an external host on the external network. The only exception was the VMWare test where the topology consisted of two VMs on the internal virtual network and one physical host on the external network. The external host was hosting an FTP server and a TFTP server which had access one large movie file to transfer. Only one NAT device from one vendor would be tested every time. Each vendor device would run through the test 5 times to ensure the validity of the data.

Before the attack began, all of the hosts issued pings to each other to verify connectivity. This also allowed for verification that the NAT translations were occurring correctly between the inside network and the external network. After initial connectivity and translation functionality were established, an internal host tested TCP and UDP functionality by using FTP and TFTP between the inside and outside. After full functionality was determined, a unidirectional JPerf test was run between an inside host and the outside host. This ensured that the NAT device could not be brought down by a full load of traffic from a single device. After all of the testing had finished, commands would be issued to the NAT device to clear the translation table and reset the NAT statistics when applicable.

After the verification process, the attacker would kick off the script and a timer (on a stopwatch) would be started. Each attack period would last for 5 minutes and then the NMAP processes would be killed. Every minute during the attack, internal hosts would attempt to ping the internal gateway, the external gateway, and the outside host. Additionally, every minute, the internal hosts would attempt to establish an FTP and a TFTP session and attempt to transfer a file. Also, every minute the relevant commands would be issued to the NAT device (when applicable) to determine the number of translations and to verify that the correct translations were being put into the translation table.

After the attack had completed, internal hosts would then attempt to ping each other, the internal gateway, the external gateway, and the external host. Additionally, internal hosts would then attempt to establish FTP and TFTP sessions with the external hosts in order to transfer the movie file. This process would be repeated until all attempts were successful or until 15 minutes elapsed.

These tests would be run initially using the default timers on the devices. However if the attacker could not fill the translation table for a particular device with any of the attacks, then tests would be run on that specific device using unlimited timers. This would allow us to observe the consequences of filling the translation table on that device.

### 4.1 Problems Encountered

During the course of our experiments, we did run into a few issues during the testing. Unfortunately some of the devices had very little or no visibility into what was happening during the translation process. The Linksys device was the worst for this since it did not have any relevant means of logging NAT Translations. This problem was exacerbated by the fact that all management for the Linksys device was in-band, which became unreachable during the attack. The VMWare NAT process also had no means to debug the translations while they were occurring. This meant that all we had to go on were the results of the experiments to determine what was actually going on during the attack.

Another issue was the processing limitations of the devices used to execute the attacks. In most cases it was a netbook that had very little processing power, which limited the number of translations we could do within a certain period of time. In the other cases it was a virtual machine with very limited resources (in the VMWare tests). We originally wrote the scripts to loop 1000 times to maximize the number of IMAP instances running, but ultimately had to scale it down to 300 concurrent sessions to keep the attacking machines from locking up.

### 4.2 Metrics

In order to determine whether connectivity was affected or not, every minute of the attack we would run: ICMP echo requests to the outside, a TFTP transfer to the outside, and an FTP transfer to the outside. We would also look at the current size of the translation table along with the CPU utilization on the device in question. The final thing we would watch for was whether or not the packets the outside device was receiving were being translated (by looking at the source address of the traffic).

## 5 Results and Discussion

The results of this experiment show that a small number of compromised nodes with the right software can severely impact enterprise grade equipment in a short period of time. While almost all platforms tested were negatively impacted by the TCP and UDP attacks, the effectiveness of certain protocols for filling the translation tables varied from vendor to vendor. Additionally, filling the translation tables also had varied effects on the devices performing the NAT translations depending on what vendor created the device.

On the Cisco device, the translations filled the available memory on the device which not only prevented new entries from being entered into the table (and therefore being translated) it also caused memory allocation errors which negatively impacted other processes that the router was performing. The Cisco device fared better against the TCP attacks due to its session tracking capability and the aggressive timeouts on the SYN packets. The UDP attack on the other hand was able to fill the table on this device and deny translation capabilities to all other devices on the network.

The Extreme device on the other hand, limited the number of NAT translations allowed on one address instead of letting it consume all of the other memory. This allowed for smaller number of total entries, but preserved the integrity of the device when the table was filled. The Extreme device fared very well against both TCP and UDP attacks due to its very aggressive default timers. However, these timers are a double edged sword and by defying best practices and standards laid down in the RFCs, could cause problems on a network with higher latency times. For example, in the case of the Extreme device, the two minute default TCP timer could cause certain TCP applications to fail because it expires so quickly. TCP logical connections can last for a very long time even if connectivity is temporarily disrupted or there isn't any traffic to send for a period of time. By expiring these TCP entries, the internal host will have to establish a new session to the outside and most likely start the conversation over.

The Linksys device by far fared the worst of all the devices that were tested. It was unable to handle the processing load of translating such a large amount of requests and therefore became completely unreachable during the attack. TCP and

UDP attacks were equally effective against this device. Unfortunately, the lack of accessibility during the attack (in-band management only) and limited logging made any more insight into the operation of this device all but impossible.

While the VMWare device fared well against the UDP attack (unlike the Cisco device) it was severely impacted by the TCP attack. Service was denied during the attack and unlike the rest of the devices tested, TCP and UDP connections could no longer be established even after the attack had ended. The only way to restore full connectivity was to restart the VMWare NAT service. This would pose a significant problem to network administrators since outside connectivity would appear to be restored (ICMP echo requests would work), but any service related applications would cease to function. Unfortunately, the VMWare NAT process has limited configuration and debug capabilities so determining the exact cause of this disruption is also next to impossible.

The Vyatta device fared extremely well during these tests. These attacks didn't appear to have any impact despite having several thousand translations entered into the table. Due to the resources at my disposal, the only device to test this platform was markedly more powerful than the other devices tested in this series of experiments. Interestingly enough, Vyatta also defies RFC recommendations by not allowing for the administrator to adjust any of the timers that are used in the translation process. Additionally, while I was unable to do so in my experiments, the default TCP timer is extremely long and the Vyatta device doesn't appear to do any session tracking (unlike the Extreme or Cisco devices). This could possibly be exploited on a device that is less powerful or being actively utilized on an enterprise network.

It is interesting to note however, that unlike when the CAM (Content Addressable Memory) table, which holds the MAC address / port entries, becomes full on a switch, the NAT device does not "fail open" and pass all traffic. When the translation table became full (in all cases where this was possible), no new translations could occur and the NAT devices would drop any new traffic not already in the table. Another point to note is that the NAT devices did not FIFO any translation entries. While expired entries were FIFO'd out, if the table became full, the entries remained in the table until they expired. Therefore, new entries could not be added until the old entries expired (which could be a significant amount of time).

Another interesting conclusion that can be drawn from these experiments is that the differing implementations of NAT between the vendors (especially in the case of timers) could cause interoperability issues if a connection has to traverse a NAT device on the source end and the destination end. While TCP connections are generally given a duly long amount of time before expiration, UDP connection timeouts varied greatly between the vendors. With VMWare and Vyatta allowing only 30 seconds before the UDP entry times out, significant application issues could occur over high latency networks that span long distances.

The relatively small number of compromised nodes required to execute the attack paired with the fact that NAT generally causes a single point of failure for a large number of users makes this a relatively easy and effective type of denial of service attack to execute. Furthermore, while it may be difficult to fill up a translation table on a device with a significant amount of memory, the effects of the attack on the processing capabilities of the NAT devices makes this attack scalable to almost any network size. Therefore, we can conclude that this attack is viable in an enterprise setting and against several mainstream vendor devices that are currently deployed.

## 5.1 Mitigation

The main ways that the NAT denial of service attack takes advantage of the NAT device is to either fill up the translation table to prevent new entries or to create so many entries that the processor on the device cannot keep up. Therefore the most obvious way to mitigate the effects of this attack is to limit the number of translations that hosts are allowed to make. Making a rule that limits the number of overall translations allowed is common, and while this may solve the issue of high CPU utilization, this technique makes it easier for the attacker to fill the table and deny service to legitimate users service. Because of this, limiting the number of translations must be applicable to individual hosts.

Cisco has features in place in their IOS to accomplish this goal relatively well. Most platforms support rate limiting of some kind, but the Cisco platform supports rate limiting by either the total number of translations, the number of translations that can be created by a list (that includes a number of hosts), and the number of translations a specific host can make [6]. Obviously limiting the total number of translations would prove counter-productive, and while limiting the number of translations a specified host can use is fine on a smaller network, this mitigation technique cannot scale to a large network.

The Extreme device on the other hand supports what the vendor calls "Auto-constraining," which limits the amount of ports a single internal host can use at one time [7]. This works by evenly distributing the amount of port space between each internal user evenly [7]. While this is easy to configure and can be effective in smaller networks, this feature could end up preventing legitimate users from making new connections in a scenario where there are a large number of internal users and a very small number of external addresses to map to.

Unfortunately, the Linksys, VMWare, and Vyatta platforms don't appear to support any of these types of features at all which makes stopping this type of attack much more difficult when using these platforms.

Another feature that is built into most NAT platforms is the ability to control the amount of time that passes before an entry in the table expires. This is best evidenced by the table on the Extreme device that couldn't be filled in this experiment using the normal timers. Unfortunately the Vyatta and Linksys platforms do not allow the administrator to configure these timers, which makes them less able to prevent the NAT DOS attack. However, administrators need to be careful when setting these timers since they could time out legitimate sessions prematurely and cause applications to malfunction, especially in higher latency networks with a high traffic volume.

## 6 Conclusions

The purpose of this paper is to closely examine infrastructure security risks when using NAT with Port Translation. Additionally, this paper analyzes the default behavior of NAPT devices from different vendors. The results of these experiments show that a single device on the internal network is capable of overwhelming the translation table. These experiments also show that the default behavior when implementing NAPT varies significantly between vendors. Default timers, entry behavior, and configurable settings are all vendor specific. We believe that the almost universal use of NAPT justifies further investigation and standardization of its use. The reality is that most consumers utilize NAPT and do not consider that the very mechanisms that allow it to work can be taken advantage of to the detriment of all internal hosts. As the various experiments outlined in the paper show, many vendors use default behaviors that leave their devices open to exploitation when implementing NAPT. Furthermore, some vendors do not even offer the ability to change crucial settings that could be used to mitigate this type of attack.

# 7 References

[1] Smith, M.; Hunt, R.; , "Network security using NAT and NAPT," Networks, 2002. ICON 2002. 10th IEEE International Conference on , vol., no., pp. 355- 360, 2002 doi: 10.1109/ICON.2002.1033337. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1033337&isnumber=22194

[2] Hartpence, B.; Johnson, D.;, "A Re-examination of Network Address Translation Security," RIT Network Security and System Administration. SAM 2010.

[3] Srisuresh, P, & Holdrege, M. (1999, August). Rfc: 2663 ip network address translator terminology and considerations. Retrieved from http://tools.ietf.org/html/rfc2663

[4] Hain, T. (2000, November). Rfc 2993: architectural implications of nat. Retrieved from http://www.ietf.org/rfc/rfc2993.txt

[5] Srisuresh, P, & Egevang, K. (2001, January). RFC 3022: traditional nat. Retrieved from http://www.ietf.org/rfc/rfc3022.txt

[6] Cisco Systems Incorporated. (2003). Rate Limiting NAT Translation. Retrieved from http://www.cisco.com/en/US/docs/ios/12_3t/12_3t4/feature/guide/gt_natrl.html

[7] Extreme Networks Incorporated. (2003, June). Summit 200 Series Switch Installation and User Guide. Retrieved from http://www.extremenetworks.com

[8] Lebovitz, G.; Thaler, D.; Zhang, L. (2010, July). RFC 5902: IAB Thoughts on IPv6 Network Address Translation. Retrieved from http://tools.ietf.org/html/rfc5902#section-1

# Toward a More Perfect Scoring System for the NECCDC

George Markowsky
School of Computing and Information Science
University of Maine, Orono, Maine, USA

**Abstract –** *The NECCDC, and its parent contest the CCDC, are tremendously challenging and motivating cyber defense exercises for college level students. We feel that their usefulness as a teaching tool is limited by defects in the scoring system. We discuss the attributes that make a scoring system useful from an educational perspective and consider the NECCDC scoring system from this perspective. This leads us to suggest improvements in the NECCDC scoring system which we hope to implement for the 2013 NECCDC.*

**Keywords:** NECCDC, Scoring System, Cybercompetition, CCDC, Educational Competitions

## 1 Introduction

The CCDC (Collegiate Cyber Defense Competition) [1] and its various regional contests such as the NECCDC (Northeast Collegiate Cyber Defense Competition) [2] are wonderful events that greatly motivate students to study cybersecurity. We believe that the educational value of the NECCDC would be greatly enhanced by development of a better scoring system.

There is no doubt that humans enjoy competition. There is even evidence of athletic competitions going back to ancient Egypt [3] and ancient China [4]. Certainly, the Olympic Games [5], which date to 776 BC, are well known to everyone in the world. One feature of most competitions is the keeping of some sort of score that is used to declare the winner. We contend that a good scoring system can contribute to the excitement of a competition. In the case of competitions staged for educational benefit, a good scoring system can greatly enhance the educational benefit of the competition.

## 2 Criteria for a Good Scoring System

To be useful, scoring systems must have the following general properties:

1. The scoring system must be understood by all participants.

2. The scoring system must be real time. In particular, scores must be available as the competition is proceeding and must be timely enough so that competitors can respond to the scores while the competition is in progress.

3. Scores must suggest strategies and tactics at various points in the competition.

For education events, such as the NECCDC, the scoring system must have at least the following additional properties:

4. The scoring system must suggest to the participants what they need to learn in order to be successful in the competition.

5. The scoring system must not embarrass participants.

6. The scoring system must provide a balanced view of the competition.

Unfortunately, in our opinion the current scoring system as evidenced by the 2012 NECCDC receives a failing grade on all six criteria. We will describe the current system and suggest how the system can be improved for the 2013 NECCDC.

## 3 System is not Clear

This has been a problem in the CCDC and the NECCDC which basically copied the policy of the CCDC. Here is a complete description of the scoring system from Section 9 of the CCDC Team Preparation Guide [6]

*9. Scoring*

*a. Scoring will be based on keeping required services up, controlling/preventing unauthorized access, and completing business tasks that will be provided throughout the competition. Teams accumulate points by successfully completing injects and maintaining services. Teams lose points by violating service level agreements, usage of recovery services, and successful penetrations by the Red Team.*

*b. Scores will be maintained by the competition officials and may be shared at the end of the competition. There will be no running totals provided during the competition. Team rankings may be provided at the beginning of each competition day.*

*c. Any team action that interrupts the scoring system is exclusively the responsibility of that team and will result in a lower score. Should any question arise about scoring, the scoring engine, or how they function, the Team Captain should immediately contact the competition officials to address the issue.*

*d. Teams are strongly encouraged to provide incident reports for each Red Team incident they detect. Incident reports can be completed as needed throughout the competition and presented to the White Team for collection. Incident reports must contain a description of what occurred (including source and destination IP addresses, timelines of activity, passwords cracked, access obtained, damage done, etc), a discussion of what was affected, and a remediation plan. A thorough incident report that correctly identifies and addresses a successful Red Team attack may reduce the Red Team penalty for that event – no partial points will be given for incomplete or vague incident reports.*

No point assignments for anything are indicated in the section on scoring. There is no indication of what a good strategy might be. Notice that (b) explicitly disallows real-time scoring. Certainly, from this description is it not clear how the scoring system works.

The NECCDC is unusual for a competition because coaches are not permitted to offer technical or strategic advice during the three day competition. This means that coaches do not have any independent information for understanding the score report.

## 4 System is not Real-Time

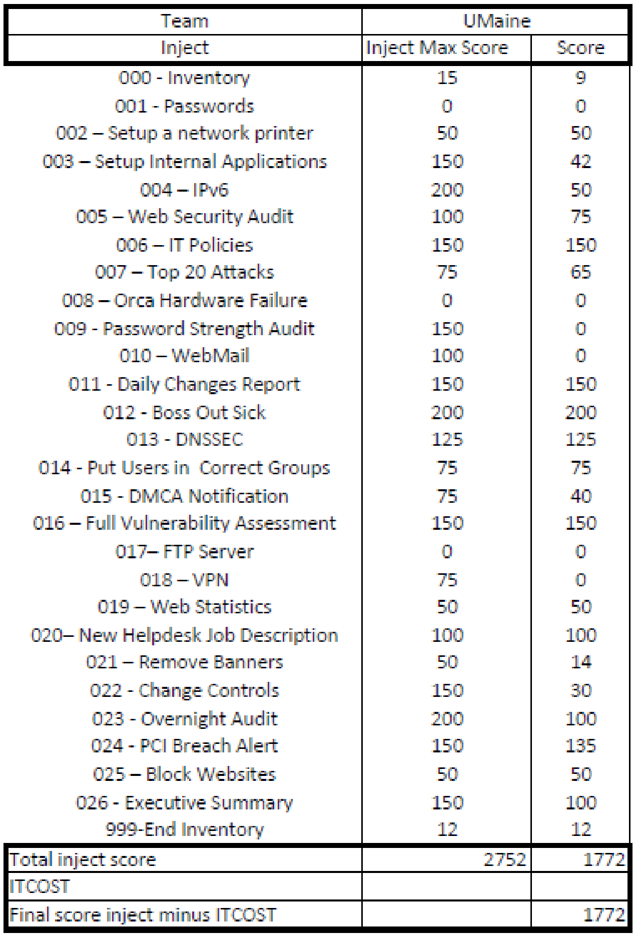
As noted in Section 3, the NECCDC scoring system is not real-time. In fact, it is not even allowed to be real-time. At most, organizers may share standings at the beginning of every day. At the 2012 NECCDC only the top three schools were mentioned every day.

## 5 System in not Useful in Competition

As you can see the NECCDC scoring system is not useful during the competition since the teams have no idea of how they are doing. They certainly have no basis for judging tradeoffs. Even in real life, people often know what their tradeoffs are, and base plans on that knowledge.

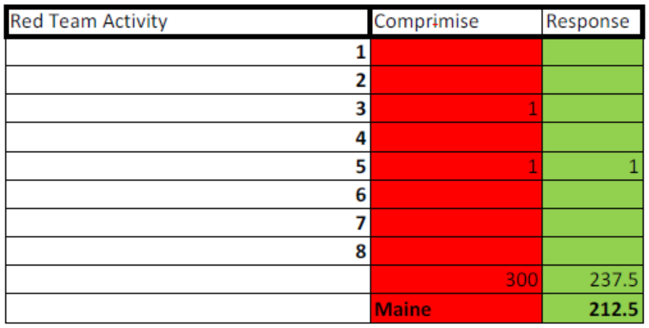
## 6 System Does not Guide Learning

After the competition, our team got its score report which is reproduced in Figures 1, 2 and 3. We should note that no other information was supplied with the grade report.



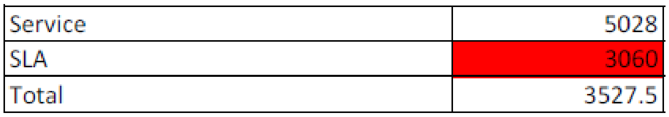
| Team | UMaine | |
|---|---|---|
| Inject | Inject Max Score | Score |
| 000 - Inventory | 15 | 9 |
| 001 - Passwords | 0 | 0 |
| 002 – Setup a network printer | 50 | 50 |
| 003 – Setup Internal Applications | 150 | 42 |
| 004 – IPv6 | 200 | 50 |
| 005 – Web Security Audit | 100 | 75 |
| 006 – IT Policies | 150 | 150 |
| 007 – Top 20 Attacks | 75 | 65 |
| 008 – Orca Hardware Failure | 0 | 0 |
| 009 - Password Strength Audit | 150 | 0 |
| 010 – WebMail | 100 | 0 |
| 011 - Daily Changes Report | 150 | 150 |
| 012 - Boss Out Sick | 200 | 200 |
| 013 - DNSSEC | 125 | 125 |
| 014 - Put Users in Correct Groups | 75 | 75 |
| 015 - DMCA Notification | 75 | 40 |
| 016 – Full Vulnerability Assessment | 150 | 150 |
| 017– FTP Server | 0 | 0 |
| 018 – VPN | 75 | 0 |
| 019 – Web Statistics | 50 | 50 |
| 020– New Helpdesk Job Description | 100 | 100 |
| 021 – Remove Banners | 50 | 14 |
| 022 - Change Controls | 150 | 30 |
| 023 - Overnight Audit | 200 | 100 |
| 024 - PCI Breach Alert | 150 | 135 |
| 025 – Block Websites | 50 | 50 |
| 026 - Executive Summary | 150 | 100 |
| 999-End Inventory | 12 | 12 |
| Total inject score | 2752 | 1772 |
| ITCOST | | |
| Final score inject minus ITCOST | | 1772 |

**Figure 1. Part I of the Score Report**



| Red Team Activity | Comprimise | Response |
|---|---|---|
| 1 | | |
| 2 | | |
| 3 | 1 | |
| 4 | | |
| 5 | 1 | 1 |
| 6 | | |
| 7 | | |
| 8 | | |
| | 300 | 237.5 |
| | Maine | **212.5** |

**Figure 2. Part II of the Scoring Report**



| | |
|---|---|
| Service | 5028 |
| SLA | 3060 |
| Total | 3527.5 |

**Figure 3. Part III of the Scoring Report**

Since coaches are kept at arm's length during the competition we had little insight into what the scores meant. Figure 4 shows a questionnaire and responses to it. The questionnaire was circulated to the members of the team that competed in the NECCDC this past March.

---

Attached is the scoresheet that we got from the NECCDC. I am writing an article about how the competition should be scored and I was hoping that some of you could help me with interpreting the scoresheet. If any of you can answer any of the questions below that would be great.

1. I noticed that some of the injects were worth 0 points: 001, 008, and 017. Does that mean that they were not worth anything or were they omitted or does it mean something else? Did you know they were worth 0 points during the competition?

*I believe 008 was when they took out the DNS box for "Hardware Failure," but I thought the box removed was Blue (could be a simple mistake). In that case, there was no option to do anything but give them the box. The FTP server may have been worth 0 points because it was scored as a service after it was implemented.*

2. There were some items that the team got 0 points on, but which were worth some points such as 009, 010 and 018. Did the team not do these or was there some other reason the team got 0? Did you not do them because of time pressure or was there some other factor involved.

*All injects where we got 0 points I believe were injects that we didn't complete enough to implement. I know that WebMail wasn't implemented at all because we wouldn't have had time to secure the new webmail client properly in the required time frame, and I believe the same was true for the other two.*

3. Is it clear to you where you lost points on the injects on which you lost points? Does the scoresheet provide enough information so you could prepare for the contest better next year?

*I would say it definitely does not provide enough information in this area. For example, on inject 021, I know everyone thought they successfully removed/changed the banners in minutes, yet we only got 14 out of 50 points. This current score sheet gives no information on where points were lost.*

4. Do you know what the ITCOST item refers to? Does it refer to reimaging or other IT services?

*I don't know what ITCOST refers to. We didn't have any boxes reimaged during this year's competition, so we wouldn't have lost any points there anyways.*

5. Do you understand the Read Team Activity scoring? How did Maine end up with a score of 212.5 based on the numbers displayed there? I am not arguing about the score, just trying to figure out how it was computed? Do you have any idea?

*Sorry, no idea on that end. My only guess is that it's based on how many points we got back for detecting the attack, but I have no idea how they determined this score.*

6. It looks like the final score is the sum of Service and Inject Score - (SLA and Red Team scores). Does SLA stand for Service Level Agreement or something else? Did you have any idea during the competition that you were losing or had lost 3060 points or that the service was worth 5028 points?

*While I don't know what exactly it stands for off of the top of my head, the SLA score is from when all services are down. This happened to us when we unplugged the router at the start of the competition while we locked down boxes. We knew we were losing a good deal of points, but we figured it would be made up for by the added security we were putting in place. What we didn't know, however, was that the red team was down for the first 30 mins or so anyways. So while we were playing it safe, we could have left everything up, been safe from attacks, and actually gaining points. It had been our plan from the start to unplug right away, but no one could anticipate the red team would be down. Had they been up from the start, I feel that the SLA penalty would be justified by the number of beginning attacks that were stopped.*

If you have any other comments about the usefulness of the scoresheet I would be happy to receive them.

*One thing I was concerned with was the red team scoring. They said we were compromised once and that we responded to it. I believe this was in error: the one time they thought they compromised us was when we downloaded the Windows Server image that they backdoored, but we never installed it. We had just been humoring them to distract them from future attacks.*

*I just remember in the red team briefings they mentioned compromising a team with a backdoored Windows installation, which is what makes me think they mistakenly thought they got us when they didn't. We did write an incident report about them giving us the infected server image though.*

---

**Figure 4. A Questionnaire Circulated to the Team**

It is interesting to note that on the basis of the information we have, we cannot tell whether the student's conjecture in the final paragraph is correct or not.

## 7 System Embarrasses Schools

After not revealing scores during the competition, the information in Figure 5 was posted on Twitter. We have replaced some of the school names by XXX.

| | | |
|---|---|---|
| 1 | RIT | ( 6,956) |
| 2 | UNH | ( 4,184) |
| 3 | UM | ( 3,515) |
| 4 | XXX | ( 2,906) |
| 5 | XXX | ( 2,486) |
| 6 | XXX | ( 1,665) |
| 7 | XXX | ( 1,523) |
| 8 | XXX | (-2,238) |
| 9 | XXX | (-3,520) |
| 10 | XXX | (-3,583) |
| 11 | XXX | (-5,601) |
| 12 | XXX | (-6,070) |

**Figure 5. The Scores that Went to Twitter**

While we believe in the full disclosure of scores in a contest, we also believe that the scores should not embarrass the participants. In particular, we feel that it is difficult for most people to interpret a negative score correctly. The negative scores result because teams are penalized per item in some instances. Thus, if there is a data breach, a team might lose a certain number of points for each credit card number stolen. If a large database is broken into, a very large number of points could be lost negating a fine performance in other areas.

In general, most people think that a score of 0 is the worst possible score and that if a team gets 0 points it is because they did not do anything worthwhile. It is easy to imagine that many people would think that a negative score might reflect some egregiously bad performance.

Figure 5 shows that of the 12 competitors, 5 received negative scores. We feel that competitions should not have negative scores. We will discuss later how to address the problem of not having negative scores and yet allow a penalty to be applied an unbounded number of times.

Finally, it is not immediately obvious how the final score is computed from the scores presented in Figures 1-3. It appears that the score is computed using the following formula:

Score = (Service + Injects) – (SLA + Red Team Activity)

= (5028 + 1772) – (3060 + 212.5) = 3527.5

We note that the value for UMaine in Figure 5 is slightly different from the value Figure 3.

## 8 System Lacks Balance

From Figure 5 and the scoring formula presented in Section 7, it appears that negative scores can only result from Red Team activity. It appears that the worst one can do from the injects and services is 0. There are upper bounds on the inject portion of the score (2752) and the services part of the score (5028). The Red Team activity seems to be unbounded in the negative direction.

It appears that the scoring algorithm must subtract a certain number of points for each instance of a failure. For example, if a database is compromised, a number of points are deducted for each credit card number stolen. If a small database is compromised, you might lose many fewer points than for a large database. The same error might be involved, but the number of points might be very different.

## 9 A Loophole

This analysis suggests a strategy for placing in the middle of the pack at the NECCDC – don't do anything for the entire competition. If you leave your computers powered down for the entire competition, as best we can tell, you will get a score of 0. That would be better than the score achieved by 5 of the 12 teams who competed. We submit that there is a problem with a scoring system that would place a team that did absolutely nothing and demonstrated no knowledge of cybersecurity above 5 teams that put in 3 grueling days defending their systems.

## 10 Recommendations

The University of Maine will be hosting the 2013 NECCDC. In putting together the competition we hope to address some of the issues discussed in this paper. We expect that this would require some negotiation with the CCDC. We would like to establish the following principles for the 2013 NECCDC.

1. Establish a clear grading system that can be explained to all participants before the competition begins.

2. Have a valuable real-time score display that informs participants of how they are doing so they can make adjustments during the competition.

3. Break the score into a number of categories. Some possibilities are given in Figure 6. Select top finishers in the various categories to recognize the hard work that so many students have put in. Categories will also help teams better understand their strengths and weaknesses and will aid coaches in preparing their teams.

| | | |
|---|---|---|
| Service Uptime | Inject Report Writing | Professionalism |
| Incident Response | Incident Reports | E-mail Services |
| Networking Skill | DNS Services | Web Services |
| Firewalling | Intrusion Detection | Forensics |

**Figure 6. Some Possible Scoring Categories**

4. Sum scores from the various categories to determine the overall winner. The scores can be weighted to reflect relative importance.

5. There will be no negative scores in any category.

## 11 Sample Report

We hope that teams can receive reports that look like the following. Of course, the final form of the score report should not put too much additional effort on the Red, White and Black teams.

**Red Team Attack**
Description
Point Value
Blue Team Response
Points Lost
Improvements

• • •
**DNS Services**
Point Value
Points Earned
Issues
Improvements

## 12 Dealing with Unbounded Penalties

There are reasons why you might want to have penalties that are worse as long as some parameter keeps increasing. In other words if one team loses a 1000 credit card numbers and another team loses 1001 credit card numbers, the penalty for the second team should be strictly greater than the penalty for the first team. At the same time just a simple linear model will lead to the loophole described in Section 9. We call the formula shown in Figure 7 the penalty bounding formula. It has the property that the score it produces is always between 0 and K as long as the Penalty is between 0 and ∞, yet as the penalty increases the score decreases.

$$Score = K - \frac{2 \times K \times \tan^{-1}(Penalty)}{\pi}$$

**Figure 7. The Penalty Bounding Formula**

Figure 8 shows the values produced by the Penalty Bounding Formula for K = 100.

```
Penalty =   0 Adj Penalty = 100.000
Penalty =   1 Adj Penalty =  50.000
Penalty =   2 Adj Penalty =  29.517
Penalty =   3 Adj Penalty =  20.483
Penalty =   4 Adj Penalty =  15.596
Penalty =   5 Adj Penalty =  12.567
Penalty =   6 Adj Penalty =  10.514
Penalty =   7 Adj Penalty =   9.033
Penalty =   8 Adj Penalty =   7.917
Penalty =   9 Adj Penalty =   7.045
Penalty =  10 Adj Penalty =   6.345
Penalty =  11 Adj Penalty =   5.772
Penalty =  12 Adj Penalty =   5.293
Penalty =  13 Adj Penalty =   4.887
Penalty =  14 Adj Penalty =   4.540
Penalty =  15 Adj Penalty =   4.238
Penalty =  16 Adj Penalty =   3.974
Penalty =  17 Adj Penalty =   3.741
Penalty =  18 Adj Penalty =   3.533
Penalty =  19 Adj Penalty =   3.348
Penalty =  20 Adj Penalty =   3.180
Penalty =  30 Adj Penalty =   2.121
Penalty =  40 Adj Penalty =   1.591
Penalty =  50 Adj Penalty =   1.273
Penalty =  60 Adj Penalty =   1.061
Penalty =  70 Adj Penalty =   0.909
Penalty =  80 Adj Penalty =   0.796
Penalty =  90 Adj Penalty =   0.707
Penalty = 100 Adj Penalty =   0.637
Penalty = 200 Adj Penalty =   0.318
Penalty = 300 Adj Penalty =   0.212
Penalty = 400 Adj Penalty =   0.159
Penalty = 500 Adj Penalty =   0.127
Penalty = 600 Adj Penalty =   0.106
Penalty = 700 Adj Penalty =   0.091
Penalty = 800 Adj Penalty =   0.080
Penalty = 900 Adj Penalty =   0.071
```

**Figure 8. Values of the Penalty Bounding Formula**

It seems reasonable to us that penalties be bounded especially from an educational perspective. After a team loses 100 credit card numbers, it will get the point. Killing its performance because it lost another 900 credit card numbers adds little educational value to the exercise.

## 13 Final Thoughts

We believe that a good scoring system will make the NECCDC more interesting for both students and spectators. It will also provide teams with information that will help them to prepare better for cyber competitions, and to better learn how to study cybersecurity. It will also help coaches work with their teams to understand cybersecurity better. With a variety of categories and prizes, a good scoring system can reward those people who worked hard.

## 14 References

[1] www.nationalccdc.org/. Website of the National CCDC.

[2] The URLs for the three most recent NECCDC competitions: www.ccs.neu.edu/neccdc2012; www.ccs.neu.edu/neccdc2011; neccdc2010.umcs.maine.edu .

[3] Wolfgang Decker, *Sports and Games of Ancient Egypt*, Yale University Press, 1992.

[4] Steve Craig, *Sports and Games of the Ancients*, Greenwood Publishing, 2002.

[5] Judith Swadding, *The Ancient Olympic Games: Second Edition*, The University of Texas Press, 2008.

[6] The CCDC Team Preparation Guide is available at http://nationalccdc.org/files/CCDC%20Team%20Prep%20Guide.pdf.

# Browser Web Storage Vulnerability Investigation

## HTML5 localStorage Object

**Daniel Bogaard**[1], **Daryl Johnson**[2], **and Robert Parody**[3]
[1]Information Technology, Rochester Institute of Technology, Rochester, NY, USA
[2]NSSA, Rochester Institute of Technology, Rochester, NY, USA
[3]CQAS, Rochester Institute of Technology, Rochester, NY, USA

**Abstract -** *Along with the introduction of HTML5 a new data storage technique, Web Storage, has been added to browsers. This technique stores larger amounts of data for an extended period of time on a client system. This technology does not (as of this writing) have a fully implemented interface to support end user control.*

*The authors interest is modeling the use of Web Storage to store illicit data. The authors built a web application that would take a file, encrypt it, split it into multiple parts and distribute it to as many clients as possible. At a later time, the system could then watch for return visits and retrieve data parts as clients interact with a host website. The recidivism rate of clients returning to the host website and the number of copies of each distributed part needed to achieve a reliable recovery rate of the entire file are under study.*

**Keywords:** browser security, Web Storage, evasion, forensics, obfuscation

## 1 Introduction

Suppose a nefarious user has a file of incriminating material (credit card number, account number, username/password or Personally Identifiable Information, drug client list…) that the user does not want to be apprehended with but needs access to from time to time. The users goal would be to store the file somewhere that can be reliably retrieved but does not reside locally (for very long) and is not usable or discernable for what it is if found where stored.

The authors propose a solution – Web Storage or localStorage. If the nefarious user has access to a domain (simple Internet Service Provider will suffice) they could hide parts of any incriminating file on various client systems without keeping a local copy that he/she might be caught with. At a later time, when the information is needed, the user could get the parts back from the clients and reconstitute the original data.

To explore this scenario, the authors have split the experiment into 3 parts. The first part (testing phase) of this study has been completed. A web application was built that proves the hypothesis that localStorage can be used for such a purpose. The second part of this study is to install the application on a working production site and statistically determine how many copies of the parts need to be disseminated in order to ensure retrieval – both over the short term and long term (would there be a difference between trying to get the data back in 10 days versus 90 days?). Phase 2 has been initiated and 67 days of data have been collected as of this writing and preliminary findings will be presented herein. Potentially, the effects of the choice of the number of segments to divide the original file could be studied, but for now they are held constant. The third part of the study will look at possible detection characteristics for this sort of behavior and the development of tools and techniques for defense.

## 2 Problem Examined

The illicit users have the same needs for information management and security that the rest of the world has, if not greater. The needs can be broken into two classes. The first class would be one shared by all digital users, Confidentiality, Integrity and Availability or CIA[1], and the second would be one that is not so common, evasion. Each of these classes is addressed in the proposed solution.

Confidentiality is the limiting of access to data to authorized or intended users. The data in this case is encrypted and then segmented into many sections. The sections are then separated, encrypted and dispersed to disassociated unaware clients. If any piece or subset of the collection is discovered and reassembled it is unusable.

Integrity is knowing if the data is trustworthy, or in this case, were all of the pieces retrieved and reassembled correctly? In this proposed solution, the individual pieces have a checksum or digest calculated and appended to the end before delivery to the client systems. Upon retrieval the checksum is recalculated and verified to ensure that the chunk of data has returned intact. Once the pieces have been reassembled, the original message is decrypted. A final checksum for the entire original message is verified assuring that the message has been retrieved intact.

Availability is being able to access the data when and where needed. In this situation, the concept of availability relates to the reliability of future access to the data. This is currently being studied as phase two of this project. The trade off is speed of access for deniability or “it’s not on my drive!” The file is available to the owner with an access time of hours, days or months depending on many factors. The benefit is that the file is unavailable to anyone else.

The last issue is evasion. Evasion is an act of subterfuge, avoiding or eluding detection. The idea here is to hide the data from an examination of the local system. Once the pieces are distributed, the local system and web database can be forensically cleaned and all evidence of the data eradicated. Even if it were suspected that the web clients might be involved, a moderately trafficked web site could have hundreds, thousands or even millions of individual clients to investigate. Since the clients are not owned by the illicit user being investigated, possible jurisdiction problems arise investigating any potential involvement of the clients.

# 3 HTML5 and Web Storage

With the advent of HTML5 and its subsequent adoption in all modern web browsers (to varying degrees[2]), programming for a browser based internet experience recently turned to the better. HTML as a standard has been around since 1990 and was standardized as HTML 4 in 1997. HTML5 is still under development (as of November 2011) and is meant to subsume not only HTML4, but XHTML1 and DOM2 HTML (JavaScript) as well[3].

Some of the advantages of HTML5 (ubiquitous coding APIs, numerous new media types, embedded semantic meanings) while a boon to both developers and users alike, are outside the scope of this paper. The area of the HTML5 improvements that the authors are planning on exploiting is the advanced data storage, or Web Storage[4]. While Web Storage is not directly part of HTML5, it has been repeatedly attributed to being part of HTML5 enough that most sources currently attribute it to HTML5. Many developers may think that Web Storage includes cookies, various browser dependent client side databases, as well as storage objects. However, by the specification, the term Web Storage is limited to the storage objects – specifically localStorage and sessionStorage.

Since Web Storage includes both localStorage and sessionStorage, both needed to be considered. Upon a quick examination it was found that sessionStorage matched its name – it is storage that exists solely for a browser session (sessions expire when the browser is closed and the data automatically cleared). Because sessionStorage is implemented effectively, it is of little use to the user for our purpose. localStorage, on the other hand, works perfectly for what is needed. From a developer’s point of view, localStorage is an associative array or hash – a name=value pair that can hold any textual content.

To understand the need for a localStorage object, a little history is needed. Since the inception of the HTTP protocol, it has been stateless and anonymous, so a mechanism had to be created to make the tracking of state possible. The ‘HTTP State Management Mechanism’ proposal was created to fill this void[5]. The outcome of which is commonly known as cookies. The cookie mechanism is a name value pair that is served up from the client to the server inside of the HTTP Request phase (based upon various criteria: path on the server, domain to be served to, protocol to be served up to – http or https). Cookies have been used in various ways through the history of the web, more often than not they are used to hold a session identifier or token. Server frameworks (.Net, PHP, JSP) often implement these identifiers but occasionally they are created by hand by the developer.

Historically, cookies were the sole means web browsers had for long-term storage capabilities. They had limited length (4096 bytes) and a limited number could be written per domain (20) for a total of 81,920 bytes of storage space[4]. Today, localStorage, as a storage mechanism, is limited to 5Mb per origin (domain)[6], or 655,360 bytes of storage (8 times larger). If the browser manufacturers maintain the size of the specification (currently IE9 allows more - 10Mb per origin), the possibility of using various client’s hard drives for illicit storage becomes tempting.

As often happens with newer technologies, they are implemented before they are fully tested. localStorage works flawlessly in the modern browsers, but the tools that the end user has to allow, view, update or delete them is very limited (see TABLE 1). Combining the amount of storage space with a lack of user control makes this an invisible attach vector for illicit users to exploit. At the time of this paper, there is no unified user interface for localStorage. If a user wants to find out what is stored on their various browsers there is no easy way. An advanced user would have to visit the domain they are interested in and then run a bit of code to see if they had any localStorage recorded.

```
for (i=0; i<localStorage.length; i++) {
  key = localStorage.key(i);
  pairs += "key:"+key+" value:"+localStorage.getItem(key);
}
console.log(pairs);                                              (1)
```

Adding to the problem of knowing if your localStorage is being used, there is no clear way for a common or average user to turn it off. Additionally, once it is written it doesn’t have an easy affordance to remove or review the data. As an example of how this can be confusing, for Firefox’s DOM Storage (Firefox’s moniker for Web Storage) can be cleared via the menus “Tools -> Clear Recent History -> Cookies” ONLY when the range is “Everything”[7].

TABLE I. BROWSER COMPARISONS

| **Current Browsers** | **Access to Web Storage** | | |
|---|---|---|---|
| | ***Disable Storage*** | ***Clear Storage*** | ***Examine Storage*** |

| **Current Browsers** | **Access to Web Storage** | | |
|---|---|---|---|
| | ***Disable Storage*** | ***Clear Storage*** | ***Examine Storage*** |
| Firefox 10.0.2 | Yes, by turning cookies off in preferences(but does not clear old values) or in about:config | Select "Tools" » "Clear Recent History", open "Details", check "Cookies" and select "Everything" as time range. | Not without an external extension |
| Safari 5.1.2 | No, with cookies turned off, localStorage is still set | Select "Safari" » "Reset Safari…" » Remove all website data | Have to go into preferences and turn on developer menu, then navigate to domain where it was set |
| Chrome 17 | Yes, turn off cookies in preferences (but does not clear old values) | Select "Tools" » "Clear browsing data…", check "Delete cookies and other site data", select "Everything" from "Clear data from this period" and click on "Clear browsing data". | Developer Tools » Tools » Developer Tools - can see localStorage, but only for domain I'm currently visiting |
| Opera 11.61 | Yes, opera:config Persistent»Storage turn off global quota (then have to turn it on on a per/domain basis) | Preferences » advanced » Storage (can delete one at a time) | Preferences » advanced » Storage See domain and size, not content |
| IE 9 | Yes, internet options»Advanced»unclick Enable DOM Storage | Select "Tools"» "Internet Options" »"General" » check "Delete browsing history on exit", click on "Delete", check "Cookies" and click on "Delete" once more. | can see being set by running profiler in Developer Tools » Profiler |

A more universal interface is needed. While it might not be necessary to split localStorage out from other data storage capabilities, listing it under Cookies may not be intuitive for average users. Also, the ability to clear stored data in a more chronologically granular way would be useful.

# 4 Problem Exploited

To exploit this possible weakness, the authors devised a web application that would take any textual file, calculate and attach a checksum, encrypt it, split it into a some number of parts (26 in our testing), give each part an identifier (both for the part of the whole and an identifier for which file it came from), calculate a checksum for the part and append it to the string then re-encrypt it. It was found that from this formula it was possible to hide the parts on different clients and on subsequent visits those parts could be retrieved and reconstituted into our original data. Should a non-textual file be the target, a simple binary to text translation tools such as base64 or uuencode would suffice.

## 4.1 Web Environment

For the implementation of the web application, the authors chose the open source LAMP architecture for it's ubiquitous nature. LAMP is an acronym for Linux, Apache HTTP Server, MySQL database, and PHP server-side scripting environment.

## 4.2 Web Software

From a top-level view, the implementation of the application via web browsers consists of an interface to take a textual file and use the processed described above to split the file and insert the parts to a database. When the authors were ready to populate the parts to the visitors that come to our site, a small client-side script that communicates covertly (via AJAX - Asynchronous JavaScript and XML) with a server-side script. The result of the server-side script is stored in the client's localStorage. Once the illicit user decides there are enough copies distributed for his purposes, he can wipe out his file, the database AND all traces of the information.

Some time later, when it is decided it is time to reconstitute the data, a different client-side script is inserted that checks return visit clients for our data. If any data was found, be it a piece that hadn't gotten back yet or one already recorded, it was decrypted, the checksum checked and stored. After a period of time, the entire file was retrieved.

For a deeper explanation, there are two sets of scripts that execute this process. One set is used to distribute the parts out to various clients and the other set is used to retrieve the data back. Each set has both a client and a server script used to access the database for storage or retrieval as is applicable.

The first small client-side script (2) can be injected into any html page. It tests if localStorage is implemented on the particular browser. Next, if the browser doesn't already have a piece of the text file from our domain, a jQuery AJAX call is triggered to the server for the part of the file that has been distributed to the fewest clients. The part is then written to the browsers localStorage under a commonly used token name (we used 'uid') to help hide our data and intentions.

```
if(window.localStorage) {
  if(localStorage.getItem('uid')==null){
    $.getJSON('localStorageSet.php',function(data){
      localStorage.setItem("uid",data.uid);
    });
  }
}                                                        (2)
```

The localStorageSet.php file that the AJAX call is hitting goes into the database of encrypted parts, finds the part that has been copied to the least number of browsers and sends it back to the client to be injected into the localStorage with ‘uid’. While the script is doing this, it also updates the total disseminated count on the part that it just served up and logs the visit to the database.

Once the authors are confident that a sufficiently large enough number of targets have been populated, the original nefarious file and the database table holding the parts were destroyed. For the truly paranoid a forensic wipe of the drive and the user would be worry free of being searched.

The second small client script (Algorithm 3) can be employed at a later date, when the data is to be reconstituted. For this, a jQuery AJAX call is employed to send the contents of the specific localStorage data back to the server.

```
$.post(‘localStorageBack.php’,{
  d:localStorage.getItem(‘uid’)
});                                                        (3)
```

The data this sends back to the server is decrypted, checksum is checked and split into our original encryption, part and file identification. The data is then populated in a database table by its part identifier and filename for future reference. Once all of the parts are recovered, the entire file is reconstituted, decrypted to the original state and the checksum verified.

# 5 Proof of Concept Testing Environment

The laboratory proof of concept testing environment is simple and easily duplicated. VMware Workstation 7.1.0 was the foundation for the test environment installed on a Lenovo T61p laptop with 6Gb of memory. The target web server was a stock BackTrack5 virtual machine image[8]. Apache 2.2.14, MySQL 14.14 and PHP 5.3.2 were used to support the testing environment on the server.

## 5.1 Configuring the Web Server

The server application used was the default install that came with BackTrack5. The only addition to this was an installation of phpMyAdmin, an open source tool for simple database access (http://www.phpmyadmin.net/). Starting Apache and MySQL was all that was necessary (no specialized settings like .httaccess was needed).

In the testing environment, there was no reason to hide what was being attempted – so therefore two separate html files, one to set the localStorage, setData.php and one to get the localStorage back, getData.php. setData.php had the client-side code that executed the AJAX call (Algorithm 2). The AJAX call triggered the server side localStorageSet.php to get the least distributed part of the file and send it back in JSON (JavaScript Object Notation) format.

getData.php had the client-side code that used AJAX to send the contents of the localStorage.getItem(‘uid’) (Algorithm 3). The server-side code this executed, localStorageBack.php, decrypts the data and checks the checksum. If the checksum was accurate the data was stored.

In both cases, localStorageSet.php and localStorageBack.php all calls were logged and recorded for future study.

## 5.2 The client setup

To emulate the Internet client population at large, additional virtual machines were employed. For the initial test, a Windows XPpro base image was constructed with no service packs installed. This was not a necessary insecurity but established a baseline. A stock install of Firefox 4.0.1 was done with no add-ons. No special configuration of Firefox was performed. Two scripts were added to the C:\ directory of this initial configuration to aid in the automation of the test case: setData.bat and getData.bat.

First, the scripts make sure that the browser is not still running by executing a taskkill. This was necessary to ensure that localStorage was not preserved for only a single browser session. By terminating Firefox the session was stopped.

## 5.3 Assembling the masses

Once the Windows XPpro client is prepared, it is shut down and only used as a master for cloning. The algorithm requires at least 26 clients to hold all of the pieces of the message. The following scripts automated the process of construction utilizing VMware’s vmrun tool[9]. The tool can issue instructions to several of VMwares virtualization tools including Workstation. The following script creates 26 clones of the master Windows XPpro virtual machine.

```
set VMRUN="C:\Program Files (x86)\VMware\VMware
     VIX\vmrun.exe"
set SRCVM="C:\LocalStorage\Masters\WinXPpro\winXPPro.vmx"
set CLONE=C:\LocalStorage\CLONES\WXP

for /L %%i IN (101 1 126) do (
  %VMRUN% -T ws clone %SRCVM%
     %CLONE%%%i\WXP%%i.vmx linked
  %VMRUN% -T ws start %CLONE%%%i\WXP%%i.vmx gui
  timeout -T 60 /NOBREAK >NUL
  %VMRUN% -T ws suspend %CLONE%%%i\WXP%%i.vmx hard
)                                                          (4)
```

Vmrun is utilized to instruct VMware Workstation to clone the base Windows XPpro virtual machine 26 times. After starting the VM a delay of 60 seconds allows the client to fully boot before the client is suspended. Suspending allows for a faster cycle time for client visits to the web site.

## 5.4 Occupy localStorage

The next phase of the test is to have each of the 26 Windows XPpro clients start a browser, surf to the web server, and run the code to cause data to be deposited in the client’s localStorage area. It is important for the browser to be started and stopped to assure that localStorage has persistence beyond the current session. The following scripts are run on the host of the virtual machines to first set or download the data chunk to the client and second to get or retrieve the chunk from the client.

```
set VMRUN="C:\Program Files (x86)\VMware\VMware
      VIX\vmrun.exe"
set CLONE=C:\LocalStorage\CLONES\WXP
set FIREFOX="C:\Program Files\Mozilla Firefox\firefox.exe"
for /L %%i IN (101 1 126) do (
   %VMRUN% -T ws start %CLONE%%%i\WXP%%i.vmx
   %VMRUN% -T ws -gu dgj -gp "ATest4LocalStorage!"
      runScriptInGuest %CLONE%%%i\WXP%%i.vmx -nowait ""
      "cmd.exe /k C:\setData.bat
   timeout -T 60 /NOBREAK >NUL
   %VMRUN% -T ws suspend %CLONE%%%i\WXP%%i.vmx hard
)                                                          (5)
```

The MakeGetVisits script differ from the MakeSetVisits script in (Algorithm 5) only in the target script that is run locally on the client system: getData.bat. This structure is only necessary in this test environment to ensure that the browser is successfully started and stopped and that sufficient time is given to the client and browser to complete the operations. Typically the setData.bat script is run first followed by the getData.bat script. The set/get operation takes about an hour to complete. The entire environment starting from making the clones to retrieving the data set takes about 2 hours. The use of linked clones keeps the storage requirements down to under 40GB for entire environment.

# 6 PHASE 2 - DATA

## 6.1 Seeding

After proving that the authors could hide and retrieve information in a client's Web Storage in a controlled environment, our task was to discover what would happen in the wild. Interesting questions surfaced, such as how many copies of our user encrypted and obfuscated parts are needed to disseminate in order to ensure recovery and feel confident that the parts could be retrieved intact after 5 days, 30 days, or even 1 year.

In order to begin answering these questions, permission was obtained to use two of the author's departmental web presences (http://www.ist.rit.edu and http://www.nssa.rit.edu). To make the results of this testing more accurate, the decision was made to remove all visitors from the 129.21.0.0/16 domain (RIT's domain). This decision was made because most of the visitors to these sites from that domain are the universities' lab machines that are forced to visit those sites on browser launch and are re-imaged at startup. Since the set data on the lab machines would be removed at startup and the machines visit these sites multiple times a day, using results from these machines would skew the results in an unfavorable way.

## 6.2 Limits

The testing and data collection phase went live on December 17th, 2011. While the data setting and collecting is still ongoing, for this paper it was decided to cap the data analysis on February 22nd , 2012 – so the preliminary data in this paper is from 67 days. While this may be a small data set, interesting trends are already being seen.

## 6.3 Observations

An observational study was run with input variables number of sets available and the number of days until retrieval. The sample size for this study was $n$ 3804 - full sets of data seeded. The response was measured as the number of days until a full set was received. Figure 1 contains the plot of the response and the number of sets available.

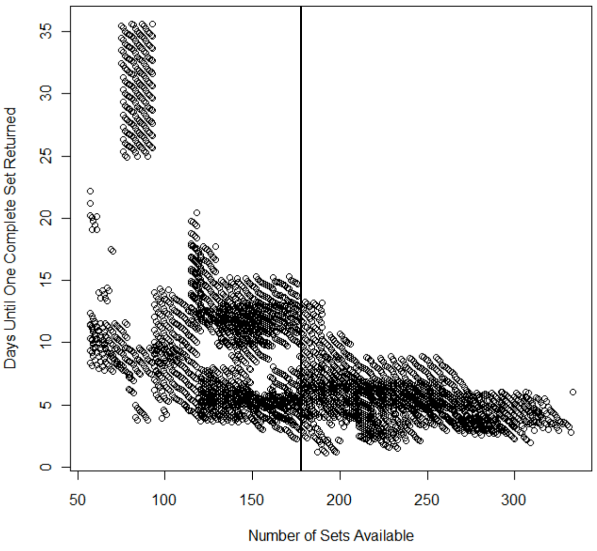


Figure 1. Days until one complete set returned against the number of sets available.

The vertical line on the plot occurs at 178 sets complete. This cut-off point was chosen since set 178 was the last set available for a total of 35 days. This is important because from the limited timeframe of our data collection, waiting 35 days for retrieval was determined to be our upper limit. As our data collection grows and ages, the authors look forward to seeing what the revisit rates will be for longer periods of time.

From the plot, a relationship is apparent. The relationship seems to be a slow decay then level off as the number of sets increase. This relationship is anticipated since it is logical to expect to get a full set back faster with more sets available. There is also a set of points between 75 and 93 sets that seem to be an anomaly as compared to the rest of the data. These data points represent 5% of the overall data and seem to occur for responses larger than 25 days.

Figure 2 contains the plot of the response and the number of days before retrieval.

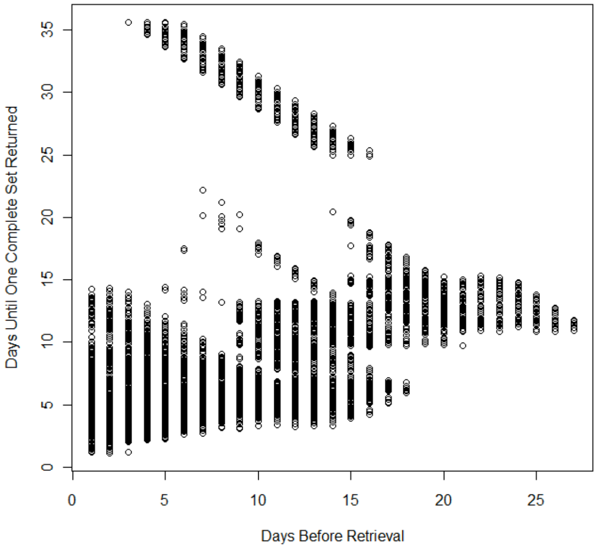


Figure 2. Days until on complete set returned over the number of days before retrieval

Based on Figure 2, overall there seems to be some sort of increasing relationship occurring. The same group of data points from Figure 1 does not seem to fit with the rest of the data on this plot as well.

Table II includes the information on the rank correlations between the response and each input variable.

TABLE II. CORRELATION TABLE

| Input Variable | Estimate | p-value |
|---|---|---|
| Sets Available | -0.451 | < 2.2E-16 |
| Days Until Retrieval | 0.361 | < 2.2E-16 |

Based on the results in Table II, there is a significant correlation between the response and both inputs at the $\alpha$ 0.05 level. From the estimates, the relationship between the response and the sets available seems to be decreasing and the relationship between the response and the days until retrieval is increasing. This matches what is in the figures above.

### 6.4 Next Steps

From here the authors would like to increase the size and age of the study to assess whether or not the anomaly data that we observed is repeatable and allow for the all of the sets to be available for an equal length of time. We would like to create a model that can be used to predict the response based on the inputs and run this study in an experimental setting with other variables such as different websites, different configurations on the break-up of the set, etc.

### 6.5 Exclusions

From the capped data (as of 2/22/12), a total of 10263 initial visit parts have been set, with 24873 re-visits. Of the total 35136 visits to our sites, 1825 (5.19%) either had Web Storage turned off or their browser was unable to implement it. For these visits, the database tracked the User Agent to determine the reasons. Some of the browsers were simply too old (Internet Explorer 7) or aren't equipped to handle local storage (Opera Mini). Of the 35136 total visits, 328 had browser versions that have Web Storage implemented, but had it turned off (0.933 %), while the remaining 1497 visits were by browsers or devices that were incapable of Web Storage.

The authors postulate that the very low number of visitors who had disabled Web Storage (<1%) may be due to many factors, including the standard being new and unfamiliar or a less than standard confusing interface. While this is an area that could use future study, it only adds to the viability of Web Storage as being a useful tool for our purposes.

## 7 Phase 3 – Detection

Once a greater data set has been accumulated, the authors are interested in studying the future application and usage of localStorage. The goal is looking for possible ways of monitoring and controlling localStorage activity, and identifying potential misuse. Intrusion Detection System tools such as Snort examine network traffic looking for digital signatures indicate that potential malicious activity is present. The development of signatures and other tools will be of primary interest during this phase.

## 8 Preventative Measures

The history of software interfaces is littered with examples of poorly designed and implemented user facing controls. The current state of the different browser interfaces to control Web Storage is lacking to say the least. The only preventative measure for not allowing something like this to happen on a client is to completely disable cookies. It should be noted that on all modern browsers there are different levels of cookie blocking (1st-party and 3rd party). However, since most trust the site they are visiting and 1st-party is what is being used, this number is relatively small. The number of visitors blocking 1st-party cookies varies greatly from one site to the next. Reports of 25% for sites about security and 1% for sites about general health are abundant. To know for sure one would need to test for their specific kind of site.. It should also be noted that once a localStorage value has been set, turning off cookies will not remove it, just make it inaccessible.

## 9 Conclusions

The authors hope these findings motivate browser architects to realize what they are making possible with their implementations and web application developers to think about the attack vectors they are creating. The need for a new storage capability in web browsers is not in question. The need to have the storage be easy to use for both developers and users alike is not in question. Although it may be a good idea to often hide implementation details from users, not giving them simple and intuitive controls that provide the ability to at least see what is being stored on their machines is in question.

## 10 References

[1] M. Stamp, "Information Security: Principles and Practice". John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/0471744190

[2] N. Leenheer, sights "The HTML5 Test." Last modified April 2012 – version 3.0, accessed April 22, 2012. http://html5test.com/

[3] I. Hickson, World Wide Web Consortium, "HTML5, A vocabulary and associated APIs for HTML and XHTML, Editor's Draft." Last modified November 04, 2011. Accessed November 05, 2011. http://dev.w3.org/html5/spec/Overview.html.

[4] I. Hickson, World Wide Web Consortium, "Web Storage, Editor's Draft." Last modified October 04, 2011. Accessed November 05, 2011. http://dev.w3.org/html5/webstorage/.

[5] D. Kristol, and L. Montulli. Netscape Communications, "HTTP State Management Mechanism." Last modified February, 1997 . Accessed November 04, 2011. http://www.w3.org/Protocols/rfc2109/rfc2109.

[6] I. Hickson, World Wide Web Consortium, "Web Storage, Editor's Draft." Last modified October 04, 2011. Accessed November 05, 2011. http://dev.w3.org/html5/webstorage/#disk-space.

[7] Mozilla Developer Network, "DOM Storage." Last modified October 23, 2011. Accessed November 08, 2011. https://developer.mozilla.org/en/DOM/Storage.

[8] back|track-linux.org, "Downloads : BackTrack Linux – Penetration Testing Distribution." Accessed November 14, 2011. http://www.backtrack-linux.org/downloads/.

[9] VMware, "Using vmrun to Control Virtual Machines." Last modified 2009. Accessed November 14, 2011. www.vmware.com/pdf/vix162_vmrun_command.pdf.

# A Multi-Disciplined Security Engineering Education Approach

William D. Casper
Lockheed Martin
Southern Methodist University, Dallas, TX
295 Niki Road, Paradise, TX 76073
william.d.casper@lmco.com
sdsu1989@yahoo.com

Stephen M. Papa
Lockheed Martin
Southern Methodist University, Dallas, TX
901 Bosque Court, Fort Worth, 76108
steve.m.papa@lmco.com

**SAM'12 – Track: Security Education**

***Abstract*—Security design for a system takes on the personality of the design team. Unfortunately these design teams are often focused within a single engineering discipline, for instance the design team is either composed of mainly software engineers or hardware engineers. The systemic problem is that not only is the design team single discipline focused but they are also normally single discipline trained and educated. Most design engineers have a degree with a single emphasis, such as a software engineer, mechanical engineer, electrical engineer, etc. The engineers are not usually trained in-depth in all areas of software, electrical/electronic, chemical, and mechanical engineering, especially as these areas pertain to security and secure processing systems. A multiple disciplined team with a more systems engineering approach is proposed as a near-term solution to this problem, but the proposed long-term solution is a Bachelor/Masters combination degree program with significant advanced course work in multiple engineering discipline areas. Broader education and engineering training in multiple disciplines would further aid a multi-discipline team resulting in more comprehensive security design. This paper describes the types of protection mechanisms individual disciplines offer, describes the added protection robustness achieved as a result of using multiple disciplines, and describes the types of discipline education necessary for a more complete Security Engineering degree.**

***Keywords- embedded security, secure hardware, secure software, security engineering, security education***

## I. INTRODUCTION

Security engineers are like most other engineers in that they will design a system utilizing their specific skills and areas of expertise. Software security engineers will want to focus the security of a system by utilizing software and software security mechanisms and hardware engineers will believe only proper hardware design can make a system secure. The usual argument against software security is that the software can be replicated as many times as necessary and this helps in the attack on the system since it allows multiple attackers to analyze, identify security flaws, and manipulate the software simultaneously. Hardware security may focus more on the fact that while you can create multiple copies of software to attack you may only have one piece of hardware to attack, and therefore only one attacker at a time can be attempting to manipulate the hardware. The penalty for attacking a software security mechanism may not result in loss of the software while some hardware attacks may result in damaged and irreparable hardware, making future attacks against that same hardware more difficult if not impossible.

Design projects and product features are normally a compilation of hardware, software, mechanical, and other engineering discipline team efforts. Mechanical engineering gets involved with the physical component structures of the housing and is often the expertise utilized for thermal and shock/vibration survivability. Systems engineers are involved with the specifications and design of how a product should operate and what functionality the product needs to support. Even management gets involved in making sure these design teams are able to function together as a team and keep the project moving forward towards completion. The overall value of the system can often be measured by the success of the team integration of the various disciplines more than the success or skill of any particular engineering discipline involved. A security design benefits the most from a similar mentality in that the overall product is stronger, more attack resilient, and more fundamentally comprehensive from a security measure when multiple engineering disciplines are utilized in developing the security solution.

The design team is composed of the multiple engineering disciplines because the majority of engineers have detailed knowledge and training in only one engineering discipline. Electrical engineers may receive some education in mechanical engineering and software engineering, but that is not the focus of their education and this additional training is usually not as comprehensive in these additional engineering discipline areas. Software engineers may receive some background on digital electronics and electronic components, but again not to the level as the electrical engineer. That is to be expected and that is what should be expected with the separation of engineering disciplines. Security engineering is a different discipline though and should be treated differently.

Security engineering is often thought of as a specialization in computer engineering, but security engineering needs to be much more than that. The prime focus of security engineering degrees is on computer engineering as it pertains to network security and information assurance. This leads to serious gaps in the mentality and abilities of the security engineer, and this hole in ability results in significant holes in the armor of the security design of a system. This education assumption needs to change if security engineers are going to be able to design a robust security solution that is more than protecting against network attacks and network intrusion.

A protection scheme designed around the strengths of a single discipline team while ignoring the protection attributes available in the other disciplines results in a design with inherent weaknesses. A more robust approach resulting in stronger protection would be a systems approach that combines the strengths of multiple disciplines. This systems approach can combine the security attributes of both secure hardware and software while adding protection mechanisms utilizing interactions between these secure components. Systems design and component utilization between secure processing hardware and software would result in a level of protection that is greater than the sum of the security benefits of the individual protection components. This system design protection concept can be used to extend security beyond the normal security boundaries for embedded systems, and can use strengths of some protection mechanisms to help compensate for the vulnerabilities or weaknesses of other protection mechanisms.

This paper discusses various protection mechanisms and categorizes them to either specific engineering disciplines or emphasizes how these protection mechanisms are actually a result of cohesive multi-discipline teamwork and design. It discusses how these disciplines interact in embedded computer system security designs and describes the interaction between the hardware protection elements and the software protection elements to accomplish the protection goals for the system. It also discusses the current state of security engineering degree programs and proposes a more robust, comprehensive approach towards the development and acquisition of a security engineering degree.

## II. Single Discipline Protection Mechanisms

### A. *Overview*

Protection mechanisms to improve system security exist in many disciplines. Some protection mechanisms are entirely composed of software features while others utilize more mechanical and chemical engineering components. Still other mechanisms are based on secure hardware processing components. Each of these mechanisms has its strengths and weaknesses, and all have their advocates and detractors. Security solutions based on a single discipline's mechanisms benefit from the strength of those mechanisms, but also suffer from the weaknesses associated with that type of solution.

A basic necessity for security of a system is a Root of Trust [1]. This Root of Trust is used as a building block for trusting operations or components in the rest of the system. The Trusted Computing Group advocates that this Root of Trust is a hardware element built and designed in a trustworthy manner, but software only protection advocates would state that a Root of Trust can be formed in a software element. Further discussion on both hardware and software Roots of Trust are described in the following paragraphs.

### B. *Software Protection Mechanisms*

Software protection mechanisms are those that use software to protect software or system functionality. These are protections such as supporting trusted boot, software hashing algorithms ensuring integrity of the executable software, stack overflow canaries, and software monitoring functions like debugger identification code [6][7].

Since the software does execute on a hardware component, usually a processor or microcontroller, one could argue that there is no such thing as software-only protection mechanism. That is probably a true statement from a purist point of view, but for this paper it is proposed that if the hardware does not aid in the protection design with specific hardware protection elements and if the software mechanism was not there that one could argue there was no protection on the system then the protection mechanism must be considered a software-only mechanism.

### C. *Hardware Protection Mechanisms*

Hardware protection mechanisms are often centered on the use of a trusted compute module (TCM) defined by the Trusted Computing Group (TCG) [1]. A TCM may also be considered a Trust Anchor since the TCM is the "anchor" the system is reliant on to embed the security design into the overall design.

One type of TCM is a secure processor. Secure processors are processing devices that allow for protection of the software that executes on them [3][5]. Secure processors usually have some level of self protection features also [8]. These features should include a trusted boot method to protect against an attack from the moment power is applied to the device. Trusted boot is a boot process that starts up in multiple stages, each stage performing some type of self-check and analysis prior to allowing execution of the next stage. Any error conditions in the boot sequence should result in the boot process execution halting. One does not want to continue attempting to overcome error conditions in a trusted boot process since this is an avenue of attack that should be prevented [2][4]. One could make the case that a trusted boot process is not a pure hardware protection mechanism but is likely a combination of hardware and software, but for now we will assume this secure boot code is resident in firmware associated with the TCM and therefore is more likely to be developed by a hardware engineer. Therefore this mechanism will be placed in the Hardware Protection Mechanisms section of this paper.

Another type of TCM is a secure Field Programmable Gate Array (FPGA). Several manufacturers now produce FPGAs that contain the ability to have an encrypted configuration file or that function with a configuration file that is burned into flash [11][12][13]. The benefit of these types of devices is that the configuration file can be developed by the development engineering team and kept programmable until prior to final release of the product. Then this configuration file can either be encrypted or burned into flash and the read-back feature on the devices can be disabled. Once the read-back feature is disabled and if the device's configuration file is encrypted if it is stored external to the FPGA, or if it is only stored in internal flash then an attacker cannot replace the FPGA configuration with their own configuration during the attack process. The configuration is secure and unavailable to the attacker and the device will function as intended, or at least as programmed. There is always the chance that the original FPGA developer allowed some bugs to remain in the final release configuration, either knowingly or unknowingly, and these bugs may be able to be exploited with the result of unintended operation of the FPGA.

### D. *Mechanical Protection Mechanisms*

In the context of this paper a mechanical protection mechanism will refer to a protective coating, tamper seal, or mechanical enclosure with security constructs. These protection mechanisms have some value, but depending on the resilience of the seal or coating or depending on the impenetrability of the mechanical housing the level of protection value is arguable.

Coatings are required on cryptographic modules utilized in some government systems [9]. While this methodology has been improved over time and various coating materials have been utilized there has also been an increased awareness of the susceptibility of these coatings to a chemical attack [10]. There is continuing research and commercial investment into coating technologies though and the challenge exists to develop a coating of “impervium”, which this paper defines as an impenetrable layer of some substance that withstands all chemical or physical attacks. If “impervium” can ever be developed this will lead to an increase in credibility for a security coating approach, but skeptics are numerous on the true viability and achievability of such an approach.

Tamper seals are a protection mechanism utilized to detect someone has opened a container. These seals are often used in the context of warranty seals and can be found on many instances of commercial electronics. The challenge with adhesive-based seals is how good is the adhesive that is applied to the seal and can replacement seals be easily acquired [10]? If the adhesive for a seal is susceptible to a chemical attack or physical attack it may be possible to remove the seal, acquire the desired design information, and then replace the seal to remain undetected. Mechanical seals are vulnerable if they can be removed and replaced without a permanent indication of this removal [21]. If there are many outside sources from which to acquire a replacement seal it may not matter how hard it is to remove the seal since the solution will be to replace the seal with a new one prior to a warranty return. The attack will not have to be concerned if the attack mechanism destroys the seal since they will be able to apply the replacement seal that was readily available. Tamper seals are also of limited value if the attacker has no intent on returning the item. If the original equipment manufacturer (OEM) does not receive a returned item then the OEM does not have the opportunity to inspect the tamper seals placed on the equipment to determine if they were removed.

## III. Multi-discipline Protection Mechanisms

### A. *Overview*

Multi-discipline protection mechanisms are those that combine the functionalities of more than a single engineering discipline. These protection mechanisms may utilize hardware and software, hardware and mechanical, chemical, electrical, and mechanical, or any of many other combinations of protection mechanisms. The key to these protection mechanisms is that they usually need a multi-disciplined engineering team to design, develop, and integrate. No longer is the team only made of up mechanical, software, or electrical engineers. Now the design team uses the knowledge and strengths of the individual contributors of the design team to develop a solution that integrates the strengths of the various engineering disciplines. This solution may result in layers of protection that allow the weaknesses of one discipline’s solution to be protected by solution components in another discipline. Thus the strength of the solution is actually greater than the sum of the components for the solution since the attack vectors now have to deal with different attack mechanisms and different attack vectors while not triggering the detection mechanisms that are likely very different between the solution components.

### B. *Multi-discipline Protection Mechanisms*

These are protections such as utilizing a TCM or FPGA with monitoring of the execution of valid software on an embedded processor [14]. This type of mechanism uses the secure processing attributes of the TCM to extend the trust anchor envelope over software executing outside the trusted boundaries. These protections may also include developing and embedded the hardware protection features to assist the monitoring of software on secure processors [15]. This now mixes the hardware and software protection mechanisms by closely coupling the hardware protections with the software protections, strengthening the overall protection solution by utilizing the strengths of multiple team disciplines.

Other multi-discipline protection approaches may combine the mechanical and electrical components instead of only hardware and software. These protection approaches include designs that protect secure processors with security coatings. If coatings are not an option another multi-discipline approach may be to house the security design inside an enclosure with tamper switches to detect enclosure penetration. Either of these types of solutions makes the penetration and attack on the security processor even harder than relying solely on the self-protect features of the security processor.

## IV. Security awareness and Usability

### A. *Overview*

Society at large is becoming more cognizant of the importance of security. However awareness of the need does not always lead to understanding and utilization of the security mechanisms provided. Sometimes using the security mechanisms is so onerous that people will find a way to work around the security mechanisms, thus weakening or even defeating the original intent of the security design. The importance of the security mechanisms and the ramifications of failed security systems are often not fully comprehended by those circumventing the security due to usability constraints. Designing security with usability as a design input would help eliminate some of this behavior. Users of security systems that are easier to use will be less likely to bypass the security since the utilization of the security mechanization could be more intuitive and less constraining.

### B. *Usability Education*

Security systems are often designed by engineers for use by non-engineers. The intuitiveness and ease of usability of a system is always measured by the end user, but it is envisioned by the design engineer. However, the end users of systems may vary greatly in areas of education, skills, and computer literacy compared to the design engineers. These variations may result in differences of opinions in the usability of the security products.

For instance, challenges in the security of systems in the healthcare industry have been studied and determined to have a significant amount of failures due to usability difficulties [24]. Personal data stored in security-controlled databases were replicated in Microsoft Excel spreadsheets due to the perceived complexity of using the database software. This resulted in numerous data leaks of patient personal information, from diagnosis, treatments, insurance information, social security numbers, etc. Other medical data storage servers were incorrectly configured resulting in significant data breaches.

Improper data storage configuration has been the blame for numerous data breaches discussed in various forums. Weak passwords have also been frequently blamed and proven to be the root cause of many data breaches. Strong passwords are now recommended for most systems, but given the number of different accounts and systems that today's users interact with they often institute password usage variations that bypass the benefits of multiple strong passwords [25]. The resultant difficulty in terms of security usability encourages the users to look for workarounds to ease the burden of the security. One could ask is the problem with the users or with the design, but clearly usability or more accurately usability limitations play a significant part in the end user behavior.

## V. SECURITY ENGINEERING EDUCATION

### A. *Overview*

Separate engineering disciplines have been identified and developed over many years and the associated education at universities has followed. New technologies and disciplines have evolved as specialization areas in more overarching engineering disciplines. Once a new discipline evolves and achieves a larger amount of recognition it may move from a specialization focus area to its own discipline. Years ago Computer Science was not a recognized discipline and it also started as a specialization area, but today most universities offer degree programs in Computer Science or Computer Engineering.

Security Engineering is usually considered a specialized field in Computer Science. As such, most programs that offer degree programs in Security Engineering or with a Security Engineering emphasis have these programs offered by the Computer Science or Computer Engineering departments at the associated university [17][18][19][20][21].

### B. *Current Programs*

Various universities offer Bachelor's or Master's degrees in Security Engineering or with a Security Engineering emphasis. Since these universities typically offer these degrees as specializations in Computer Science or Computer Engineering these programs often focus on security associated with Information Assurance (IA) and IP Network Security. What is meant by this is that the security education often assumes the engineering knowledge needed for adequate security engineering design can be achieved with only Computer Science curriculum. If you look at degree programs as offered by Southern Methodist University (SMU), DePaul University, Virginia Tech, FH Joanneum, and most others the focus is usually on items you would consider as IA security [17][18][19][20][21]. This includes but is not limited to secure network operation, honeypots, virus detection, intrusion detection, cryptography, and others.

Virginia Tech's graduate program that includes security education is the Harry Lynde Bradley Department of Electrical & Computer Engineering. As such they teach a class titled Secure Hardware Design that does reach past the IA security mentality and starts approaching a more multi-disciplined approach to security [21]. This class utilizes both the Electrical Engineering discipline and the Computer Engineering discipline to jointly define a mechanism and process to achieve a more robust secure hardware solution. Dr. Patrick Schaumont of Virginia Tech has proposed co-designing hardware and software when doing embedded system architecture education so there is a start to the recognition of a multi-disciplined approach to the engineering education for embedded systems [16].

Southern Methodist University's graduate program that includes security education is the Bobby B. Lyle School of Engineering Department of Computer Science and Engineering. This program is available on campus, but is also extensively taught on weekends to industry professionals as a MS program focused on security engineering. SMU has recently added a class on hardware security including steps that could be followed to increase the trust of integrated circuits produced by an external foundry [19]. This class utilizes both the Electrical Engineering discipline and Computer Science discipline to jointly understand the security concerns and vulnerabilities systems may have due to Trojan hardware components. SMU has a research lab called the High Assurance

Computing and Network Labs (HACNet), led by Dr. Suku Nair, that focuses on security research for government and industry solutions [23]. These efforts by SMU recognize the benefits of multi-disciplined approaches for security engineering education.

### C. *Proposed Education Solution*

The authors of this paper propose that Security Engineering should not be considered a single discipline emphasis. A multi-disciplined design team can design a more robust and attack-resistant solution. Similarly a multi-discipline trained Security Engineer would possess the vital skills to design a more robust security design. Industry and academia should recognize this need and evaluate how design teams have become integrated between engineering disciplines. The complex solutions of today require skills from several types of engineers, and the complex solution for a robust security design requires the skills from many engineering disciplines.

The challenge of training Security Engineers in more engineering disciplines is how to evaluate the depth of training needed from each of the focus areas. Each Bachelor of Science engineering degree program strives to give the engineer both breadth and depth in a particular engineering discipline. This is vital to the education strength of the degree and the individuals obtaining that training. Removing some of that depth to achieve a wider breadth of knowledge would seem to lessen the overall ability of the Security Engineer to perform their design tasks properly. It is proposed that a jack-of-all-trades training with no technical depth in any specific area would result in engineers having such a shallow level of specific knowledge or training that they would not be effective. The knowledge gained by strong education foundations is critical in unlocking the potential of an engineer's mind, allowing them to use their experience and training along with new thoughts of their own to further advance the technology the world depends on. A shallow level of education may not plant the necessary seeds of knowledge into the engineer's mind that need to grow to advance the engineering potential of society.

The well-trained Security Engineer should have significant knowledge in the areas currently covered by degrees focusing on IA, but should also possess in-depth knowledge in other areas of Computer Engineering, Systems Engineering, Electrical Engineering, Mechanical Engineering, Materials Science, Cryptography, System Usability, and more. It is likely difficult if not unachievable to obtain this level of knowledge in the normal curriculum encountered during most Bachelor of Science degrees. In order to achieve this desired knowledge base it is proposed that a Security Engineering degree should be either a five year degree to allow enough coursework to be learned or potentially should be a combined Bachelor/Masters degree with a projected duration of six years for normal completion. Coupling multiple areas of engineering discipline into a particular degree could be done by carefully utilizing various credits that are often random technical elective credits and focusing these credits on specific areas. This would enable the Security Engineer to obtain sufficient breadth and depth in many of the focus areas that contribute to robust security designs. This type of trained Security Engineer would be able to perform in a multi-disciplined manner similar to the result from having a design team of several single discipline trained engineers. Creating a better trained Security Engineer should result in the creation of better security designs.

Even with this type of cross-training though, a team of these better trained Security Engineers may not be able to replace a well-developed team with experts from many individual disciplines. A strong team of experts who can utilize the various skills of the design team can probably design the better security system. There are proven reasons for having teams with disciplines from multiple backgrounds. However having the skills for some of the team members to better understand the capabilities and strengths of various disciplines would be beneficial. Strong multidiscipline teams also sometimes clash due to the individual strengths of each team member. As was mentioned previously the strong electronic hardware team member may believe the best solution can only be achieved through hardware, while the strong software team member may believe software is the only answer. Cross-training in multiple disciplines may help break down these barrier walls while also providing the mechanism for developing a more secure and safe society.

## VI. CONCLUSIONS

Multi-disciplined design teams are able to develop more secure designs by utilizing the strengths and knowledge of the individual team members. A similar approach should be taken when developing the security solution for a computer system design. A multi-disciplined team is better positioned to develop a more robust security solution than one that is developed by individuals from a single discipline.

Security Engineering is more than Information Assurance. A good security design encompasses the knowledge from multiple engineering disciplines. Current degree programs offered for Security Engineering emphasize focus on Information Assurance from a Computer Engineering point of view. It is proposed that a new degree program be developed that provides multiple engineering discipline breadth with sufficient depth to create the Security Engineer needed for the future. Whether this degree program is an extended duration Bachelor of Science degree or some type of combination Bachelor/Masters degree may be debatable, but the necessity of such a degree program and such a trained engineer becomes more and more apparent as security designs increase in complexity to combat the ever-increasing abilities of those who strive to exploit designs and defeat their security mechanisms.

## REFERENCES

[1] TCG-Mobile-Phone-Work-Group,, "TCG Mobile Trusted Module Specification Version 1 rev 1.0", June 12, 2007

[2] Kurt Dietrich, Johannes Winter, "Secure Boot Revisited", 9th International Conference for Young Computer Scientists, 2008

[3] Ross Anderson, Mike Bond, Jolyon Clulow, Sergei Skorobogatov, "Cryptographic Processors—A Survey", Proceedings of the IEEE, Vol. 94, No. 2, February 2006

[4] Bo Zhao, huangguo Zhang, Zhede Li, "A trusted start-up based on embedded System", IEEE Ninth International Conference on Computer and Information Technology, IEEE Computer Society 2009

[5] Eugen Leontie, Gedare Bloom, Bhagirath Narahari, Rahul Simha, Joseph Zambreno, "Hardware Containers for Software Components: A Trusted Platform for COTS-Based Systems", 2009 International Conference on Computational Science and Engineering, 2009 IEEE Computer Society

[6] Divya Arora, Srivaths Ravi, Anand Raghunathan, Niraj K. Jha, "Architectural Support for Run-Time Validation of Program Data Properties", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 15, No. 5, May 2007

[7] Ilkka Uusitalo, Kaarina Karppinen, Pasi Ahonen, Heimo Pentikäinen, "Towards Evaluation of Security Assurance during the Software Development Lifecycle", 2009 International Conference on Availability, Reliability and Security, 2009 IEEE Computer Society

[8] Siddhartha Chhabra, Brian Rogers and Yan Solihin, "SHIELDSTRAP: Making Secure Processors Truly Secure", IEEE 2009

[9] Federal Information Processing Standards Publication FIPS PUB 140-2, "Security Requirements for Cryptographic Modules", January 11, 1994

[10] Carol Cantlon, "Potential Chemical Attacks on Coatings and Tamper Evident Seal Adhesives, National Institute of Standards & Technology (NIST), September 6, 2005

[11] http://www.xilinx.com/

[12] http://www.altera.com/

[13] http://www.actel.com/

[14] Huaqiang Huang, Chen Hu, Jianhua He, "A Security Embedded System Based on TCM and FPGA", 2009 2nd IEEE International Conference on Computer Science and Information Technology, IEEE Computer Society, Aug 2009

[15] Divya Arora, Srivaths Ravi, Anand Raghunathan, Niraj K. Jha, "Hardware-Assisted Run-Time Monitoring for Secure Program Execution on Embedded Processors", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 14, No. 12, December 2006

[16] Patrick Schaumont, "Hardware/Software Co-design is a starting point in Embedded Systems Architecture Education", Workshop on Embedded Systems Education, WESE 2008

[17] http://www.cdm.depaul.edu/academics/Pages/MSinComputerInformationandNetworkSecurity.aspx

[18] http://sse.stevens.edu/academics/graduate/systems-security-engineering/courses-schedules/

[19] http://www.smu.edu/Lyle/Departments/CSE/Programs/MS_SEC

[20] http://www.fh-joanneum.at/aw/~a/home/?lan=en

[21] http://graduateschool.vt.edu/graduate_catalog/courses.htm

[22] Gary R. Cook, "User's Guide on Security Seals for Domestic Cargo", United States Department of Defense Lock Program, United States Department of Homeland Security, 2007

[23] http://hacnet.smu.edu/#MissionStatement

[24] M. Eric Johnson, Nicholas D. Willey, "Usability Failures and Healthcare Data Hemorrhages", IEEE Security & Privacy, March/April 2011

[25] Bryan D. Payne, W. Keith Edwards, "A Brief Introduction to Usable Security", IEEE Internet Computing, May/June 2008

[26] Dirk Balfanz, Glenn Durfee, D.K. Smetters, Rebecca E. Grinter, "In Search of Usable Security: Five Lessons from the Field", IEEE Security & Privacy, September/October 2004

[27] Ka-Ping Yee, "Aligning Security and Usability", IEEE Security & Privacy, September/October 2004

# SESSION

# INFORMATION ASSURANCE, SECURITY, AND MANAGEMENT

## Chair(s)

### TBA

# Application of Criterion-Based Multilayer Access Control to XML Documents

**Leon Pan**
Department of Computer Information Systems, University of the Fraser Valley, Abbotsford, BC, Canada

**Abstract -** *An approach is proposed to address the fine-grained multilayer access control requirements of XML files. The system is based on a set of predefined security criteria and security criterion expressions which serve as keys and locks respectively. The security criterion expressions are defined in the XML schema and embedded into individual XML files automatically. The XML files are enhanced to be secure XML files by embedding a set of security criterion expressions while users are enhanced to be secure users by associating with security criterion subsets. The embedded security criterion expressions are evaluated by the elements in the user's security criterion subset to determine the accessible part(s).*

**Keywords:** Criterion-Based Multilayer Access Control, Security Criterion Expression, Security Criterion Subset, XML security, Secure XML files, Secure users

## 1 Introduction

XML [17] is becoming increasingly popular in describing, managing, storing, and sharing data. It provides means of defining a vocabulary of element tags and attributes to structure information of interest. Anyone can define his/her own markup tags and attributes which describe the data and are combined with data to form XML documents. Because sharing and distributing data over Internet becomes more and more common, XML document security becomes more important and the needs of the effective and efficient XML security systems arise.

The author proposes a method of introducing security attributes to every element of the XML documents and applying the criterion-based access control to support XML document security. The values of the security attributes, which are defined in XML schema as the fixed values, are in the form of security criterion expressions. These security criterion expressions are evaluated to determine whether a user has the access to the XML document as a whole and its corresponding elements.

The rest of the paper is organized as follows. Section 2 briefly overviews previous works of the XML security. In section 3, the criterion-based access control is reviewed. Many important concepts of this model are discussed as well. Section 4 presents the details of embedding security information into XML documents and defining the security attributes through XML schema. Section 5 provides the further discussions and conclusion.

## 2 Overview of XML security

XML security has attracted increasing attention of researchers and many approaches and models [1, 3-8, 12-15, 18-20] has been proposed and developed, including encryption, digital signature, XML key management, authentication and authorization, and access control. Among these approaches, the access control method is very popular because it is convenient to take advantage of the characteristics of XML document when enforcing the authorization rules.

Some XML access control approaches [4, 6, 7, 8, 14, 19] are based on views that are created to enforce the authorization rules. A specific view created for a user specifies the portion of XML document that is accessible to that user. The approach proposed in [13] supports the access control by filtering out the queries that are not compatible to the access control policy. In [1, 4], queries are rewritten by combining with the access control policy in order to enhance the performance.

## 3 Criterion-Based Multilayer Access Control

### 3.1 Basic Concepts

The criterion-based access control approach was first proposed to integrate with role-based access control model to deal with multilayer security of multimedia applications [10, 11]. In this approach, security criteria, security criterion expressions, and security criterion subsets are introduced. Security criterion expressions and security criterion subsets serve as locks and keys, respectively. Each object or sub object is embedded into a lock and each user (subject) is assigned a set of keys. The user's keys are used to actuate the locks and the state of the locks determines whether the user has access to an object or sub object.

A security criterion is a criterion used to both specify the user's security attributes and define the object's (and the sub object's) security attributes. Each security criterion is represented by a symbol $s_i$. Security criteria are abstracted from authorization rules. From the whole set of authorization rules, a set of security criteria $s_1, s_2, \ldots, s_n$ in an application domain can be abstracted. The collection of all security criteria, their complement counterparts ($\overline{s_j}$), constant false F and true T form a set which is called the security criterion set.

A user may have more than one security attributes. So, several security criteria are often required to specify the user's security attributes. The set composed of these security criteria is called a security criterion subset (SCSS). When a user is associated with a security criterion subset (SCSS), he/she is enhanced to be a secure user (SU).

To precisely reflect authorization rules, objects (and the sub objects), as well as their security attributes, are defined by security criterion expressions. A security criterion expression (SCE) is a Boolean expression in terms of security criteria. A Boolean expression is considered to be a security criterion expression iff it reflects one or more authorization rules. Following are legal security criterion expressions:

(1) A constant true, T, or false, F

(2) A Boolean expression derived directly from an authorization rule

(3) Logical "OR" of (1) and (2)

The constant true, T, or false, F, represents special cases. When an (sub) object needs unconditional protection, its corresponding security criterion expression should be the constant true, T. On the other hand, when the security criterion expression is the constant false, F, the related (sub) object is accessible in any circumstances.

To support fine-grained multilevel access, in an object, each part (i.e. sub object) with different security attributes and thus of different security levels has an embedded security criterion expression to specify its security attributes. A sub object with an embedded security criterion expression is a secure sub object.

## 3.2 Security criterion abstraction and secure object and secure user generations

A systematic method has been developed to abstract security criteria from authorization rules, to transform authorization rules into security criterion expressions, and to generate security criterion subset based on the authorization rules [11]. For details, please reference [11].

Example 1: There is a XML database which contains XML documents of patients' medical history records. Each XML document includes insensitive information such as recuperation information and sensitive information, including personal information, nursing information, diagnosis information, and treatment information. Suppose the following authorization rules are applied to the XML documents.

1. Non-professions who are not registered have no access to the files

2. Non-professions don't have access to the personal information and professional information

3. Non-professionals and nurses have no permission for diagnosis and treatment information

It is easy to generate the components following the method provided in [11]. Tables 1, 2, and 3 show the generated security criteria, security criterion expressions for the sensitive sub objects, and the subset for each group of users, respectively.

Table 1 Generated security criteria

| Security criterion symbol | Security criterion meaning |
|---|---|
| $s_1$: | Non-professionals |
| $s_2$: | Not registered users |
| $s_3$: | Nurses |

Table 2 Security criterion expressions, corresponding authorization rules, and the applied (sub) objects

| Security criterion expression | Authorization rules | (Sub) Object |
|---|---|---|
| $s_1 \cap s_2$* | Authorization rule 1 | Whole file |
| $s_1$ | Authorization rule 2 | Personal information |
| $s_1$ | Authorization rule 2 | Professional information |
| $s_1 \cup s_3$* | Authorization rule 3 | Diagnosis information |
| $s_1 \cup s_3$* | Authorization rule 3 | Treatment information |

*Note: The authorization rules determine whether intersection or union is used. For example, authorization rule 1 indicates that non-professionals ($s_1$) who are not registered ($s_2$) do not have access. On the other hand, $s_1 \cup s_3$ means either non-professionals ($s_1$) or nurses ($s_3$) have access to the diagnosis information, which reflects authorization rule 3 accurately.

Table 3 Security criterion subset and the corresponding user groups

| Security criterion subset | User group |
|---|---|
| $\{S_1\}$ | non-professionals |
| $\{S_1, S_2\}$ | Not registered non-professionals |
| $\{S_3\}$ | Nurses |

## 3.3 Achieving fine-grained access control

In the Criterion-Based Access Control model, an (sub) object's security attributes and security level are implied by indicating users who do not have access rather than explicitly defining them. The system becomes simpler because one mechanism is used to define both the user's security attributes and the (sub) object's security attributes. The security criterion expressions embedded in a secure object can be regarded as locks, while the security criteria in the security criterion subset can be considered as keys. When a secure user accesses a secure object, he/she uses the available keys to actuate the locks. Whether the secure user is allowed to access the secure sub object depends on the state of the corresponding locks.

A security criterion expression is evaluated in the following two steps. First, substitute all the security criteria in the security criterion expression with true, T, or false, F. If replace those security criteria in the security criterion expression with true if they also appear in the secure user's security criterion subset. Otherwise, replace them with false. Second, the security criterion expression is evaluated according to the normal evaluation procedure in Boolean algebra. The evaluation value T of a security criterion expression implies that users with security attributes specified by these security criteria are not allowed to access the corresponding secure sub object. On the contrary, a false evaluation value, F, of the security criterion expression implies that the security criterion expression does not prevent these secure users from accessing this sub object.

# 4 Criterion-Based fine-grained XML document access control

An XML file forms a tree structure that starts at the root element and branches to the leaf elements. Each file has a unique root element which is the parent or grandparent of all other elements. To support fine-grained access control, security information must be embedded into every element. This paper proposes a novel idea to achieve fine-grained access control by making use of the properties of XML.

## 4.1 Introducing security attributes

Security information can be expressed in form of security criterion expressions which reflect the relevant security policy. Security criterion expressions are embedded into every element to specify the security features of the element and sensitive attributes (Sensitive attributes access control will be discussed in another paper). If an element and its descendants are accessible to public, a constant false, F is embedded. Otherwise, a specific security criterion expression is generated and embedded according to the security policy. The embedded security criterion expression specifies the users who do not have access to this element.

Figure 1 shows a simplified medical file which includes both sensitive and insensitive information about a patient's medical records. A specific security criterion expression, which extracted from relevant authorization rules, is embedded into each node in the file. When a user proposes a request to the file, his/her security criterion subset is used to evaluate these embedded security criterion expressions. If the evaluation value is true, the protection conditions for the corresponding element are satisfied. Thus, the user has no access to the node. If the evaluation is false, the user has access to the current element. However, the descendants' security criterion expressions are still need to be evaluated to decide the accessibility of each descendant element.

For example, if a user's security criterion subset is $\{s_3\}$ (Therefore, he/she is a nurse), only the evaluation values of diagnosis and treatment are true, which results in inaccessibility of these nodes. This conforms to the authorization rule 3.

Although it is possible to embed the security criterion expressions manually, it is very time-consuming, labor-intensive, and error prone to do so. To enforce the security policy more efficiently and consistently, an approach is proposed in this paper which defines the XML documents' security attributes by using enhanced XML schema.

## 4.2 Defining security attribute in XML schema

An XML schema [2, 16, 21] describes the structure of an XML file. It defines the legal building blocks of an XML file, including elements, attributes, child elements of an element, the number and order of the child elements, element with or without content, data types for elements and attributes, and default and fixed values for elements and attributes. In an enhanced schema, a security criterion expression can be defined similarly for every element or attribute as a security attribute with the fixed value which is inserted into corresponding element of an XML file automatically when the XML file is created. Again, only the method of defining security attributes for elements will be discussed in this paper.

Therefore, each element of the enhanced schema is accompanied with a security attribute whose value is defined as fixed value of the security criterion expression extracted from the relevant authorization rules. This security attribute specifies the protection condition for the current element and its descendants. A special security criterion expression, constant false F, is embedded if the corresponding element does not contain any sensitive information and thus is accessible to everyone according to the security policy.

```
<?xml version=”1.0”?>
<smr:medical_history security="s1 ∩ s2"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://www.secure.medical.records/redords smredords.xsd"
   xmlns:smr="http://www.securemedicalrecords">
   <smr:personal_info security="s1">
     <smr:patientname security="s1">John Smith</smr:patientname>
     <smr:address security="s1">1234 66th Ave. Hammond IN</smr:address>
     <smr:phone security="s1">219-555-5555 </smr:phone>
   </smr:personal_info>
   <smr:medical_records security=""s1 ∩ s2" >
     <smr:medical_record security=""s1 ∩ s2">
       <smr:non_professional_info security=""s1 ∩ s2">
         <smr:visitdate security=""s1 ∩ s2">02-28-2012 </smr:visitdate>
         <smr:recuperation security=""s1 ∩ s2">Exercise regularly</smr:recuperation>
       </smr:non_professional_info>
       <smr:professional_info security=""s1" >
         <smr:diagnosis security="s1 ∪ s3" >diagnosis info</smr:diagnosis>
         <smr:treatment security="s1 ∪ s3"  >treat info</smr:treatment>
         <smr:nursing security="s1" >nursing info</smr:nursing>
       </smr:non_professional_info>
     </smr:medical_record>
   </smr:medical_records>
</smr:medical_history>
```

Figure 1 An XML document of a patient’s medical records with security information embedded

The schema shown in figure 2 defines the elements and their security attributes for XML files of simplified patient’s medical records discussed in 4.1 subsection. By using this schema, the required security criterion expressions are embedded into every element for any relevant XML documents.

```
<?xml version=”1.0”?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
  targetNamespace="http://www.secure.medical.records/records"
  xmlns="http://www.secure.medical.records"
  elementFormDefault="qualified">
<xs:element name="medical_history">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="personal_info>
          <xs:complexType>
            <xs:sequence>
              <xs:element name="patientname" type=”string1Type”/>
              <xs:element name="address" type=”string1Type”/>
              <xs:element name="phone" type=”string1Type”/>
            </xs:sequence>
          </xs:complexType>
          <xs:attribute name=”security” type=”xs:string” fixed="s1"/ >
        </xs:element>
        <xs:element name="medical_records">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="medical_record" maxOccurs="unbounded">
                <xs:complexType>
                  <xs:sequence>
                    <xs:element name="non_professional_info">
                      <xs:complexType>
                        <xs:sequence>
                          <xs:element name="visitdate" type="visitdateType"/>
                          <xs:element name="symptom" type="string2Type"/>
                          <xs:element name="recuperation" type="string2Type"/>
                        </xs:sequence>
                      </xs:complexType>
                  <xs:attribute name=”security” type=”xs:string” fixed="s1 ∩ s2"/>
                    </xs:element >
                    <xs:element name="professional_info">
                      <xs:complexType>
                        <xs:sequence>
                          <xs:element name="diagnosis" type=”string3Type”/>
                          <xs:element name="treatment" type=”string3Type”/>
                          <xs:element name="nursing" type="string1Type"/>
                        </xs:sequence>
                        <xs:attribute name="security" type="xs:string" fixed="s1"/>
                      </xs:complexType>
                    </xs:element >
                  </xs:sequence>
                </xs:complexType>
            <xs:attribute name=”security” type=”xs:string” fixed="s1 ∩ s2"/>
              </xs:element>
            </xs:sequence>
          </xs:complexType>
      <xs:attribute name=”security” type=”xs:string” fixed="s1 ∩ s2"/>
        </xs:element >
      </xs:sequence>
       <xs:attribute name="security" type="xs:string" fixed=""s1 ∩ s2"/>
    </xs:complexType>
  </xs:element>
  <xs:complexType name=”string1Type”>
    <xs:simpleContent>
```

```
      <xs:extension base=”xs:string”>
        <xs:attribute name="security" type="xs:string" fixed="$s_1$"/>
      </xs:extension>
</xs:simpleContent>
  </xs:complexType>
  <xs:complexType name=”visitdateType”>
    <xs:simpleContent>
      <xs:extension base=”xs:date”>
    <xs:attribute name="security" type="xs:string" fixed="$s_1 \cap s_2$"/>
      </xs:extension>
</xs:simpleContent>
  </xs:complexType>
  <xs:complexType name=”string2Type”>
    <xs:simpleContent>
      <xs:extension base=”xs:string”>
    <xs:attribute name="security" type="xs:string" fixed="$s_1 \cap s_2$"/>
      </xs:extension>
</xs:simpleContent>
  </xs:complexType>
  <xs:complexType name=”string3Type”>
    <xs:simpleContent>
      <xs:extension base=”xs:string”>
        <xs:attribute name="security" type="xs:string" fixed=”$s_1 \cup s_3$"/>
      </xs:extension>
</xs:simpleContent>
  </xs:complexType>
</xs schema>
```

Figure 2 An XML schema defining the elements and (security) attributes of a patient’s medical records

# 5 Discussions and Conclusion

This paper proposes an approach to support fine-grained XML document access control by using a set of predefined security criteria. Security criterion expressions are extracted from the system authorization rules and are embedded into every element of the document. These security criterion expressions serve as locks. On the other hand, every user is assigned a security criterion subset the elements of which serve as keys. Keys are used to actuate the locks and the accessibility for a user to an XML document as well as its elements is determined by the status of the locks. To enforce the security policy systemically and consistently, the author developed a method in which security attributes are defined in enhanced XML schema and the security criterion expressions are embedded into XML documents automatically.

There are several advantages to define the security attributes in an enhanced schema. First, by embedding security criterion expressions automatically, security policy can be enforced consistently and errors are prevented. Secondly, security officers are relieved from the heavy burden of embedding the security information for every XML file and thus system performance is improved. Finally, the security level of the XML documents is further improved because any malicious modification to the security attributes (security criterion expressions) will be discovered when the documents are validated. As a result, the XML databases enhanced by this method enjoy both high level of security and high performance.

# 6 References

[1] B. Luo, D. Lee, W.C. Lee, and P. Liu QFilter: Fine-Grained Run-Time XML Access Control via NFA-Based Query Rewriting. In CIKM '04: Proceedings of the thirteenth ACM international conference on Information and knowledge management, November 2004, Pages 543-552

[2] D. Fallside and P. Walmsley XML Schema Part 0: Primer Second Edition. http://www.w3.org/TR/xmlschema-0/

[3] D. Lee, W. C. Lee and P. Liu. Supporting XML Security Models using Relational Databases: A Vision. In

XML Database Symposium (XSym), Berlin, Germany, 2003.

[4] E. Bertino and E. Ferrari. Secure and Selective Dissemination of XML Documents. ACM TISSEC, 5(3):290–331, Aug. 2002.

[5] T. Bray, J. Paoli, and C. M. Sperberg-McQueen (Eds). Extensible Markup Language (XML) 1.0 (2nd Ed.). W3C Recommendation, Oct. 2000.

[6] E. Damiani, S. De Capitani di Vimercati, S. Paraboschi, and P. Samarati. A Fine-Grained Access Control System for XML Documents. ACM TISSEC, 5(2):169–202, May 2002.

[7] E. Damiani, S. De Capitani Di Vimercati, S. Paraboschi, and P. Samarati. Design and Implementation of an Access Control Processor for XML Documents. Computer Networks, 33(6):59–75, 2000.

[8] G. Kuper, F. Massacci, and N. Rassadko Generalized XML security views. In SACMAT '05 Proceedings of the tenth ACM symposium on Access control models and technologies, Pages 77-84

[9] H. Thompson, D. Beech, M. Maloney, and N. Mendelsohn XML Schema Part 1: Structures Second Edition. http://www.w3.org/TR/xmlschema-1/

[10] L. Pan and C. N. Zhang A Criterion-Based Role-Based Multilayer Access Control model for multimedia Applications. In IEEE International Symposium on Multimedia (ISM2006). San Diego, Pages: 145-152

[11] L. Pan and C. N. Zhang A Criterion-Based Multilayer Access Control Approach for Multimedia Applications and the Implementation Consideration. In ACM Transactions on Multimedia Computing, Communications and Applications, Vol. 5, No. 2, Article 17, Publication date: November 2008

[12] M. Kudo and S. Hada. XML document security based on provisional authorization. In ACM CCS, 2000.

[13] M. Murata, A. Tozawa, and M. Kudo. XML Access Control using Static Analysis. In ACM CCS, Washington D.C., 2003.

[14] N. Rassadko Query rewriting algorithm evaluation for XML security views. In SDM'07 Proceedings of the 4th VLDB conference on Secure data management, Pages 64-80

[15] R. Rizvi, A. Mendelzon, S. Sudarshan, P. Roy. Extending Query Rewriting Techniques for Fine-Grained Access Control. In ACM SIGMOD 2004.

[16] P. Biron, K. Permanente and A. Malhotra XML Schema Part 2: Datatypes Second Edition. http://www.w3.org/TR/xmlschema-2/

[17] T. Bray, J. Paoli, C. M. Sperberg-McQueen, E. Maler, S. Microsystems, and F. Yergeau Extensible Markup Language (XML) 1.0 (Fifth Edition). http://www.w3.org/TR/xml/

[18] T. Yu, D. Srivastava, L. V.S. Lakshmanan, and H. V. Jagadish. Compressed Accessibility Map: Efficient Access Control for XML. In VLDB, Hong Kong, China, Aug. 2002.

[19] W. Fan, C.Y. Chan, and M. Garofalakis Secure XML querying with security views. In SIGMOD '04 Proceedings of the 2004 ACM SIGMOD international conference on Management of data, Paris, France — June 13 - 18, 2004, Pages 587-598

[20] Y. Diao and M. J. Franklin. High-Performance XML Filtering: An Overview of YFilter. IEEE Data Eng. Bulletin, Mar. 2003.

[21] Ekelhart, S. Fenz, G. Goluch, M. Steinkellner, and E. Weippl XML security - A comparative literature review. In Journal of Systems and Software, Volume 81 Issue 10, October, 2008

# Weaknesses in the EU Countries Cyber Strategies

**Kimberly Lukin**
Department of Information Technology, University of Turku, Turku, Finland

**Abstract –** *European Union countries are poorly prepared for cyber war and have not understood the benefit of centralized management in cyber security. If a country´s cyber preparedness is not directed from one centralized ministry, it is difficult to have adequate situation awareness and it is hard to put recovery actions into practice; such a country would be vulnerable in exceptional situations. Instead of countries making immediate large scale investments, their military performance could be increased in the short term by researching the weaknesses of other countries national IT security systems and improving defensive and offensive cyber attack methods. In this paper new recommendations for national cyber strategies for EU countries are presented. The collected data are analyzed to assess the existing weaknesses in cyber strategies and how the weaknesses can be and have been exploited. Based on this analysis, cyber strategy recommendations meeting modern requirements have been developed.*

**Keywords:** cyber security, critical infrastructure protection (CIP), cyber war, cyber domain, national cyber security strategy

## 1 Introduction

Warfare methods are evolving but new methods have rarely led to entire defense reform, as has been the case for cyber war countermeasures. Cyber war should be considered as a new weapon system [1]. Alongside the traditional methods of warfare, cyber war provides superiority through the destruction of the other party’s information infrastructure or by causing interruption in the information flow to their operative intelligence systems. Cyber war might increase the speed of initiation of war because there is no time for long-term mobilization; therefore one of the recommendations of this study is that governments should put into place robust cyber war management structures. These structures are also currently lacking in most of the EU member state strategies. Only 7 EU countries of the 27 member states have deployed cyber security strategies [2]. The cyber domain has not been defined in any of these strategies; neither has it been defined as an integral part of any other domains such as air, space, ground and sea forces. This raises the question of how countries could use cyber war alongside other war fighting methods, if it is not perceived as an independent weapon system and if it is considered to be just the protection of critical infrastructure. The slow reactions to cyber security issues in Europe has evoked questions such as whether EU nations are waiting for the EU to take a lead in cyber security and provide regulations for it, or whether the countries do not understand the risks associated with cyber attacks.

In related work McAfee´s cyber defense study claims that, along with Israel, Finland and Sweden both EU´s member states, are the leading “cyber readiness” countries in the world [3]. McAfee´s study reviews firewalls and antivirus protection, and well-informed governance and education. However the measures investigated by McAfee are not adequate to evaluate the IT security levels of governments and their critical infrastructure organizations. Also, the report does not evaluate the cyber security management structures of the countries´, which is the most critical part of cyber preparedness. The Economist Intelligence Unit cyber hub survey investigates countries´ ability to withstand cyber-attacks and deploy critical digital infrastructure, paying attention to strong digital development of the countries which according to their survey increases cyber power potential. However survey did not asses the security aspects of industrial applications, investments and practices in IT security and recovery processes, which should be part of cyber security metrics [4]. The main point of the I3P´s study was introducing critical interdependencies in the supply chain and process control metrics [5]. This approach is of practical importance since definitions are lacking in most of the EU countries cyber strategies. According to the ITU´s guide, governmental cyber security should be led by a cyber security coordinator [6]. However, this paper proposes a different approach, which will be discussed in Section IV.

This study also analyzes the weaknesses of EU states cyber strategies´ and puts forward recommendations on how protection levels could be raised. The recommendations are based on empirical studies and data collected from the EU´s official websites and research papers during the period 2011-2012. The basis of this research was the documentation of the EU´s security strategies, directives and the role of the EU institutions. It became obvious that during the past years the institutions have tried to take a leadership role in IT and cyber security in the EU without success. Other relevant documents that were reviewed in this study were the cyber security strategies of EU countries and ENISA´s country reports. Finnish governmental IT security projects and management structures were used when comparing the benefits of centralized and decentralized IT security. The rest of the paper is organized as follows: in Section II is a survey of the EU´s role in cyber security management. Section III reviews national cyber strategies in the EU. Section VI develops a specification and model for national cyber strategies based on the extensive review of existing strategies and their weaknesses given in Section III. The section also describes what strategies needed for good cyber preparedness and compares these strategies with existing ones. Section IV contains concluding remarks about this research.

## 2 Cyber security role of EU institutions

The EU has been slow to put together a comprehensive approach to cyber security and therefore countries have developed cyber security strategies independently. Also, the EU´s cyber and IT security responsibilities are decentralized between different agencies. However the EU has a security strategy [7] and a strategy for a Secure Information Society [8] both of which stress the importance of cyber security and the prevention of cyber crime while referring to “Network and Information Security: proposal for a European Policy approach” [9]. This document puts forward what additional steps should be adopted to improve network and information security. Use of this strategy is not yet mandatory for the EU countries since there are international security standards. The European Defense Agency (EDA), which was established in 2004, tried to start EU level cyber security cooperation in 2007 but this initiative did not succeed because most of the EU member states worked out particular questions with NATO. Therefore EDA continued its work on Network Enabled Capabilities and have developed a network capability vision and network operations risk analysis which should support the EU member countries defense development work [10]. The 2001 Council of Europe´s Convention on Cybercrime has been the basis for the development of the fight against cyber crimes. This agreement has received broad international support, and it is open to all countries, e.g. The United States has signed and ratified the agreement, but Russia has not [11].

The European Council proposed that the Commission prepare an overall strategy to protect critical infrastructure, this gained impetus after 9/11 in the United States. The Commission adopted the European Programme for Critical Infrastructure Protection [12] in which terrorist threats were a priority and it concentrated on the energy and traffic sectors but cyber threats not included. The Commission and the European Network and Information Security Agency ENISA are the chief interlocutors for cyber policy and they work closely with European member-states [13]. ENISA conducted the first pan-European cyber security exercise [14] which was meant for EU member states to test their cyber security preparedness in the fields of communication and coordination in relation to other states. According to the ENISA Executive Director, EU countries need a holistic cyber security strategy because the private sector and government need to share mutual information about cyber threats [15].

European Commission game up with the following two directive proposals: Comprehensive Reform of the Data Protection Rules [16] and a EU Privacy Framework [17], which EU countries have to fulfill if they are accepted. This is the first concrete step in the EU to control and lead cyber preparedness with legislation. The Directive insists that there should be a 24-hour data breach notification, mandatory security assessments and systematic security analyzes should be undertaken. When it formally became part of the European Union External Action, in 2008, the European Union Military Staff (EUMS) started preparations for cyber security development work. Plans for the EU leading cyber military operations [18] were published in 2010. The next level might be the EU planning to establish a European cyber security unit in the future.

## 3 EU countries cyber security strategies

This study examines some EU- countries´ official cyber security strategies and country reports that ENISA collects and publishes centrally, as well as the OECD´s reports, and other relevant research and articles. None of the strategies describe how an adequate preparedness level of cyber security is measured.

### 3.1 France

In France the French Network and Information Security Agency is responsible for national defense information systems, it also is responsible for reacting to cases of computer attacks aimed at the infrastructure of the state. The Operational Centre for Information Systems Security coordinates governmental issues in cyber crises [19]. One of the cyber strategy objectives is that France should be able to create cooperation at a new level [20]. To fulfill this objective France created a historical defense deal with the UK known as the Anglo-French Treaty. Counter-terrorism was named as one of cyber security objectives [21]. The French Government considers cryptography as the most important way of preserving strategic independence. It is fairly interesting that the strategy does not pay attention to other methods such as architectural protection. Furthermore, the strategy does not mention anything about cyber management even though one of the strategic objectives is that France must safeguard its ability to make decisions, this refers to the well-protected information exchange methods between governmental offices. According to its strategy, France aims to become a world power in cyber defense, but the execution program should be more comprehensive, as discussed in section IV, if this aim is to be fulfilled.

### 3.2 The United Kingdom

The UK is the only country that has clear sub-programs for the execution of cyber security measures. However, it has been claimed that UK cyber security strategy implementation is slow and needs concrete timetable, metrics and an implementation guideline. The strategy implementation should be ready by the end of 2015 [22]. It is remarkable that 59 percent of Cyber Strategy Programme investments come from the Single Intelligence Account since intelligence has a big role in cyber preparedness. Even cyber domain and control structures are missing from the strategy; the UK has invested in new cyber security units and re-organized its governmental structures. However, cyber and IT security management has been centralized between different offices. The Cyber Security Operations Centre is one of the three UK´s Intelligence Agencies and was established by the Cyber Security Strategy in 2009 [23]. The Cyber Security Strategy does not clarify how the new Joint Forces Command will manage development of cyber warfare and integrate it with other war domains. However, the implementation program itself is very ambitious and its sub programs are the most concrete of any of the EU country´s cyber strategies [24].

### 3.3 The Netherlands

The National cyber security strategy of the Netherlands is directed by the Ministry of Security and Justice. However, the Ministry of Economic Affairs, Agriculture and Innovation is responsible for developing network and IT security [25]. The strategy stresses cooperation with other countries, governments, NATO and the EU. The role of the EU is especially highlighted since the strategy itself says the Netherlands is a proponent of the broad ratification and implementation of the Cyber Crime Convention of the Council of Europe. This implies that the Netherlands is waiting for the EU to start supervising and coordinating cyber preparedness. In the Netherlands IT security and cyber security leading are decentralized between different ministries. According to the strategy they will establish a new Cyber Security Board, which will consist of public and private parties and National Cyber Security Center. The Netherlands aim to be amongst the world leaders in terms of use and deployment of ICT in society and ensuring the security of the digital society.

### 3.4 Norway

Norway published its National Cyber Defense Strategy draft in 2010. This pays attention to the fact that civilian information infrastructures that directly or indirectly support military operations, systems need to be protected [26]. This clause concretely guarantees that government is willing to manage critical infrastructure protection, including those of the private sector. The National Cyber Defense Center which is under the national security agency will be fully operational in 2013-2014 and it spent 100 million USD on cyber-related defense in 2010 [27]. It was noteworthy that the Norwegian government has a new approach to IT security; solutions are made permanent only after piloting them with the aim of preventing loss of investments [28].

### 3.5 Germany

According to the cyber security strategy of Germany effective IT security requires powerful structures in all federal authorities [29]. However responsibilities for IT security and cyber security management are widely decentralized. Even the Federal Ministry of the Interior is responsible for information security policy and the protection of critical information infrastructures; the Federal Office for Information Security is the central IT security service provider for the government. In contrast, information security policy and ICT policies are part of other ministries. Moreover, cyber security strategy brings new functions and offices. The cyber strategy highlights shared responsibilities between the state, industry and society but also with other multinational organizations such as the United Nations, the EU and NATO. Germany supports the idea of increasing cyber security coordinator role for the European Network and Information Security Agency [30].

### 3.6 The Czech Republic

The Czech cyber security strategy is the only strategy which emphasizes that IT security is not a separate part of cyber security-, preparedness or war [31]. The cyber strategy does not highlight the governmental management structures and their role in cyber strategy Cyber and Informational Security Department is under guidance of the Ministry of the Interior. The Ministry of the Interior is also responsible for information security coordination in public administration which includes coordination of critical infrastructure protection. The cyber strategy does not present any execution programs because they are still being developed.

### 3.7 Estonia

Estonia has clear sub-programs to raise cyber capability and it is one of the few countries in the EU which have the ability to conduct cyber operations against other countries, because it can mobilize a volunteer Cyber Defense League which would work under military command during war and Estonia has experience of large scale cyber attacks. The League consists of programmers, computer scientists and software engineers [32]. Estonia is also the only country developing a system of security measures that would provide action plans when it is under cyber attack. Estonia also has a NATO CCOED center that was created after the Bronze Solder incident [33].

## 4 Recommendations

This Section puts forward recommendations for a national cyber security strategy model. The EU countries cyber strategies are focused on strategic defense rather than prior offensive methods, which has resulted in the absence of a definition for a cyber domain. The fact remains that a cyber attack is a method of warfare and should not be considered just an issue of security and preparedness. This Section introduces the requirements for cyber security domain and action plans all of which are absent in the EU countries´ cyber strategies.

## 4.1 The cyber domain

A national cyber security strategy model should be based on the cyber domain, which according to this study should consist of six sub-sections. Without these six sub-sections the cyber domain will not have full war fighting capability. The functionality of each unit in the domain will define its success on the battlefield. If one unit were not working properly success and time span on a battlefield would be reduced. A cyber domain should provide a robust platform for cyber warfare. The Pentagon, the headquarters of the United States Department of Defence, classifies the Internet as a war domain [34]. A cyber domain should consist of:

1) A centralized cyber security management structure. None of the EU´s cyber strategies define the importance of this structure. There might be for historical reason that; decentralization and local governance are recognized as basic components of democratic governance [35]. According to this study the Finnish government decentralized IT management has led to the development of separate systems, which are difficult to unify. That will prevent forming real time situational awareness in recovery situations. One example is the Finnish State Treasury´s VALDA project which cost the Finnish taxpayers 9 million Euros; but because of the problems in integrating it to disparate systems it was never taken into extensive use [36]. VALDA was supposed to enable organizations to manage a variety of digital processes securely.

In agreement with the findings of this study, Hellman´s study showed that without centralized management many enterprises have found that their security programs have evolved into a complex patchwork of disparate systems that generate an overwhelming flood of data but offer little visibility of true threats and attacks [37]. Cyber preparedness should be led by one ministry that has the knowledge of information security levels of all state institutions and is capable of improving situation awareness, analyzing and sharing information and leading recovery operations. The military should only be responsible of executing cyber warfare operations. Without centralized management, government and critical infrastructure organizations will lose situational awareness which is crucial during large scale cyber attacks and in recovery situations.

2) An active defense through sensor network for governmental and critical infrastructure organizations will recognize the first waves of cyber attacks such as abnormal network flow and attack vectors. Protection of electronic equipment, the electromagnetic spectrum, networks and data which is handled stored and transferred by them.

3) A command-and-control system, which gathers information to provide situation awareness and support decision making on the basis of combined and analyzed data.

4) Cyber warfare should be tied into other war fighting methods and information warfare operations. Cyber war might blur traditional borders of different military branches in the future and unify them because many military weapon systems are based on IT systems. Combining military branches may have an important cost saving effect as has been the case for unified the AirSea battle (ASB) concept which was introduced in 2010 [38].

5) Offensive cyber warfare methods. There is a need to find methods to conduct counter attacks against attackers systems, sensors, information and to cause interference.

6) Telecommunication interception and signal intelligence will produce information about security issues relating to other parties' arguments and behaviours; these could be used for recognizing possible targets in cyber war.

## 4.2 Required changes

The action plans introduced here absent from the EU countries cyber strategies.

1) Ensure that telecommunication connections nationwide, including rural areas, function correctly even under extreme conditions. In the EU, ENISA stresses the importance of cooperation between governments and their public sectors. However the process has been slow because of the historical burden when there was a certain hierarchy between a government, companies and the public sector. Also the EU´s strict corrupt legislation might restrict cooperation between governments and the private sector. Furthermore, the European Commission frequently invites public sector representatives to give their opinions but they are not established participants in the EU´s decision making procedures. Whether real cooperation is replaced with artificial actions or not it will be difficult to commit organizations to work together in recovery situations. Furthermore, naval studies have claimed that: “transfer of data and situational awareness across the “civil-military boundary” was a huge problem because civil organizations do not respect any higher authority” [39]. However, this problem would not occur if government classified critical civil functions and critical infrastructure organizations received benefits such as financial support for recovery actions or technical auditing help. Also there should be exercises to rehearse the required recovery actions and binding contracts should be drawn between the government and civilian organizations for their implementation in times of war.

2) A cyber security strategy enforcement program should raise a country’s cyber warfare capability in the short term. Cyber threats are evolving fast so reliable defense requires operational short-term projects and strategic long-term development plan.

3) Government should increase in-house technical expertise instead of obtaining it by outsourcing. The latter method of

obtaining expertise makes it difficult to commit people to work in a crisis situation or get the needed level of service. State organizations with preparedness role should secure their own technical expertise, especially in the case of confidential IT systems.

4) Governments should support the cyber security industry and increase investments in cryptographic research and the overall capability for national intelligence. There should be technical level conversations at the expert level regularly since governmental institutions need an external point of view.

5) The EU should create an electronic defense unit in the field of cyber security. This would allow other member states to engage in cyber warfare if any member state is under a cyber attack. Game theory should define responsibilities of other countries, e.g. roles could involve recognizing and isolating botnets. Nordic countries already share a Nordic resource network to protect against cyber attacks that is based on information sharing, but it does not have role as defence unit offering support for countries under cyber attack. The Lisbon Treaty [40] which was signed in 2009 by EU members addresses the mutual assistance of countries during crisis, but not yet at the level of cyber security.

6) Recognize critical infrastructure, critical processes and raise their protection levels. In the EU countries 80 percent of critical infrastructure is in the hands of the civil sector, the government is responsible for 15 percent and the remaining 5 percent is taken care of by the military. However, governments are responsible for leading their entire countries´ recovery actions; this of course includes the civil sector. This may cause problems because critical infrastructure organizations are not ready to invest in duplications or recovery actions. Another challenge for the EU countries is how to raise critical infrastructure protection levels after their rundown following the World War II. In most cases countries have sold or outsourced to other countries a major part of their critical infrastructure, this would prevent commitment to recovery actions [41]. The EU does not yet have a critical infrastructure protection directive for ICT. The EU´s approach to critical infrastructure classification is that if a country has an organization or function which affects other member countries then it is part of the EU´s critical infrastructure. Interdependence of the EU countries also causes risks, e.g. the EU´s Single Europe Payments Area (SEPA) money transfer system belongs to this category. There is no consensus yet on whether the EU will economically support these organizations in the future. Another challenge is that a critical infrastructure protection directive might be interpreted differently in each EU country because of the different legal systems in each member state, and this may cause variations in the enforcement of the directive [42]. Furthermore, critical processes should be recognized, such as how to ensure that supply chain components of critical systems are available in the event of failure and supply chain should be guaranteed by contracts.

## 4.3 Technical recommendations

In cyber security and war, even though it is important to give IT security guidance for the people the main focus should be on technical protection methods, surveillance and developing of attack methods. These technical counter actions are absent from EU countries cyber strategies.

1) Cyber warfare could be conducted on several levels. Supportive cyber attacks could be done via an operative system or website which would hide the attack structure and the used methods, such as use of botnets in DDoS attacks. This kind of method was used in the Georgian – Russian conflict; A Russian website offered the possibility for everyone to participate in targeted DDoS attacks against Georgia by downloading free software that made computer part of the botnet [43]. This kind of attack could be conducted via non-specialists to give support to other warfare actions. Technical experts in a cyber unit could undertake more sophisticated cyber warfare methods and the developing of computer viruses.

2) Technical actions for protecting operational networks and the electromagnetic spectrum are needed. Attacks against embedded systems are increasing. Since they control aircraft engines, weapon systems, networked sensors and industrial systems they will be major cyber attack targets in the future. There should be self-protection modes that isolate part of a system when it´s not working as it was designed to, it should have inner authentication methods that continuously recognize deviations. The active protective measures counteract threats at the early stages of an attack.

3) Application and crypto security play a major role in cyber security. Governmental institutions and critical infrastructure organizations should have a baseline for software development standards or give recommendations for reliable cryptographic software to be used in high security level information systems. Governments should pay attention to the quality of cryptographic methods used in classified systems. In Russia, a British subsidiary company was banned from developing cryptographic solutions on grounds of national defense and security [44].

4) There is a need for a command and control management system where situation awareness and distributed information is analyzed and shared with critical infrastructure organizations, ministries and intelligence units. If cyber attack is aimed at the critical infrastructure organization, other organizations should receive information of the attack without delay because the same attack could affect them e.g. if their applications have the same vulnerability. In the Georgian conflict Russian aimed DDoS attacks against Georgian government websites before its troops and tanks crossed over the Georgian border. Cyber attacks caused information

distribution problems between government and people in Georgia and other countries [45]. Information of the cyber attacks did not reach all the critical infrastructure organizations fast enough. The Georgian government could have prepared against the Russian invasion if they had shared situational awareness in properly protected telecommunication networks and also if actions had been lead from a centralized command center.

5) There is a need for research into the weaknesses of other countries IT security vulnerabilities such as assessment of the likelihood of success of different attack vectors. Governments should examine changes in the security environment and find methods affect to other countries´ information systems, sensors and information exchange. Furthermore, governments should be aware of failure points in their most critical systems and ensure their functionality in every situation.

6) Governments should publish standards for application security, indicating whether to use open source or COTS applications and concentrate on application security to protect their high security level and embedded systems.

7) Network and telecommunication connections between most the important cities should be ensured, encrypted and recovery actions rehearsed in order to ensure connections work in every situations e.g. weather conditions, interference and in a war. The European power grid was not constructed to survive to cross-border power exchanges; this could cause serious bottlenecks [46]. The EU has focused more on trading than on reliability which has lead to vulnerable in the electricity grid. The situation is about to change because a single European energy grid is under development and ready in 2020 [47].

8) Telecommunication interception and signal intelligence would offer possibilities to create automated monitoring and tracking methods in order to prevent cyber terrorism. Collected and analyzed intelligence information could be used as target acquisition in cyber attacks. Also information exchange between intelligence agencies would help to form situation awareness.

9) There should be a common architecture both at the network and application levels which would counteract the success of attack vectors. Governmental operative systems, application security standards, architectural and platform solution and encryption methods should differ from those used in from non-classified systems. Security hardening should be based on preventing the success of different attack vectors.

10) According to Russia´s Cyber Strategy, Information warfare is an integral part of cyber war [48]. In cyber war information superiority is needed, which means that countries should rehearse scenarios where they control the Internet and social media. Misinformation might lead to riots or civil war as happened in the Arab spring [49]. Information warfare and signal intelligence should be basic components of cyber war because they offer information superiority.

# 5 Conclusions

In this study the absence of centralized cyber and IT security management model of the EU countries cyber strategies is revealed. Decentralized management jeopardizes a country´s ability to recover from a large scale cyber attacks. Moreover the definition of cyber domains and methods to integrate cyber action with other war fighting methods is missing from the cyber strategies. If domain functions are not recognized and one of them is not working properly, it will affect country´s capabilities to fight a war on the battlefield. The EU countries have spent great amounts of money on cyber security and have been forced to make cuts in other governmental functions in order to create new cyber units. Preparing against new threats requires resources that small countries usually do not have. Cyber security defense capability should be built up gradually, instead of with immediate large-scale investments. Military cyber performance could be raised in the short term by research into the weaknesses in countries´ national IT security systems and by simulating offensive cyber attack methods. If the EU institutions are willing to lead cyber security they cannot avoid the question of whether there will be unified military forces for the EU, since cyber is a new warfare domain and it needs to be tied to other war fighting methods. This will cause other challenges such as how the EU can lead cyber war actions against other nations, how EU member states should engage into cyber war and how legal basis for such actions should be prepared. If the EU takes the lead in ensuring the cyber security and preparedness of its member countries it would be a new angle of approach toward a military alliance. This study attempts to define the missing structures of EU countries cyber security strategies. Recommendations given in this study will raise countries´ capabilities to conduct successful cyber warfare operations and enhance their ability to cooperate in the case of severe cyber attacks.

# 6 References

[1] The Economist, *War in the fifth domain*, July 1st 2010. [Online]. Available: http://www.economist.com/node/16478792

[2] ENISA, European Cyber Security Strategies, 6th March 2012, [Online]. Available: http://www.enisa.europa.eu/search?SearchableText=cyber+strategy

[3] BBC, *Israel tops cyber-readiness poll but China lags behind*, 30th January 2012, [Online]. Available: http://www.bbc.co.uk/news/technology-16787509

[4] The Economist Intelligence Unit, Cyber Hub survey, [Online]. Available: http://www.cyberhub.com/CyberPowerIndex

[5] Institute for Information Infrastructure Protection, *National Cyber Security Research and Development Challenges*, 2009, http://www.cyber.st.dhs.gov/docs/i3pnationalcybersecurity.pdf

[6] ITU NATIONAL CYBER SECURITY GUIDELINE, 2009, http://www.itu.int/ITU-D/cyb/cybersecurity/docs/ITUNationalCybersecurityStrategyGuide.pdf

[7] European Security Strategy, Brussels 12Th December 2003, [Online]. Available: http://www.consilium.europa.eu/uedocs/cmsUpload/78367.pdf

[8] A Strategy for a Secure Information Society – "Dialogue, partnership and empowerment, 2006, [Online]. Available: http://ec.europa.eu/information_society/doc/com2006251.pdf

[9] Network and Information Security: *Proposal for a European Policy Approach*. [Online]. Available: http://ec.europa.eu/information_society/eeurope/2002/news_library/pdf_files/netsec_en.pdf

[10] The EU concept of computer network Operations, EDA 16th August 2010, Functional and Technical Specifications 10-CAP-OP-37 [Online]. Available*:* www.eda.europa.eu/WebUtils/downloadfile.aspx*?*

[11] Convention, 2001, http://conventions.coe.int/Treaty/Commun/QueVoulezVous.asp?NT=185&CM=8&DF=6/6/2006&CL=ENG

[12] Critical Infrastructure Protection, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:345:0075:0082:EN:PDF

[13] Bertelsmann Foundation, Euro wire, July 2011, [Online]. Available: http://www.bfna.org/sites/default/files/publications//EuroWire%20April%202012_0.pdf

[14] ENISA Cyber security exercise, http://www.prnewswire.com/news-releases/interim-findings-of-cyber-europe-2010-the-first-pan-european-cyber-security-exercise-a-successful-cyber-stress-test-for-europe-107032268.html

[15] ENISA, 24th February, 2010, *Member States need holistic cyber security strategies*, [Online]. Available: http://www.enisa.europa.eu/media/news-items/v2member-states-need-holistic-cyber-security-strategies

[16] EU NEC VISION, 2009, *Extract from the NECK Vision*, [Online]. Available: http://www.eda.europa.eu/Libraries/Documents/Extract_from_NEC_Vision_Report.sflb.ashx

[17] European Commission, Brussels 25.1.2012, *Regulation of the European Parliament and of the Council*, [Online]. Available: http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf

[18] EU Data Protection Directive, 2012, [Online]. Available: http://epic.org/privacy/intl/eu_data_protection_directive.html

[19] France country report, [Online]. Available:http://www.enisa.europa.eu/activities/stakeholder-relations/files/country-reports/France.pdf

[20] ENISA, 2011, *Information Systems Defense and Strategy*, [Online]. Available: http://www.enisa.europa.eu/media/news-items/Information_system_security_France_strategy.pdf

[21] The Offcial Site of British Prime Minister´s Office, *UKFrance summit 2010 declaration on defence and security co-operation*, 2th November 2010, [Online]. Available: http://www.number10.gov.uk/news/uk%E2%80%93france-summit-2010-declaration-on-defence-and-security-co-operation/

[22] Computerworld, *UK cyber security strategy implementation is too slow*, 8th February, 2012, [Online]. Available: http://www.computerworlduk.com/news/security/3335972/uk-cyber-strategy-implementation-too-slow-says-former-security-minister/

[23] ENISA, *United Kingdom Country Report*, 2011, [Online]. Available: http://www.enisa.europa.eu/activities/stakeholder-relations/files/country-reports/UK.pdf

[24] The Cabinet Office, *the UK cyber Security Strategy, Protecting and promoting the UK in a digital world*, 25th November 2011, [Online]. Available: http://www.cabinetoffice.gov.uk/sites/default/files/resources/uk-cyber-security-strategy-final.pdf

[25] The Netherlands country report, [Online]. Available: http://www.enisa.europa.eu/activities/stakeholderrelations/files/country-reports/Netherlands.pdf

[26] Høringsuttalelse fra INI - Forslag til strategi for Cybersikkerhet (English: *Proposed strategy for cyber space),* http://www.regjeringen.no/pages/2534053/Cybersikkerhet_svar-med-merknader_Forsvarets-informasjonsinfrastruktur.pdf

[27] Defense News, *Norway Massacre adds Urgency to Cyber Strategy*, 15th August 2011, http://mobile.defensenews.com/story.php?i=7388379&c=FEA&s=SPE

[28] OECD Studies in Risk Management, Norway Information Security, 2006, http://www.oecd.org/dataoecd/36/16/36100106.pdf

[29] Germany country report, 2011, [Online]. Available: http://www.enisa.europa.eu/activities/stakeholder-relations/files/country-reports/Germany.pdf

[30] Cyber Security Strategy for Germany, 25th 2011, [Online]. Available: www.enisa.europa.eu(media/news-items/german-cyber-security-strategy-2011-1

[31] ENISA, Czech, [Online]. Available: http://www.enisa.europa.eu/media/news-items/CZ_Cyber_Security_Strategy_20112015.PDF

[32] The Daily Mail, 2008, [Online]. Available: http://www.dailymail.co.uk/sciencetech/article-1344402/Estonia-trains-army-experts-protect-cyber-attacks.html

[33] Ministry of Defence of Estonia,: *Cyber Security Strategy*, [Online]. Available: http://www.mod.gov.ee/files/kmin/img/files/Kuberjulgeoleku_strateegia_2008-2013_ENG.pdf

[34] Bennet John. T, Pentagon declares The Internet a war domain, 15th July 2011, http://www.globalresearch.ca/index.php?context=va&aid=25645

[35] Nikolov, Dimce, St.Petersburg, 28-30 September, 2006. *Decentralization and decentralized governance for enhancing delivery of services in transition conditions*, [Online]. Available: http://unpan1.un.org/intradoc/groups/public/documents/un/unpan025134.pdf

[36] Tietoviikko, 30th March 2012, *Valtio tuhlasi 9 miljoonaa euroa tietojärjestelmään jolle ei löytynyt käyttäjiä* (in English: Government squandered 9 million to computer system which did not gain enough users, [Online]. Available: http://www.tietoviikko.fi/kaikki_uutiset/valtio+tuhlasi+9+miljoonaa+euroa+it+jarjestelmaan+jolle+ei+loytynyt+kayttajia/a795414

[37] Hellman, Gretchen, *From Logs to Logic, Best practices for security Information Management* [Online]. Available: http://www.infosectoday.com/Articles/logstologic.htm

[38] Wall Street Journal, *Battle Plans Tempt Chill in U.S.-China Relations* 10th November 2011, http://blogs.wsj.com/chinarealtime/2011/11/10/battle-plans-tempt-chill-in-u-s-china-relations/

[39] Denning J. Peter, Hayes-Roth Rick, *Decision making in Very Large Networks,* USA, Volume 14, Issue 11, November 2006. [Online]. Available: http://dl.acm.org/citation.cfm?id=1167852&bnc=1

[40] EU Watch, *Lisbon Treaty: Common Defense, Neutrality & Interventionism*, [Online]. Available: http://www.ffeud.eu/uploads/file/euwatch_11.pdf

[41] Interview, State IT Director Benson Yrjö, Ministry of Defense, 2012.

[42] Interview, Senior Advisor Kari Ojala, Ministry of Transport and Communication, 2012.

[43] The Jamestown Foundation, *The Cyber dimension on Russia´s attack on Georgia*, http://www.jamestown.org/single/?no_cache=1&tx_ttnews%5Btt_news%5D=33936

[44] Россия: ИТ-компании запретили управлять собственной "дочкой" ради "безопасности страны" (In English: Russia: IT companies banned their own "daughter" company for the sake of national security, Online]. Available: 30.3.2012, http://www.cnews.ru/news/top/index.shtml?2012/03/30/483814

[45] CCDCOE, Cyber *Attacks Against Georgia: Legal Lessons Identified*. Eneken Tikk, Kadri Kaska, Kristel Runnimeri, Mari Kert, Anna-Maria Talihärm, Liis Vihul. 2008 [Online]. Available: www.carlisle.army.mil/DIME/documents/Georgia%201%200.pdf. Verified 2012-02-25

[46] ABB, EU Directorade General for Energy and Transport, 2000. *The European power Grid – the need for regulatory changes and advanced technology*, [Online]. Available: http://www02.abb.com/GLOBAL/GAD/GAD02181.NSF/viewunid/820203085E6C3B2BC1256DB7003EF8A6/$file/Grid+Issues+in+Europe.pdf

[47] Climate Change Corp, 8th May 2008, *The single European energy group*, [Online]. Available: http://www.climatechangecorp.com/content.asp?ContentID=5309

[48] Conceptual Views Regarding the Activity of the Armed Forces of the Russian Federation in the Information Space, 2012. [Online]. Available: http://www.ens.mil.ru/science/publications/more.htm?id=10845074@cmsArticle

[49] Oxford Islamic Studies [Online]. Available: http://www.oxfordislamicstudies.com/Public/focus/essay0611_social_media.html

# A Multi-attribute Evaluation of Information Security Controls in Organizations Using Grey Systems Theory

**A. Ejnioui**[1], **A. R. Otero**[2], **G. Tejay**[2], **C. E. Otero**[1] **and A. A. Qureshi**[3]
[1]Information Technology, University of South Florida Lakeland, Lakeland, Florida, USA
[2]Computer and Information Sciences, Nova Southeastern University, Fort Lauderdale, Florida, USA
[3]Mathematics & Computer Science, University of Virginia's College at Wise, Wise, Virginia, USA

**Abstract –** *Although the protection of information is critical, organizations have endured information losses that affected the values of their assets. As a result, organizations are currently driven to find ways to implement effective information security controls to protect their critical and sensitive information. Methods to implement these controls such as risk analysis or best practice guidelines have been proposed in the past. However, these methods tend to be limited due to their subjective nature and fail to take into account specific constraints such as implementation costs, resource availability, and scheduling limitations. This paper proposes a new approach where the problem of implementing information security controls is formulated as a multi-attribute decision making using grey system theory. The problem is addressed by devising a utility function for ranking different possible implementations of information security controls. This approach is applied in a case study where it is clearly shown how effective and highly customizable this approach is in evaluating the quality of information security controls with regard to the security needs of an organization.*

**Keywords:** information security; information security controls; risk analysis and management; baseline manuals; best practice frameworks; grey systems; multi-attribute decision.

## 1 Introduction

Protecting information is of utmost importance in most organizations. Many organizations have endured substantial information losses that impacted significantly the value of their assets and information. In fact, losses related to information security will continue to occur with a devastating effect on organizations [1]. In 2004, the CCI/FBI Computer Crime and Security Survey stated that total losses in the United States attributable to computer security breaches reached \$141,496,560. These alarming figures point to an inadequacy in today's information security practices and serve as motivation for finding new ways to help organizations improve their capabilities for securing valuable information.

In today's organizational culture, the use of security tools and technologies, such as encryption, firewalls, and access management are used to address security challenges [2, 3]. Although tools and technologies are an integral part of organizations' information security plans, it is argued that they alone are not sufficient to address information security problems [4]. To improve overall information security, organizations must evaluate (and thus implement) appropriate information security controls (ISC) that satisfy their specific security requirements [5-7]. However, due to a variety of organizational-specific constraints (e.g., cost, schedule, resources availability), organizations do not have the luxury of selecting and implementing all required ISCs. Therefore, the selection, adoption, and implementation of ISCs within organizations' business constraints become a non-trivial task.

This paper proposes a novel approach for evaluating and identifying the most appropriate ISCs based on organization specific criteria. The proposed approach uses an attribute-based utility measure in grey systems theory to quantify the desirability of each ISC taking into account benefits and penalties (restrictions) associated with implementing the ISC. This provides Management with a measurement that is representative of the overall quality of each ISC based on organizational goals. The derived quality measurement can be used as the main metric for selecting ISCs. The remainder of the paper is organized as follows. Section 2 provides a summary of previous work on ISC selection. Section 3 briefly describes the proposed solution approach. Section 4 provides detailed explanations of grey systems theory and how it is used in multi-attribute decision making. Section 5 presents the results of a case study. Finally, Section 6 provides summarized conclusions and highlights of the proposed approach.

## 2 Related Work

Various reasons have been put forth for explaining the lack of effectiveness in the evaluation, selection, and implementation process of ISCs. Sometimes, the implementation of ISCs in organizations may constitute a barrier to progress [8]. Employees may view ISCs as interrupting their day-to-day tasks and therefore, may ignore implementing them in order to be effective and efficient with their daily job tasks [9]. According to [10], organizations are required to identify and implement appropriate controls to ensure adequate information security. Others place emphasis on the fact that "different organizations have different security needs, and thus different security requirements and objectives" [11]. In

[12], the authors stress that there is no single information security solution that can fit all organizations. As a result, ISCs must be carefully selected to fit the specific needs of the organization.

In [5], the authors claim that the process of identifying and selecting the most effective ISCs in organizations has been a challenge in the past, and plenty of attempts have been made to come up with the most effective way possible. Risk analysis and management (RAM) is just one example. RAM has been recognized in the literature as an effective approach to identify ISCs [5]. RAM would list the information security requirements as well as the proposed ISCs to be implemented to mitigate the risks resulting from the analyses and assessments performed. RAM, however, has been described as a subjective, bottom-up approach, not taking into account organizations' specific constraints [13]. Management must, therefore, explore new ways to determine/measure the relevancy of these ISCs considering the various constraints experienced by organizations.

Another option that has been considered by many organizations for introducing security controls is to adopt best practice frameworks [5]. Based on [10], best practice frameworks assist organizations in identifying appropriate ISCs. Some best practices include: Control Objectives for Information and related Technology (COBIT); Information Technology Infrastructure Library (ITIL); Operationally Critical Threat, Asset and Vulnerability Evaluation (OCTAVE); International Organization for Standardization (ISO) / International Electrotechnical Commission (IEC) 177995 and ISO/IEC 27001; PROTECT; Capability Maturity Model (CMM); and Information Security Architecture (ISA) [6].

The process of selecting the most effective set of ISCs from these best practice frameworks can be challenging. According to [13], best practice frameworks leave the choice of controls to the user, while offering little guidance in terms of determining the best controls to provide adequate security for the particular business situation. Additionally, frameworks do not take into consideration organization specific constraints, such as, costs of implementation, scheduling, and resource constraints. Other less formal methods used in the past, such as, *ad hoc* or random approaches, could lead to the inclusion of unnecessary controls and/or exclusion of required/necessary controls [5]. Identifying and selecting ISCs based on the above may result in organizations not being able to protect the overall confidentiality, integrity, and availability of their information [10]. In order to increase the effectiveness of the selection and prioritization process for ISCs, new methods need to be developed that save time while considering major factors (e.g., constraints, restrictions, etc.) that undoubtedly affect the selection of ISCs.

From the reviewed literature, it is evident that the selection of ISCs is mostly driven by cost, scheduling, and resource availability. In other words, ISCs at organizations will be selected by management when the benefits of implementing them surpass the costs of establishing the control. Equally important, scheduling issues may affect whether ISCs should be selected. Implementation of ISCs may require specific scheduled times, not necessarily planned by the organization. Finally, availability of personnel often determines whether ISCs can be selected or not. Effective information system security implementation requires the identification and adoption of the most appropriate and effective set of ISCs taking into account the issues presented above [13].

# 3 Solution Approach

To properly evaluate the quality, importance, and priority of ISCs in organizations, management must follow a methodology that takes into consideration the quality attributes of the ISCs that are considered relevant. The methodology must provide capabilities to determine the relative importance of each identified quality attribute. This would allow the methodology to provide an ISC selection scheme that represent how well these ISCs meet quality attributes and how important those quality attributes are for the specific organization. To achieve this, a methodology using grey systems theory to solve a multi-attribute decision problem is proposed. First, a set of quality attributes is identified as evaluation criteria for all possible ISCs. These attributes are defined in terms of different features, where the importance of each feature is expressed as a grey number. ISCs that satisfy the highest number of features would expose a higher level of quality for that particular quality attribute. After all ISCs are evaluated and measurements computed for all features, the proposed approach uses an additive weighted utility function to fuse all measurements into one unified value that is representative of the overall quality of the ISC. This unified value is computed by using a set of utility values that take into consideration the importance of each quality attribute. Therefore, the resulting ranking of each ISC is derived based on the goals and specific needs of the organization. This results in an ISC ranking approach based on how well ISCs meet quality attributes and how important those quality attributes are for the organization.

# 4 Multi-attribute Decision Making in Grey Systems Theory

Multi-attribute decision making problems occur in situations where a finite set of alternatives need to be evaluated according to a number of criteria or attributes. The evaluation consists of selecting the best alternative or ranking the set of alternatives based on those attributes. However, many decision problems present data that is imprecise or ambiguous leading to conflicting situations in which the evaluation of alternatives becomes difficult. This is the case when implementing ISCs in organizations. In the past, this information uncertainty has been modeled using fuzzy sets [14] or grey numbers [15]. While the former has been around for some time, only recently has interest been growing in the latter, since uncertainty can be modeled and manipulated in more flexible ways using grey number systems than fuzzy sets [15].

### 4.1 Grey Numbers and Grey Systems Theory

In practical applications, a grey number represents an indeterminate number that takes its possible value from an interval or a set of numbers. The symbol $\otimes$ denotes a grey number. The most basic types of grey numbers are [15]:

- Grey numbers with only a lower bound: $\otimes \in [\underline{a, \infty]}$ or $\otimes(\underline{a})$, where $\underline{a}$ is a fixed number representing the lower bound.
- Grey numbers with only an upper bound: $\otimes \in [-\infty, a]$ or $\otimes(\overline{a})$ where $\overline{a}$ is a fixed number representing the upper bound.
- Interval grey numbers: $\otimes \in [\underline{a}, \overline{a}]$ where $\underline{a}$ and $\overline{a}$ are the lower and upper bounds respectively.
- Continuous and discrete grey numbers: The former numbers can take any values within an interval while the latter can take only a finite number of potential values.
- Black and white numbers: When $\otimes \in [-\infty, \infty]$, that is when $\otimes$ has neither an upper nor lower bound, it is knows as a black number. On the other hand, when $\otimes \in [\underline{a}, \overline{a}]$ and $\underline{a} = \overline{a}$, it is knows as a white number.

After three decades of research, grey systems theory emerged as new discipline with contributions in [15]:

- Grey algebraic systems, grey equations, grey matrices, etc.
- Sequence operators and generation of grey sequences.
- System analysis based on grey incidence spaces and grey clustering.
- Grey prediction models.
- Decision making using grey target decision models.

Optimization models using grey programming, grey game theory and grey control.

### 4.2 Selection of Information Security Controls

The first step involves identifying a set of ISCs that could be implemented in the organization. These ISCs can be obtained from the best practice frameworks mentioned in Section II. For instance, the ISO/IEC 177995 standard has over 127 ISCs available according to the organizations' specific needs [10]. Once selected, the results of these ISCs are captured in the ISC vector as:

$$I = \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_n \end{bmatrix} \qquad (1).$$

### 4.3 Information Security Attributes

When planning to implement ISCs, it is often necessary to address important aspects of security. These aspects can be viewed as quality attributes in the decision problem. For instance, one can use the following quality attributes that have been defined in the ISO/IEC 177995 standard [10]:

- *Restrictions* – there are restrictions that Management must take into account before selecting and implementing ISCs. These may include whether the costs involved in the selection and implementation of the ISC are high, whether resources are not available, and whether there are scheduling constraints associated with implementing the ISCs. The presence of any of the above will negatively affect the specific quality attribute. That is, ISCs with all features present will result in a lower priority; conversely, ISCs with all features missing will result in a higher priority. A high priority scenario will be one where the implementation cost of the specific ISC is considered adequate and/or manageable (i.e., within budget), resources are available to implement the particular ISC, and there are no restrictions in terms of scheduling the ISC (i.e., the ISC can be scheduled anytime during the year). Restrictions are defined as: Costs (C), Availability of Resources (AoR), and Scheduling (T).
- *Scope* – This quality attribute assesses the impact of the ISCs on the organization. ISCs that provide security of information in many systems have a higher priority than ISCs that address security of information in a minimal number of systems. Scope is defined as: System 1 (S1), System 2 (S2), …, System n (Sn).
- *Organization's Objectives* – the number of information security objectives the ISC satisfies. As the number of objectives the ISC satisfies, so does its suitability. Organization's objectives are defined with the following features: Objective 1 (O1), Objective 2 (O2), …, and Objective n (On).
- *Physical Access* – ISCs will prevent and/or record unauthorized access to the organization's building facilities, including data centers where information processing takes place, the finance/accounting department, human resources department, etc. As the number of physical locations addressed by the ISC, so does it suitability for selection. Physical access is defined as: Location 1 (L1), Location 2 (L2), …, Location n (Ln).
- *Access Controls* – implementation of an ISC for this quality attribute will promote appropriate levels of access controls to ensure protection of the organization's systems/applications against unauthorized activities. Organizations may implement network access controls (N), operating systems access controls (O), and application controls (A) based on their specific needs.
- *Human Resources* – implementation of an ISC supports reductions of risk of theft, fraud, or misuse of computer resources by promoting information security awareness (Aw), training (Tn), and education of employees (E). Depending on the particular situation, costs involved, and availability of personnel, organizations may select which of these to employ.
- *Communications and Operations Management* – ISCs will ensure the correct and secure operation of information processing facilities, which includes addressing for adequate segregation of duties (SoD), change management (CM), and network security (NS).

Organizations may select ISCs to address all of these or just some depending on their particular needs.

- *Systems Acquisition, Development, and Maintenance* – ISCs will support security related to the organization's in-house and/or off-the-shelf systems or applications (e.g., ensuring personnel with authorized access can move changes into production environments, etc.). As the number of systems or applications addressed by the ISC increases, so does the suitability of selecting the ISC. Systems Acquisition, Development, and Maintenance is defined as: Systems or Applications 1(SoA1), Systems or Applications 2 (SoA2), …, and Systems or Applications n (SoAn).
- *Incident Management* – ISC will ensure that security-related incidents (e.g., attempts to change/manipulate financial data, etc.) identified within the organization's processing of information are communicated in a timely manner, and that corrective action is taken for any exceptions identified. Incident management may apply to online processing and/or batch processing. Incident Management is defined as Processing 1 (P1), Processing 2 (P2), …, and Processing n (Pn).

The above quality attributes can be represented in the following vector:

$$A = [q_1\, q_2\, \cdots\, q_m] \qquad (2)$$

for $j = 1, 2, \ldots, m$.

### 4.4 Feature Aggregation

Once the ISC vector is identified, each ISC implementation can be evaluated against a set of $m$ quality attributes $q_1$, $q_2$, … $q_m$. The evaluation process takes place as follows. First, each attribute is defined in terms of $f$ features, where $f > 1$. Because of the uncertain nature of data, the evaluation of each feature is represented as a grey number. For example, the implementation of an ISC plan must consider possible restrictions as described in the Restrictions attribute above. These restrictions can be defined in terms of costs (C), availability of resources (AoR), and scheduling and timing constraints (T). As such, the number of features in the restrictions attributes is $f = 3$. With these features in place, evaluating the importance of the Restrictions attribute $q_j$ for an ISC implementation $I_i$ can be computed as follows:

$$d_{ij} = [l_{ij}, u_{ij}] = \left[\frac{1}{f}\sum_{k=1}^{f} l_k, \frac{1}{f}\sum_{k=1}^{f} u_k\right] \qquad (3)$$

where $k$ is the number of features identified for attribute $q_j$. This computation is the arithmetic mean of the values assigned to the features in attribute $q_j$. As this mean increases, so does the importance of the attribute. This computation is performed for all features in each attribute. The overall assessment of the $n$ ISCs based on all $m$ quality attributes is captured using the following decision matrix $D$:

$$D = \begin{bmatrix} [l_{11}, u_{11}] & [l_{12}, u_{12}] & \ldots [l_{1m}, u_{1m}] \\ [l_{21}, u_{21}] & [l_{22}, u_{22}] & \ldots [l_{2m}, u_{2m}] \\ \vdots & \vdots & \vdots \\ [l_{n1}, u_{n1}] & [l_{n2}, u_{n2}] & \ldots [l_{nm}, u_{nm}] \end{bmatrix} \qquad (4)$$

where the rows represent alternatives considered in ISC implementation while the columns represent the attributes of the same problem. Note that the $l_{ij}$ and $u_{ij}$ represent respectively the lower and upper bounds of grey number $d_{ij}$ for $i = 1, 2, .., n$ and $j = 1, 2, .., m$.

### 4.5 Attribute Weights

In general, an ISC attribute will also be characterized by very specific goals. For example, the goals of an alternative may consist of minimizing restrictions while maximizing the rest of the attributes listed in the ISCs above. Optimization goals consist mostly of minimizing or maximizing one or more attributes associated with a given decision problem. However, these goals may not have the same importance in some cases. To assess the relative importance of each quality attribute, the following weight vector $W$ is created:

$$W = [w_1\, w_2\, \cdots\, w_m] \qquad (5)$$

where $w_j$ represents the importance of attribute $q_j$. These weights can be decided by one or more experts in a subjective manner or synthesized objectively from the matrix $A$. In this paper, weights are synthesized from the decision matrix using the concept of statistical variance. In contrast to other approaches for synthesizing weights such as the entropy method [16, 17], statistical variance is effective and easy to implement [18]. Unlike statistical analysis where focus is placed on the extremes, variance examines how data points are scattered around the mean. As such, variance provides useful information about how important an attribute is to a decision problem.

**Definition 1.** Let $d = [l, u]$ be a grey number with $l < u$. If $d$ is continuous, then,

$$\hat{d} = \frac{1}{2}(l + u) \qquad (6)$$

is the core of $a$ [15]. The cores of all grey numbers in the matrix $D$ can be used to compute the weights from $D$ using statistical variance as follows:

$$v_j = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{d}_{ij} - \overline{d}_{ij}\right)^2 \qquad (7)$$

where $\hat{d}_{ij}$ is the core of grey number $d_{ij}$ while $\overline{d}_{ij}$ is the statistical mean of the cores of all grey numbers in attribute $q_j$. The synthetic weight of attribute $q_j$ can be computed as follows:

$$w_j = \frac{v_j}{\sum_{k=1}^{m} v_k} \qquad (8)$$

for $j = 1, 2, \ldots, m$.

### 4.6 Normalization of the Decision Matrix

Grey numbers in the matrix can be normalized by using the sum of the cores in each matrix column as follows [15]:

$$\overline{l_{ij}} = \frac{l_{ij}}{\frac{1}{2}\left(\sum_{i=1}^{n} l_{ij} + \sum_{i=1}^{n} u_{ij}\right)} = \frac{2l_{ij}}{\sum_{i=1}^{n}\left(l_{ij} + u_{ij}\right)} \qquad (9)$$

$$\overline{u_{ij}} = \frac{u_{ij}}{\frac{1}{2}\left(\sum_{i=1}^{n} l_{ij} + \sum_{i=1}^{n} u_{ij}\right)} = \frac{2u_{ij}}{\sum_{i=1}^{n}\left(l_{ij} + u_{ij}\right)} \qquad (10)$$

for $i$ = 1, 2, .., $n$ and $j$ = 1, 2, .., $m$ where $l_{ij}$ and $u_{ij}$ are as defined in equation (3) and (4). The resulting normalized matrix is $\overline{D}$.

### 4.7 Weighting of the Normalized Matrix

The normalized matrix can be weighted by multiplying the bounds of each grey numbers in the matrix by the weight of its attribute. Let $\overline{d_{ij}} = [\overline{l_{ij}}, \overline{u_{ij}}]$ be a grey number in the normalized matrix. Each grey number in the matrix is multiplied by its attribute weight as follows [19]:

$$\widehat{d_{ij}} = \overline{d_{ij}} \times w_j = [w_j\overline{l_{ij}}, w_j\overline{u_{ij}}] = [\widehat{l_{ij}}, \widehat{u_{ij}}] \qquad (11)$$

for $i$ = 1, 2, .., $n$ and $j$ = 1, 2, .., $m$. The resulting weighted normalized matrix is $\widehat{D}$.

### 4.8 Benefits and Costs in the Weighted Normalized Matrix

A simple weighted additive approach, similar to the COPRAS-G method, can be used to compute the benefits and costs of the attributes for each ISC implementation in $\widehat{D}$ as follows [19, 20]:

$$P_i = \frac{1}{2}\sum_{j=1}^{k}\left(\widehat{l_{ij}} + \widehat{u_{ij}}\right) \qquad (12)$$

$$R_i = \frac{1}{2}\sum_{j=k+1}^{m}\left(\widehat{l_{ij}} + \widehat{u_{ij}}\right) \qquad (13)$$

assuming that the first $k$ attributes are benefits while the remaining ($m$–$k$) attributes are costs in $\widehat{D}$.

### 4.9 Relative Weights of Each ISC Implementation

The importance of each ISC implementation in the weighted normalized matrix can be calculated as follows [19, 20]:

$$Q_i = P_i + \frac{\sum_{i=1}^{n} R_i}{R_i \sum_{i=1}^{n} \frac{1}{R_i}} \qquad (14).$$

### 4.10 Utility of Each ISC Implementation

The utility degree of each implementation can be calculated based on its relative weight as follows [19, 20]:

$$U_i = \frac{Q_i}{\max\limits_{1 \le i \le n} Q_i} \qquad (15)$$

for $i$ = 1, 2, .., $n$. The implementation with the highest utility degree is considered the best ISC choice given the $m$ security attributes.

## 5 Case Study

This section presents the results of an ISC evaluation case study using the proposed approach. The case study evaluates 10 ISCs based on the quality attributes identified in section 4.3. Using synthetic data for the identified quality attributes, an input matrix is generated for the features of the 10 ISCs listed above. After the features of all attributes have been aggregated using equation (3), the input matrix is reduced to a decision matrix represented by equation (4). Table 1 shows this reduced decision matrix. Next, the method based on statistical variance is applied on the decision matrix to synthesize attribute weights using equations (7) and (8). These weights ($w_j$) are shown below the decision matrix in Table 1. Next, the decision matrix is normalized using equations (9) and (10), after which the matrix is weighted using equation (11). Table 2 shows the weighted normalized matrix. Among the nine attributes considered in this case study, the importance of the features in the restrictions attribute ought to be minimized while that of the features in the remaining attributes ought to be maximized. As such, the restrictions attribute can be viewed as a cost while the remaining attributes can be viewed as benefits. Based on these considerations, equation (12) is applied to all attributes with the exception of the restrictions attribute on which equation (13) is applied. The obtained results are shown in the columns labeled $P_j$ and $R_j$ in Table 3. The computations of determining costs and benefits of the appropriate attributes are used to compute the relative weights and utility of each ISC implementation as the columns labeled $Q_j$ and $U_j$ in Table 3 show.

As Table 3 shows, the most desirable ISC implementations are ISC 4 (100%), followed by ISC 2 (99.1%) and ISC 9 (98.4%). It is important to note that the evaluation of ISCs using this approach is fully dependent on the particular organization and its security objectives. This approach is highly customizable since it can accommodate different features and attributes for ranking ISC implementations. This is possibly the most meaningful contribution from this research. In addition, this approach can be easily implemented in a spreadsheet or software tool to help management select the right ISC implementation.

## 6 Conclusion

The research presented in this paper develops an innovative approach for evaluating the quality of ISCs in organizations based on a multiple quality evaluation criteria. Specifically, it presents a methodology that uses grey systems theory to create a unified measurement that represent how well an ISC implementation meets quality attributes and how important these attributes are for organizations. Through a case study, the approach is proven successful in providing a way for measuring the quality of ISCs for the security objectives of an organization.

Table 1. Decision matrix and synthesized weights after feature aggregation.

<table>
<tr><th rowspan="2">ISC</th><th colspan="6">A1</th><th colspan="6">A2</th><th colspan="6">A3</th><th colspan="6">A4</th><th colspan="6">A5</th><th colspan="6">A6</th><th colspan="6">A7</th><th colspan="6">A8</th><th colspan="6">A9</th></tr>
<tr><th colspan="2">C</th><th colspan="2">AoR</th><th colspan="2">T</th><th colspan="2">S1</th><th colspan="2">S2</th><th colspan="2">Sn</th><th colspan="2">O1</th><th colspan="2">O2</th><th colspan="2">On</th><th colspan="2">L1</th><th colspan="2">L2</th><th colspan="2">Ln</th><th colspan="2">N</th><th colspan="2">O</th><th colspan="2">A</th><th colspan="2">Aw</th><th colspan="2">Tn</th><th colspan="2">E</th><th colspan="2">SoD</th><th colspan="2">CM</th><th colspan="2">NS</th><th colspan="2">SoA1</th><th colspan="2">SoA2</th><th colspan="2">SoAn</th><th colspan="2">P1</th><th colspan="2">P2</th><th colspan="2">Pn</th></tr>
<tr><th></th><th colspan="3">l</th><th colspan="3">u</th><th colspan="3">l</th><th colspan="3">u</th><th colspan="3">l</th><th colspan="3">u</th><th colspan="3">l</th><th colspan="3">u</th><th colspan="3">l</th><th colspan="3">u</th><th colspan="3">l</th><th colspan="3">u</th><th colspan="3">l</th><th colspan="3">u</th><th colspan="3">l</th><th colspan="3">u</th><th colspan="3">l</th><th colspan="3">u</th></tr>
<tr><td>1</td><td colspan="3">3.66</td><td colspan="3">10.05</td><td colspan="3">3.57</td><td colspan="3">13.69</td><td colspan="3">4.46</td><td colspan="3">11.78</td><td colspan="3">7.50</td><td colspan="3">10.39</td><td colspan="3">4.71</td><td colspan="3">10.15</td><td colspan="3">2.09</td><td colspan="3">7.42</td><td colspan="3">7.87</td><td colspan="3">11.06</td><td colspan="3">2.69</td><td colspan="3">19.10</td><td colspan="3">4.46</td><td colspan="3">9.92</td></tr>
<tr><td>2</td><td colspan="3">4.25</td><td colspan="3">13.41</td><td colspan="3">6.39</td><td colspan="3">10.64</td><td colspan="3">6.95</td><td colspan="3">16.35</td><td colspan="3">5.04</td><td colspan="3">15.23</td><td colspan="3">4.84</td><td colspan="3">7.96</td><td colspan="3">6.20</td><td colspan="3">15.73</td><td colspan="3">5.12</td><td colspan="3">14.44</td><td colspan="3">7.29</td><td colspan="3">11.47</td><td colspan="3">3.24</td><td colspan="3">8.97</td></tr>
<tr><td>3</td><td colspan="3">5.70</td><td colspan="3">13.73</td><td colspan="3">3.94</td><td colspan="3">12.29</td><td colspan="3">2.74</td><td colspan="3">13.93</td><td colspan="3">4.87</td><td colspan="3">9.30</td><td colspan="3">2.42</td><td colspan="3">12.95</td><td colspan="3">5.34</td><td colspan="3">12.54</td><td colspan="3">3.83</td><td colspan="3">14.06</td><td colspan="3">6.24</td><td colspan="3">12.46</td><td colspan="3">5.15</td><td colspan="3">12.61</td></tr>
<tr><td>4</td><td colspan="3">3.17</td><td colspan="3">9.49</td><td colspan="3">6.77</td><td colspan="3">10.60</td><td colspan="3">6.98</td><td colspan="3">15.99</td><td colspan="3">5.38</td><td colspan="3">13.73</td><td colspan="3">6.20</td><td colspan="3">12.54</td><td colspan="3">3.81</td><td colspan="3">16.42</td><td colspan="3">4.46</td><td colspan="3">15.54</td><td colspan="3">1.43</td><td colspan="3">10.60</td><td colspan="3">5.24</td><td colspan="3">13.34</td></tr>
<tr><td>5</td><td colspan="3">5.81</td><td colspan="3">8.81</td><td colspan="3">4.20</td><td colspan="3">11.45</td><td colspan="3">3.87</td><td colspan="3">13.54</td><td colspan="3">5.19</td><td colspan="3">16.46</td><td colspan="3">2.39</td><td colspan="3">11.44</td><td colspan="3">3.84</td><td colspan="3">8.88</td><td colspan="3">8.00</td><td colspan="3">17.36</td><td colspan="3">3.16</td><td colspan="3">16.40</td><td colspan="3">5.00</td><td colspan="3">9.57</td></tr>
<tr><td>6</td><td colspan="3">6.34</td><td colspan="3">13.72</td><td colspan="3">5.45</td><td colspan="3">12.22</td><td colspan="3">3.61</td><td colspan="3">9.74</td><td colspan="3">8.11</td><td colspan="3">16.23</td><td colspan="3">3.39</td><td colspan="3">16.55</td><td colspan="3">4.27</td><td colspan="3">14.23</td><td colspan="3">4.60</td><td colspan="3">12.28</td><td colspan="3">6.20</td><td colspan="3">12.21</td><td colspan="3">2.67</td><td colspan="3">8.30</td></tr>
<tr><td>7</td><td colspan="3">3.52</td><td colspan="3">16.53</td><td colspan="3">2.53</td><td colspan="3">10.50</td><td colspan="3">2.20</td><td colspan="3">11.46</td><td colspan="3">3.48</td><td colspan="3">16.45</td><td colspan="3">3.61</td><td colspan="3">6.81</td><td colspan="3">4.58</td><td colspan="3">11.30</td><td colspan="3">3.03</td><td colspan="3">9.02</td><td colspan="3">3.43</td><td colspan="3">11.39</td><td colspan="3">2.85</td><td colspan="3">12.90</td></tr>
<tr><td>8</td><td colspan="3">4.12</td><td colspan="3">16.44</td><td colspan="3">6.49</td><td colspan="3">12.12</td><td colspan="3">5.98</td><td colspan="3">16.94</td><td colspan="3">3.56</td><td colspan="3">7.91</td><td colspan="3">7.01</td><td colspan="3">13.87</td><td colspan="3">5.67</td><td colspan="3">17.18</td><td colspan="3">3.00</td><td colspan="3">9.31</td><td colspan="3">2.86</td><td colspan="3">12.00</td><td colspan="3">4.35</td><td colspan="3">9.04</td></tr>
<tr><td>9</td><td colspan="3">3.95</td><td colspan="3">10.92</td><td colspan="3">4.80</td><td colspan="3">15.80</td><td colspan="3">2.56</td><td colspan="3">14.56</td><td colspan="3">4.34</td><td colspan="3">10.43</td><td colspan="3">6.06</td><td colspan="3">14.91</td><td colspan="3">5.91</td><td colspan="3">18.49</td><td colspan="3">7.03</td><td colspan="3">15.28</td><td colspan="3">6.97</td><td colspan="3">13.18</td><td colspan="3">2.69</td><td colspan="3">16.23</td></tr>
<tr><td>10</td><td colspan="3">8.59</td><td colspan="3">15.22</td><td colspan="3">5.85</td><td colspan="3">14.01</td><td colspan="3">4.34</td><td colspan="3">12.08</td><td colspan="3">4.48</td><td colspan="3">10.23</td><td colspan="3">7.64</td><td colspan="3">14.24</td><td colspan="3">6.68</td><td colspan="3">13.91</td><td colspan="3">6.60</td><td colspan="3">11.27</td><td colspan="3">7.04</td><td colspan="3">10.95</td><td colspan="3">3.78</td><td colspan="3">8.53</td></tr>
</table>

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Wj | 0.111 | 0.039 | 0.118 | 0.134 | 0.136 | 0.183 | 0.139 | 0.073 | 0.067 |

Table 2. Weighted normalized matrix.

<table>
<tr><th rowspan="2">ISC</th><th colspan="6">A1</th><th colspan="6">A2</th><th colspan="6">A3</th><th colspan="6">A4</th><th colspan="6">A5</th><th colspan="6">A6</th><th colspan="6">A7</th><th colspan="6">A8</th><th colspan="6">A9</th></tr>
<tr><th colspan="2">C</th><th colspan="2">AoR</th><th colspan="2">T</th><th colspan="2">S1</th><th colspan="2">S2</th><th colspan="2">Sn</th><th colspan="2">O1</th><th colspan="2">O2</th><th colspan="2">On</th><th colspan="2">L1</th><th colspan="2">L2</th><th colspan="2">Ln</th><th colspan="2">N</th><th colspan="2">O</th><th colspan="2">A</th><th colspan="2">Aw</th><th colspan="2">Tn</th><th colspan="2">E</th><th colspan="2">SoD</th><th colspan="2">CM</th><th colspan="2">NS</th><th colspan="2">SoA1</th><th colspan="2">SoA2</th><th colspan="2">SoAn</th><th colspan="2">P1</th><th colspan="2">P2</th><th colspan="2">Pn</th></tr>
<tr><th></th><th colspan="3">l</th><th colspan="3">u</th><th colspan="3">l</th><th colspan="3">u</th><th colspan="3">l</th><th colspan="3">u</th><th colspan="3">l</th><th colspan="3">u</th><th colspan="3">l</th><th colspan="3">u</th><th colspan="3">l</th><th colspan="3">u</th><th colspan="3">l</th><th colspan="3">u</th><th colspan="3">l</th><th colspan="3">u</th><th colspan="3">l</th><th colspan="3">u</th></tr>
<tr><td>1</td><td colspan="3">0.005</td><td colspan="3">0.013</td><td colspan="3">0.002</td><td colspan="3">0.006</td><td colspan="3">0.006</td><td colspan="3">0.016</td><td colspan="3">0.011</td><td colspan="3">0.253</td><td colspan="3">0.008</td><td colspan="3">0.016</td><td colspan="3">0.004</td><td colspan="3">0.263</td><td colspan="3">0.017</td><td colspan="3">0.017</td><td colspan="3">0.002</td><td colspan="3">0.016</td><td colspan="3">0.004</td><td colspan="3">0.009</td></tr>
<tr><td>2</td><td colspan="3">0.005</td><td colspan="3">0.017</td><td colspan="3">0.003</td><td colspan="3">0.005</td><td colspan="3">0.009</td><td colspan="3">0.022</td><td colspan="3">0.008</td><td colspan="3">0.308</td><td colspan="3">0.008</td><td colspan="3">0.013</td><td colspan="3">0.012</td><td colspan="3">0.353</td><td colspan="3">0.022</td><td colspan="3">0.022</td><td colspan="3">0.006</td><td colspan="3">0.009</td><td colspan="3">0.003</td><td colspan="3">0.008</td></tr>
<tr><td>3</td><td colspan="3">0.007</td><td colspan="3">0.017</td><td colspan="3">0.002</td><td colspan="3">0.006</td><td colspan="3">0.004</td><td colspan="3">0.018</td><td colspan="3">0.007</td><td colspan="3">0.240</td><td colspan="3">0.004</td><td colspan="3">0.021</td><td colspan="3">0.011</td><td colspan="3">0.319</td><td colspan="3">0.021</td><td colspan="3">0.021</td><td colspan="3">0.005</td><td colspan="3">0.010</td><td colspan="3">0.005</td><td colspan="3">0.011</td></tr>
<tr><td>4</td><td colspan="3">0.004</td><td colspan="3">0.012</td><td colspan="3">0.003</td><td colspan="3">0.005</td><td colspan="3">0.009</td><td colspan="3">0.021</td><td colspan="3">0.008</td><td colspan="3">0.291</td><td colspan="3">0.010</td><td colspan="3">0.020</td><td colspan="3">0.008</td><td colspan="3">0.361</td><td colspan="3">0.024</td><td colspan="3">0.024</td><td colspan="3">0.001</td><td colspan="3">0.009</td><td colspan="3">0.005</td><td colspan="3">0.012</td></tr>
<tr><td>5</td><td colspan="3">0.007</td><td colspan="3">0.011</td><td colspan="3">0.002</td><td colspan="3">0.005</td><td colspan="3">0.005</td><td colspan="3">0.018</td><td colspan="3">0.008</td><td colspan="3">0.322</td><td colspan="3">0.004</td><td colspan="3">0.018</td><td colspan="3">0.008</td><td colspan="3">0.279</td><td colspan="3">0.026</td><td colspan="3">0.026</td><td colspan="3">0.003</td><td colspan="3">0.013</td><td colspan="3">0.004</td><td colspan="3">0.009</td></tr>
<tr><td>6</td><td colspan="3">0.008</td><td colspan="3">0.017</td><td colspan="3">0.002</td><td colspan="3">0.005</td><td colspan="3">0.005</td><td colspan="3">0.013</td><td colspan="3">0.012</td><td colspan="3">0.319</td><td colspan="3">0.005</td><td colspan="3">0.027</td><td colspan="3">0.008</td><td colspan="3">0.337</td><td colspan="3">0.019</td><td colspan="3">0.019</td><td colspan="3">0.005</td><td colspan="3">0.010</td><td colspan="3">0.002</td><td colspan="3">0.007</td></tr>
<tr><td>7</td><td colspan="3">0.004</td><td colspan="3">0.021</td><td colspan="3">0.001</td><td colspan="3">0.005</td><td colspan="3">0.003</td><td colspan="3">0.015</td><td colspan="3">0.005</td><td colspan="3">0.322</td><td colspan="3">0.006</td><td colspan="3">0.011</td><td colspan="3">0.009</td><td colspan="3">0.305</td><td colspan="3">0.014</td><td colspan="3">0.014</td><td colspan="3">0.003</td><td colspan="3">0.009</td><td colspan="3">0.003</td><td colspan="3">0.012</td></tr>
<tr><td>8</td><td colspan="3">0.005</td><td colspan="3">0.021</td><td colspan="3">0.003</td><td colspan="3">0.005</td><td colspan="3">0.008</td><td colspan="3">0.022</td><td colspan="3">0.005</td><td colspan="3">0.224</td><td colspan="3">0.011</td><td colspan="3">0.022</td><td colspan="3">0.011</td><td colspan="3">0.369</td><td colspan="3">0.014</td><td colspan="3">0.014</td><td colspan="3">0.002</td><td colspan="3">0.010</td><td colspan="3">0.004</td><td colspan="3">0.008</td></tr>
<tr><td>9</td><td colspan="3">0.005</td><td colspan="3">0.014</td><td colspan="3">0.002</td><td colspan="3">0.007</td><td colspan="3">0.003</td><td colspan="3">0.019</td><td colspan="3">0.007</td><td colspan="3">0.253</td><td colspan="3">0.010</td><td colspan="3">0.024</td><td colspan="3">0.012</td><td colspan="3">0.383</td><td colspan="3">0.023</td><td colspan="3">0.023</td><td colspan="3">0.006</td><td colspan="3">0.011</td><td colspan="3">0.002</td><td colspan="3">0.015</td></tr>
<tr><td>10</td><td colspan="3">0.011</td><td colspan="3">0.019</td><td colspan="3">0.003</td><td colspan="3">0.006</td><td colspan="3">0.006</td><td colspan="3">0.016</td><td colspan="3">0.007</td><td colspan="3">0.251</td><td colspan="3">0.012</td><td colspan="3">0.023</td><td colspan="3">0.013</td><td colspan="3">0.334</td><td colspan="3">0.017</td><td colspan="3">0.017</td><td colspan="3">0.006</td><td colspan="3">0.009</td><td colspan="3">0.003</td><td colspan="3">0.008</td></tr>
</table>

Table 3. Relative weights and utility values of all ISC implementations.

| ISC | Pj | Rj | Qj | Uj |
|---|---|---|---|---|
| 1 | 0.324 | 0.009 | 0.338 | 0.806 |
| 2 | 0.405 | 0.011 | 0.416 | 0.991 |
| 3 | 0.352 | 0.012 | 0.362 | 0.863 |
| 4 | 0.405 | 0.008 | 0.420 | 1.000 |
| 5 | 0.375 | 0.009 | 0.376 | 0.897 |
| 6 | 0.399 | 0.013 | 0.408 | 0.972 |
| 7 | 0.368 | 0.013 | 0.377 | 0.900 |
| 8 | 0.367 | 0.013 | 0.376 | 0.897 |
| 9 | 0.400 | 0.009 | 0.413 | 0.984 |
| 10 | 0.365 | 0.015 | 0.373 | 0.889 |

There are several important contributions from this research. First, the approach is simple and can be easily implemented. This can promote usage in practical scenarios, where highly complex methodologies for ISCs selection are impractical. Second, the approach fuses multiple evaluation criteria and features to provide a holistic view of the overall ISC quality. Third, the approach is easily extended to include additional quality attributes not considered within this research. Finally, the approach provides a mechanism to evaluate the quality of ISCs in various domains. Overall, the approach presented in this research proved to be a feasible technique for efficiently evaluating the quality of ISCs in organizations.

# 7 References

[1] M. Schwartz, "Computer security: Planning to protect corporate assets," *Journal of Business Strategy*, vol. 11, no. 1, pp. 38-41, January-February 1990.

[2] L. Volonino and S. R. Robinson, *Principles and Practice of Information Security*, Pearson Prentice Hall, Inc., New Jersey, 2004.

[3] E. Vaast, "Danger is in the eye of the beholders: Social representations of information systems security in healthcare," *Journal of Strategic Information Systems*, vol. 16, no. 2, pp. 130-152, June 2007.

[4] T. Herath and H. R. Rao, "Encouraging information security behaviors in organizations: Role of penalties, pressures, and perceived effectiveness," *Decision Support Systems*, vol. 47, no. 2, pp. 154-165, May 2009.

[5] L. Barnard and R. Von Solms, "A formalized approach to the effective selection and evaluation of information security controls," *Computers & Security*, vol. 19, no. 2, pp. 185-194, February 2000.

[6] A. Da Veiga and J. H. P. Eloff, "An information security governance framework," *Information Systems Management*, vol. 24, no. 4, pp. 361-372, 2007.

[7] M. Karyda, E. Kiountouzis, and S. Kokolakis, "Information systems security policies: A contextual perspective," *Computer Security*, vol. 24, no. 3, pp. 246-260, May 2004.

[8] C. Wood, "An unappreciated reason why security policies fail," *Computer Fraud and Security*, vol. 2000, no. 10, pp. 13-14, October 2000.

[9] G. V. Post and A. Kagan, "Evaluating information security tradeoffs: Restricting access can interfere with user tasks," *Computers & Security*, vol. 26, no. 3, pp. 229-237, May 2007.

[10] R. Saint-Germain, "Information security management best practice based on ISO/IEC 17799," *The Information Management Journal*, pp. 60-66, July-August 2005.

[11] R. Baskerville and M. Siponen, "An information security meta-policy for emergent organizations," *Journal of Logistics Information Management*, vol. 15, no. 1, pp. 337-346, 2002.

[12] M. E. Whitman, A. M. Towsend, and R. J. Alberts, "Information systems security and the need for policy," in G. Dhillon, Eds. *Information security management: Global challenges in the new millennium*, pp. 9-18, Hershey, Pennsylvania, Idea Group Publishing, 2001.

[13] H. Van der Haar and R. Von Solms, "A model for deriving information security controls attribute profiles," *Computers & Security*, vol. 22, no. 3, pp. 233-244, April 2003.

[14] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, NJ, 1995.

[15] S. Liu and Y. Lin, *Grey Systems: Theory and Applications*, Springer-Verlag, Berlin Heiderlberg, 2011.

[16] D. H. Jee and K. J. “A method for optimal material selection aided with decision making theory,” *Material Design*, 21, no. 3, pp. 199–206, June 2000.

[17] A. Shanian and O. Savadogo, “TOPSIS multiple-criteria decision support analysis for material selection of metallic bipolar plates for polymer electrolyte fuel cell,” *Journal of Power Sources*, vol. 159, no. 2, pp. 1095–104, September 2006.

[18] R. V. Rao and B. K. Patel, “A subjective and objective integrated multiple attribute decision making mehtod for material selection,” *Materials and Design*, vol. 31, no. 10, pp. 4738-4747, December 2010.

[19] E. K. Zavadskas, A. Kaklauskas, Z. Turskis, and J. Tamosatitiene, “Multi-attribute decision-making model by applying grey numbers,” *Informatica*, vol. 20, no. 2, pp. 305-320, 2009.

[20] E. K. Zavadskas, Z. Turskis, J. Tamosaitiene, and V. Marina, “Selection of Construction Project Managers by Applying COPRAS-G Method,” *International Conference on Reliability and Statistics in Transportation and Communication*, Riga, Latvia, pp. 344-350, October 2008.

# Management — An Achilles Heel of Information Assurance Security: A Case Study of Verizon's Data Breach Reports

**Pedro A. Diaz-Gomez**[1], **Alfonso Valencia Rodriguez**[2], **and Luis E. Gomez H.**[2]
[1]Computing & Technology Department, Cameron University, Lawton, OK, USA
[2]Ingenieria de Sistemas, Universidad Piloto de Colombia, Bogota, Colombia

**Abstract**—*As society becomes increasingly interconnected electronically and new avenues for getting data and information become available, the assurance and security of data and information in businesses are more critical, not only as a legal requirement or business compliance, but also as a responsibility to customers and as a first step in pursuing business continuity. Attacks on data and information are a continual threat, but it has been shown that basic countermeasures can detect some of those at early stages of penetration or misuse. This paper focuses on managerial principles that help organizations prevent security breaches on data and information, and it presents a systemic view of Information Security Management.*

**Keywords:** Data breach, information assurance, information security, security management.

## 1. Introduction

Every year, security institutions like *Verizon* and *FBI* report data breaches, and security recommendations in order to prevent and/or mitigate information assurance/security incidents [1], [2], [3], [14], [15], [16], [17]. The goal of this paper is to highlight those recommendations that are simple to implement and could have the potential to address many data breaches (See Section 3). At the same time, the basic recommendations proposed in this paper could be seen as steps in the ladder of quality improvement suggested by managerial principles. The information assurance/security management system of planing, doing, evaluating and updating as a continuous process must be present [5], [18].

The basic controls that appear in this article do not pretend to minimize the security problem organizations face nowadays. Attackers of computer resources are developing new techniques that allow sophisticated penetrations and anti-forensics [16], [17]. In response, security policies, procedures, standards and computer and network countermeasures have been proposed; however, a fault in management has allowed computer penetrations to permeate organizations without notice for hours, days, months and even years [15], [16], [17].

The order of this paper is as follows: Section 1.1 summarizes the *Payment Card Industry Security Standards (PCI)*, Section 2 presents the *Verizon* statistics concerning data breaches from 2004 to 2010, Section 3 relates to the analysis of the statistics presented, Section 4 describes the information assurance/security management system, Section 5 proposes security architecture in a systemic way, Section 6 proposes basic controls, and Section 7 presents conclusions.

### 1.1 The *PCI* Security Standards

*PCI* security standards are private and mandatory for institutions and their partners that offer electronic card payment. Not being compliant overcome monetary sanctions or revocation of service and loss of prestige [13].

*PCI* security standards are general and simple, which does not mean easy to implement and/or maintain. Ultimately one of the most important *PCI* mandates is to maintain a policy that addresses information security for all personnel. Security policy is the road map of data and information assurance/security, and all personnel must be committed to it (see Section 4). A security policy for an organization belonging to the pay card industry must begin by including the *PCI* security standards, which could be summarized as physical security, backups and encryption of data, security mechanisms like firewalls and antiviruses, tracking and monitoring, the develop and maintenance of secure systems and applications, and the regularly testing of them. *PCI* is proposing a proactive approach through security testing of computer systems, applications, networks and security mechanisms that can anticipate the discovery of vulnerabilities and weaknesses that could be re-mediated depending on the risk and benefit/cost of countermeasures.

## 2. Statistics Verizon

This paper has focused its analysis in the reports provided by *Verizon* because those reports reflect forensic investigations of security data breaches. It is not an easy task to find consistent data about violations of information security [6], or statistics presented in a consistent way as *Verizon* does.

As the *Verizon Team* points out, their study is focused on real data breaches, not network activity, attack signatures, vulnerabilities, or public disclosures. It needs to be emphasized that the economic sectors presented in *Verizon's* reports are those in which *Verizon* has done investigations, and those are probably not a statistical sample carefully selected to make inferences that could be applied to other organizations. However, communities can learn from such experiences, or, unfortunately the bad experience of others.

Table 1: Percentage of Data Breaches by Sector.

| *Year* | *Financial* | *Retail & Hospitality* | *Others* |
|---|---|---|---|
| 04 − 07 | 14% | 37% | 49% |
| 2008 | 30% | 37% | 33% |
| 2009 | 33% | 38% | 29% |
| 2010 | 35% | 56% | 9% |

Table 2: Target Organizations & *PCI* Compliant.

| *Year* | *Opportunistic* | *Targeted* | *Compliant* | *Not Compliant* |
|---|---|---|---|---|
| 04 − 07 | 85% | 15% | − | − |
| 2008 | 72% | 28% | 19% | 81% |
| 2009 | 74% | 27% | 21% | 79% |
| 2010 | 83% | 17% | 11% | 89% |

*Verizon's* statistics are presented in this paper following the *Five W's*: what, who, where, when, and why, + *one H* how.

## 2.1 What Organizations & # of Records

Table 1, that represents the percentage of breaches by sector analyzed by Verizon from 2004 to 2010, shows a steady rate arount the $30^{th}$ percentile for the financial sector and an increase in breaches for retail and hospitality in the 2008 - 2010 period. However, it seems this is not the result of an increase in the number of breaches to targeted organizations, but most likely of opportunistic attacks, as shown in Table 2, left half. *Targeted organizations* are those the attacker first chose and then attacked. *Opportunistic organizations* are those in which the attacker first finds or knows a weakness or vulnerability and thereafter decides to attack, exploiting the vulnerability [14], [17].

## 2.2 Who Made and Who Discovered

Table 3 shows the categories of penetrators: *External* which corresponds to hackers and malware; *Internal*, which corresponds to employees, independent contractors and interns with some privileges; *Business Partners*, which corresponds to third parties who share some kind of business relationship, like vendors, suppliers and customers; and *Multiple Parties*, which corresponds to multiple agents like employees in coalition with external parties. It is clear the *External* category is the one prevalent in all periods presented (See Table 3 for statistics with no intersections, and Table 4 with percentages that overlap, i.e., a data breach committed by more than one agent is counted in more than one category).

Table 5 shows who discovered the data breaches. The *External* category which includes customers and law enforcement is the one that discovers more data breaches. The *Internal* category which discovers fewer breaches, has two subcategories: *active* that includes countermeasures in place to detect intrusions, and *passive* that includes detections that occur as a matter of happenstance.

Table 3: Who Made Data Breaches. Exclusives.

| *Year* | *External* | *Internal* | *Business Partners* | *Multiple Parties* |
|---|---|---|---|---|
| 2008 | 43% | 11% | 7% | 39% |
| 2009 | 45% | 27% | 1% | 27% |
| 2010 | 83% | 7% | < 1% | 9% |

Table 4: Who Made Data Breaches. With Inclusions.

| *Year* | *External* | *Internal* | *Business Partners* | *Multiple Parties* |
|---|---|---|---|---|
| 04 − 07 | 73% | 18% | 39% | 30% |
| 2008 | 74% | 20% | 32% | 39% |
| 2009 | 70% | 48% | 11% | 27% |
| 2010 | 92% | 17% | < 1% | 9% |

Table 5: Who Discovered Data Breaches.

| *Year* | *External* | *Internal* | | *Unknown* |
|---|---|---|---|---|
| | | *Active* | *Passive* | |
| 04 − 07 | 75% | 7% | 18% | − |
| 2008 | 69% | 7% | 24% | − |
| 2009 | 61% | 16% | 23% | − |
| 2010 | 86% | 6% | 5% | 3% |

## 2.3 Where Data Breaches Occurred

Table 6 shows where data was compromised. *Servers* corresponds to data available in places like online databases, files or authentication/directory information on servers; *User Devices* include laptops and workstations; *Off-line Data* includes resources like backup tapes; *People* was included in the 2010 report because information can be stolen directly from people [16]; and *Networks* includes network devices like routers.

The category of servers has been the one that prevails in all the periods presented, but the category of user devices has increased from 7% in 2004 − 2007 period, to 56% in 2010.

## 2.4 How Data Breaches Occurred

Table 7 shows how data breaches are performed. *Hacking* corresponds to techniques that involve penetration of systems without authorized credentials; *Malware* is unauthorized code that is injected in computers in order to compromise or harm information assets; *Privilege Misuse* corresponds to authorized users who misuse their privileges; *Physical* relates to intrusions on physical facilities in order to steal computers or information assets; *Social Tactics* is the use of social engineering in order to commit a penetration; and *Error* is the category of human error or omission.

*Human error* was the highest category in the period from (2004 − 2008), followed by hacking, malware, privilege misuse, physical and social tactics with corresponding averages

Table 6: Where Data Breaches Occurred.

| *Year* | *Servers* | *User Devices* | *Off-line Data* | *People* | *Networks* |
|---|---|---|---|---|---|
| 04 − 07 | 93% | 7% | 7% | − | 5% |
| 2008 | 94% | 17% | 2% | − | 0% |
| 2009 | 50% | 36% | 25% | 4% | 1% |
| 2010 | 57% | 56% | 12% | 10% | 2% |

Table 7: How Data Breaches Occurred.

| *Year* | *Hacking* | *Malware* | *Privil. Misuse* | *Physical* | *Social Tac.* | *Error* |
|---|---|---|---|---|---|---|
| 04 − 07 | 59% | 31% | 22% | 15% | 10% | 62% |
| 2008 | 64% | 38% | 22% | 9% | 2% | 67% |
| 2009 | 40% | 38% | 48% | 15% | 28% | − |
| 2010 | 50% | 49% | 17% | 29% | 11% | − |
| *Ave.* | 53.25% | 39.0% | 27.25% | 17.0% | 10.25% | − |

Table 8: Difficulty of Data Breaches.

| *Year* | *High* | *Moderate* | *Low* | *None* |
|---|---|---|---|---|
| 04 − 07 | 17% | 28% | 52% | 6% |
| 2008 | 17% | 31% | 42% | 10% |
| 2009 | 15% | 44% | 28% | 13% |
| 2010 | 8% | 49% | 37% | 6% |

of 53.25%, 39.00%, 27.25%, 17.00%, and 10.25% (See Table 7).

The difficulty of performing the data breach is presented in Table 8. Categories go from highly skilled attackers to script kiddies, and Table 9 presents the difficulty of the countermeasures to mitigate data breaches. The labeling is more than self explanatory.

### 2.5 How long Data is Compromised Without Discovery

The length of time an intrusion was in a system before it was discovered needs a specific subsection because of its importance in the goal of this paper to highlight the lack of management in information assurance/security. Table 10 shows the cases from 2004 to 2010. *Point of Entry to Compromise* is the period of time between when the attacker began scanning and the penetration. *Compromise to Discovery* is the elapsed time between the intrusion and the discovery of the breach; and *Discovery to Containment* is the time between when the breach was discovered and its containment [14], [15], [16], [17].

### 2.6 Why

Table 2 right part, shows the percentage of organizations that suffered data breaches and are or are not *PCI* compliant. The year that shows the lowest compliant percentage with *PCI* is 2010, and at the same time, 2010 has one of the higher percentages of opportunistic organizations attacked. It can be infered that the opportunity to commit a data breach (See Table 2, left part), as well as the ease of its commitment (See Table 8), can certainly be among the causes of data breaches.

Table 9: Difficulty of Countermeasures.

| *Year* | *Simple & Cheap* | *Intermediate* | *Difficult & Expensive* | *None* |
|---|---|---|---|---|
| 04 − 07 | 52% | 28% | 17% | 3% |
| 2008 | 53% | 34% | 13% | − |
| 2009 | 64% | 32% | 4% | − |
| 2010 | 63% | 33% | 4% | − |

Table 10: Time Elapsed - Point of Entry and Containment.

| | 2004 − 2007 | | | | | |
|---|---|---|---|---|---|---|
| *Category* | *Min.* | *Hours* | *Days* | *Weeks* | *Months* | *Years/Never* |
| *P. Entry-Compromise* | 11% | 36% | 28% | 18% | 7% | 0% |
| *Compromise-Discovery* | 0% | 3% | 14% | 18% | 63% | 2% |
| *Discovery-Containment* | 0% | 10% | 27% | 48% | 14% | 1% |
| | 2008 | | | | | |
| *P. Entry-Compromise* | 27% | 21% | 29% | 17% | 6% | 0% |
| *Compromise-Discovery* | 0% | 8% | 16% | 25% | 49% | 1% |
| *Discovery-Containment* | 0% | 6% | 37% | 42% | 15% | 0% |
| | 2009 | | | | | |
| *P. Entry-Compromise* | 31% | 8% | 20% | 20% | 20% | 2% |
| *Compromise-Discovery* | 5% | 6% | 22% | 24% | 37% | 7% |
| *Discovery-Containment* | 4% | 9% | 32% | 24% | 29% | 3% |
| | 2010 | | | | | |
| *P. Entry-Compromise* | 33% | 14% | 44% | 5% | 4% | < 1% |
| *Compromise-Discovery* | < 1% | 4% | 17% | 38% | 36% | 5% |
| *Discovery-Containment* | < 1% | 11% | 23% | 49% | 15% | 2% |

## 3. Analysis

It is important to talk primarily about statistics as presented in percentages. The reader has to take into account that the percentages of one year or period give different numbers of cases [6]. For example, in 2009, there were 141 breaches, and in 2010 there were 761 breaches investigated [17]. If a statistic shows, say 10% in 2009, that would be 14.1, but in 2010, that would be 76.1. The other issue to take into account is that usually, as statistics are presented, percentages constitute more than 100% because of intersections. For example, Table 6 shows where data breaches occurred, but certainly a data breach could compromise more than one asset.

Retail and hospitality, as well as financial organizations, suffered a higher percentage of data breaches than other organizations (See Fig. 1 where 2008 shows the statistics from 2004 to 2007, and 2011 is the report of statistics corresponding to 2010 [14], [17]). Financial records with debit and credit card information have been appealing to attackers because of their easy conversion to money. However, the way the black market moves and/or later prosecutions that took place could make them look at other venues of data and information assets like authentication credentials (log-in and passwords), personal information (social security numbers, date of birth, name, address), and intellectual property [17].

Certainly there are specific targeted organizations, but the majority of attacks happen because attackers found or knew a weakness and/or a vulnerability (See Table 2, left part). The size of the organization is irrelevant as long as there is something to exploit. However, as table 11 shows, organizations of small and medium size (those from 11 to 10,000 employees) are the ones that have suffered higher percentages of data breaches.

Summing up column *Simple & Cheap* as in Table 9, and finding the average, gives 58.0% of simple and cheap countermeasures that could be applied for preventing data breaches. This result is conservative when compared with

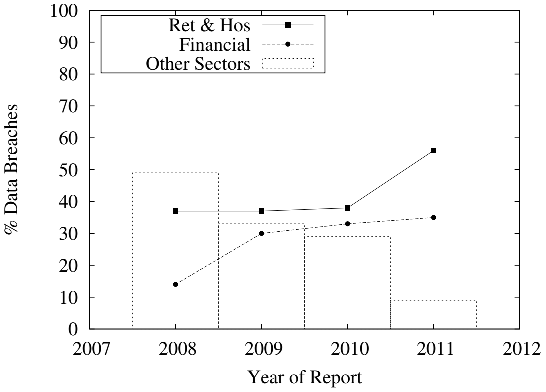


Fig. 1: Percentages of Data Breaches. *Verizon's* Reports.

Table 11: Size of Organizations with Data Breaches.

| *Num. Employees* | $04-07$ | 2008 | 2009 | 2010 | *Ave.* |
|---|---|---|---|---|---|
| $1-10$ | 2% | 7% | 9% | 6% | 6.0% |
| $11-100$ | **30%** | 26% | 18% | **57%** | **32.8%** |
| $101-1,000$ | 22% | 17% | 23% | 10% | **18.0%** |
| $1,001-10,000$ | 26% | **27%** | **26%** | 6% | **21.3%** |
| $10,001-100,000$ | 14% | 18% | 20% | 8% | 15.0% |
| $Over 100,000$ | 3% | 6% | 2% | 7% | 4.5% |

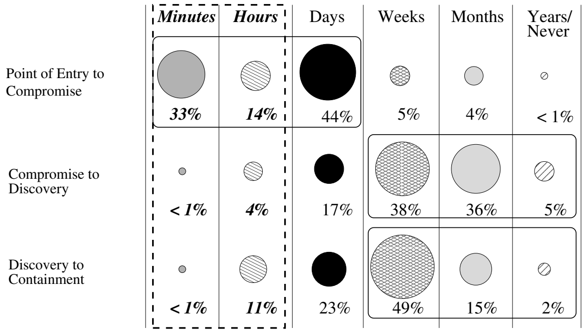


Fig. 2: Time Elapsed since Point of Entry to Containment. *Verizon's* 2011 Report [17].

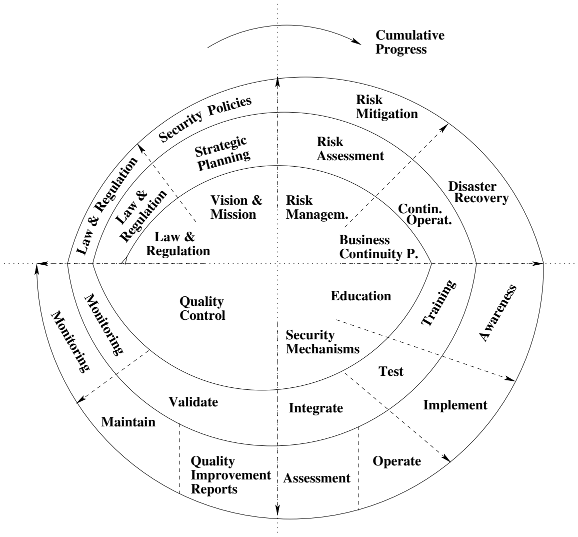


Fig. 3: Information Assurance & Security Management System.

the 85% of mitigation that can be addressed with the top four mitigation strategies proposed by the Australian Goverment Department of Defense [4] (see Section 6).

Tables 3 and 4 show the principal agents of data breaches. No expert view is needed in order to infer that the highest threat is external. Table 6 shows where data breaches occurred; usually, the two sides of the communication (servers and user devices) are the ones with higher percentage of data breaches. However, it is important to see a shift of data breaches occurring from *Servers* to *User Devices*.

Table 10 shows the time elapsed between point of entry and containment. The time from the point of entry to compromise is usually quick; the majority take place in less than a few days. The majority of discovery and containment take place during weeks and months, and a few during years (See Figure 2). A bit atypical was 2009 with one digit percentages in discovery and containment in minutes [14], [15], [16] (See Table 10).

The interested reader can look at Figure 2 [17], which shows the time slots for 2010, as in Table 10, and graphically describes the pattern stated previously. What one can infer is that there is not a straight relation when looking at columns. For example, if one looks at columns *minutes* and *hours*, one can say that 47% of breaches took minutes and hours from point of entry to compromise, and of that 47%, fewer than 1% were discovered in minutes and 4% were discovered in hours. Of that 47%, fewer than 1% were contained in minutes and 11% were contained in hours. However, if just fewer than 5% were discovered in hours, how many fewer than 12% were contained in hours or less? In other words, if 4% was discovered in hours, how could it be that 11% is contained in hours? Certainly some discrepancies can be inferred, but one cannot miss the lesson taught in Figure 2: from the point of entry to compromise time is fast, but from compromise to discovery, as well as containment, could take weeks, months or years.

## 4. Information Assurance & Security Management System

Sections 2 and 3 reported a lack in management in preventing data breaches. Facts like the discovery of data breaches by third parties and, that the ones discovered by internals are mostly by chance; that the majority of data breaches discovered take months and that the majority of containment take weeks; and, that the application of the majority of countermeasures are simple and cheap are spot lights that corroborate a lack in management in the information assurance process.

Assurance and security of information is an integral component of any business. It must be originated from the *CEO* and *Executive Officers*, and it must permeate the entire hierarchy of the organization [11], [18]. A commitment from shareholders guaranties resources to carry out the information/assurance system; and well defined processes with responsibilities and accountability to stake holders, guaranties its success.

Figure 3 suggests a quality improvement spiral of planing, doing, controlling and updating as a continuous system to help in addressing the lack of management as stated previously. The spiral can be seen along its concentric path or following the arrows [7]. For example, the *Vision & Mission* is framed in *Law & regulation*. Changes in one process could affect all other processes (see Section 5). Looking in the arrow direction, *Strategic Planning* is framed in the *Vision & Mission*. Likewise, *Security Policies* are framed in *Strategic Planning*. The spiral suggest a domino effect. Changes occurring in one process affect other processes.

The spiral shows an out-circular approach, where the big picture is in its center, and it spreads out detail processes. Processes permeate all along the spiral. However, at each out ladder, the process is refined or redefined based on the previous process and on the stage it is on. Each "ring" gives a specific level of abstraction.

*Training & Education* are part of the systematic process of information assurance and security. It is common to think that the use of security mechanisms is enough, but part of the system is the human factor [9], [19] (see Section 5).

*Quality Control* is the checking process. *Contingency & Recovery* plans, and *Security Mechanisms* are tested, implemented and validated; based on the metrics defined, *Operations* take *Assessment* data that is used to produce *Quality Improvement Reports*. These reports are analyzed, looking at higher risks, so actions can be taken through *Maintenance* in order to minimize risks.

# 5. Security Architecture as Systemic Approach

Figure 4 proposed by Whitman [19] gives another view of Figure 3. The core is *Data & Information*, i.e., it is the asset to assure and secure. On the left part are *People*, and on the right part is *Technology*; that includes media to access data and information like *Systems*, *Networks*, and the *Internet*, as well as assurance and security mechanisms like *Access Controls*, *Encryption* and *Backups* [19].

Figure 4 shows an interconnected net between processes and security mechanisms proposed by the authors,. A systematic approach (see Section 4) is suggested: for example, *Law & Regulation* could affect security mechanisms like *Monitoring* (at different levels), and/or it could affect other processes like *Security Policies*.

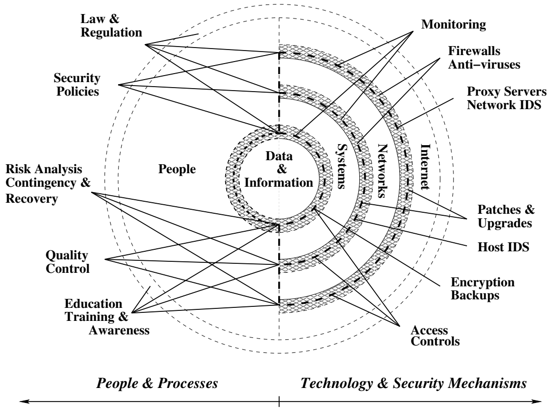


Fig. 4: Security Architecture as Systemic Approach.

Data and information are intangible assets [11], and technology has made them available in places that do not necessarily belong to the organization; so the idea to access data and information through networks and computer systems that belong only to an organization, does not always hold. Figure 4 (dash points) suggest some risks to data and information: *People* [9], [18], the interconnection to the *Internet* [12], and vulnerabilities in networks, systems and applications. However, *Processes*, and *Security Mechanisms* must be in place to help mitigate those risks.

## 5.1 Security Mechanisms

*Security mechanisms* are countermeasures that help to protect data and information. Figure 4 shows some examples which are mandatory for the *Pay Card Industry* like firewalls, encryption, backups and access controls.

*Monitoring* is a process quite related with quality control (see Section 4). Based on metrics previously defined, monitoring alerts deviation from "normality" and/or "misuse". Processes like intrusion detection systems, revision of logs, auditing and certifications can be considered part of monitoring.

*Firewalls* are security mechanisms standard de facto that help to protect networks and systems monitoring internal and external traffic. Their importance is established by being gate keepers to access data and information that could implement secure protocols like IP Security (IPSec) [12]. Categories of firewalls range from packed filtering devices to application level firewalls like application servers.

In the chain of security mechanisms, *Patches & Upgrades*, *Encryption*, *Backups* and *Access Controls* close the loop of protecting one of the most valuable assets in organizations: data and information.

# 6. Recommendations

Information assurance security is a systematic system that must focus on continual quality and improvement. Security mechanisms like the ones proposed by *PCI* (interested readers can look at [10]) are one step on the ladder of this system (see Sections 4 and 5). Once security mechanisms have been implemented, maintenance and monitoring are a must. This step is probably an Achilles heel of information assurance security; a huge percentage of organizations are not compliant at the time of a breach (See table 2, right part), the opportunity of being attacked is on the rise (See Table 2, left part), and the discovery of data breaches is accomplished by third parties (See Table 5).

This paper accentuates that organizations with simple and cheap countermeasures, like the top four mitigation strategies found by the Australian Government of patching operating systems and applications, the assignment of least priviledge, and the prevention of malware [4], plus monitoring of events logs, passwords, firewalls configurations, anti-viruses, physical and logical accesses, backups, and the encryption of sensitive data, could avoid or at least make data breaches more difficult (See Tables 8 and 9 concerning the difficulty of data breaches and countermeasures).

## 7. Conclusions

This paper summarizes data breaches investigated by *Verizon* from 2004 to 2010, following the guidelines of the *Five W's + one H*, focusing on basic countermeasures that can be on place to mitigate them and proposing a systemic view of security management.

The *PCI* security standard is focused on the protection of card-holder data, which is the way information assurance/security should be addressed in the protection of the most valuable institutional assets: data and information [8]. However, data and information are intangible assets, and as such they are difficult to track and secure. The definition of who is the *owner*, who is the *custodian*, who is the *user* of data; as well as the data baselines and the countermeasures to protect them, are part of the management of information assurance/security. Data and information is stated here as the overall information resource of organizations: that includes users, employees, financial, technical and systems data, from general in scope to a specific record or a transaction view [11]. Once each baseline has been defined, the corresponding countermeasures to protect it must be defined and implemented. This paper has highlighted the fact that basic countermeasures certainly can mitigate data breaches. Organizations have to come back to the basics: a focus on where the data is, what networks, computer systems, applications and processes access it, how to protect it, and how to guarantee that information assurance/security is a continuous improvement system; this is a direct responsibility of information security managers.

## 8. Thanks

Thanks to Dr. Carolyn Lindsey Kinslow and Mr. Ellis Hooley for proofreading this paper. Thanks to the University of Oklahoma (Dr. Dean Hougen at the Computer Science Department), Norman, USA, for allowing us to use computer resources.

## References

[1] CSI Computer Crime and Security Survey Report, "CSI/FBI Computer Crime and Security Survey," 2000, accessed June 2011. [Online]. Available: http://www.pbs.org/wgbh/pages/frontline/shows/kackers/risks/csi-fbi2000.pdf.

[2] ——, "8th CSI/FBI Computer Crime and Security Survey," 2003, accessed November 2010. [Online]. Available: http://www.citadel-information.com/library/4/2003-fbi-csi-survey.pdf.

[3] ——, "14th CSI Computer Crime and Security Survey," 2009, accessed November 2010. [Online]. Available: http://www.personal.utulsa.edu/ james-childress/cs5493/CSISurvey/CSISurvey2009.pdf.

[4] Defense Signals Directorate (DSD), "Top 35 mitigation strategies," Australian Government. Department of Defense. Intelligence and Security, 2011, accessed Dec. 16 2011. [Online]. Available: http://www.dsd.gov.au/infosec/top35mitigationstrategies.htm.

[5] W. E. Deming, *The New Economics for Industry, Government, Education*. MA: Cambridge, MIT Press, 2000.

[6] P. A. Diaz-Gomez, G. D. ValleCarcamo, and D. Jones, "Internal Vs. External Penetrations: A Computer Security Dilemma," in *Proceedings of the 2010 International Conference on Security & Management*, 2010.

[7] S. Heim, *The Resonant Interface*. MA: Addison Wesley, 2008.

[8] J. Maloney, "Data Protection 2.0: It's not Just Names and Numbers Anymore," 2010, accessed July 2011. [Online]. Available: http://www.tripwire.com/register/?resourceId=706684

[9] K. D. Mitnick and W. L. Simon, *The Art of Deception*. IN: Wiley Publishing, Inc., 2002.

[10] PCI Security Standards Council LLC, "Payment Card Industry (PCI) Data Security Standard Navigating PCI DSS. Understanding the Intent of the Requirements. Version 2.0," 2010, accessed June 2011. [Online]. Available: http://www.-pcisecuritystandards.org/documents/navigating_dss_v20.pdf.

[11] C. Schou and D. Shoemaker, *Information Assurance for the Enterprise. A Roadmap to Information Security*. USA: McGraw Hill, 2007.

[12] W. Stallings and L. Brown, *Computer Security. Principles and Practice. Second Edition*. USA: Pearson Prentice Hall, 2012.

[13] Tripwire, "PCI Basics: What it takes to be Compliant," 2010, accessed July 2011. [Online]. Available: http://www.tripwire.com/en/apac/register/981526?cat=PCI&type=wp

[14] Verizon Business Risk Team, "2008 Data Breach Investigations Report," 2008, accessed June 2011. [Online]. Available: http://www.-verizonbusiness.com/resources/security/databreachreport.pdf.

[15] ——, "2009 Data Breach Investigations Report," 2009, accessed June 2011. [Online]. Available: http://www. verizonbusiness.com/resources/security/reports/2009_databreach_rp.pdf.

[16] ——, "2010 Data Breach Investigations Report," 2010, accessed June 2011. [Online]. Available: http://www.verizonbusiness.com/resources/reports/rp_2010-data-breach-report_en_xg.pdf.

[17] ——, "2011 Data Breach Investigations Report," 2011, accessed June 2011. [Online]. Available: http://www.verizonbusiness.com/resources/reports/rp_data-breach-investigations-report-2011_en_xg.pdf.

[18] M. E. Whitman and H. J. Mattord, *Management of Information Security*. Canada: Thomson, 2004.

[19] ——, *Principles of Information Security. Third Edition*. Canada: Cengage Learning, 2009.

# Impact Evaluation and Prediction Model for Clustering Social Groups

**Yanping Zhao** [1], **Senlin Luo** [2], **and BaoShan Bu** [1]

1 School of Management and Economics, Beijing Institute of Technology, Beijing 100081, China

2 School of Information and Electronics Science, Beijing Institute of Technology, Beijing 100081, China

**Abstract** - *The impact evaluation and prediction for the unusual clustering events of social groups will help to identify some of signs before major security incidents happen, to provide indicators for early warning, to prevent or intervene for the major security events and network eruption group events or social damages. This paper presents impact evaluation and evolution detection framework based on empirical knowledge for searching dense groups and tracking their evolutions, and prediction based on Latent Space Model to monitor the relations of key actors to major changes of influential groups, conduct experimental research on the data set from the U.S. Enron scandal event, and graphical visualization exhibition, found some interesting evolution modes of the network dense groups, proved effectiveness and practical value of our model.*

**Keywords:** Impact evaluation, Latent Space Model, Cluster , Cohesion Social Group, Content and relational security.

## 1 Introduction

Many social security events, natural disasters, or big accidents would cause problems and have big influences on our society, such as UK- London riots on the street, Libya - the Rebel fighters in the Tripoli, and China - High-speed rail accidents and so on. There are many social groups taken intention to the events, and if not treated well they will lead severe damages, fights, deaths or more serious problems.

While in the Internet there are virtual social groups actively reflecting the events in reality. Since the network is so huge, the information is in varieties and spread so fast and widely, and the dimension is too high to detect and track, these characteristics of the groups really make the research for monitoring the events a challenge job. Even the general method of Social Network Analysis(SNA) involves many connections to deal with, it is difficult to determine which group of members subordinate to another and the decomposition of overlapped subgroups are especially difficult, so as to the prediction of important trends of their relations, etc. These are important issues to the public safety, and also to the security of the assets of enterprises or organizations.

This paper discusses the model of how to remove and filter out unimportant groups, and reduce the irrelevant nodes; meanwhile taking Latent Space Model theory to excavate and predict on the important actors' effect to the merging events of increasing social groups; finally make an empirical analysis on the real Enron data, and found a number of important and interesting signs of groups aggregation and that has an important meaning to the future prediction and prevention of the organizational scandal incidents.

## 2 Related research

The study for the impact of the clustering events of both virtual and real social groups begins from recent years, the laws of the evolution and its impact assessment studies is still relatively small and mainly in case studies and news. Especially the evaluation for the impact of the relations between cohesion sub-groups to the mass outbreak events of real world and virtual network is even rare, but attracting a lot of attention.

Dynamic social network model obtained remarkable result in the mining of 9/11 members of the terrorist organization. After 2001, the method gained rapid development. In 2002, led by the Office of Naval Research, major U.S. research groups held dynamic network analysis workshop [1], further contribution to the development of dynamic social network research. In 2005, on SIGKDD '05 conference, Purnamrita Sarkar and Andrew W.Moore [2] proposed a model named Dynamic Social Network Analysis using Latent Space Models (DSNL), the model successfully converted static model into a time-varying dynamic model. In 2006 IEEE held a Conference on Web Intelligence, Falkowski and Bartelheimer[3] et al published a paper presented two ways of mining community Subgroups. In KDD'08 conference Lei Tang, Jianping Zhang [4] first proposed a multi-mode network and its advantages. These models give in general the dynamic social network and community mining models for understanding their structures and natures; we use and extend them for event related clusters detection and evolution trends prediction. But they have not given methods to identify overlapping groups and trends for their expansions. For mining the clusters of social groups, the k-means method used in [2] must provide pre-specified number of the clusters before clustering, our revised ant colony clustering algorithm by introducing the priori background knowledge [5] to that more direct and more efficient principal in the ant colony algorithm [6], not only to suit any shape of the clusters structures and automatically detect the actual number of them, also get better quality clustering results.

The research of social network cohesion subgroup in China began in 2004. June 2008, Zhang Lipu [7], published "Quantifying analysis of condensing subgroup potency in social network analysis". In May 2008, Ren Yike, Du Haifeng [8] et al. make a empirical study of the social network of Chinese migrant workers, found the social networks of Chinese migrant workers exhibit cohesive subgroup structure, and has large overlap between subgroups. In May 2009 Wang Lili etc., used "Carrefour incident" as an example to analyse the characteristics of the network group events and the disposal principle [9]. Most of these studies are on the generating mechanism and characteristics of the social groups, since the massive Internet information, they have not yet provide dynamic tracking and evaluation of clustering social group events. If these strategies are implemented manually, its workload will be too heavy to the security and regulatory authorities.

# 3 The dynamic impact evaluation model for the clustering events

In order to observe the changes of harmful groups of network over time, we introduce the dimension of time in traditional Social Network Analysis, and define the Moving time window and the concept of group $G_t$.

**Group $G_t$**: is a connection diagram observed at time step t, which is composed of a series of connected individual component.

**Dense Subgroups**: or cohesive clusters, is a subgroup of closely related individuals (nodes) and has special structure, or has strong tendency of closer relationship detected significant bring certain influence.

**Dynamic Monitoring:** is based on the moving time window, and flexibly choosing an hour, day, or month… to be the time interval as $\varDelta t$ . For example, in research on the Enron data set, using a quarter as a time interval unit, depending on the significance of the situation, which is the amount of data to show some meaningful results and financial report frequency.

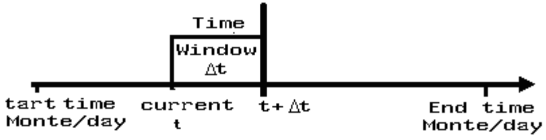


Fig.1. Moving time window

As shown in Fig. 1, the current month t (or days) is the start and within a $\varDelta t$ time interval is a window. When the window moving forward, contacts and structure will also change, ultimately it will be able to get the evolution of group structure and trends.

**Dense subgroups evolution model**: for dense subgroups of network community, analysis of two adjacent time of different subgroups evolution, discovers the evolution examples of the growth, reduce, merge, split, generate and dissolve of the network community subgroups, as shown in Fig. 2.

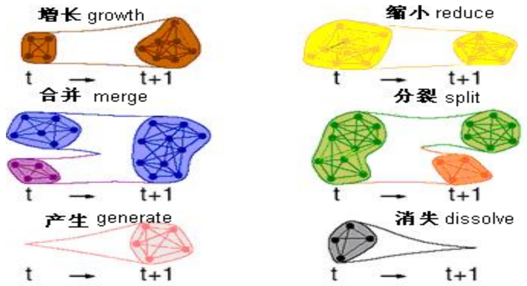


Fig. 2. Evolution of cohesion social subgroups

## 3.1 The dynamic impact evaluation model

**Evolution stability indicator:**

$$Stability\ \ of\, a\ Clust\ er\ \varPhi_{stability}(t) = Eg_{in}(t) \big/ (Eg_{in}(t) + Eg_{out}(t)) \quad (1)$$

where $Eg_{in}(t)$ is the number of internal edges at time t, $Eg_{out}(t)$ the number of edges in the group outwardly connected to other groups. The greater stability of a group possess, the stronger the group's cohesion, the more difficult to split, and the stronger external pressure to withstand, and vice versa.

**Evolution merger indicators:**

$$EvolutionOfCluster: \phi_{sim}(t) = \sum_i Sim(A_i(t), B(t+\Delta t)) I_i(\beta) \quad (2)$$

where $\Sigma_i$ in the formula above, represents that B is merger of significant evolution of every $A_i$; at time t there is A(t), at the next time t+Δt, exist a group B (t+Δt), define similarity Sim (A (t), B (t +Δt)) and a threshold of $\beta$ in the time period, if Sim (A (t), B (t +Δt))> $\beta$ , called group B is an evolution of group A, else there has not evolutionary relationship between two groups. $I_i(\beta)=1$ if $Sim(A_i, B)> \beta$ , else $I_i(\beta)=0$.

The similarity or overlap[10] between group A and B is：

$$\mathrm{Sim(A,B)=Sim(B,A)} = \frac{|A(t) \cap B(t+\Delta t)|}{\min\ (|A(t)|, |B(t+\Delta t)|)} \quad (3)$$

where | A | is the number of elements in group A, and ∩ represents the intersection of A and B.

## 3.2 Mining and visualization of cohesion subgroups

Our researches on cohesion subgroup are roughly as follows: condensation subgroup mining, the characteristics of relationship among members within the subgroup; characteristics of the relationship between subgroups; the characteristics of a relationship between members of one subgroup and members of another sub-group.

### 3.2.1 Dynamic data acquisition

Our model allows effectively monitoring of network micro-blog, Blog, BBS, or E-mail messages and several other data [11]. BBS or micro-blog has public content, E-mail may not get its content, and we will describe our data acquisition method. We use robots technique to obtain multiple broadband multimedia network features set or online Abuse Reports to provide clues (such as network behavior feature, speaker's semantic features, pictures or videos, audio features, network link modes, etc.), filtering out the relevant groups according to event to do real-time data acquisition by *Δt* time period or by significant change detected, and make processing and restoration. In order to minimize the interference of the normal network operation and improve the system efficiency.

### 3.2.2 Elimination of overlaps of cohesive subgroups

Members of social network or community may appear in different subgroups, which are caused by the multiple correlations of social network subgroups. Since the virtual network is an epitome and embodiment of the real world, people could either join subgroups through a topic of same interest, or to form another cohesive subgroup to discuss business problems, that makes the overlap between the subgroups appear, also means an entity could be a number of different subgroups due to different relations among members. Now the classical clustering algorithms seldom meet this requirement. This paper uses the multiple semantic analysis technology based on ontology and rules[11] to find different semantic subgroups, then using our revised ant colony clustering algorithm based on prior experience knowledge[5] in the semantic subgroup to mine the cohesive subgroups, which maximally eliminate the overlap influence of the subgroups.

### 3.2.3 Calculate the relationship strength matrix

First, we focus on establishing a time window for calculating relationship strength matrix of a cohesive subgroup and establish the contact strength matrix.

For convenience, we first assume that the social network subgroup has five staffs: A, B, C, D, E. They setup a social network through sending messages or emails to each other, in which, we called "contact strength" or "relationship strength" between A and B as the sum of the text messages (mails) that A sends to B, here we kept the sum but deliberately combine the both directional data together, for the sake of simplification and the significance of the relation, so the contact matrix is symmetric. In which, if there is a message sent to oneself, the contact strength is 0. Table 1 shows the contact strength matrix of five people A, B, C, D, E.

Table 1
THE CONTACT STRENGTH MATRIX

| | A | B | C | D | E |
|---|---|---|---|---|---|
| **A** | 0 | 3 | 2 | 1 | 0 |
| **B** | 3 | 0 | 1 | 5 | 4 |
| **C** | 2 | 1 | 0 | 3 | 0 |
| **D** | 1 | 5 | 3 | 0 | 2 |
| **E** | 0 | 4 | 0 | 2 | 0 |

where the cell of row A and column B is 3, means they exchange 3 messages.

### 3.2.4 Map the group with a specific topic to the Latent Space

The Latent Space Model (LSM) [2] was first proposed by American scholars in 2004. It is a technique reducing the complexity of the original data. This paper presents a new latent space model to analyze the strength between individuals, and use physical "distance" in the latent space to conduct efficient mining and prediction of harmful groups.

Because the distance is usually calculated by the entities' coordinates, however the social network has not existed coordinates of the points, but only has the contact strength, the original algorithm for LSM is not suitable to this situation. This paper uses relationship strength matrix as prior knowledge, to transform the real nodes contact matrix to Latent Space by using a new version of multidimensional scaling analysis (MDS[12]) as a map of cohesive subgroups, with semantic similarity, to the latent space with coordinates, and matrix singular value decomposition to obtain low-dimensional reconstruction of the data.

Table 2 is the obtained physical coordinates of the entities in Table 1, by means of new MDS, the first line of the Table (X, Y) = (0.4365, 0.7109) is the coordinates of entity A in the in Latent Space.

Table 2
THE COORDINATES IN THE LATENT SPACE

| | X | Y |
|---|---|---|
| A | 0.4365 | 0.7109 |
| B | -2.0929 | -0.0968 |
| C | -0.7382 | -0.4165 |
| D | 1.7676 | 0.4752 |
| E | 0.627 | -0.6728 |

Next according to the data in different time windows, we make subgroups evolution analysis.

### 3.2.5 Clustering and evolutionary analysis

By using our efficient ant-colony clustering algorithm by introducing priori knowledge[5], Our approach is different from [2] in that it gives the natural results of the number of densely related subgroups after clustering. It gets all dense subgroups in the group of the same time window, calculating the subgroups' stability, closeness and other measures after clustering analysis, and draw graphs of them (Fig.3 in IV).

## 3.3 Model of group merge and prediction

Merge prediction model is designed for predicting whether the connection between two harmful groups will occur, to prevent the vicious expansion of the mass incidents both in network and society.

To achieve this, we made a preliminary exploration in connection probabilities based on revision of the Purnamrita Sarkar's observation model [2]. Taking into account the entities in the social interaction are though direct and indirect contact of friends in the central nodes, we put forward the following hypothesis, two individuals, not yet had a direct connection, would have a direct link possibility proportional to the active level of an individual to contact to others; and proportional to the individuals connections.

Therefore, we define the radius of entity i, it expresses the social interaction range of an entity in the real social network, and set $r_i = l \times (\varphi_i + 1)$ , 1 in formula to ensure the radius is non-zero, *l* is a constant used to adjust the physical radius *r*<sub></sub>$r_i$,

$\varphi_i$ express the degree of node $i$ (by addition of out-degree and in-degree, it is a dynamic value change over time window).

Radius between the two entities is defined as $r_{ij}$, whose value is the larger of the radius $r_i$ and $r_j$, intuitively, the higher the physical degree, the bigger number of contacts in social interactions. So we define the physical radius as $r_{ij} = l \times (max(\varphi_i, \varphi_j) + 1)$ to replace the former.

We introduce a symmetric Kernel function as:

$$K(d_{ij}) = \begin{cases} (1-(d_{ij}/r_{ij})^2)^2 & 当 d_{ij} \le r_{ij} \\ 0 & \text{else} \end{cases} \tag{4}$$

and to ensure the Kernel function is continuous and differentiable when $d_{ij} = r_{ij}$.

By using this function we define a formula to predict the probability of connection between nodes i and j in time window $t$ to $t + \Delta t$ as follows.

$$p_{ij}^{\,t+\Delta t} = \frac{1}{1+e^{d_{ij}-r_{ij}}} K(d_{ij}^{\,t}) + \rho(1 - K(d_{ij}^{\,t})) \tag{5}$$

where a constant value $\rho$ is for the probability remains constant when the distance is far between two nodes.

# 4 Model testing and discussions

To verify our model's effectiveness and practical value, we use the email data set from U.S Enron scandal event [13] in order to obtain all the contents of interactions and check found facts.

## 4.1 The test data set

Enron data set is composed of all e-mail communications of the Enron employees in September 1998 to June 2002. It was uploaded on-line by the U.S. Federal Energy Planning Board in October 14, 2003, since the Board found Enron's financial problems by the message of that data set and put it on the Internet for researchers to use, the website is http://www-2.cs.cmu.edu/~enron/. The entire data set contains 30,329 valid e-mail and 151 e-mail accounts, time span of 46 months. Each message contains a recipient, sender, CC, BCC, message subject, body parts, and attachments and so on, which covers all aspects of business and personal relationship in the company. The data set downloaded is a .sql file, after import it into Mysql database, there contains four tables.

The first table is employeelist. It contains the employee's first and last name, as well as their e-mail addresses. The second is message table, it contains Senders，Subjects，Email body，Date etc. The third is recipientInfo table, it contains the transmit information such as send, CC, or BCC. The fourth is referenceInfo it contains original and replied messages.

We establish an improved ant colony clustering algorithm in Microsoft Visual Studio 2008 using C# coding, it plays very good effect in display the dense subgroups after the clustering and, easy to compare their structures. Next we transplant an open source packages Java Universal Network Graph(JUNG)[14], produce our own DrawMap to display the dense sub-groups. We developed impact evaluation tools under JDK1.5[15] for group similarity, closeness and stability and other measures.

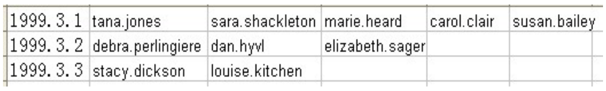
## 4.2 Discovery and analysis of problem groups

To avoid overlapping of the groups, we extracted all data sets related to financial from the Enron e-mail messages, by key words: finance, credit, capital, energy, financial, bank, account and so on. Thus, the groups formed in the same period are all of financial issues, they made illegal operations of the company and give a profound impact into bankruptcy. Therefore, we make evolution and prediction analysis in Enron e-mail dense groups focus on this issue.

After excluding overlapping points of the data, we find that if the window size on the aggregation of subgroups was month, the amount of data was too small to analyze, the link between each point is at most 1. Then by taking into account of the financial report experts experience and significance considerations, we use a unit of quarter, and adjust the time unit to a quarter, this won't affect the subgroups excavated.

After getting the contact strength matrix, we make the cohesion subgroup mining by using our efficient ant colony clustering algorithm, and store them in Excel table shown as below.

Table 3
COHESION SUBGROUPS OF THE 3RD QUARTER, 1999



| | | | | | |
|---|---|---|---|---|---|
| 1999.3.1 | tana.jones | sara.shackleton | marie.heard | carol.clair | susan.bailey |
| 1999.3.2 | debra.perlingiere | dan.hyvl | elizabeth.sager | | |
| 1999.3.3 | stacy.dickson | louise.kitchen | | | |

On the Table 3, the first column represents the data's time period; For example, 1999.3.1 indicates that the data is in 3rd quarter of 1999, .1 followed denotes the 1st subgroup of the dense subgroups, and the names behind is the members in the subgroup. In order To confirm our mining is in accordance with the high strength of contacts in the same group, and mining is effective, we draw the chart of the subgroups and with their original contacts directions we kept in the period of time for detailed checking, the data in the 3rd quarter of 1999 is as follows:

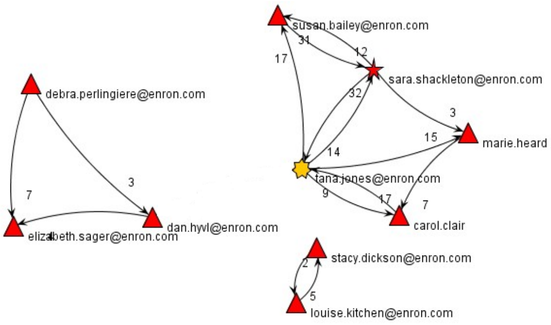


Fig. 3. Dense subgroups of 3rd quarter of 1999

In Fig. 3, each node at a polygon representing its subgroup, the higher the degree, the more the number of edges connects to it; Arrows indicate the direction of sending e-mail, and its data indicate contact strength of the nodes. We find on the top right subgroup, each arc has greater than 3 of strength, so they can be regarded as a cohesive subgroup. However the three in the lower left corner can form a cohesive subgroup too, and there is only two people left, and the two are not link to any others. It shows that they are a separate subgroup. Through these results we can roughly determine there are three cohesive subgroups with quarterly data of the Group, it proves our priori knowledge of the result from Table 3 of ant colony cluster algorithm.

Then we calculate the cohesive subgroups of the Group of Kenneth Lay, Chairman and CEO of Enron, in the entire time windows sequences, get the structure, closeness, stability and changes over time (by formula 1, 2, 3) to and show their curve in Fig. 4.

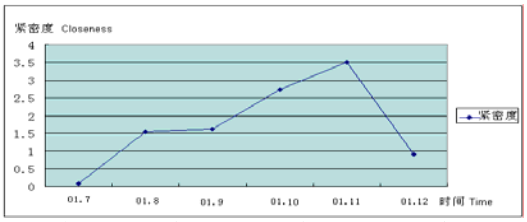


Fig. 4. Kenneth •Lay group closeness versus time

From Fig. 4, Kenneth Lay has been in the groups where closeness increases greatly, and from July to August, and October to November the two curves rises very steeply. Check with the literature [16], we easily find between July and August, people started to pay attention to Enron, some well known entrepreneurs, economists have raised the questions of Enron, it could be said that at the beginning, in order to better cope with external pressure of public opinion, Kenneth • Lay's group would surely has closer ties within the group than in the past, which is consistent with the results of our curve. Between October-November, Enron admitted false accounting, when Enron was undoubtedly in the most difficult time, as the company's key leaders Kenneth Lay, the group he was in connected much more closer than usual, which is in consistent with the facts that the company declared bankruptcy in December, the closeness of the entire group dropped down quickly to a minimum, so the curve is steep.

## 4.3 Factors affect the evolution analysis

The filtered out Enron e-mails have 101 cohesive subgroups and 40 pairs of them are identified as similar subgroups. We classify the subgroups with analysis of the evolution of which are growth, merge, generate, reduce, split, and dissolve, and find that there are 8 pairs of groups evolution of merge, 6 of them split, and the other 25 growth.

### 4.3.1 The growth mode of the subgroups evolution

By using our improved ant colony clustering algorithm, in the entire 13 quarters of the data set we've got a total of 101 dense subgroups. According to the evolution and similarity formula (5), we calculate the groups' similarity in two adjacent time periods, for example, between the 2nd quarter of 1999 and the 3rd quarter of 1999, and the 4th quarter of 1999, the similarity of two subgroups with each of them in one period and their evolution trends as following Fig. 5.

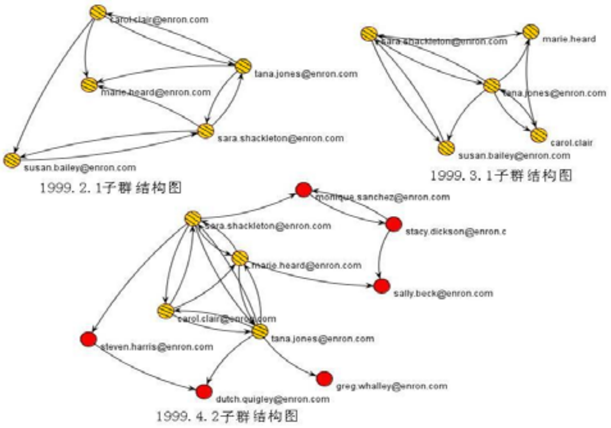


Fig. 5. Growth of a dense subgroup

Fig.5 shows the growth evolution we have dug out from all cohesive subgroups which have evolutionary relationship. In the Fig.5, striped nodes exist in the evolution of each time period. The upper left one of Picture is the 1st subgroup in the 2nd quarter of 1999(1999.2.1). We can see that the dense subgroup is made up of five nodes, and inner connection of sub-groups is relatively close. The upper right is the network structure of the 1st subgroup in 3rd quarter of 1999(1999.3.1). By comparing the two groups, we can see that the names and number of the two nodes in the Fig.3 are not changed, just the strength of links between nodes changed. By simple calculation we find the similarity of these two groups is almost 1, which can determine 1999.3.1 is the result of evolution of 1999.2.1.

Between the second quarter of 1999 and the third quarter of 1999, the amount of the members in the group is not changed. The only change is just the strength of links in the group. The bottom part is the 2nd subgroup of the 4th quarter of 1999 (1999.4.2). Compared with the former 2 subgroups, we find that four nodes appeared in the previous two groups, and the similarity is 0.37, which means that we can determine that these three groups are similar, the 3rd one is merely the growth evolution of the last ones.

### 4.3.2 The split mode of subgroup evolution

Fig.6 below is a structure of dense subgroups that show the

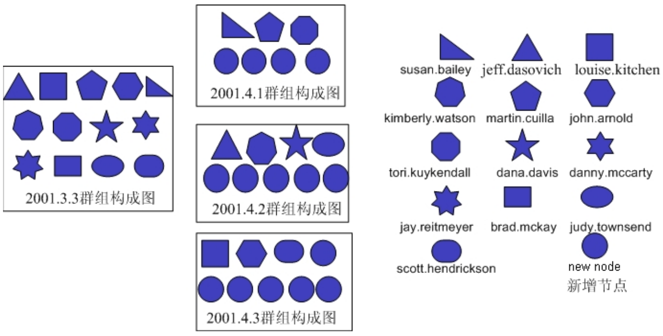


Fig. 6. Instances of the subgroup split

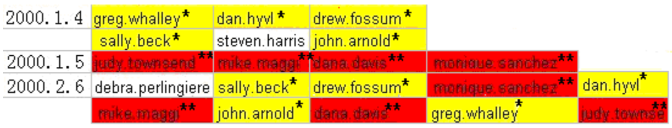
splitting evolution mode dug out from the 3rd to the 4th quarter of 2001.

In the right part of Fig.6 there are 13 different nodes. Letters below the icons are the names of the specific people involved. The round node indicates the persons are not in the 3rd subgroup of the 3rd quarter of 2001.

From the Fig.6 above we can see that from the 3rd quarter of 2001 to 4th quarter of 2001, except 2 nodes are not in the first three subgroups in the 4th quarter(there are more than three subgroups, only these three have the relation), the other nodes are almost equally distributed into the 1st three subgroups of the 4th quarter, and the similarity values of the 3rd group in the 3rd quarter and the first three groups in the 4th quarter are almost the same, so we can determine that the 3rd group of the 3rd quarter split into the first three groups of the 4th quarter. Although during the whole process of the splitting, still other people joined, alone for the 3rd group of 2001, it shows a tendency to split.

### 4.3.3 The merge mode of Subgroups

During the analysis of the evolution of dense subgroups of the entire data of Enron e-mail, we find that due to the high frequency change of group members and the complexity of virtual network, the merge evolution of subgroups not appear clearly. Fig. 7 gives the instances of the subgroup merge.



| | | | | | |
|---|---|---|---|---|---|
| 2000.1.4 | greg.whalley* | dan.hyvl * | drew.fossum * | | |
| | sally.beck* | steven.harris | john.arnold* | | |
| 2000.1.5 | judy.townsend** | mike.maggi** | dana.davis** | monique.sanchez** | |
| 2000.2.6 | debra.perlingiere | sally.beck* | drew.fossum* | monique.sanchez** | dan.hyvl* |
| | mike.maggi** | john.arnold* | dana.davis** | greg.whalley* | judy.townse** |

Fig. 7. The merge of subgroups

In Fig. 7, the first two lines show the six members with five in yellow(shade*) except one would disappear from the dense subgroups in 2000.1.4; and the third line of four dense members (red**) in 2000.1.5, and the forth two lines are the members of subgroup of 2000.2.6 which is from both first subgroup of 2000.1.4 and second subgroup of 2000.1.5. Therefore we can draw the conclusion that the subgroup of 2000.2.6 was a new group made of the first two groups.

### 4.3.4 The dissolution of subgroups

As for the dissolution of dense subgroups, it doesn't mean there is no evolution of the dense subgroup. On the basis of careful analysis of the data, we find that the similarity of these subgroups with their previous subgroups and all subgroups in the next time period are very low. If we infer according to 0.1428, the similarity threshold, and it is not similar to any subgroups in the next time period, then we conclude that the group died out. But that doesn't mean all the members died--perhaps all the members dissolved into many small units and joined in other subgroups. In a sense, it's a more thorough dissolution. For example, the similarities of the 7th groups of the 3rd quarter of 2001 and the 5 groups of the 4th quarter of 2001 are 0.0967, 0.0571, 0.0657, 0.0667, 0. According to the threshold of 0.1428 we can determine that the group has dissolved, but after the analysis of the reason of low similarity values, we find two reasons: one is that part of the data appeared in the 7th subgroup in the 3rd quarter did not show up in the ones of the 4th quarter; the other one is that the amount of members of the 7 dense subgroups in the third quarter is relatively less and the amount of the members in each subgroup of the fourth quarter is relatively more, thus resulting in the low similarity values.

We took all the statistical analysis of the stability and closeness of emerge, growth, split, dissolve and so on, and contrast between them is shown in Fig.8 below.

It can be seen that the stabilities of emerged subgroups are above 0.45, and the stabilities of the growth subgroups are higher, about 0.55. Thus, we can judge that the criteria of the closeness for the merge and growth are 0.45 and 0.55 respectively, and similarly, the stability of them is 0.69 and 0.83 respectively.

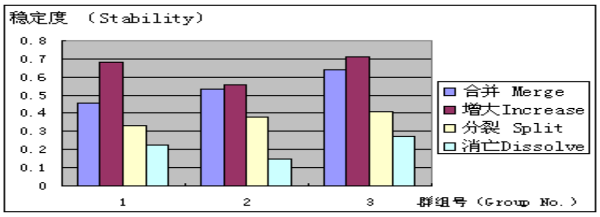


Fig. 8. Comparison of the evolution of models

## 4.4 Prediction of the key role in the merge situation

In the study about the evolution of dense subgroups, we found we could judge the trend of dense subgroup in the future by monitoring the closeness, and other value, but it is not enough to just judge the trend of dense subgroup in the merge subgroups. Like the network group events, if two dense subgroups had merged, it means it would happen in several harmful groups, which could lead an expansion and worse situation for security events, which is not allowed for the security department. So, if we could predict which key role would lead merge before it happens, the monitor would probably tack action to prevent or control them. This method could be used in any two nodes in the net, especially for the key role in process of subgroups' merge.

We choose the dense subgroups which have a trend to

Table 4
PREDICT THE PROBABILITY OF CONNECTIONS BETWEEN
2000.1.4 AND 2000.1.5

| 2000.1.4<br>2000.1.5 | **greg.whalley** | **dan.hyvl** | **drew.fossum** |
|---|---|---|---|
| judy.townsend | 0.2697 | 0.1623 | 0.3017 |
| mike.maggi | 0.1556 | 0.1479 | 0.1398 |
| dana.davis | 0.3173 | 0.1962 | 0.2896 |
| monique.sanchez | 0.1786 | 0.1989 | 0.1592 |
| 2000.1.4<br>2000.1.5 | **sally.beck** | **debra.perlingiere** | **john.arnold** |
| judy.townsend | 0.1145 | 0.0987 | 0.2592 |
| mike.maggi | 0.1589 | 0.2579 | 0.9832 |
| dana.davis | 0.1872 | 0.1338 | 0.1678 |
| monique.sanchez | 0.1327 | 0.0783 | 0.1763 |

merge in the first quarter in 2000 to calculate the probability (according to formula 5) of having links in the 2nd quarter in 2000, as shown in the Table 4.

The first row in the Table 4 stands for 2000.1.4, the first column stands for the 2000.1.5. The number in the cells shows the probabilities of the corresponding two entities having links in the next period. For instance, 0.2697 in the cell of the second column and the second row is the probability of the two person greg.whalley and judy.townsend may have links. If we use threshold 0.2 getting from the training before, to judge whether the entities would have links. However, when we check Table 4, there is no significant key role much more than others among the subgroups, we see several entities from each subgroup may have link probability greater than 0.2 which means merge between two groups does not depend on one key relationship, and it is very interesting! To verify the correctness of our discovery, we drew the net structure graph of these groups in the future(2000.2.6), shown as Fig. 9.

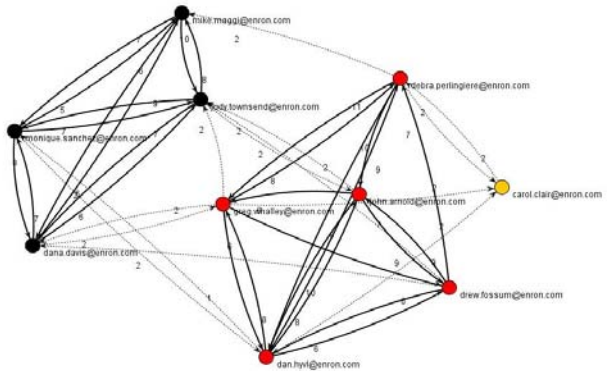


Fig. 9. The 6th group network structure in the 2nd quarter of 2000 (merge of 2000.1.4 and 2000.1.5)

We can see from Fig.9 that the subgroups 2000.1.4 and 2000.1.5 did have several links before merge happened, instead of only one key role for the merge. This means that the merge may not always predicted by take attention to a key link, sometimes there may be several links happened at the same time for the subgroups merge. So we did not include the impact model with prediction formula for the consideration of further study.

## 5 Summary

This paper provides a dynamic evaluation model for analysing and monitoring social network groups. And using Enron data to do empirical verifications, we discover some interesting fact in term of prediction of potential harmful group event, which play an important role in study and perception on network group events for social security, in responses, and effectively provide prevention solutions in the early stages.

## 6 Acknowledgment

This paper is supported by the major cultivating special plan of Beijing Institute of Technology, the project No. 2011CX01015. And the Application Innovation Project of the Ministry of Public Security of People's Republic of China, No.2011YYCXGADX124.

## 7 References

[1] Kathleen M. Carley†, Jana Diesner, Jeffrey Reminga, "Toward an Interoperable Dynamic Network Analysis Toolkit," *Decision Support Systems*, vol. 43, no.4, pp. 1324-1347, Aug. 2007.

[2] P.Sarkar, A. W. Moore, "Dynamic social network analysis using latent space models," *ACM SIGKDD Explorations Newsletter*, Vol.7, no.2, pp. 31-40. 2005.

[3] Falkowski T, Bartelheimer J, Spiliopoulou M, "Mining and Visualizing the Evolution of Subgroups in Social Networks," *Web Intelligence*, IEEE/WIC/ACM International Conference, pp.52-58, 2006.

[4] LeiTang, HuanLiu, JianpingZhang, "CommunityEvolution in Dynamic Multi-Mode Networks," in *Proc. KDD'08*, Las Vegas, Nevada, USA. 24-27, Aug. 2008.

[5] Bu Baoshan, "Research on Evolution Model of Condensation Subgroup In Social Networks," *Master's Dissertation*, Beijing Institute of Technology, pp. 17-20, June, 2010.

[6] Ling Chen, Xiao-hua Xu, Yi-xin Chen, "An adaptive ant colony clustering algorithm," In *Proc. of the Third International Conference on Machine Learning and Cybernetics*, pp. 1387-1392, 2004.,

[7] Zhang Lipu, "The Quantifying Analysis of Condensing Subgroup Potency in Social Network Analysis," *Journal Of Adult Education School Of Hebei University Of Technology*, vol.23,no.2,pp.45-49, 2008.

[8] Ren Yike, Du haifeng, Yu Xiao, "An Analysis of the Cohesive Subgroup of Chinese migrant workers in social network," *Society*, vol.28, no.5, pp. 20-42, 2008.

[9] Wang Lili, "Characteristics of the network group communication," *Oriental Media*, no.5, pp. 92-95, 2009

[10] Falkowski T, Bartelheimer J, Spiliopoulou M, "Mining and Visualizing the Evolution of Subgroups in Social Networks," *Web Intelligence*, IEEE/WIC/ACM International Conference, pp. 52-58, 2006.

[11] Zhao Yanping, Lu Wei, "An Efficient Algorithm for Content Security Filtering Based on Double-Byte," In *Proc. 2007 IEEE Intelligence and Security and Infomatics* (IEEE ISI-2007), New Brunswick, New Jersey, US. 23-24, pp.300 – 307, May 2007.

[12] Borg, I., Groenen, P, *Modern Multidimensional Scaling: theory and applications (2nd ed.)*. New York: Springer-Verlag. pp. 207–212, 2005.

[13] Jitesh Shetty , Jafar Adibi, "The Enron Email Dataset Database Schema and Brief Statistical Report," Technical report, Information Sciences Institute, pp. 44-45, 2004.

[14] Java Universal Network/Graph(JUNG) project, [Online]. Available: http://jung.sourceforge.net.

[15] Java Development Kit, [Online]. Available: http://java.sun.com/.

[16] Bethany McLean, Peter Elkind, *The Smartest Guys in the Room: The Whole Story of the Enron bankruptcy case*, translation, China: Social Sciences Press, 2007.

# EC- RBAC Model: secured access control model for Exigent Scenarios in Defense Systems.

Abdullah Sharaf Alghamdi
Software Engineering Department,
CCIS, King Saud Univeristy,
Riyadh, Saudi Arabia,
ghamdi@ksu.edu.sa

Syed Shah Amanullah Quadri
Software Engineering Department,
CCIS, King Saud Univeristy,
Riyadh, Saudi Arabia,
aquadri@ksu.edu.sa

## Abstract:

**With the advent of the exponential enterprise system growth in defense systems two areas have emerged as points of serious concerns viz., the abuse of privileges and control of the system as a whole with a super access that overrides the underlying access controls especially in exigent and emergency scenarios. In this paper we propose a RBAC model for exigency control (EC-RBAC) in emergency scenario which is expected to ensure a secured super access by a defense authority in the defense system, using Rational Software Modeler by IBM, a UML Case tool. The EC-RBAC not only controls the whole system but also ensures that such access is competently secured.**

***Keywords: C4I Systems, Security, RBAC, UML.***

## 1. INTRODUCTION

Role based access control has been enjoying the fame and acceptance as a one stop solution to the business, enterprise and e-commerces' information security and access control requirements since its inception. The NIST model of RBAC incorporates reference model and functional specifications defined at the core. This is mainly intended for software engineers and product development manager for designing of access control features (1).

RBAC functionalizes using basic elements like users, roles, permission, operations and objects and is divided into four model components – Core RBAC, Hierarchical RBAC, Static Separation of Duties (SOD) and Dynamic Separation of duties (D-SOD) (1).

Core RBAC defines the element sets and is responsible for activation of user session. This model component is compulsory entity of any system that implements access control. However, the rest of the three model components are independent of each other and can be implemented separately. Hierarchical RBAC is a collection of mapping sets which map the inheritance of permission among the roles and is best represented using mathematical representations. Static Separation of Duties (SOD) component is key element to eradicate inconsistencies among the role permissions which in turn contributes to exclusivity to user assignment. Finally, the Dynamic Separation of Duties (D-SOD) concerns the precision of the relations of roles activated in the users' session (2).

## 2. BACKGROUND

Duoqiang Wang et al., in their work [3] pointed out that the D-SOD policies do not enforce SOD policies as they don't prevent users from activating mutually exclusive roles across multiple sessions. Thus, D-SOD has been assumed as requirement qualifying policy wherein users can carry out their steps in a sensitive task which requires to be verified for quantitative requirements check [3].

## 3. PROPOSED EC-RBAC MODEL

In this paper we present a RBAC mechanism that can efficiently handle the exigencies and its adverse effects on the defense system. Having noted that in exigent scenarios in defense system, there is bright window possibility of the authority over lapse and which in turn can provide access in the defense system and its critical levels of security. This is not necessary an allowed access therefore the authority and the roles can be wrongfully utilized. In the EC-RBAC model we assume that only a few yet trustworthy high ranking defense authorities are in function, which can be trusted with the whole command of the complete defense functionalities.

The immediate task at hand is to allow access to proper defense authority into the system and let the RBAC know that : (1) the authority logging on is to be treated with critical superiority, (2) all the other critical accesses in the defense system will be rolled back and overridden by the current superior access, (3) The superior authority logging in has raised a EXG flag to his logon and his logon is confirmed using a special security mechanism and (4) Once this authority has assumed control over the defense system no other logon is with the exigency flag or superiority is allowed to access the system.

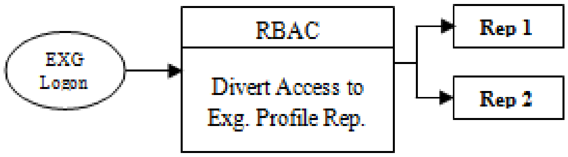


Fig. 1: Logon using exigent flag to RBAC Repository

In Fig.1 we represent that, firstly the defense authority logs in to the defense system using a login which is attached with an exigent flag (Exg), this user login details are pre-available in the RBAC system for verification in a remote storage facility for security purposes. The instance the RBAC identifies the login is an exigent logon it will divert the user login to two different consoles of pass key repositories (Rep 1 and Rep 2)to acquire the pass keys which are stored in a remote facility again for security purposes invoked only by RBAC. The user will have to logon to the consoles using two sets of usernames and passkeys to acquire the single passkey to access the exigent permission repository of the RBAC.

Secondly, once the user has successfully acquired the passkeys the RBAC will be informed the same and a copy of the passkeys will be mirrored to main RBAC pass key repository. Then, RBAC will divert the user console back to main logon of the exigent permission repository. Once user logs on using the passkeys acquired, RBAC will verify the passkeys using its mirrored copy. If the keys match RBAC will proceed further else the login will be cancelled and the Exigent Logon details i.e., the passkeys and login id will be purged.

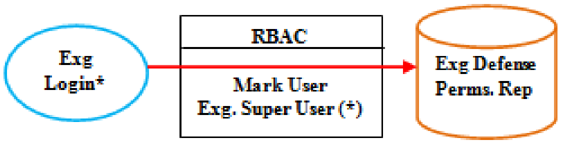


Fig. 2: RBAC initiating Exg. Defense Permissions Repository

Further, In Fig. 2 we show that once the user has been successfully authenticated as an exigency super user by RBAC the user will be granted access to the Exigent Defense Permission Repository which will be dynamically instantiated, invoked and granted access to the Super user by RBAC.

The Exigent Defense Permission Repository (ExDPR) is not a static one, this ensure that no access in the regular functioning of the defense system will have control of the critical operations of the different verticals of the Defense system. The ExDPR will be dynamically created by RBAC by following the sequence of the operations as follows:

### A. *Formation of Repository*

From the permission repository copies of the RBAC all operations of all verticals of the defense system marked critical will be shortlisted and assigned to super user represented as:

SupU→ SLPerm.

### B. *Revoke Access*

Revoke access of all the logins in the RBAC from the shortlisted operations.

### C. *Alert Issuance*

Issue system wide alert of exigent super user access in the system.

### D. *ExDPR Formulation*

Formulate dynamically the ExDPR repository with the shortlisted operations.

### E. *Accessing ExDPR*

Allow access to the ExDPR only by the Exigent Super User.

### F. *Restriction further access*

Disallow any further login to the system using exigent flag.

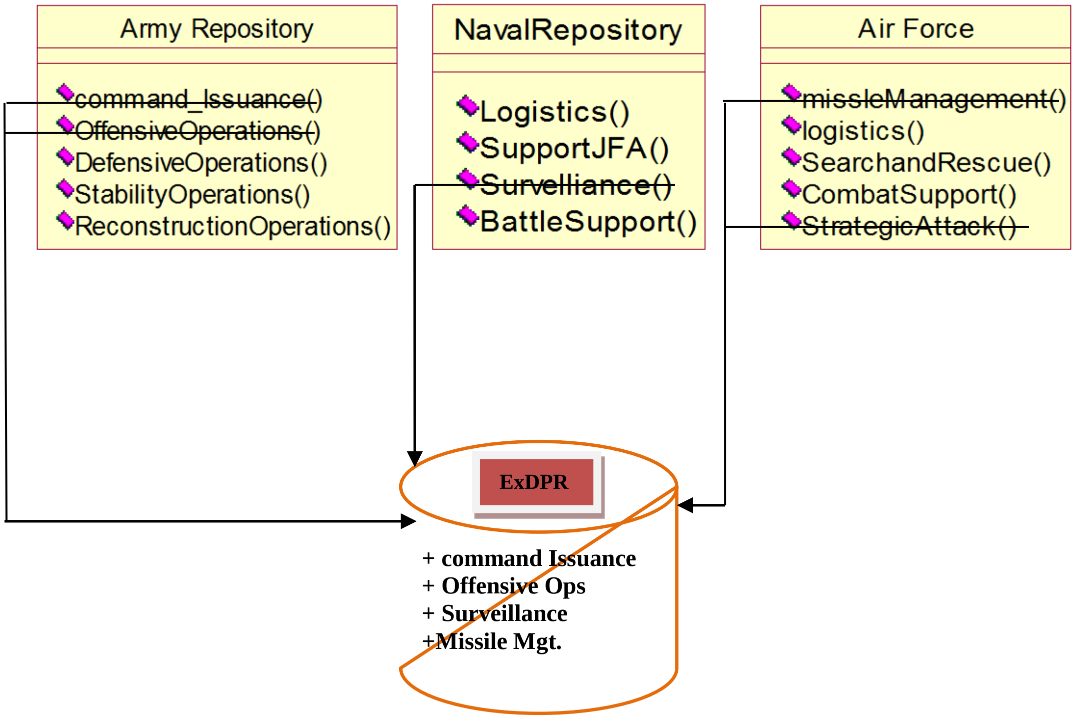


Fig. 3: Dynamic Formulation of ExDPR using Defense Critical Operations

### *3.1 Specification and Verification*

For the practical Specification, implementation and verification we are making use of UML as it is convenient to express constraints, sequential flow of functionality and also has the ability to depict the pre and post conditions along with Alternatives, Artifacts and Exceptions.

### *3.2 Dynamic ExDPR Formulation*

The most important and challenging issue of the model is dynamic compilation of the ExDPR whose definition and design should be a result of sound Software Engineering techniques in general and the pivotal focus being the compilation process modeling in particular. The main focus should be modeling the auto-generation of views to compile the permission repository under emergency situation. This is to ensure that exigent permission repository is not susceptible to intruders and the dynamic compilation of ExDPR will ensure that the permission repository will exist in the system only after the Exigent Login has been authorized successfully. Thus, even if somehow the intruders using specialized intrusion techniques have compromised the remote logon mechanism, until the second level RBAC has not cleared the entry using passkeys the ExDPR will not be available in the system anywhere. Only a genuine exigent login acknowledged by second level RBAC will run the view in an executable format to assemble the critical operations from the permission repositories to formulate the ExDPR only after satisfying the preconditions as represented in figure 4.

### *3.3 Formalization of EC-RBAC Policies*

The formalization of the EC-RBAC Policies needs to ensure the maintainability, efficiency and scalability of the overall functionality of the Defense system. This implies that the presence of the Exigent access in the system which overrides and controls all the critical functions of the system, should ensure that the critical functions in process are not forced into deadlocks. Thus, formalization of policies by combining authorization constraints to avoid above mentioned side effects. The formal verification techniques like theorem proofer, proof assistants and like thereof, though cumbersome yet can contribute great help to guarantee the avoidance of such issues

up to a competent level of control and acceptance.

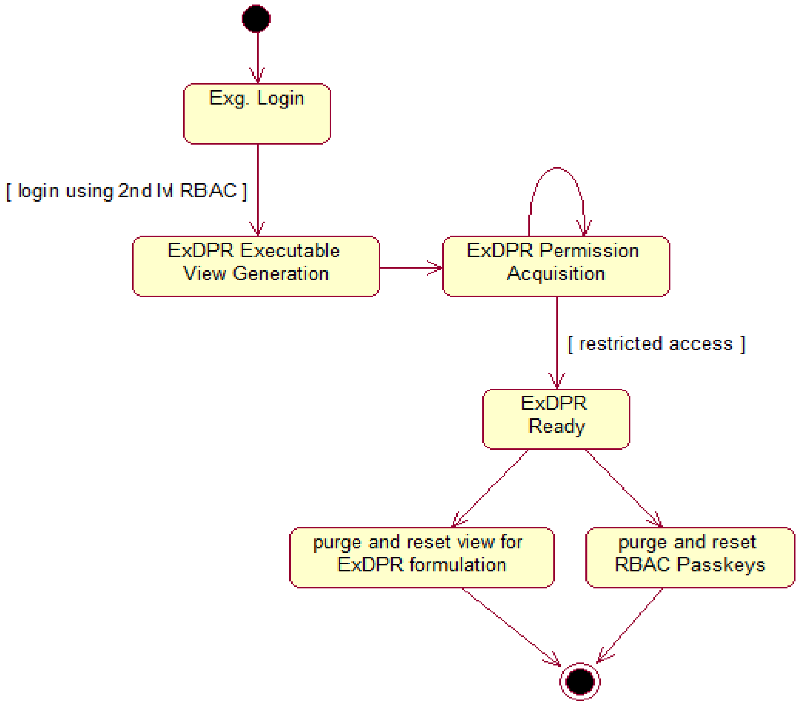

Exg. Login
[ login using 2nd lvl RBAC ]
ExDPR Executable View Generation
ExDPR Permission Acquisition
[ restricted access ]
ExDPR Ready
purge and reset view for ExDPR formulation
purge and reset RBAC Passkeys

Fig. 4: State diagram of the ExDPR formulation process.

In Figure 4 we have shown the state diagram as an activity model reflecting the stepwise flow of the operations in the following steps:

*A. Authorization*

Authorization of the exigent login including the RBAC execution at 2$^{nd}$ level of the repository view generation process.

*B. Formulation*

Formulation of ExDPR, repository acting on itself by acquiring core system functions critical to the defense system.

*C. Revocation*

Revoking access on any current core system functions being acquired by the ExDPR.

*D. Post security measures*

*D.1. Purging and Resetting View*

The current settings of the operational view which formulated the ExDPR to be purged and reset.

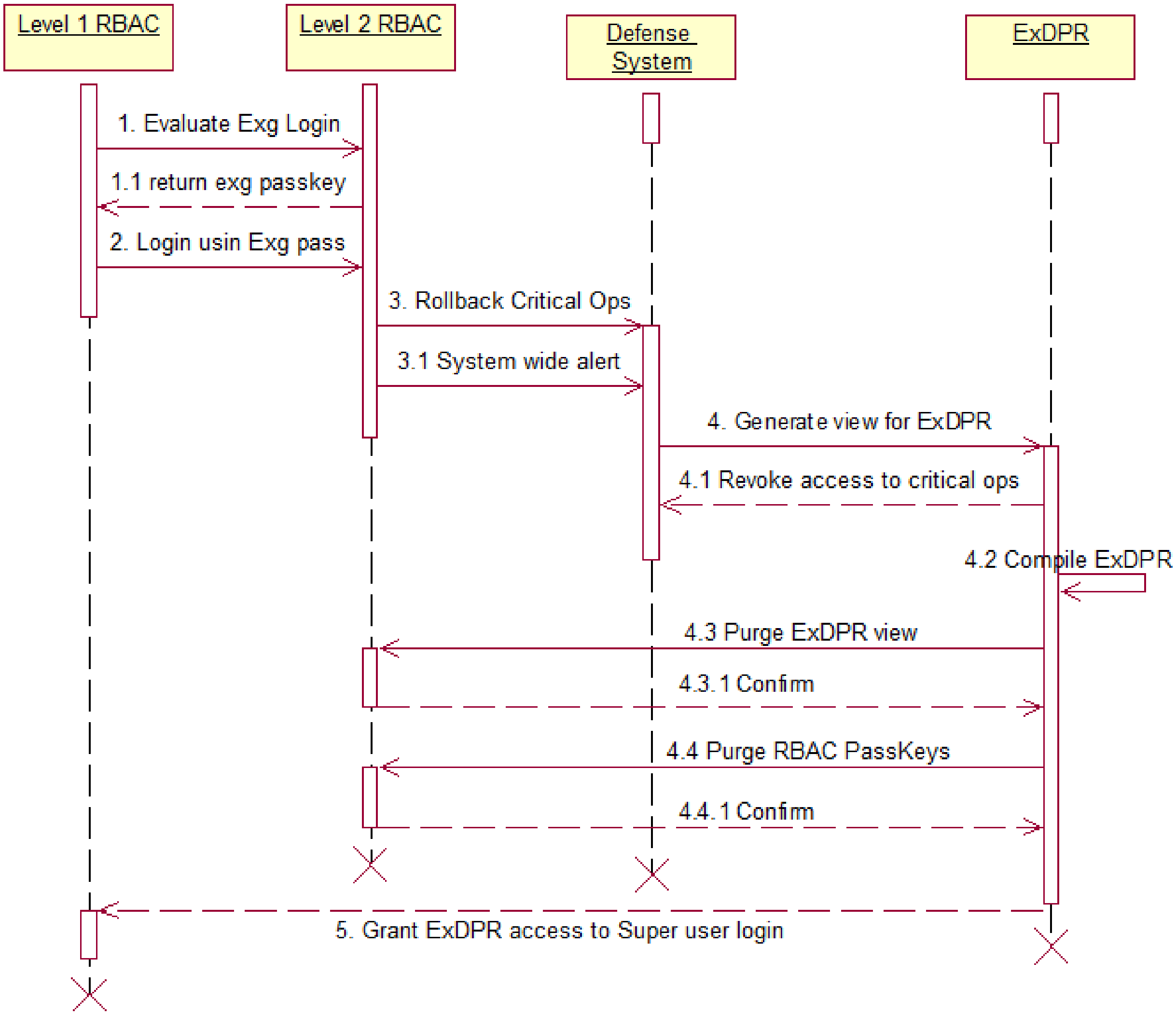
*D.2. Purging and Resetting Pass Keys*

The current pass keys that granted access to the ExDPR generating view and resetting the RBAC pass keys.

The Operations carried out in the Post Security measures are ensure that in case of system compromise, the original view settings and passkeys are never available once the ExDPR has been generated and pass keys have been utilized to gain access to the core system functions of the defense architecture.

The whole scenario includes objects also known as participants in Unified Modeling Language, represented as life lines which communicate among each other. This can be via object referencing or via interface.

To demonstrate the participant wise understanding of the EC-RBAC model elaborating the focus of the detailed description of the each participant involved, EC-RBAC can be represented using sequence diagram shown in figure 5.



## IV. Conclusion And Future Work

In this paper we presented a model for Exigency control in defense system which is capable of securing the critical operations of the defense system by a secured exigent super login contrivance and assigning the permissions to control those core functions as a super sole user.

This mechanism can help the defense system in two ways. (i) as a super control overriding the control on the core operations in the defense circumference and (ii) can also serve as a monitoring tool for the critical operations to ensure the confidentiality, integrity and security of the defense system.

The future work for this research is panoramic, in our forth coming works we intend to proceed on a modular basis. Initially, mainly focusing the policy designing of the contents of the ExDPR and then a secured mechanism, implementing the first and the second level of RBAC as discussed in the paper. At the secondary and tertiary levels we will look into implementation of the mechanism of the dynamic generation of the ExDPR repository via the available technologies that can aid the purpose.

## References

1) Role Based Access Control, American National Standard for technology, Draft 4/4/2003,http://*csrc.nist.gov/rbac/rbac-std-ncits.pdf.*

2) Zhichao Wen; Bo Zhou; Di Wu; "Three-Layers Role-Based Access Control Framework in Large Financial Web Systems," Computational Intelligence and Software Engineering, 2009. CiSE

2009. International Conference on , vol., no., pp.1-4, 11-13 Dec. 2009, doi: 10.1109/CISE.2009.5362682.

3) The Department of Defense Architecture Framework (DoDAF),version 2.0, *Volume 01: Introduction, Overview & Concepts, Manager's guide*, May 2009 doi: 10.1109/EBISS.2009.5137873.

4) Ministry UD (2008). UK Ministry of Defense Architecture Framework MODAF revision version 1.2. Retrieved 15 March 2011, from www.mod.org.uk

5) D. W. Chadwick A. Otenko, RBAC Policies in XML for X.509 Based Privilege Management, Security in the Information Society: Visions and Perspectives: IFIP TC11 17 th Int. Conf. On Information Security (SEC2002), page 7-9, May 2002.

6) Alghamdi, A.S.; Siddiqui, Z.; Quadri, S.S.A.; , "A Common Information Exchange Model for Multiple C4I Architectures," *Computer Modelling and Simulation (UKSim), 2010 12th International Conference on* , vol., no., pp.538-542, 24-26 March 2010 doi: 10.1109/UKSIM.2010.104

7) Duoqiang Wang; Wengfang Liu; Jianfeng Lu; Xiaopu Ma; , "A History-Based Constraint for Separation-of-Duty Policy in Role Based Access Control Model," *E-Business and Information System Security, 2009. EBISS '09. International Conference on* , vol., no., pp.1-5, 23-24 May 2009

8) Sohr, K.; Drouineaud, M.; Ahn, G.-J.; Gogolla, M.; , "Analyzing and Managing Role-Based Access Control Policies," *Knowledge and Data Engineering, IEEE Transactions on* , vol.20, no.7, pp.924-939, July 2008 doi: 10.1109/TKDE.2008.28.

9) Command Control Communication and Computer Intelligent (C4I) organization, official definition, http://www.C4I.org.

10) Fei Hu and Neeraj K. Sharma, Security considerations in ad hoc sensor networks, Ad hoc Networks, volume 3, issn 1570-8705, page 69-89.

# Reconfigurable Steganographic System with Hidden Double Laered Authentication

**G. Margarov, V. Markarov, A. Khachaturov**
Information Security and Software Development Department,
State Engineering University of Armenia,
105 Teryan str., Yerevan, 0009, ARMENIA
gmargarov@gmail.com, vmarkarov@yahoo.com, ashotian@gmail.com www.seua.am

**Abstract** - *This work proposes a reconfigurable steganographic system offering mechanisms for secret sharing, reception of new and restoration of lost parts of the secret data, validity test for the parts of divided secret data, restoration of divided data employing a certain number of divided parts which provides long-term data storage. Finally this work introduces a mechanism for hidden double layered authentication, which hides the existence of steganographic system.*

**Keywords:** Steganography, data hiding, secret sharing, graphical authentication, behavioral authentication, click point.

## 1. Introduction

Nowadays, due to the widespread availability of the Internet and other technologies, the threat of data loss, compromise or an attack on sensitive data located on the Internet, increases simultaneously. Security is an essential constituent of the process of incorporating new information technologies in all the spheres of social life. That is why wide-scale use of information and telecommunication technologies results in qualitatively and constantly evolving new possibilities of an unsanctioned access to the sources and data of various information systems, thus exposing them to ever growing threats and vulnerability.

Steganography is one of the most effective tools for protection of digital information. Modern steganographic systems provide features such as hiding secret data in visual images or in digital objects, which can be located on a local computer, in a local network or the Internet. The latter could be considered as the most appropriate environment for storage (and transfer) of secret data. The biggest downfall of this type of storage or transfer of secret data could be considered the risk of damaging or deleting steganographic files in the result of a targeted hostile attack or of ‘natural’ causes, e.g. such as closure or reorganization of the relevant web pages (mail boxes), etc. A solution to the above-mentioned problems could be the application of reconfigurable steganographic systems [1].

## 2. Reconfigurable stganographic system

The presented reconfigurable steganographic system has the following functionality:

- Secret sharing and recovering
- Checking validity of shares
- Embedding into steganographic containers
- Extracting from staganographic containers

Sharing and recovering of secret data is implemented basing on the Shamir threshold method of data division based on the principle of interpolation [2]. For the division of secret data $S$ using the threshold method $(M, N)$ it is necessary to choose polynomial $\phi(x)$ of $M-1$ with a free term of $S$ value and random coefficients of $a_1,...,a_{M-1}$. With the help of this polynomial, separate parts $S_i$ of secret $S$ can be calculated with the following formula.

$$S_i = S + a_1 i + a_2 i^2 + ... + a_{M-1} i^{M-1} \quad (1)$$

Each of the $S_i$ parts is then hidden in container files $\alpha$, $(|\alpha| = N)$.

To recover the secret data $S$ participation of at least $M$ number of $S_i$ parts is necessary. These $M$ numbers are extracted from the relevant containers $\beta$, $(|\beta| = M)$, and through the application of Lagrange’s method of interpolation [3] the secret $S$ is then being recovered.

$$S = \sum_{i \in \beta} S_i \prod_{j \in \beta, j \neq i} \frac{j}{j-i} \quad (2)$$

Thus we have $N$ containers with embedded secret parts and by using $M$ containers we can recover the whole secret.

To check the validity of shares $p$ and $q$ numbers are randomly selected so that $p = 2 * q + 1$. A number $g$ is selected so that there is a congruence of $g^q \bmod p = 1$. The following equations are calculated: $r_i = g_i^{a_i} \bmod p, i = 1,2,...,t$. For the verification of the validity of this or that component of the secret data the congruence of the following equations are tested:

$$g^{S_j}(\bmod p) = r_0 * (r_1)^j * \ldots * (r_t)^{j^t}(\bmod p) \quad (3)$$

Therefore, if the congruence of the equation (3) works, we can be sure that the tested component, has not been modified or generally replaced by other data, and hence can be used in the process of recovering the secret data [4].

Embedding into steganographic containers as well as extracting, is implemented in the steganographic system with the help of the known algorithms. As steganographic container can be considered any digital file, for which embedding and extracting algorithms have already been developed. In the reconfigurable steganographic system new staganographic methods can be added as addons or plugins.

The staganographic system with a dynamically reconfigurable structure increases security, integrity and availability of the secret information.

Existence of steganographic software on the computer system is suspicious therefore the application of steganography is secret by itself. Thus steganographic software should be hidden and not obviously present in the computer system.

Any security system should have authentication means, to be sure, that only authenticated users can access the secret data. Regular authentication mechanisms show the existence of secret information or system. That's why regular authentication is not acceptable for steganographic system and we need some hidden mechanisms of authentication in this case.

## 3. Hidden Double Layered Authentication

To ensure high secrecy and at the same time providing steganographic (hidden) authentication we propose a method based on the behavior (a set of actions) in an ordinary program, under which a steganographic system itself is hidden. In this paper, a graphic editor is considered as an ordinary program.

Let's denote operations of graphical editor as $d_i(a_{i1}, a_{i2}, ..., a_{in})$, where $a_{ij}$ is the parameter $j$ of operation $i$ in graphic editor. We also define two main operations, $d^*$ as the first operation of the steganographic key and $d^m$ as the final one. The secret steganographic authentication key to the behavioural authentication is the following sequence of operations:

$$D{=}d^*, d_1(a_{1,1}, a_{1,2}, \ldots a_{1,n_1}), d_2(a_{2,1}, a_{2,2}, \ldots a_{2,n_2}), \ldots$$
$$d_k(a_{k,1}, a_{k,2}, \ldots a_{k,n_k}), d^m \quad , \quad (4)$$

where $k$ is the number of operations in the steganographic key and $n_i$ is the number of parameters in the operation $i$.

The user will be authenticated to the steganographic system when the set of actions performed by the user is exactly the same as in case of the registration. Only after a successful authentication a steganographic system will be executed.

To calculate the probability of an accidental authentication during normal operation with the graphical editor we limit the number of possible operations involved in construction of the steganographic authentication key by 60 and the maximum length of the steganographic key by 10. In this case all the possible combinations of the steganographic key will be calculated as follows:

$$\binom{\hat{60}}{10} = \binom{60+10-1}{7} = \binom{69}{7} = \frac{69!}{7!*62!} \approx 10^9 \quad (5)$$

If we assume, that for performing each operation of steganographic key user needs 10 seconds, then the resulting value can be transferred in time complexity, which is $\approx 342$ years.

In fact, different authentication schemes or the second layer of authentication might be used. Nevertheless, use of alphanumeric passwords has certain drawbacks, e.g. passwords with high level of entropy are difficult to remember whereas passwords with low level of entropy are not secure. Biometric authentication is more secure but the weak point is that special hardware is needed. Nowadays schemes based on graphical passwords are alternative and developing trends in authentication systems. .

The main idea of graphical passwords is that the majority of people remember better natural images and pictures rather than artificial words. For example, we can recognize familiar faces among thousands of strangers, and this phenomenon is used in the development of graphical authentication schemes. User can select a sequence of points in the image. This leads to a large number of candidate points if the image is big and complex and has high resolution (number of pixels).

In the proposed scheme the password consists of a sequence of mouse click points, that user selects in the image. The image itself is not a secret and bears no information for the scheme, it only helps the user to remember the click points positions.

To authenticate the user must click close to the points selected during the registration in the appropriate order. Since it is practically impossible for users to click several times on the same point, in the scheme we provide an error admission rate r in click locations (ie the disk with radius r = 10 or 15 pixels). This is implemented by discretization of click locations using three different square grids [5]. The distance between the gridlines of each grid (horizontal or vertical) is 6r. Each of the three grids shifted in a staggered order by 2r vertically and 2r horizontally:

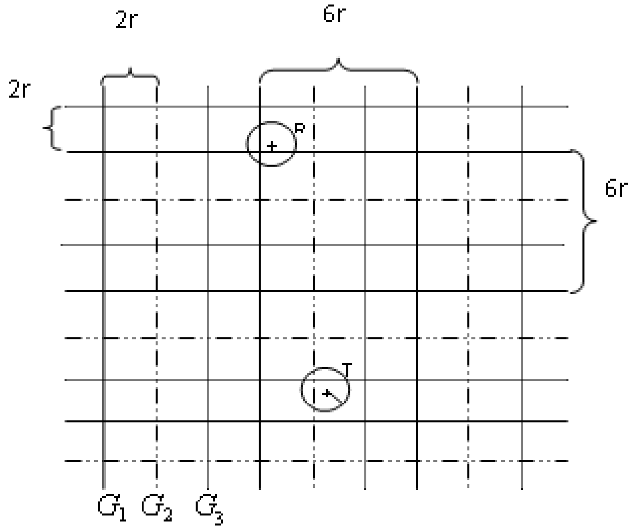


Fig. 1 Three grids map. $G_1$, $G_2$, and $G_3$. Point T is in grid $G_1$. Point R is in $G_2$ and $G_3$

If there were only one discretization grid, the mouse click point might be close to the grid line and slight variations in the click points might result in an incorrect password. On the other hand, it is proved that in case of three staggered grids, the distance of any point in a two-dimensional picture from the grid lines doesn't exceed r [5].

Because of the simultaneous use of multiple grids the selected points are "resistant" to the inevitable minor uncertainties in the point selection; therefore this method of discretization is called "stable discretization" or "stable quantization". Selected positions (click points) are applied to the grids map. The sequence of the selected points is the sequence of grids and grid squares, i.e. $G_1, G_2$ *and* $G_3$ values are calculated for each mouse click,, for example, the value of $R$ is $G_1(-4,-2)G_2(1,2)G_3(1,2)$. A negative value indicates that the circle crosses the relevant horizontal or vertical grid line. To ensure a secure storage of passwords in the system the sequence of grid squares is encrypted using cryptographic hash function.

One of the main attributes of the proposed scheme is that for graphical password creation can be used complex pictures of the real world as well as users can enter their own images. Natural images help users better remember complex passwords. Thus, in the "human context" (i.e. with the participation of users), conditional entropy of password depends on the underlying image.

In 2007 scientists from the University of Carleton, proved using ANOVA analysis tool that the image has a significant influence on the result of authentication [6]. In this sense "quality picture" is a picture on which the user can find a lot of parts (focal points) differing from the background and from each other, which may be potential points for the password. We also tried to determine the optimal size of these focal points by statistical experiment. Two groups of people were involved in the experiment: students with computer skills and people without such. Such a selection is to provide an opportunity to determine the optimal focal point size for users with different skills.

In the first (registration) phase of the experiment, participants were asked to choose any complex image, select three points in this image and remember their location and selection order. In the second (authentication) phase participants should repeat these points 20 times.

As a result, during the experiments the error of the selected points (phase 2) was calculated in comparison with the original ones (points selected in the first phase). Based on the results of 838 qualitative efforts out of 1000, we determined the optimal value of the focal points. Fig. 2 presents the outcoming experimental data (abscissa corresponds to the distance between the registered point and the authentication point, the ordinate shows the frequency of this distance).

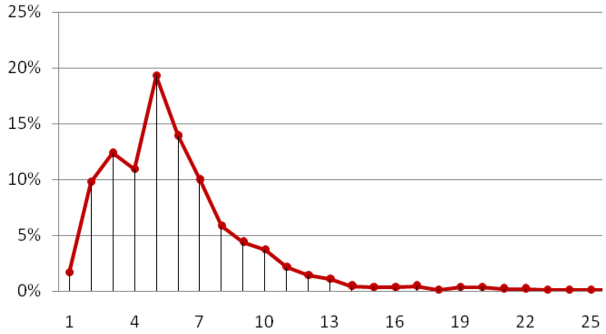


Fig. 2 The result of experiment.

The experiment revealed that for 97% of authentications, distances between the registered point and the authentication point lay within the range of 1 to 14 pixels, therefore, to reduce the number of authentication failures we choose r = 14.

To calculate the probability of accidental authentication, we limit the authentication image size by 1204x768 pixels and the number of points selected as password by 6. In case of such limitations, the maximum number of the possible passwords for graphical and alphanumeric passwords are calculated as follows:

- for alphanumeric passwords with upper and lower case letters, special symbols and numbers maximum number of the possible passwords equals 966
- for graphical passwords the maximum number of the possible passwords is calculated as follows:

$$\begin{pmatrix} \widehat{k} \\ \dfrac{H\cdot W}{(2r)^2} \end{pmatrix} = \begin{pmatrix} k \\ \dfrac{H\cdot W}{4r^2}+k-1 \end{pmatrix} \approx 10^{18} \quad (6),$$

where k is the number of points in the password, H and W are the height and width and r is the error admission rate.

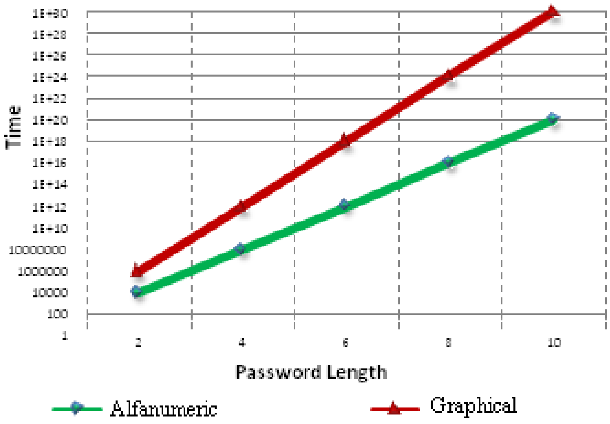


Fig. 3. Stability assessment

In Fig. 3 the stability of alphanumeric and graphical passwords are compared which shows the dependence of the brute force time on the number of characters in a password.

## 4. Conclusion

The paper describes reconfigurable steganographic system which includes methods for secret sharing and recovering, checking validity of shares, embedding into steganographic containers extracting from staganographic containers. For the described system double layered hidden authentication mechanism has been proposed. Behavioral authentication based on the user's behavior in graphical editor has been proposed as the first layer of authentication. A mechanism based on graphical passwords has been proposed for the second layer of authentication. The paper also points out the effectiveness of the proposed authentication mechanisms.

## 5. References:

[1] G. Margarov, V. Markarov, A. Khachaturov, Steganographic system with dynamically reconfigurable structure, Proceedings of the 2009 International Conference on Security & Management, SAM'09, Vol 1, 2009, pp. 43 – 45.

[2] A. Shamir, How to Share a Secret, Communications of the ACM, Vol 22, Issue 11, 1979,pp. 612-613.

[3] J. Berrut, L. Trefethen, Barycentric Lagrange Interpolation, SIAM REVIEW Vol. 46, No. 3, 2004, pp. 501–517.

[4] V. Markarov, A. Khachaturov, Recovery and Verification in Steganographic System with Reconfigurable Structure, Applications of Information Theory, Coding and Security, 2010, pp. 39-42.

[5] J. Birget, D. Hong, N. Memon, Graphical passwords based on robust discretization, IEEE Transactions on Information Forensics and Security, Vol. 1, No.3, 2006, pp. 395-399.

[6] S. Chiasson, R. Biddle , P.C. van Oorschot, A Second Look at the Usability of Click-Based Graphical Passwords, Symposium On Usable Privacy and Security (SOUPS), 2007, pp. 1-12.

# Global QoS Framework for Cloud Security:
# A Paradigm Shift toward a New Trust Concept

**Nabil EL KADHI**
Associate Professor
Ahlia University, Bahrain

nelkadhi@ahlia.edu.bh

**Dana AL Themazi**
System Andministrator
Bahrain Internet Exchange
Student at Ahlia University, Bahrain
system.admin@bix.com

**Abstract: This paper review the classical security services and requirement in an open environment and introduce a Trust Based framework for cloud security. After reviewing deeply cloud computing and cloud challenges we suggest a QoS Framework for a trust based security service.**

*Index Terms*—**Information Security, Security services, cloud computing, Trust, Qos for security**

## 1. Introduction

Looking to the Humanity history and achievement, it is nowadays notable what a huge impacts does technological progress and inventions have on market places. As we have seen those last decades, Computers and mainly Personal Computers have transformed the human behavior, capabilities of data processing and computing. The nineteenth have been clearly marked by the democratization of the large and wide area networks and mainly the Internet technologies.

From one step to another the interest of security features and their importance increased progressively

In fact from a mainly physical protection of the huge size computers we gradually emphasis on data protection and data access control (mainly on 70th and 80th) to gradually focus more on communication protection and distant application management and finally e-processing of crucial data through a wide set of applications.

E-Banking, E-Ticketing, E-Commerce and E-Anything lead to the migration of sensitive and crucial data from a relatively controlled and sizable environment to an almost non controlled unbounded environment. Those changes and progression create a new vague of security tools and services ranging from the data protection technique [19, 63], to the cryptographic protocol design, implementation and verification [6,13,14,9].

A deep analysis of security levels and action domains allow us to distinguish today the following:

- Security tools and protocols for data and communication protection [9, 42]
- Security tool for prevention [38]
- Security tools for attacks detection and logging ([44, 26, 28])

Among all those changes and improvement the security services kept almost the same basic definitions with slight adaptation from one environment to another. As described section 2, five security services are required and aim in any complete scurried information system.

In this paper, we argue that the newly Cloud computing and Cloud applications are introducing today a quite important paradigm shift in information security. In fact, Cloud computing as described section 3 relay on virtualization and virtual environment that may be seen as a drastic change when dealing with security tools and services. Basically, and for the first time, we need to define and specify security features independently from the applied technique, the physical layers and tools as well as the intended goals.

What is the classical view of security services? What cloud computing emergent features? What are the cloud security challenges and how we are claiming answering chose recent conditions? Those are the major questions addressed in this paper and leading to a newly defining Qos Framework for a Trust Based cloud security.

Our paper is structured as following. **Section 2** is a global overview of the security services and their requirements, **section 3** brows the cloud computing features and properties, **sectio**n 4 review the cloud security challenges and particularities. **Section 5** introduce the paradigm shift leading to a trust based security and introduce our QoS framework.

## 2. Security services overview

Specialized literature and among the various classifications of security services and tools mainly agreed on the five following required services in any Information system security solution.

- **Access control and Integrity:** One of the common definition of integrity is that the content (of a file or a message) is as it is supposed to be after the last known authorized access. Such constraint may easily be linked to access control rules and logging when dealing with physical data protection and file and database use. It is quite different in a communication context since it refers more to the integrity and conformity between the initial (in general sent) version and the actually checked (commonly the received) version of an exchanged message over a non private network. Various technique are commonly applied for access control [18, 48, 50] and Integrity check values based on Hash functions [17, 60, 16]
- **Authenticity:** As well as Authentication refers to either having the proof of the identity of the communicating parties or the proof that the analyzed (received) data is as it has been sent by an already authenticated user. As we can easily notice it authenticity leads, implicitly, to integrity. A large set of cryptography based techniques are applied for authentication verification [59, 37]
- **Secrecy:** Despite the differences between Data secrecy and Information secrecy, we usually intend to ensure that a crucial content is analyzed and being interpreted in a usable way only by the legitimate users. As it has been widely discussed in a large set of research and industrial worksheet, cryptography still the main stone when dealing with data and information security [51].
- **Data flow secrecy:** Commonly known as passive attacks, the data flow analysis is in fact the set of intruder actions allowing data collection regarding the attacked organization or networks. It consist in general in a variety of static analysis of received/sent data volume, language frequency and so on. It dos in fact include any useful information that may be used later on to build active attacks. Globally there is no efficient data flow secrecy technique. The goal here is to lead the intruder to somehow wrong information or a non realistic environment such as done by the honey pot project [20].
- **Non repudiation:** The non repudiation is generally based on certification techniques and/or acknowledgement. The main goal is to have formal proof that a sender issued a content or a message when dealing with sender non repudiation. It also means that any receiver will not be able to deny the reception of an acknowledged or certified received data.[25 ]
- **Availability**: Availability may be defined in different context and in various manners depending on the offered services and activities. Independently, availability, in a security service context, means basically that the offered/required service is active and accessible in way ensuring that all the previous security services are respected and ensured. This means basically that we are able to ensure all the previously mentioned services or to detect, and clearly notify, any miss use or defection in any security service. Availability is no more than a result of the combination of the main described security services.

## 3. Cloud computing: A global overview

After introducing the five security services elements, let's now describe how these elements are applied in Cloud computing. To start first we should define the term of Cloud Computing? What are the Cloud services? And do we need the cloud computing in our life?

Cloud computing has been called the 5th utility in the line of electricity; water, telephony and gas. The reason why cloud has been called with such a name is that the cloud computing has been changing the way computer resources have been used up to now [41].

Cloud computing has brought a standard change in how computing resources have been worked. And that's works as the following; Cloud providers host their resources on the internet on virtual computers and make them available to multiple clients. Multiple virtual computers can run on one physical computer sharing the resources such as storage, memory, the CPU and interfaces giving the feeling to the client that each client has his own dedicated hardware to work on. Virtualization thus gives the ability to the providers to sell the same hardware resources among multiple clients.

Cloud computing is the computing equivalent of the electricity as "Miller" says in his world, it's like the "revolution of a century ago" [34] . Previously the advent of electrical utilities, in every farm and business produced its own electricity from freestanding generators. But after the electrical grid was created, all the farms and businesses shut down their generators and bought electricity from the utilities, at a much lower price. (and with much greater reliability) than they could produce on their own. Cloud computing came with huge revolution to ensure that cloud computing takes hold. The desktop-centric notion of computing that we hold today is bound to fall by the wayside as we come to expect the universal access, 24/7 reliability, and ubiquitous collaboration promised by cloud computing. It is the way of the future.

### 3.1 Definitions:-

In a world that sees new technological trends bloom and fade on almost a daily basis, one new trend promises more longevity. This trend is called cloud computing, and it will change the way you use your computer and the Internet. Cloud computing portends a major change in how we store information and run applications. Instead of running programs and data on an individual desktop computer, everything is hosted in the "cloud"—a nebulous assemblage of computers and servers accessed via the Internet. Cloud computing lets you access all your applications and documents from anywhere in the world, freeing you from the confines of the desktop and making it easier for group members in different locations to collaborate.

Cloud computing has many meaning and definitions, but the most definitions are depends on the understanding of the concept of cloud computing technology and how its work, in sample way "Cloud computing is a general term for anything that includes using hosted services over the Internet everywhere". [48].

In other word, Cloud computing enables a move from the computer to the user, from applications to tasks, and from isolated data to data that can be accessed from anywhere and shared with anyone. On the other hand "The user no longer has to take on the task of data management; he doesn't even have to remember where the data is. All that matters is that the data is in the cloud, and thus immediately available to that user and to other authorized users." [34]. Nevertheless, "The Cloud Computing is regarded as an evolutionary rather than a revolutionary step. In other words, Cloud Computing hasn't drastically altered existing technologies, but rather it has succeeded as a result of the collaboration of several existing technologies. The actual definition of cloud computing is frequently contested. Most will agree that any computing model that qualifies as cloud computing must at minimum have these criteria: Elasticity, Multitenancy, Economics, & Abstraction."[4].

Cloud computing provides a widely more efficient, flexible, Elasticity, QoS, and cost-effective way and more than that for IT to meet escalating business needs. Cloud computing has three main services which are "Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), Software-as-a-Service (SaaS)". [12,13,14,15,16]. All these services are presented on the three types of the cloud, First is "Privet Cloud", Second is "Public Cloud", Third is "Hybrid

Cloud". [10]. These types coming under the sizing of the cloud.

### 3.2 Sizing:-

Sizing the cloud computing, it's mean the capacity of the Cloud in general, the cloud is virtual space where we can find the three types of the cloud. Each type has the ability to serve the data or information's which is in the cloud, also it has the specifications for each one of them, and for all types we should use Virtualization technologies.

**Firstly**, In public cloud we have to use enterprises cloud functionality from to respectively offer their own services to users outside of the company. That's by providing the user with the actual capability to exploit the cloud features for his / her own purposes also allows other enterprises to outsource their services to such cloud providers, thus reducing costs and effort to build up their own infrastructure. Example of that: Amazon, Google Apps, Windows Azure. [54].

**Secondly**, In Private Clouds we are typically owned by the respective enterprise and / or leased. Functionalities are not directly exposed to the customer, though in some cases services with cloud enhanced features may be offered – this is similar to (Cloud) Software as a Service from the customer point of view. Private clouds are of two types: On-premise private clouds and externally hosted private clouds. Externally hosted private clouds are also exclusively used by one organization, but are hosted by a third party specializing in cloud infrastructure. Externally hosted private clouds are cheaper than On-premise private clouds. Example: eBay. **Thirdly**, In Hybrid Clouds though public clouds allow enterprises to outsource parts of their infrastructure to cloud providers, they at the same time would lose control over the resources and the distribution /management of code and data. In some cases, this is not desired by the respective enterprise. Hybrid clouds consist of a mixed employment of private and public cloud infrastructures so as to achieve a maximum of cost reduction through outsourcing whilst maintaining the desired degree of control over e.g. sensitive data by employing local private clouds. There are not many hybrid clouds actually in use today, though initial initiatives such as the one by IBM and Juniper already introduce base technologies for their realization [53] . In other words, the organizations may host critical applications on private clouds and applications with relatively less security concerns on the public cloud. The usage of both private and public clouds together is called hybrid cloud. A related term is Cloud Bursting. In Cloud bursting organization use their own computing infrastructure for normal usage, but access the cloud for high/peak load requirements. This ensures that a sudden increase in computing requirement is handled gracefully. [63].

In our paper we will focus on (PaaS), since this service become a middleware platform. Pre integrated and, in many cases, simplified platforms for the development of general-purpose business applications will become a serious alternative for developing custom applications.

We found some suggested article [34], to subdivisions cloud sizing by grouped them and offering them in order to make forecasts for cloud computing more virtual In particulars:-

1. **Cloud Access Devices (CADs)** – Grouped the devices to which Cloud Computing is delivered, taking PCs, smart phones, smart tablets gaming consoles, peripherals and set top boxes; by estimated only the expenditure on those devices accessing the relevant applications and excluding the proportion of the time these machines are involved with traditional computing; iPhones downloading music from iTunes are excluded, those using Google maps are included for instance.
2. **Cloud Delivery Platforms (CDP)** – For CDP happen should conform the trust relation between the clouds and that is by achieve the five security services elements which shown in section 2. Plus the Service-Level-Agreement (SLA) between the trusted cloud. When all these components available and passed successfully, then deliver the

Cloud Computing Platform to a company's own employees or consumers; elements including servers, storage systems, networking equipment, operating systems, infrastructure software and applications as well as hardware and software maintenance will be legally insure the trust between the clouds.

3. **Cloud Telecom Connections (CTC)** – It will estimate the costs of Telecoms service (fixed line and wireless) used to access and deliver Cloud Computing applications to users. [5, 39].

4. **Cloud Development Services (CDS)** – In (CDS) it's also depend on CDP, these two components are capturing the amount being spent on developing Cloud Computing by users and aggregators when the trust is not there. The components include implementation service, as well custom development and test software. We'll descript it more in section5.

5. **Retail Cloud Services (RCS)** – This is the third candidate for (CDP, and CDS) which insure that the trust relations between the clouds are meet the SLA agreement also to capture the majority of Cloud Computing spending; and the two existing 100% classes are the **Broadband and Internet Service**, they joined by the newer services in the cloud.

6. **Cloud Location (CL)** – Cloud location is designed to identify the physical place of the cloud, and to insure that the services are hosted in safe and secured place. All this to make sure that the business between selling services and Internet-Service-Provider ISP, are happening in the virtual privet cloud.

Finally, the major component in the Cloud is to build the trust between the clouds for exchanging or getting the data and other IT factors. In other hand, there are other services in the cloud we didn't go through them, such as (Storage-as-a-Service), (Database-as-a-Service), (Information-as-a-Services), (Security-as-a-Services), etc. [63]. For all these services and our main three services (PaaS, IaaS, and SaaS) there are aspects and crucial applications, we will see them in the next point.

## 4- Aspects and crucial application:-

Over the last few years we've been seen remarkable growth of the aspects and crucial applications for the Cloud Computing, as witnessed by the many popular Web applications used today, including VoIP applications, social applications, media services applications, content distribution applications , financial applications, E-mails applications, Backup services applications, Monitoring services applications. [11, 8] One of the most popular applications in the cloud is the Customer Relationship Management (CRM), Microsoft recently released their on-line "cloud" based CRM (Customer Relationship Management) at an introductory price of $42/month per user. As a user and seller of Microsoft CRM solutions, this is a welcome entry for businesses, especially small to medium businesses. This financial benefit to businesses is that you can centralize all your client contacts in one area for a low operational cost per user. The alternative of paying for an on-premise solution is still available, but the cloud based solution is more cost and time effective. CRM is one of the original "killer applications" for cloud computing where you pay as you go and only pay for what you use. In fact, the cloud based CRM leader, SalesForce.com, has been pushing the on-line model for many years. The big difference with Microsoft announcement: The Price and Efficiency with Microsoft look and feel. Microsoft has designed their products to work together and have a common look. This makes it easier for CRM to be adopted by users already familiar with Microsoft. [11, 8, 22].

Before ending the aspect's and crucial applications, we should reminds that there are many of applications, but we chose the fames one which is CRM; and that is to insure that the CRM cloud application are suitable with all the clouds, but first we must insure the trust relations as we showed it in section 3.2.

### 4.1 Cloud Security Challenges:-

There are many benefits of cloud computing by virtue of abstraction, prevents the consumer from having the same level of influence over the computing resource. Great concern is the ability of consumer to assert quality of service [5]. QoS refers to aspects of a service that are not functional but are important considerations, This is leads to some of the following challenges with public cloud computing. “One of the key challenges in cloud computing is data-level security” [24]. Starting with the most important challenges which are:

- (Availability)
- Then ,( Data Residency),
- And,( Multitenancy),
- With,( Performance),
- (Data Evacuation),
- Ending with Supervisory Access & Privacy**.**

Even large enterprises with significant resources face considerable challenges at the network level of infrastructure security. Are the risks associated with cloud computing actually higher than the risks enterprises are facing today? Consider existing private and public extranets, and take into account partner connections when making such a comparison. For large enterprises without significant resources, or for small to medium-size businesses (SMBs), is the risk of using public clouds (assuming that such enterprises lack the resources necessary for private clouds) really higher than the risks inherent in their current infrastructures? In many cases, the answer is probably no—there is not a higher level of risk. In other hand, the virtualization technologies enable multitenancy cloud business models by providing a scalable, shared resource platform for all tenants. More importantly, they provide a dedicated resource view for the platform’s consumers. From an enterprise perspective, virtualization offers data center consolidation and improved IT operational efficiency. Today, enterprises have deployed virtualization technologies within data centers in various forms, including OS virtualization (VMware, Xen), storage virtualization (NAS, SAN), database virtualization, and application or software virtualization (Apache Tomcat, JBoss, Oracle App Server, Web Sphere). From a public cloud perspective, depending on the cloud services delivery model (SPI) and architecture, virtualization appears as a shared resource at various layers of the virtualized service (e.g., OS, storage, database, application) [22,24].

The simplicity of self-provisioning new virtual servers on an IaaS platform creates a risk that insecure virtual servers will be created. Secure-by-default configuration needs to be ensured by following or exceeding available industry baselines. Securing the virtual server in the cloud requires strong operational security procedures coupled with automation of procedures. Here are some recommendations:-

• Use a secure-by-default configuration. Harden your image and use a standard hardened image for instantiating VMs (the guest OS) in a public cloud. A best practice for cloud based applications is to build custom VM images that have only the capabilities and services necessary to support the application stack. Limiting the capabilities of the underlying application stack not only limits the host’s overall attack surface, but also greatly reduces the number of patches needed to keep that application stack secure.

• Track the inventory of VM images and OS versions that are prepared for cloud hosting. The IaaS provider provides some of these VM images. When a virtual image from the IaaS provider is used it should undergo the same level of security verification and hardening for hosts within the enterprise. The best alternative is to provide your own image that conforms to the same security standards as internal trusted hosts.

• Protect the integrity of the hardened image from unauthorized access.

• Safeguard the private keys required to access hosts in the public cloud.

• In general, isolate the decryption keys from the cloud where the data is hosted—unless they are necessary for decryption, and then only for the duration of an actual decryption activity. If your application requires a key to encrypt and decrypt for continuous data processing, it may not be possible to protect the key since it will be collocated with the application.

•Include no authentication credentials in your virtualized images except for a key to decrypt the file system key.

•Do not allow password-based authentication for shell access.

*PaaS* platforms also have functional differences from traditional development platforms and that make it one of the challenges, which include [57]:

- **Multitenant development tools:-**
  Traditional development tools are intended for a single user; a cloud-based studio must support multiple users, each with multiple active projects.
- **Multitenant deployment architecture:-**
  Scalability is often not a concern of the initial development effort and is left instead for the system administrators to handle when the project deploys. In PaaS, scalability of the application and data tiers must be built-in (e.g., load balancing and failover should be basic elements of the developing platform).
- **Integrated management:-**
  Traditional development solutions (usually) are not associated with runtime monitoring, but in PaaS the monitoring ability should be built into the development platform.
- **Integrated billing:-**
  PaaS offerings require mechanisms for billing based on usage that are unique to the SaaS world.

One of the furthermost challenges are Identity Management-As-A-Service (IDaaS) only recently emerged as an example of SaaS, in comparison to email filtering, web content filtering, and vulnerability management, which are more established as SaaS offerings. There are some significant deficiencies in today's Identity and Access Management (IAM) capabilities with regard to uses in cloud computing (e.g., scalability). IDaaS attempts to provide some IAM services in the cloud. [57].

To end with benefits of PaaS lie in greatly increasing the number of people who can develop, maintain, and deploy web applications. In short, PaaS offers to democratize the development of web applications in much the same way that Microsoft Access democratized the development of the client/server application [57]. For that we must know the exiting tools which in the cloud computing to understand it more.

**4.2 Existing Tools:-**

Cloud Computing introduces a number of new challenges around the tools and services required to build and maintain applications. In fact one of the most important existing tools in the Cloud Computing which is Application Programming Interfaces (APIs). This tool is suitable application programming interface (API) and it is another enabler for the cloud computing services delivery model . APIs empower users by enabling features such as self-provisioning and programmatic control of cloud services and resources. Depending on the type of cloud services delivery model (SPI), an API can manifest in different forms, ranging from simple URL manipulations to advanced SOA-like programming models. APIs also help to exploit the full potential of cloud computing and mask the complexity involved in extending Existing IT management processes and practices to cloud services. [57, 30].

Although there is no cloud API standard, standardization efforts are mushrooming and are driven by vendor as well as user communities. One such effort is Universal Cloud Interface (UCI), an attempt to create an open and standardized cloud interface for the unification of various cloud APIs. The UCI forum claims that the goal is to achieve a singular programmatic point of contact that can encompass the entire infrastructure stack, as well as emerging cloud centric technologies, all through a unified interface.[22]. The alternative to PaaS is to develop web applications using desktop development tools, such as Eclipse or Microsoft Access, and then manually deploy those applications to a cloud - hosting provider, such as Amazon Web Services (AWS) [57,30].

In fact the users are accessing the cloud computing every day, by using the internet network client devices interface such as laptops, desktops, smartphones, tablets, etc. some of these (cloud clients) rely on cloud computing for all majority of their applications, ether for reading form e-library or viewing images or other applications, but for that it required to use programing languages such as HTML, XML, Java and other languages, to make the interface of the user devices looking batter.

Once we introduce the challenges and the existing tools in the cloud computing, we are going to explain more about the trust and the cloud computing paradigm shift using QoS Framework in section 4.

## 5 Trust: A Cloud Paradigm shift using QoS Framework

Among the previous sections in this paper we tried to show the gap or new challenges introduced by the cloud technology when dealing with security services. In fact security services relays mainly on cryptography, hashing functions and a set of parameters strongly correlated to the implementation environment and the used architectures. This appears for example thru the SSL protocol handshake step or the LDAP and Kerberos implementations for authentications [35, 36].

Cloud technology comes mainly with a level of abstraction hiding internal details and implementations. Abstraction layers as introduced by several virtualization tools [27, 55] allow data mobility as well as platform independency use and code migration as shown in Section 3 and 4.

Many leaders in Cloud solutions [12, 61, 21] are today offering a global cloud environment supporting a quite complete range of security services. A highest cloud and virtualization level will allow user to migrate from one platform to another in a quite transparent way. Actually cloud platform seems to be non interoperable. And if a compatibility exists, it is en general limited to data interoperability and access. Migrating from one platform to another necessitate in general a new authentication, a new definition or paradigm shift of the security concepts definitions, costs and techniques as supported by the concerned platform. Such limitation is considered as brakes to the cloud use and reduces it is efficiency. In this paper we argue that a higher abstraction is in fact required for the security services in order to introduce a more flexible cloud environment.

The idea is to be detached from the classical definitions of the security services where the service verification and achievement is linked to the used techniques, the implementation method and/or the applied constraints and to support the dynamicity of the cloud environment.

The paradigm shifts we are considering here intend to merge all the security services to a single one which is a **trust relationship**. Two cloud environments will trust each other regarding any of the claimed or requested security service. The trust relation is here in fact like a certificate exchanged between the two cloud environment to ensure that the required service is offered by the two parties without being obliged to decline any specific technique tool or environment.

Even the idea seems to be quite sample and intuitive it does necessitate a deep analysis. In fact, assuming that a required service (let say authentication for example) is offered by two cloud environment. Shall we consider the authentication of one valid for the other? Shall we consider that the authentication requirements of the two parties are compatible? This may or may not be the case depending on several conditions and contexts. It is of course inacceptable to check different conditions and constraints as many times as a virtual environment is changing. The key solution here is to replace such challenging condition by an abstraction where only simple verification is required. If we look deeply to the authentication requirements (as in our actual example) we are in fact not caring much more about the used technique, the key size and the certification authority than the hardness, level or quality of the offered authentications in any environment. So if the two environment require a simple authentication or a strong one and if there is a third party to classify and certify the authentication level of each party, without deeply detailing the used techniques, the two parties will be able to 'trust' each other regarding the authentication even if they are using various virtualization platform and security techniques.

The additional challenge here is that the 'trust' relation and if we assume that it has been established initially, must be guaranteed among all the cloud dynamicity and physical environment changes. Each party will have and additional challenge which is having an adaptable dynamic certificate that guarantee that the 'trust' relation is active when changing any cloud environment component. QoS contracts are a suitable tool to ensure a continually verified set of conditions and rules.

### 5.1 Trust between PaaS Platforms in the clouds

As discussed in the earlier point that the trust relation in clouds it required a certificate to exchanged between the two cloud environment to ensure that the required service is offered by the two parties. Before that we should clarified the trust? And we should know the Components of Trust in Cloud Computing?

The Trust can be explained in a diverse fields such as psychology, sociology, and economics. Also trust can be classified in different meaning for many writers. But In dictionaries and other authors , trust is generally related to "levels of confidence in something or someone" [7,40,30].

The classified trust which is used in Peer to Peer networks based on the models used to build them. It's identified and described as the following;

1- CuboidTrust,
2- EigenTrust,
3- Bayesian Network based Trust Management (BNBTM),
4- GroupRep,
5- AntRep,
6- Semantic Web,
7- Global Trust,
8- PeerTrust,
9- comPrehensive reputAtion-based TRust mOdeL with Fuzzy subsystems (PATROL-F),
10- Trust Evolution,
11- Time-based Dynamic Trust Model (TDTM)
12- Trust Ant Colony System (TACS)

These classified trust must be in the cloud, also as its described in section (3.2) , Cloud Delivery Platforms (CDP), Cloud Development Services (CDS), & Retail Cloud Services (RCS); with these factors and Security services.

In this paper we want to assure that the trust relation, between clouds as explained in sections (5, 3.2) should achieve the QoS stander when any cloud will exchange their platform services. The main *components affecting cloud trust* are:

**1) Security:** Mechanisms (e.g. encryption) which make it extremely difficult or uneconomical for an unauthorized person to access some information.[23,40].

**2) Privacy:** Protection against the exposure or leakage of personal or confidential data (e.g. personally identifiable information (PII)).[7]

**3) Accountability:** Defined as "the obligation and/ or willingness to demonstrate and take responsibility for performance in light of agreed-upon expectations".

**4) Auditability:** The relative ease of auditing a system or an environment. Poor auditability means that the system has poorly-maintained (or non-existent) records and systems that enable efficient auditing of processes within the cloud. Audit ability is also an enabler of (retrospective) accountability "It allows an action to be reviewed against a pre-determined policy to decide if the action was compliant", and if it was not, to hold accountable the person or organization responsible for the action.[40,62]. The Framework of trust in cloud are as shown in the figure.1 [49,50].

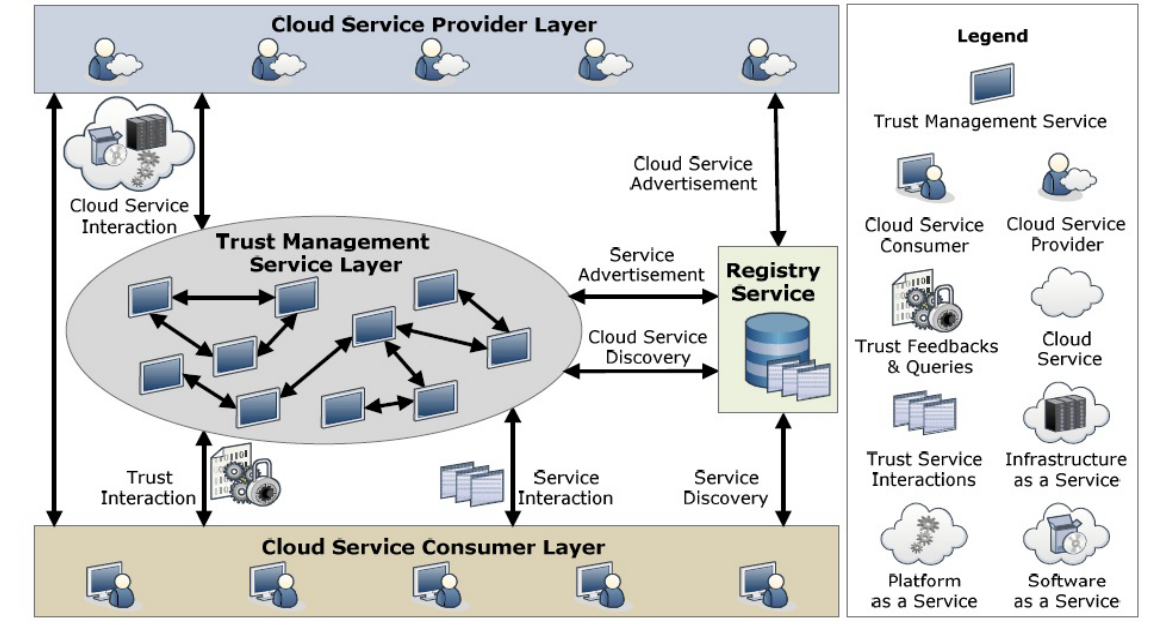


**Fig. 1.** Architecture of the Trust as a Service Framework

1. **The Cloud Service Provider Layer:** This layer consists of different cloud service providers who provide cloud services. The minimum indicative feature that every cloud service provider should have is to provide the infrastructure as a service (i.e., the cloud provider should have a data center that provides the storage, the process, and the communication).
2. **The Trust Management Service Layer**: This layer consists of several distributed TMS nodes that expose interfaces so that cloud service consumers can give their trust feedbacks or inquire about the trust results represents. [63].

## 6 Conclusions

At last we should know that the cloud computing are the next huge wave in computing world, and for success the security level in the cloud computing; it must achieve all the security services features, which we mentioned it in section1. As we knows that the cloud computing holds a lot of promises and we believe that it is likely to be a major influence on hosting and applications developments. Also it has many benefits such as hardware managements, software's, and virtualizations; since all the computers and servers are almost same from hardware side. A cloud also provides better and easier management of data security in. Furthermore, the cloud computing has three types of clouds and each type has his own services, and our main focus in this paper is about PaaS, and the way of exchanging the data or platforms from cloud to other cloud, and for that Trust relation should be there also all the trust layers must be present. When the trust relation succeed then we will achieved our QoS framework of trust based on cloud security, through the SLA's and other standers just to insure the Quality of the trust in the cloud. Finally, we have to give emphasis to the importance of PaaS as the future of cloud services and we are now seeing complete transformation of PaaS space with all the Platforms vendors focusing on multi language and multi cloud trends. In fact the way of developing the PaaS framework is by develop the way of securing the clouds with insuring the infrastructure security, data security and storage, security management, privacy and Security as a service.

**References.**

**1.** Aidan Finn, Hans Verdevoort, Patrick Lownds, and Damian Flynn, Microsoft Private Cloud Computing - (Jun 26 , 2012).

**2.** Anthony T. Velte, Toby Velte, Robert Elsenpeter, Cloud Computing A Practical Approach, Mc Graw Hill Companies, ISBN: 978-0-07-162695-8, Edition – 2010.

**3.** Barrie Sonsinsky , Cloud computing Bible , (Jan 11, 2011).

**4.** Ben Halpert, (2011), "Auditing Cloud Computing",( pp: 02); United State Of America/Library of Congress Cataloging.

**5.** Ben Halpert, (2011), "Auditing Cloud Computing",( pp: 10-12);United State Of America/Library of Congress Cataloging.

**6.** Bicarregui, J.C. and Matthews, B.M.; Formal Methods in Practice: A Comparison of Two Support Systems for Proof, SOFSEM '95: Theory and Practice of Informatics, Editions Springer-Verlag 1995,

**7.** Changhoon Lee, Jean Marc Seigneur, James J. Park and Roland R. Wagner, Secure and Trust Computing, Data Management, and Applications: STA 2011 Workshops: IWCS 2011 and STAVE 2011, Loutraki, Greece, Jun 28-30.2011, in computer information science, (Nov 4, 2011).

**8.** Charles Babcock, Management Strategies for the cloud Revolution: How Cloud Computing Is Transforming Business and Why /you Can't Afford to Be Left Behind – (Apr 16, 2010).

**9.** CISCO, http://www.cisco.com. Accessed March 2008

**10.** David Patterson and Armando Fox , Engineering Long-Lasting Software: An Agile Approach Using SaaS and Cloud Computing, Alpha Edition (Jan 12, 2012).

**11.** Don Peppers, Managing Customer Relationship: A Strategic Framework – (Jan 11,2011).

**12.** Edward Haletky, VMware vSphere and Virtual Infrastructure Security: Securing the Virtual Environment, ( Jul 2, 2009).

**13.** EL Kadhi, N and H. EL Gendy, Advanced Method for Cryptographic Protocol Verification, Journal of Computational Methods in Sciences and Engineering, Volume 6, Issue 5, Supplement 1, Year 2006, Pages: 109-119

**14.** EL Kadhi, N, Hadjar, K and EL Zant, J, A Mobile Agents and Artificial Neural Networks. For Intrusion Detection, JOURNAL OF SOFTWARE, VOL. 7, NO. 1, JANUARY 2012

**15.** Eugenio Pace, Moving Application to the Cloud on the Microsoft Azure Platform (Patterns & Practces)- (Sep 6 , 2010).

**16.** FIPS PUB 180-1, Supersedes FIPS PUB 180, SECURE HASH STANDARD, 1993 May 11

**17.** Gene Tsudik, "Message Authentication with one-way Hash Functions", ACM SIgcom 1992

**18.** Gopalakrishnan, M. & Patnaik, L.M Medium access Control Schemes for Local Area Networks with Multiple Priority" Function, the computer journal, vol 31 N°3 1988

**19.** Goscinski, Andrzej, Distributed Operating Systems, The Logical Design. Addison-Wesley 1991

**20.** Hone ypot Project: http://projecthoneypot.org/

**21.** James Beswick, Google Apps Express: The Fast Way To Start Working in the Cloud , (Mar 5, 2011).

**22.** Joel Scott, Microsoft CRM for Dummies – (Jul 25, 2003).

**23.** John Townsend , Beyond Boundaries: Learning to Trust Again in Relationships (Sep 27, 2011).

**24.** John W. Rittinghouse and James F. Ransome, (2009), "Cloud Computing: Implementation, Management, and Security", New York, Auerbach Publications.

**25.** Jose A. Onieva, Secure Multi-Party Non-Repudiation Protocols and Applications (Advances in Information Security (Dec 8, 2010).

**26.** K. Deeter, K. Singh, S. Wilson, L. Filipozzi and S. Vuong, APHIDS: A Mobile Agent-Based Programmable Hybrid , Intrusion Detection System, in Mobility Aware Technologies and Applications. LNCS, vol. 3284, (Springer, Heidelberg, 2004).

**27.** Kenneth Hess and Joe Brockmeier, Practical Virtualization Solutions: Virtualization from the Trenches , (Oct 22, 2009).

**28.** Ktata, F, EL Kadhi, N and Ghedira K. MAFID Agent IDS Based on Missuse Approach, Journal Name: Journal of Software, Volume 4 Number 6 Year 2009, Pages: 495-507, ISSN 1796-217X, ACADEMIC PUBLISHER

**29.** Lars Nielsen, The little Book of cloud computing , 2011 Edition (Apr 11, 2011).

**30.** Lars Nielsen, The little book of cloud computing security - Edition (Dec 18,2011).

**31.** Lvanka Menken, Cloud Computing PaaS Platform and Storage Management Specialist Level Complete Certification Kit – Platform as a Service Study Guide Book and Online – Certification Sepcialist – Second Edition (Aug 20, 2010).

**32.** Manuel, P., Thamarai Selvi, S., Barr, M.E.: Trust Management System for Grid and Cloud Resources. In: Proc. of ICAC 2009, Chennai, India (December 2009).

**33.** Martin Hingley, (September 21, 2011), "UK Cloud Computing Forecast Highlights", (UK Cloud Computing Forecast – Recession-Busting Growth), available at (http://itcandor.net/2011/09/21/cc-uk-q311/).

**34.** Michael Miller, (2009) ," Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online",( pp: 07,08,09); United State Of America/Library of Congress Cataloging.

**35.** MIT, Kerberos, 'Kerberos: The Network Authentication Protocol, available at (http://web.mit.edu/Kerberos/ )

**36.** MIT , Kerberos: K'Kerberos Security Advisories. Available at (http://web.mit.edu/Kerberos/advisories/ )

**37.** Needham, R. M. and Schroeder, Michael D., Using Encryption for Authentication in Large Networks of Computers, Communications of the ACM}, V 21, Number CSL-78--4}, December 1978.

**38.** Nitesh Dhanjani and Justin Clarke, Network Security Tools: Writing Hacking and Modifying Security Tools, Apr 11, 2005

**39.** Open Trust Security is about Trust, (2012), available at (http://www.opentrust.com/en/saas ).

**40.** Qing Zhang, Ting Yu, and Keith Irwin, "A Classification Scheme for Trust Functions in Reputation-Based Trust Management," in International Workshop on Trust, Security, and Reputation on the Semantic Web, Hiroshima, Japan, 2004.

**41.** R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility, Future Generation Computer Systems, Volume 25, Number 6, Pages: 599-616, ISSN: 0167-

739X, Elsevier Science, Amsterdam, The Netherlands, June 2009.

**42.** RealSecure, http://www.iss.net. Accessed March 2008.

**43.** Russell Dean Vines and Ronald L. Krutz - Cloud Security : A Comprehensive Guide to Secure Cloud Computing - (Aug9, 2010).

**44.** S. Specht and R. Lee, Distributed Denial of Service: Taxonomies of Attacks, Tools, and Countermeasures, in Proceedings of the 17th International Conference on Parallel and Distributed Computing Systems, (September 2004).

**45.** Scott Lowe, Mastering Vmware v Sphere 5 - Oct 18, 2011).

**46.** Search Cloud Computing , (December 2007), available at (http://searchcloudcomputing.techtarget.com/definition/cloud-computing ).

**47.** SHEA, G.O , Controlling the Dependencyof User access Control Mechanisms on correctness of User Identification, The Computer Journal vol 31 N° 6 1988

**48.** Sheth, A.P., Gomadam, K., Lathem, J.: SA-REST: Semantically Interoperable and Easier-to-Use Services and Mashups. IEEE Internet Computing 11(6), 84–87 (2007).

**49.** Shiu-Kai Chin and Susan Beth Older, Access Control, Security, and Trust: A Logical Approach (Chapman & Hall/CRC Cryptography and Network Security Series, 2010.

**50.** Shneier, B. Applied Cryptography: Protocols, Algorithms, and Source Code in C, Second Edition, Paperback , 1996

**51.** Sims, K. (2009), 'IBM Blue Cloud Initiative Advances Enterprise Cloud Computing' – available at http://www-03.ibm.com/press/us/en/pressrelease/26642.wss .

**52.** Stefan Ried, (April 21, 2011), "Sizing the Cloud", (For Vender Strategy Professional Blog), available at (http://blogs.forrester.com/stefan_ried/11-04-21-sizing_the_cloud ).

**53.** Stephen R Smoot , Private Cloud Computing : Consolidation, Virtualization, and Service-Oriented Infrastructure - (Oct 29, 2011).

**54.** Talal H. Noor and Quan Z. Sheng, Trust as a Service: A Framework for Trust Management in Cloud Environments, School of Computer Science,The University of Adelaide, Adelaide SA 5005, Australia.

**55.** Tim Mather, Subra Kumaraswamy, and Shahed Latif - Cloud Security and Privacy – (September 2009).

**56.** Timothy Chou, Introduction to Cloud Computing(in about 1,000 words) (Dec 27, 2010).

**57.** Todorov, D. Mechanics of User Identification and Authentication: Fundamentals of Identity Management, Dob, HardCover, June 2007.

**58.** Tony Guidici and Tejaswi Redkar, Windows Azure Platform , (Dec 7 , 2011).

**59.** Tsudik, Gene, Message Authentication with One-Way Hash Function, Computer Communication Review, 1992,

**60.** Types of Cloud Computing, available at ( www.thecloudtutorial.com ).

**61.** Vic j.R Winkler, Securing the Cloud: Cloud Computer Security Techniques and Tactics - (Apr 29.2011).

**62.** William Stalling & Laxrie Brown 'Computer Security Principals and Practices' Pearson Edition 2008.

**63.** Yong Wang, Vinny Cahill, Elizabeth Gray, Colin Harris, and Lejian Liao, "Bayesian network based trust management," in Autonomic and Trusted Computing. Berlin / Heidelberg: Springer, 2006, pp. 246-257.

# SESSION

# BIOMETRICS AND FORENSICS

## Chair(s)

### TBA

# Android Forensics: a Physical Approach

**Lamine M. Aouad, Tahar M. Kechadi**
Centre of Cybercrime Investigation
University College Dublin - Ireland

**Abstract**— *There has been an exponential growth of Android systems in the last few years. However, the capability to perform efficient and fast forensic analyses on these devices is still limited, due to the lack of standardized processes along with the wide range of variants or versions in the operating system, the file system, the data storage, in addition to the manufacturers specific customizations. In this paper, we present a generalized method for physical acquisition and analysis of memory images of Android devices. It is known that the main advantage of acquiring physical memory images is a more complete capture of the data, including deleted items. In addition, physical acquisition methods can work with damaged devices and generally make fewer alterations to the original device while being acquired. Yaffs2 (Yet Another Flash File System) used in the majority of existing devices is still not fully supported by forensic commercial tools. We aim at covering this gap by presenting an easy end-to-end procedure for the acquisition of data partitions on a range of Android systems using yaffs2, as well as the mounting and analysis of these memory images on a Linux workstation.*

**Keywords:** Android, Memory imaging, Yaffs2.

## 1. Introduction

The number of mobile phone subscriptions worldwide reached more than 5.6 billion last year (Gartner research - 2011). The technology and functionality present on these phones is continually evolving. Smart phones are now becoming widely spread, and have certainly hugely contributed to the phenomenal increase in mobile phone subscriptions (700% in the last ten years!). Android is becoming the most common platform for these phones, with 43% penetration in the US market (Q3 2011) [7].

The amount of information stored on these devices has increased dramatically. These include emails, SMSs, browser history, bookmarks, messages, chat, network passwords, personal notes, contacts, call logs, geolocation information, and much more. There is also a wealth of information in third-party applications. These are all potentially relevant in a forensic investigation. However, the growing number and variety of devices and customized systems and interfaces make it difficult to develop a single process or tool for effective data extraction and analysis.

Low-level analysis of complete memory images can offer a solution to this. The literature has also shown that the memory imaging approach is the “holy grail” of any forensic acquisition. Performing a bit-by-bit copy of the original media was indeed always ranked the highest in terms of effectiveness and accuracy, such as in [2], [3]. The work presented in this paper focuses on a generalized method for physical acquisition and memory imaging and analysis on Android devices. We specifically target the support of the native file system used by Android, namely yaffs.

The next section will present a brief state-of-the-art and overview of underlying systems. Section 3 will then present the proposed method, evaluation, and a discussion. Section 4 presents then concluding remarks.

## 2. Background

In recent years, there has been an increasing interest in mobile devices forensic, and many studies and surveys (on current methods and existing tools) have been presented, including [2] [3] [5] [6], among others. Given the huge variety of these systems and devices (the Android OS for instance is compliant with 300+ different smartphone models), it should come as no surprise that it is quite a large list of specifications. Indeed, no standardized or generalized methods exist, either software or hardware. An interesting fact to mention here is that most of the existing tools are commercial, with unspecified implementation, and no or little documentation of their architecture or the way they do either logical or physical acquisitions. In this work, we aim at setting up a general physical acquisition method and document the fundamentals of analyzing it. In the following, we will present the underlying systems and technologies.

### 2.1 Android OS

Android is an open source mobile device OS developed by Google, based on the Linux 2.6 kernel. The Linux kernel was chosen due to its proven driver model, existing drivers, memory and process management, networking support along with other core operating system services [8]. It has also developed its own Java runtime engine, optimized for the limited resources available on a mobile platform, called the “Dalvik Virtual Machine”. Lastly, the application framework was created in order to provide the system libraries in a concise manner to the end-user applications [9].

### 2.2 Yaffs filesystem

Android uses the yaffs flash file system, the first NAND optimized Linux flash file system. For mobile devices, hard

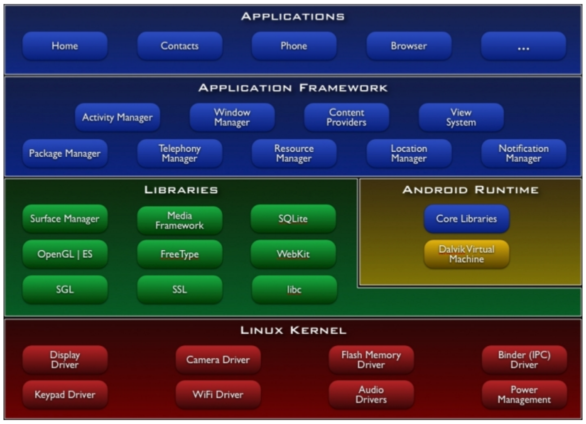


Fig. 1: Android system architecture

disks are too large in size, too fragile and consume too much power to be useful. In contrast, flash memory provides fast read access time and better kinetic shock resistance than hard disks. There are fundamentally two different types of flash memory: NOR and NAND. NOR is low density, offers slow writes and fast reads. NAND is low cost, high density and offers fast writes and slow reads. Embedded systems are increasingly using NAND flash for storage and NOR for code and execution [10].

Yaffs was developed by Toby Churchill Ltd (TCL) as a reliable filing system with fast boot time for their flash memory devices . The authors initially tried to modify existing flash file systems such as JFFS (used mainly for NOR) to add NAND support, but it turned out that the slow boot time and RAM consumption of existing flash file systems was unacceptable. Furthermore, there are too many fundamental differences between NOR and NAND to make performance optimal. For instance, since erasing NOR is much longer than for NAND, garbage collection methodologies for NOR are not suitable for NAND. This led to the development of a different flash file system especially for NAND according to its features and limitations to optimize performance and ensure robustness. Upon completion yaffs performed better than existing flash file systems and can still be used with NOR flash even though it was specifically designed for NAND. The description of yaffs is given in figure 2.

### 2.3 Memory Technology Device

Linux only understands character and block devices, such as keyboards and disk drives. With Linux on flash, however, a flash transition layer provides the system with device functionality. A Memory Technology Device (MTD) is needed to provide an interface between the Linux OS and the physical flash device because flash memory devices are not seen as character or block devices. The MTD system is simply "an abstraction layer for raw flash devices" that allows software to utilize a single interface to access a variety of flash technologies. For most Android devices, the MTD subsystem addressed NAND flash in blocks that are divided into 64 chunks with each chunk containing 2048 bytes (so blocks are 128K) plus a 64-byte out-of-band/spare area (OOB) where various tags and metadata are stored, as we will see below.

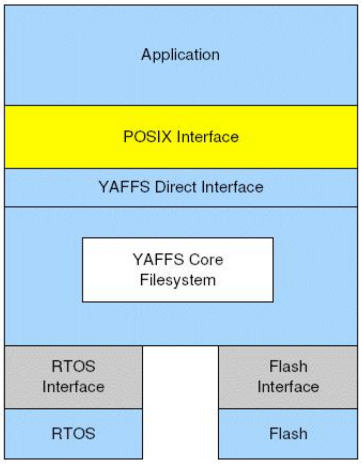


Fig. 2: Yaffs embedded structure

## 3. Process overview

The main idea here is to provide the user with a generalized method that can be carried out without the need of any specific forensic tool. We present the setting up followed by the overall process and discussion.

### 3.1 Setup

The presented method runs under Linux. The Android SDK tools are to be installed, including the Android Debug Bridge [11]. As yaffs2 is not supported by default in Linux, we had to incorporate it. Yaffs2 has been downloaded and compiled to enable kernel support. Also, the `mtd-utils` package needs to be installed. Lastly, MTD is cross-compiled to be used on the devices.

### 3.2 The method

We tested this method on a NexusOne with Android 2.1 and kernel version 2.6.29. The phone has to be rooted. In order to acquire access to the root directory, Universal Serial Bus (USB) debugging will have to be enabled on the phone. Our target partition is the $5^{th}$, and it is mounted on `/dev/block/mtdblock5`. Note that the process can be applied to any other partition, or a set of partitions.

### 3.3 Acquiring the memory image

We extract the memory contents in their entirety through the communication port. For MTD devices, `nanddump` can be used to collect NAND data independently of the higher-level filesystem deployed on the memory. For devices that do not employ MTD, other collection techniques can be employed. For instance, the `dd` utility can be used. It is also important to note that not all the data is necessarily stored in on-board memory [12]. We used an empty sdcard, with respect to the best practices in forensics. The linux shell command is the following:

```
#cd mtd-utils-arm
#adb push nanddump /sdcard
#adb push mtd_debug /sdcard
#adb shell [here we are on the phone]
#mount -o remount,rw /sdcard /sdcard
#chmod 755 /sdcard/nanddump
```

Now we have a cross-compiled copy of `nanddump` and `mtddebug`, executable on our device. We also take note of the version of yaffs that is running on our device (`cat /proc/yaffs`). As we know the mounting point of our target partition, we can collect some other important information about it (via `cat /proc/mtd`).

From this, we can see how yaffs2 and MTD organize the NAND flash structure. We can see that `totalBytesPerChunk` is equal to 2048, so now we know that for this version of yaffs, and for this device, the data page size (`Chunk` for yaffs) is 2048 bytes. The size of block (`erasesize`) is 00020000 in hexadecimal, i.e. 131072 (128 kilobytes). Usually, each block is followed by 64-byte out-of-band/spare area (OOB) where various tags and metadata are stored. We also know there are 64 chunks (2048 bytes for each) per block. This partition has 1570 blocks so the total size is: $1570 \times 131072$ bytes $= 205783040$ bytes. This is the same size that we can see in hexadecimal (running the $2^{nd}$ command mentioned above, i.e. `cat /proc/mtd`) at mtd5 (0c440000). To check that, we can use the mtd-debug command (`mtd_debug info /dev/mtd/mtd5`). The structure of our yaffs2 blocks is shown in figure 3.

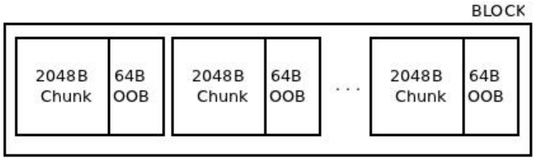


Fig. 3: Block structure

Now we are ready to use `nanddump` to make a dump of the whole userdata partition. There is a variety of options; we used `-o` as we need to dump also the OOB data to mount the image on the target machine, the `-f` option to specify the path of the file, and `-bb=padbad` to specify that we want to copy the badblock as well.

```
#cd /sdcard
#./nanddump -o -f /sdcard/userdataoobpadbad.nanddump
             /dev/mtd/mtd5 --bb=padbad
ECC failed: 0
ECC corrected: 0
Number of bad blocks: 1
Number of bbt blocks: 0
Block size 131072, page size 2048, OOB size 64
Dumping data starting at 0x00000000
and ending at 0x0c440000...
```

We also generate a dump without the OOB area as some of the techniques that we are going to use work better without OOB:

```
#./nanddump  -f /sdcard/userdatapadbad.nanddump
             /dev/mtd/mtd5 --bb=padbad
ECC failed: 0
ECC corrected: 0
Number of bad blocks: 1
Number of bbt blocks: 0
Block size 131072, page size 2048, OOB size 64
Dumping data starting at
0x00000000 and ending at 0x0c440000...
```

The resulted images are now on the sdcard:

```
#ls -l
----rwxr-x system   sdcard_rw
702360 2011-09-03 21:11 nanddump
----rwxr-x system   sdcard_rw
644505 2011-09-03 21:11 mtd_debug
----rwxr-x system   sdcard_rw
212213760 2011-11-22 18:38
userdataoobpadbad.nanddump
----rwxr-x system   sdcard_rw
205783040 2011-11-22 18:43
userdatapadbad.nanddump
```

It might appear that the size of the first dump is incorrect, however, if we consider that we took also the OOB area into consideration, now the size of each structure is:

```
Size of a chunk = (2048 bytes + 64 bytes) = 2112
Size of a block = (Size of a chunk) * 64
= 135168 bytes (132 kilobytes)
Total Size of partition
= number of block * size of block
= 1570 * 135168 bytes = 212213760 bytes
```

As we can see, both the `nanddump` with the OOB area and the `nanddump` without the OOB area correspond to the expected sizes. Finally we have to export this dump from the sdcard to the Ubuntu workstation using `adb pull`.

## 3.4 Mounting the image

We are now going to mount the userdata partition which includes the OOB data. We will use a simulated NAND device. The yaffs2 module is responsible for all aspects of the file system, while the MTD driver manages the writing of the data to the NAND flash. We have then to take into consideration an additional layer of complexity. This is a tricky step because the version of the yaffs2 module available on your system might not be compatible with the one present on the phone. The same applies to the MTD driver. In both cases you will see only the folder `LOST+FOUND`.

First we load the driver and the yaffs2 module:

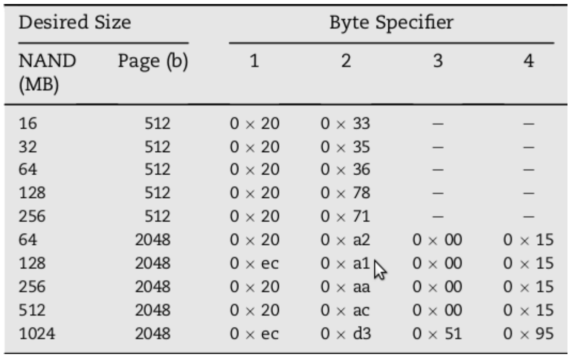
```
#sudo modprobe mtdchar
#sudo modprobe mtd
#sudo modprobe mtdblock
#insmod
#$DIR/$YOURKERNELNAME/fs/yaffs2/yaffs.ko
```

We can now build a simulated NAND device of 1GB:

```
#sudo modprobe nandsim
```

```
first_id_byte=0xec second_id_byte=0xd3
third_id_byte=0x51 fourth_id_byte=0x95
```

Figure 4 shows other parameters for different sizes of the simulated device. We can have a look inside this simulated partition with a simple `nanddump` command. Using a normal hexadecimal examiner can present some problems with the OOB area.



| Desired Size | | Byte Specifier | | | |
|---|---|---|---|---|---|
| NAND (MB) | Page (b) | 1 | 2 | 3 | 4 |
| 16 | 512 | 0 × 20 | 0 × 33 | – | – |
| 32 | 512 | 0 × 20 | 0 × 35 | – | – |
| 64 | 512 | 0 × 20 | 0 × 36 | – | – |
| 128 | 512 | 0 × 20 | 0 × 78 | – | – |
| 256 | 512 | 0 × 20 | 0 × 71 | – | – |
| 64 | 2048 | 0 × 20 | 0 × a2 | 0 × 00 | 0 × 15 |
| 128 | 2048 | 0 × ec | 0 × a1 | 0 × 00 | 0 × 15 |
| 256 | 2048 | 0 × 20 | 0 × aa | 0 × 00 | 0 × 15 |
| 512 | 2048 | 0 × 20 | 0 × ac | 0 × 00 | 0 × 15 |
| 1024 | 2048 | 0 × ec | 0 × d3 | 0 × 51 | 0 × 95 |

Fig. 4: Nandsim parameter

```
#sudo nanddump -a /dev/mtd0 | xxd | less
0000000: ffff ffff ffff ffff ...........
0000010: ffff ffff ffff ffff ...........
0000020: ffff ffff ffff ffff ...........
0000030: ffff ffff ffff ffff ...........
0000040: ffff ffff ffff ffff ...........
0000050: ffff ffff ffff ffff ...........
```

We see all groups of `ffff`, because when a block is erased in a NAND device, the entire block is overwritten with `0xFF`. The erase operation is the only mechanism by which a 0 can be changed to a 1 in NAND flash [4]. Optionally we can enable the debug mode of yaffs [1]. The next step is to use `nandwrite` to copy both the data and OOB on the simulated NAND flash:

```
#sudo nandwrite -a -o /dev/mtd0
      ~/userdataoobpadbad.nanddump

<snip>

Writing data to block 1565 at offset 0xc3a0000
Writing data to block 1566 at offset 0xc3c0000
Writing data to block 1567 at offset 0xc3e0000
Writing data to block 1568 at offset 0xc400000
Writing data to block 1569 at offset 0xc420000
```

The 1570 yaffs2 blocks are now copied on dev0. We can initially do an hex analysis of the mounted partition, and we can see that the content of the device is completely changed (see below, `nanddump -c /dev/mtd0`). Finally we can mount the image:

```
#sudo mount -t yaffs2 /dev/mtdblock0 /mnt/mtd
```

and we will see then the full `/data` filesystem of our device accessible on the Linux workstation. Simple analysis techniques can then be used. The next two sections present two examples with carving and strings analysis.

## 3.5 Yaffs2 file carving

File carving is the process of reassembling computer files from fragments in the absence of filesystem metadata. The carving process makes use of knowledge of common file structures, information contained in files, and heuristics regarding how filesystems fragment data. Fusing these three sources of information, a file carving system infers which fragments belong together [14]. There are many commercial tools to carve data file, but not many of them support yaffs2.

We used an open source tool called Scapel [15]. Scalpel is a fast file carver that reads a database of header and footer definitions and extracts matching files or data fragments from a set of image files or raw device files. Scalpel is filesystem-independent and would carve files from FATx, NTFS, ext2/3, HFS+, or raw partitions with the help of a configuration file. It is useful for both digital forensic investigations and file recovery. The parameters set up is very important. It will define the extension, the maximum size to carve, the header definition and the footer. The final result of the analysis heavily depends on these parameters. Here we used the image without the OOB area.

```
#scalpel -o ~/scalpel ~/userdatapadbad.nanddump
Scalpel is done, files carved
= 7998, elapsed = 93 seconds.
```

Note that the tested phone has been used very little, and we still recovered an extraordinary amount of information, 7998 files. All files are categorized into folders with the name of their extension. They can now be analyzed using established traditional forensic techniques usually applied to other filesystems.

## 3.6 yaffs2 strings analysis

For each file, strings prints the printable character sequences that are at least 4 characters long (or the number given with the options below), and are followed by an unprintable character. We can use this command to extract some data from the userdata image. Let us consider two examples. In the first one we would like to know all the names of wireless networks that the tested phone was connected to (note that the respective passwords are stored without encryption):

```
#strings --all --radix=x
         userdatapadbad.nanddump | grep ssid | less

  445049  ssid="eduroam"
  445119  ssid="Fitzsimons Hotel Bar"
  44515f  ssid="MPC - Fitzsimons 2nd Floor"
  4451ab  ssid="Paddy Wagon_WiFi"
  4451ed  ssid="Kinlay House"
  44522b  ssid="Wireless-Galway-1"
  44526e  ssid="www.izone.ie"
  4452ac  ssid="Stephen's Green Free WIFI"
  4452f7  ssid="eircom"
  44532f  ssid="bitbuzz"
  445368  ssid="Harcourt Hotel FREE WiFi"
  4453b2  ssid="opennet"
  4453eb  ssid="WaveLAN Network"
```

```
#nanddump -c /dev/mtd0
       | grep -v "00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00"
       | grep -v "ff ff ff ff ff ff ff ff ff ff ff ff ff ff ff ff"
       | less

<snip>

0x000197b0: 36 34 33 30 31 65 31 37 30 64 33 30 33 39 33 31  |64301e170d303931|
0x000197c0: 33 31 33 32 33 36 33 30 33 30 33 30 33 37 33 30  |3132363030303730|
0x000197d0: 33 32 35 61 31 37 30 64 33 33 33 37 33 30 33 34  |325a170d33373034|
0x000197e0: 33 31 33 33 33 30 33 30 33 30 33 37 33 30 33 32  |3133303030373032|
0x000197f0: 35 61 33 30 37 34 33 31 30 62 33 30 30 39 30 36  |5a3074310b300906|
OOB Data: ff ff a2 9e 07 30 1e a3 ce b8 0d 02 2c ed e6 a9    |.....0......,...|
OOB Data: 35 27 97 c2 7a 99 aa b4 bd 09 7c a7 b0 d1 7b d2    |5’..z.....|...{.|
OOB Data: 3c 34 00 00 0c 03 00 00 66 69 9b 95 55 9b 0f ff    |<4......fi..U...|
OOB Data: cf f3 03 cf 96 99 9b 96 55 a7 f3 ff cf aa 55 9b    |........U.....U.|

<snip>
```

The `all` option means that we want to analyze the entire file. The `radix=x` means that the offset will be printed in hex format. With the grep command we can do a search for specific keywords.

The second example shows all the places that have been searched on Google maps:

```
#strings --all --radix=x userdatapadbad.nanddump
| grep maps.google.com | less
7381305  http://maps.google.com/?q=harcourt street
8692f0f  http://maps.google.com/?q=rome
86930d3  http://maps.google.com/?q=dawson street
885e1ed  http://maps.google.com/?q=barcellona
886c083  http://maps.google.com/?q=ginevra
```

These kind of techniques represent simple, yet powerful, framework for fast and accurate memory image analysis, retrieving targeted information in few minutes. Also, these, and other basic data extraction, can be scripted-out fairly easily depending on the users needs.

### 3.7 Discussion

We proposed a method for a fast imaging and analysis of a data partition of an Android device based on the yaffs2 filesystem. With this generalized procedure, we demonstrate that a wealth of information can be recovered in a forensically sound manner in few minutes without the need of any specific tool or system.

This work can be extended in many ways, including in hex analysis that permits to recover raw and deleted data from the phone, fine-tuning configuration parameters for files analysis, and building a set of significant searching words (ssid, passwd, etc.) to be used in the analysis in a more automated way. The whole procedure was also scripted-out so it can be used without having a deep knowledge of the device or underlying technologies and commands.

## 4. Conclusion

In this paper, we present a physical acquisition method for Android data partitions based on yaffs2. The method aims to be generic and easily applicable. Also, yaffs2 is not currently well supported, including by well known forensic tools. This work introduces then a fast, yet powerful, imaging and analysis technique, along with documenting its fundamentals.

## References

[1] yaffs official website. Available at http://www.yaffs.net

[2] Hoog A. and Gaffaney K. *iPhone forensics.* Via Forensics White paper, 2009.

[3] Hoog A. *Android forensics.* Mobile Forensics World. 2009.

[4] Hoog A. *Android Forensics - Investigation, Analysis and Mobile Security for Google Android.* Elsevier, 2011.

[5] Ayers R., and Jansen W., and Moenner L, and Delaitre A. *Cell Phone Forensic Tools: An Overview and Analysis update.* NIST Technical Report. 2007.

[6] Jansen W. and Ayers R. *Guidelines on Cell Phone Forensics. Recommendations of the National Institute of Standards and Technology.* NIST Technical Report. 2009.

[7] Nielsen Research. *Android Grew Its Smartphone Marketshare; iPhone Stayed Flat.* 2011.

[8] Androidology I: Architecture. Available at http://www.android.com/

[9] Frank Maker and Yu-Hsuan Chan *A Survey on Android vs. Linux.* Department of Electrical and Computer Engineering, University of California, Davis. 2011.

[10] *NAND vs. NOR Flash Memory: Technology Overview*. Available at http://www.toshiba.com/taec/components/Generic/Memory_Resources/NANDvsNOR.pdf

[11] *Android Debug Bridge.* Available at http://developer.android.com/guide/developing/tools/adb.html

[12] Timothy Vidas, Chengye Zhang. *Toward a general collection methodology for Android devices.* In Proceedings of the 11th Digital Forensics Research Workshop (DFRWS 2011). New Orleans, LA. August 2011.

[13] *NANDSIM options.* Available at http://www.linux-mtd.infradead.org/

[14] Christiaan Beek. *Introduction to File carving.* White paper. McAfee. 2011.

[15] *Scalpel tool.* Available at http://www.digitalforensicssolutions.com/Scalpel/

# Facial Image and Voice Data based One-Time Password(FV-OTP) Mechanism for e-Financial Authentication System on a Smart Phone

You Joung Ham , Won-Bin Choi, Hyung-Woo Lee
School of Computer Engineering, Hanshin Univ., 411, Yangsan-dong, Osan,
Gyyunggi Province, 447-791, Rep. of Korea.
e-mail: you86400@hanmail.net, bindon@hanmir.com, hwlee@hs.ac.kr

*Abstract*— **The number of Internet banking transactions and e-financial services based on the publicly authenticated certifications of smart phone have increased rapidly. We can use several services such as wireless Internet access and Web surfing for information retrieval on iOS based iPhone device. Additionally, we can take advantage of using electronic banking services without the constraints of time and location on a smart phone. However, personal and financial information stored in smart phone can also be disclosed to the outside malicious attacker insecurely. Therefore, it is necessary to authenticate the smart phone user's real identity correctly on using wireless e-financial services. In this study, we proposed a new one-time password mechanism (FV-OTP) by using user-related multimedia such as image and voice data captured from the user's multimedia capture module embedded on a smart phone device. A proposed one-time password value was created securely both from a facial image obtained through the camera module and a voice information input through the microphone device mounted on smart phone for providing secure authentication services. Using proposed FV-OTP mechanism, we can construct an enhanced authentication system by providing secure e-financial transactions on a smart phone.**

***Keywords- e-Financial service, Security, Authentication, One-time password, Facial image, Voice, Smart Phone***

## I. INTRODUCTION[1]

There are ongoing studies that attempt to incorporate voice information in a mobile environment like a smart phone[1,2,3]. However, when smart phone is used for electronic banking services, the vulnerability of such banking services is revealed due to the possible leakage of personal information and malicious attacks on user authentication[4]. For these reasons, the existing e-banking services related to internet banking are recommending the use of multimedia based one-time passwords (OTP) and promoting service standardization in a way to strengthen user authentication [5,6]. Actually, since the e-banking services can be exploited through illegally authorized terminal devices, a stricter user authentication process must be provided and applied.

The pre-existing OTP approach used in smart phone based e-banking services does not provide any procedure of authenticating an OTP token as that of the actual smart phone owner and is subject to man-in-the-middle (MITM) attacks on OTP information. That's why there should be a technological approach that compensates for the problem of vulnerability [6].

In an attempt to provide a more strengthened authentication process to smart phone users, the present study comes up with a method by which a final OTP value is generated from the user's voice data in an OTP-based authentication process and used for smart phone user authentication. The proposed approach will be able to help prevent illegal users' possible bypass attacks on smart phone-based services and provide a multifactor authentication feature available on smart phones.

## II. OTP AND MULTIMEDIA BASED AUTHENTICATION

### A. *Vulnerability of User Authentication on Smart Phone*

Recently, the frequent occurrence of smart phone-related security incidents has aroused widespread social concern, especially over the leakage of personal information stored in terminal devices (handsets). Smart phone users may encounter problems caused by malicious codes, such as remote control, operational disturbance, and billing inducement related to e-banking services. Moreover, when software installed in a smart phone is running in a multitasking fashion, there is a potential for personal information to leak out and accordingly it is required to take technical action against the problem.

Security vulnerabilities regarding smart phone handsets include malicious-code infection, data leakage, phone misuse/abuse and phone loss/theft; and as potential security vulnerabilities in public or transport networks, data forgery/falsification and leakage can be caused due to attacks or hacks on AP during Wi-Fi based communications. In addition, while using the internet on a smart phone, the user is exposed to the risk of DoS/DDoS attacks by malicious codes, authentication bypassing, or account takeover[13].

Therefore, solutions to those security problems should involve an approach of strengthening smart phone user authentication. More strengthened authentication mechanisms should be used with focus on access to internal information in smart phones.

[1] Corresponding Author: Hyung-Woo Lee is with the School of Computer Engineering, Hanshin Univ., 411, Yangsan-dong, Osan, Gyyunggi Province, 447-791, Rep. of Korea. (e-mail: hwlee@hs.ac.kr)

### B. *Security of One-Time Password Mechanism*

The existing one-time password (OTP) technologies [5,6] are to generate one-time passwords. The OTP generation technologies are divided into two main approaches: synchronization and asynchronous method. The asynchronous approach for OTP generation works using a challenge-response mechanism and implements an authentication process based on the result of response to a challenge value. For this approach, there is no need for synchronization with a server, but it requires user input and sometimes causes a network overload. The OTP synchronization approach works in a mechanism of time synchronization, event synchronization or event-time synchronization and requires an accurate synchronization process between an OTP token and authentication server. The approach is designed to generate a password at specific time intervals based on time information synchronized between the server and OTP device[5,6]. However, this synchronization approach is vulnerable to MITM attacks and restricted by cool down time[13].

### C. *Facial Image and Voice Data based Authentication*

Recently there are ongoing studies that attempt to incorporate voice information in a one-time password generation process [7]. Therefore it is an multi-factor authentication approach to authenticate which requires the presentation of “two or more” of the three authentication factors[8]. Recently, image based two-factor authentication mechanism was proposed for providing confidence on mobile application[9].

In order to strengthen security and improve user authentication vulnerability in smart phone-based e-banking services which are becoming a worldwide issue, this study aims to propose a FV-OTP mechanism that grafts a person's unique facial image and voice information onto OTP technologies used in the pre-existing e-banking services. The proposed architecture is designed to make up for the vulnerabilities of smart phone user authentication and strengthen security, and in the architecture, a one-time password is generated with a user's facial image and voice information added to an existing authentication system which only replies on a user ID and password. As a result, this mechanism is designed and implemented to strengthen smart phone user authentication to a multi-factor authentication system.

## III. PROPOSED FV-OTP BASED AUTHENTICATION

### A. *Architecture of FV-OTP based User Authentication*

In clearing a user authentication process, e-banking service users read a challenge value received from a server in their own voice using a microphone in their smart phone and have an OTP generated using captured voice information. Fig. 1 shows detailed procedures of generating an OTP from both a user’s facial image and voice information for FV-OTP authentication.

As a previous step, smart phone users register their own ID and password in advance to the server, in the same way as when using the existing e-banking services. Now on the assumption that UAC's ID and password information is stored in the server, the approach designed and implemented in this study does not transmit such pre-registered password information via a network in an actual user authentication process but performs user authentication securely through a FV-OTP mechanism by using the information only in the client terminal and server.

The server generates a random challenge value and transmits the value to user device (UAC). The UAC inputs facial image in Step 1 on the challenge value via a camera module mounted on the smart phone and sends the information to the server, which then performs an authentication process for the client terminal based on the information received from the and transmits the result to the UAC using a random value generated in the server. And then the user also input his/her own voice data based on the challenge value received from the server.

Lastly, after validating the information received from the server as a mutual authentication procedure, the UAC can generate a FV-OTP value using the voice information input and challenge value received from the server and returns this FV-OTP value to the server for verification finally. The server performs the final authentication process for UAC by verifying the OTP value.

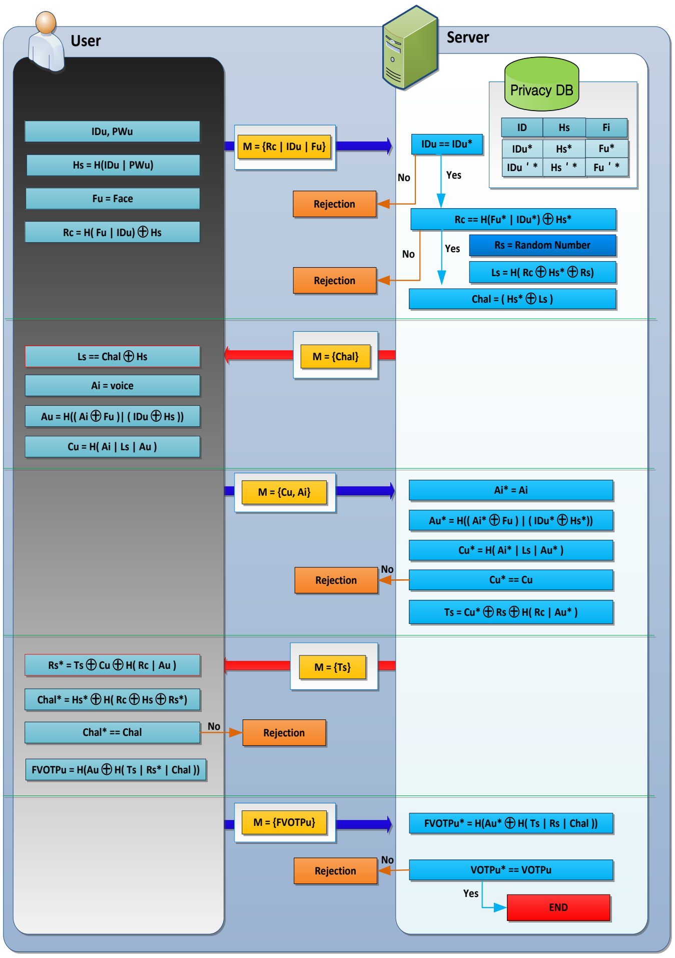


Figure 1. The Entire Architecture of Proposed FV-OTP based Authentication Mechanism

### B. FV-OTP Generation Mechanism

Detailed steps for FV-OTP-based user authentication are as follows:

Step 1: User can capture his/her own facial image (*Fu*) using the camera module on smart phone. The UAC transmits the user's ID (*IDu*) with facial image (*Fu*) and a randomly generated random number (*Rc*) to the server. In this case, the *Rc* value will be generated after XOR calculation using both the user's uniquely generated hash data *Hs* and *H( Fu | IDu )*. And user sends *Message*, as an equation (1) below, to the server. And then the server identifies its database information to determine whether the user is registered. If the user is not registered, the FV-OTP generation process is terminated. If the *IDu* received from UAC is about a registered user ID, the server generates a random number (*Rs*) and then a challenge value (*Chal*) with *Rc* received from UAC for performing facial image based sender authentication process after image comparison process stored on the server side DB previously, as an equation (2) below, and transmits the challenge value to UAC.

$$
\begin{aligned}
&IDu == IDu^*,\ PWu == PWu^*,\ Fu == Fu^* \\
&Hs = H(\ IDu \mid PWu\ ) \\
&Rc = H(\ Fu \mid IDu\ ) \oplus Hs \\
&Message = \{\ Rc \mid IDu \mid Fu\} \qquad (1)
\end{aligned}
$$

$$
\begin{aligned}
&Rc^* = H(\ Fu^* \mid IDu^*\ ) \oplus Hs^*,\ Rc == Rc^*,\ Rs == Rs^* \\
&Ls = H(\ Rc \oplus Hs^* \oplus Rs\ ) \ \ and \ \ Chal = (\ Hs^* \oplus Ls) \quad (2)
\end{aligned}
$$

The *Chal* value is given as a combination of numbers or letters and displayed on the client's smart phone terminal. Random numbers *Rc* and *Rs* are selected by the client and server, and both the values change uniquely whenever FV-OTP-based user authentication is performed, as a follow Fig. 2.

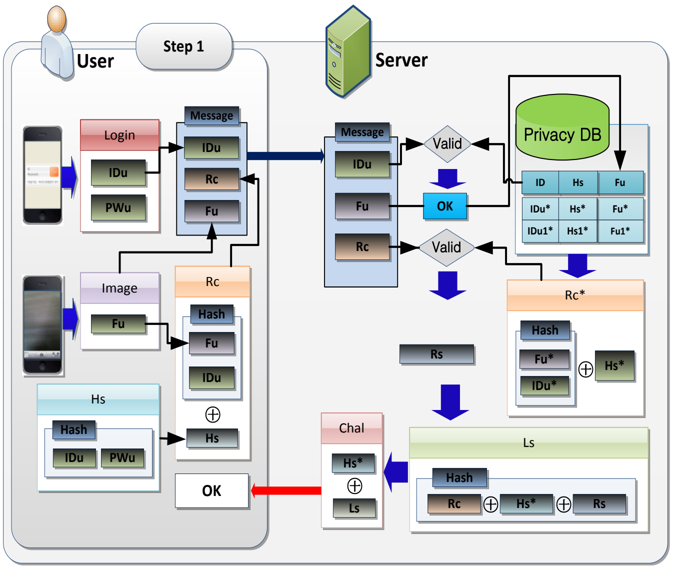


Figure 2. FV-OTP Step 1 for Image based User Authentication

Step 2: Next step is to authenticate the user using voice data. The UAC reads the *Chal* value information received from the server in a user's voice via a microphone in the smart phone handset. In first, the client can get *Ls** value after calculating XOR operation on *Chal* using *Hs* value because *Ls* is same as a *Chal ⊕ Hs*. In this study, the user is allowed to enter voice data on the *Chal* value during five seconds so that an *Ai* value is generated. As shown in equation (3), a hash value, *Au= H(( Ai ⊕ Fu ) | ( IDu ⊕ Hs ))*, is generated using both the voice information (*Ai*) input by the user and previously captured facial image (*Fu*), and the UAC user's ID information (*IDu*) and hashed value *Hs = H (IDu | PWu)* generated by the client previously; and the *Ls** value recalculated on client side using *Ls* value received from the server is used to generate a response value according to *Cu= H(Ai | Ls* | Au )* as given in equation (4). As a response, the *Cu* will be sent to the server together with the user-input voice data value (*Ai*) as a concatenated *Message = { Cu | Ai }* as a follow Fig. 3.

$$
\begin{aligned}
&Ai = voice\ data \\
&Ls^* == Chal \oplus Hs \\
&Au = H((\ Ai \oplus Fu\ ) \mid (\ IDu \oplus Hs\ )) \qquad (3) \\
&Cu = H(Ai \mid Ls^* \mid Au\ ) \qquad (4) \\
&Message = \{\ Cu \mid Ai\ \}
\end{aligned}
$$

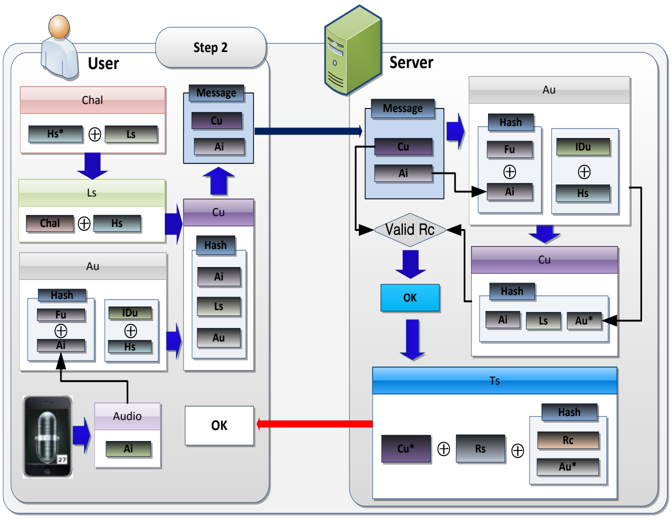


Figure 3. FV-OTP Step 2 for Audio based User Authentication

The server performs a verification process and a integrity check procedure for *Cu* and *Ai* values received from UAC. As shown in equation (5), a value for *Au* = H(( Ai* | Fu* ⊕ IDu* ⊕ Hs*))* is generated by a server using the previously stored user's ID (*IDu**) and also a hashed identity values *Hs** stored in the database, the facial image value (*Fu**) stored on a server DB, and the voice data value (*Ai*); And as given in equation (6), a value for *Cu* = H( Ai* | Ls | Au* )* is generated from a server-produced *Ls* value and compared with the *Cu* value received from UAC for providing a mutual authentication between client and server. If the results of mutual authentication are consistent, then a process for FV-OTP generation is conducted. As can be seen from equation (7) $Ts= Cu^* \oplus Rs \oplus H( Rc \mid Au^* )$, the server generates a one-time token $Ts$ for FV-OTP procedure using a server side random number $Rs$ and sends the token $Ts$ to UAC. Expressions for this step are as follows:

$$Ai^* = Ai$$
$$Au^* = H(( Ai^* \mid Fu^* \oplus IDu^* \oplus Hs^*)) \quad (5)$$
$$Cu^* = H( Ai^* \mid Ls \mid Au^* ) \quad (6)$$
$$Cu^* == Cu$$
$$Ts = Cu^* \oplus Rs \oplus H( Rc \mid Au^* ) \quad (7)$$

Therefore, we can perform both a mutual authentication procedure and a one-time token generation for which voice information is used.

Step 3: As shown in equation (8), the UAC performs the process of getting a server-selected random value $Rs^* = Ts \oplus Cu \oplus H( Rc \mid Au )$ from the $Ts$ value received from the server. In this step, we can proof the equation (8) as a $Rs^*= Cu^* \oplus Rs \oplus H( Rc \mid Au^* ) \oplus Cu \oplus H( Rc \mid Au ) = Rs$. In terms of equation (9), a server re-authentication process is conducted to determine whether $Hs^* \oplus H (Rc \oplus Hs \oplus Rs^*) = Hs^* \oplus Ls$ produces the same value as $Chal$ received from the server at Step 1 and Step 2. Further, as given in equation (10), it is possible to generate a one-time password value ($FVOTPu$) which will be sent to the server as a follow Fig. 4.

$$Rs^* = Ts \oplus Cu \oplus H( Rc \mid Au ) \quad (8)$$
$$Chal^* = Hs^* \oplus H (Rc \oplus Hs \oplus Rs^*) = Hs^* \oplus Ls \quad (9)$$
$$FVOTPu = H( Au \oplus H( Ts \mid Rs^* \mid Chal )) \quad (10)$$

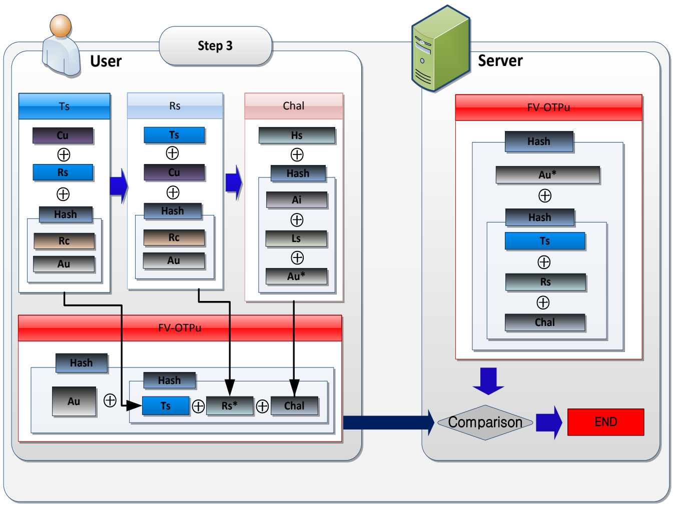


Figure 4. FV-OTP Step 3 for OTP based User Authentication

Step 4: In terms of equation (11) the server identifies and verifies the facial image and voice data based one-time password ($FVOTPu$) received from UAC and completes the process for smart phone user authentication. The $FVOTPu$ value delivered from UAC is one-time password information that is generated using the user's voice data ($Au$) made via a microphone based on an audio input $Ai$, and $Chal$ value sent from the server, token $Ts$ and $Rs^*$ values calculated by client in the mutual authentication process securely.

$$FVOTPu^* = H( Au^* \oplus H( Ts \mid Rs \mid Chal )) \quad (11)$$
$$FVOTPu^* == FVOTPu$$

In consequence, the proposed approach can be recommended as a solution to user authentication problems which occur in the existing smart phone environment, and since the approach uses a user's voice data, it provides a secure way to prevent MITM and replay attacks. The approach allows the authentication process to be performed only with a cryptographically secure hash function, a random number generator and a XOR function.

## IV. Implementation Results

### A. Implementation Results

iPhone iOS4.2 was used to test the implementation results regarding the proposed mechanism. An Apple-provided Xcode 3.2.5 development environment was employed, and a MySQL-based OTP server was implemented. As shown in the figure above, the client has a *Chal* value generated from the server using a microphone controller module in the iOS-based smart phone, and the user is allowed to read the numeric information of the value in his/her own voice for five seconds via a microphone. The proposed mechanism also allows the numeric information to be captured by the smart phone terminal, and a mutual authentication process takes place between the client and server. Once a one-time token value generated from voice information is sent back to the smart phone, the server works to generate the final one-time password from voice information. Fig. 5 shows start-up display pages with an OTP mechanism implemented.

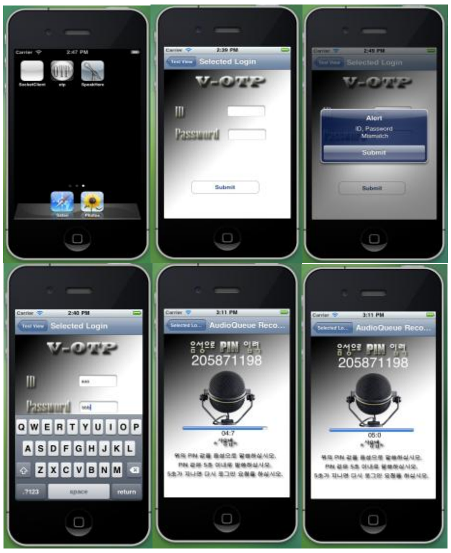


Figure 5. Implementation of Face &Voice Based OTP Authentication

The user is asked to enter his/her ID and password information registered with the server like in the existing authentication process. The user's password information is used only in his/her smart phone handset without being transmitted to the server via a network. If the user ID input is sent to the server with a random value (*Rc*) generated from the user's facial image in the client terminal, the server checks the input value first against the ID list stored in its database. If the ID is found unregistered, the subsequent process is not activated and an alert message appears on the smart phone display as shown in Fig. 5.

After the user identity is confirmed, the server generates a *Chal* value and sends it to the smart phone terminal. If the start-up screen appears, the user will see the challenge value of '205871198' received from the server and can read its numeric information in his/her own voice using a built-in microphone device. In this study, users were given a 9-digit number and asked to enter it as voice information within a time length of five seconds. With information input done, the numeric expressions presented above were used to generate a *Cu* value, which was then transmitted to the server.

The random value (*Rc*) of 16807 is generated in terms of a smarphone user's ID "aaa" and transmitted to the server, and the *Chal* value of 205871198 is received as generated and transmitted by the server. The client generates *Au* and *Cu* values by reading the user's voice information input from the buffer and sends the values to the server.

Subsequently, the client module extracts an *Rs* value using a one-time token value (*Ts*) generated and sent from the server and has the same *Chal* value obtained through the verification process above. In the end, a OTP value is generated using voice information and transmitted to the server. A user ID and random value are received from the client and the server generates the random value of 879160508. Again in the server, the PIN value of 205871198 corresponding to a *Chal* value is generated using the random value, and the server transmits the PIN value to the client terminal. The server performs a process of validating and verifying the *Cu* value sent from the client and generates and returns a *Ts* value to a smart phone. Finally, with the OTP value from the client validated and verified, OTP-based authentication in a smart phone environment is completed.

## V. Security and Performance Analysis

### A. Security Analysis

An one-time password (FV-OTP) mechanism proposed in this study has a smart phone user's facial image and voice information captured via both using a camera and a microphone device in the handset. And proposed mechanism allows an *Au* value to be generated using a *Rs** value which is randomly selected upon request for ID information (*IDu*) or password information (*PWu*) or otherwise for authentication. The *Au* value contains not only the user's identity information, but also a randomly selected hash value (*Rc*) which varies whenever there is an authentication request from a user's facial image. In the Au information transmitted via a network, the user's secret does not transmitted to the server as we can use both a hash function and XOR function, and as a result the one way primitive of hash function makes it difficult to find the original password. In addition, for voice information (*Ai*) in *Au*, since only part of information entered by a user for five seconds is made available, there is a lower possibility of password leakage due to dictionary attacks. Therefore, even if an MITM attack is launched on e-banking services, any password, once generated, can't be used later again because a *Rc* value randomly generated by the client is contained in *Au* information transmitted to the client.

The voice information in the *Au* value is generated with a PIN value received again by taking as a *Chal* value the hash value generated using a server-generated *Rs* value for the *Rc* value sent from the client to the server, and therefore it is also available for authentication or non-repudiation of the client.

In order to strengthen the security of the OTP-generating process, the proposed mechanism has allowed the client to transmit the *Chal* value received from the server in an abbreviated form with *Au* and voice information *Ai* by computing a hash value, instead of sending the *Au* value directly to the server. As a result, the server has been allowed to validate message integrity and verify delivered messages.

We can imagine the case where the server spoofs the client and performs an authentication process instead of the client after producing information on its own like at steps 2 and 3 described above. In this case, however, the server should also generate voice information on a *Chal* value, instead of the client. In other words, it should generate as much voice information as possible for a five-second time length by spoofing the client.

Accordingly, the server can launch an attack of generating voice information itself or otherwise exploit voice information (*Ai_old*) the client sends from a previously performed transaction. In this instance, however, the server should also forge or produce a client random value (*Rc*) in addition to the *Au*_fake* information that it should generate on its own. The self-generated fake value (*Cu*_fake*) is contained in a one-time token the server transmits to the client for self-authentication or V-OTP generation. A *Chal* value contained in the fake value is assumed to contain a random value sent by the client in a form of $H(Hs^* \oplus Ls\ )$. This implies that the server is not likely to launch spoofing attacks on the steps 2 and 3 posing as the client.

The client extracts or computes an *Rs** value using the one-time token (*Ts*) transmitted from the server at steps 3 where a facial image and voice based one-time password is generated. In addition, the client is allowed to perform hashing with a client-selected random value (*Rc*) and compare the random value with the *Chal* value received from the server at step 2. All transactions are organically related to each other, and a process of mutual authentication is performed for security enhancement: The client serves to authenticate the server and vice versa.

A one way hash function, an XOR function, and random values with a cryptographically safe length are only allowed for the information that is transmitted between the client and the server the authentication process. This is proposed in view of computation performance on smart phone and as a way to effectively strengthen user authentication features with a minimum of resources. In order to strengthen the security of OTP approaches that are used for user authentication in smart phone-based e-banking services, the proposed authentication mechanism asks individual users to enter a server-transmitted PIN value in their own voice via a microphone device and performs smart phone user authentication using FV-OTP information through the procedures of mutual authentication and verification. The authentication process can be done via a network without exposing users' password information.

### B. *Performance Analysis*

As shown in Table 1, the mechanism proposed in this study and the pre-existing ones were compared in terms of security and authentication performance. The present study attempted an analysis of computational complexity compared with the existing studies, in order to evaluate the performance of the proposed mechanism based on the number of unidirectional hash functions (*Th*) used at each step.

[Table 1] A Comparative Analysis of Authentication Complexity and Functionality

| Process/Step | Proposed Mechanism | Wang-Li [10] | Yoon-Yoo [11] | Khan et al. [12] |
|---|---|---|---|---|
| Registration | *4Th* | *3Th* | *3Th* | *2Th* |
| Login | *3Th* | *2Th* | *3Th* | *2Th* |
| Authentication | *6Th* | $5^{Th}$ | *4Th* | *2Th* |
| Mutual Authentication | O | O | O | O |
| Time Synchronization | X | X | O | O |
| FV-OTP Generation | O | X | X | X |
| Biometric Info. | Face & Voice | Finger-print | Finger-print | Finger-print |

The proposed mechanism has been found to have similar computation complexity to existing mechanisms. In the mechanism, human biometric data such as face and voice are used for user authentication and OTP generation on a smart phone. For biometric information available in the authentication process, the existing mechanisms [10,11,12] use fingerprint information, while the proposed mechanism allows a user's biometric information to be used for authentication. Actually, the new mechanism has been designed with considerations for the convenience of OTP users and the environment where smart phones are used. It is also applicable to fingerprint information like the existing ones.

## VI. CONCLUSION

When smart phones are used for e-banking or internet-based services, the smart phone environment needs to have more strengthened user authentication features. One reason is that such services can be wrongfully used by illegal mobile phones issued as a result of phone loss or identity theft.

In this study, the proposed mechanism asks smart phone users to send their facial image and voice information as a PIN value transmitted from the server, using a camera and a microphone device in their handset. It enables the server to perform a verification process and then generate a one-time token based on a random value and user's multimedia information. As a result, this suggests that the mechanism can make or has made a significant improvement both in security and authentication performance. The study also came up with a way to strengthen user authentication in e-banking services using image and voice based OTP value generated from the client. When employed in smart phone-based e-banking or internet services, the proposed mechanism is expected to help step up the security of mobile wireless services.

## ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (Grant # 2012R1A1A2004573)

## REFERENCES

[1] Voice Authentication: Making Access a Figure of Speech, http://www.computerworld.com/s/article/86897/Making_access_a_figure_of_speech.

[2] "Voice verification - for mobile banking security?", http://www.finextra.com/community/fullblog.aspx?id=3949

[3] Voice PIN 2.0, http://www.voiceverified.com/products.htm

[4] Jacek Lach, "Using Mobile Devices for User Authentication", CN2010, CCIS 79, pp.263-268, 2010

[5] Agnitio, "One-Time Password (OTP) Management Secured with Voice Biometrics", Voice Biometrics White Paper, 2009. http://www.banking-businessreview.com/suppliers/agnitio_voice_biometrics_for_homeland_security/whitepapers/one_time_password_otp_management_secured_with_voice_biometrics

[6] Agnitio, "One-Time Password (OTP) Management Secured with Voice Biometrics", Voice Biometrics White Paper, 2009.

[7] Helena Rif'a-Pous, "A Secure Mobile-Based Authentication System for e-Banking", OTM 2009, Part II. LNCS 5871, pp.848-860, 2009

[8] Two-factor authentication, Wikipedia, http://en.wikipedia.org/wiki/Two-factor_authentication

[9] Confident Multifactor Authentication, Two-Factor Authentication Using Images, http://www.confidenttechnologies.com/products/mobile-phone-authentication-factor.

[10] De-Song Wang, Jian-Ping Li, "A new fingerprint -based remote user authentication scheme using mobile devices", International Conference on Apperceiving Computing and Intelligence Analysis, ICACIA 2009, pp.65-68, 2009.

[11] Yoon E.J., and Yoo K.Y., "A secure chaotic hash-based biometric remote user authentication scheme using mobile devices" , APWeb/WAIM 2007, Huang Shan, pp. 612-623, June 2007.

[12] Khan M.K., Zhang J.S., and Wang X.M., "Chaotic hash-based fingerprint biometric remote user authentication scheme on mobile devices" , Chaos, Solitons & Fractals, Vol. 35, pp. 519-524, 2008.

[13] Sik-Wan Cho, Hyung-Woo Lee, "Design and Implementation of Voice One-Time Password(V-OTP) based User Authentication Mechanism on Smart Phone," KIPS Journal, No.2, Vol.18-C, pp.79-88, 2011.

# Survey on password recovery methods for forensic purpose

**Sang Su Lee[1,2], Sung Kyong Un[1], and Soon-Ja Kim[2]**
[1]Cyber Security-Convergence Research Laboratory, ETRI, Daejeon, Korea
[2]School of Electrical Engineering and Computer Science, KNU, Daegu, Korea

**Abstract -** *In this article, we introduce the password recovery methods against encrypted files like Microsoft Office, Adobe PDF, and other things. Of course, the password recovery may be recognized as finding cryptographic key with a cipher text through cryptanalysis. However, the forensic password recovery process would be more complicate than traditional cryptanalysis in general. We will explain the basic process of the forensic password recovery, and also introduce various methods to achieve it. In addition, commercial products including S/W and H/W for the same purpose are compared, too. Finally, we summarize the recent problems it is facing and suggest a brief description on countermeasures against the problems..*

**Keywords:** computer forensics, password recovery, password cracking, FPGA, GPU, CELL

# 1 Introduction

Digital forensics (sometimes known as digital forensic science) is a branch of forensic science encompassing the recovery and investigation of material found in digital devices, often in relation to computer crime[1-2]. According to Wikipedia(http://en.wikipedia.org/wiki/Digital_forensics), it has several sub-branches according to target devices, media or artifacts. Computer forensics for a computer system, storage media or electronic document, mobile device forensics focused on cell/SMS/Email data, network forensics concerned with the monitoring and analysis of computer network traffic, and DB forensics for DB contents and log files. Thus, digital forensics investigations requires a variety of IT technologies to deal actual cases.

In general, digital forensics is done with 3 stages: acquisition or imaging of data from digital evidences, analysis of data, and reporting of investigation [3-4]. In acquisition stage, the forensic duplicate of the digital data stored in evidence digital devices is created in forensically sound manner. The forensic investigators do analysis with the copy. The evidence is analyzed to reconstruct events or actions and to reach conclusions, work that can often be performed by less specialized staff[1]. When an investigation is complete the data is presented, usually in the form of a written report, in lay persons' terms[1]. The famous commercial forensic tools like EnCase and FTK can handle the entire stages of investigation.

Anti-forensic techniques are to prevent proper forensic investigation or make it much harder[5]. They include data destruction such as wiping data, data contraception not to create any evidences, and data hiding such as encryption and/or steganography. Actually, the experts to computers or cryptography could use the techniques years ago. Recently, various anti-forensic tools with easy user interface can be obtained from internet. That means time and efforts required for forensic investigations would be longer and more.

One of the greatest challenges faced by a forensic examiner must be an encrypted file or data. Recovering original data from encrypted one may be not an uncommon case in forensic investigations. The impact of encryption on a digital forensic investigation is largely determined by the type of data being encrypted and how. The extent of what is encrypted combined with the strength of encryption methodology will have the greatest impact on the level of difficulty imposed on the investigator[7].

In this paper, we describe the password recovery for forensic investigation. In section 2, the password verification procedure and issues related to it are described. For easy understand, we show an example with encrypted Microsoft Office Word 2007 file found in forensic investigation. In section 3, we proposed future issues for more effective password recovery and to come up with the future trend. Finally, conclusions of this article is given in section 4.

# 2 Password recovery for forensics

## 2.1 How to verify

The forensic investigators may handle various file programs. Let's imagine that a forensic examiner who found encrypted file of Microsoft Office Word 2007. The first thing forensic investigator must know is file encryption procedure target program is applying. Most commercial programs apply the same procedure for file encryption as shown in figure 1.

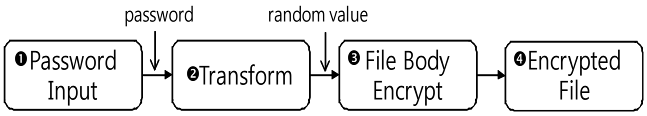


Figure 1. The general procedure of file encryption in the commercial products

The password a user selected is transformed into a string of random values, and it is used for real encryption key to the file body. The file encryption algorithm used in third step may be not useful because understanding the transform mechanism for the second step would give way to find password put into the first step. Thus, the examiner must know where the reference value is in the file, too. Each vendor applies a different transformation mechanism based on safe-proof algorithms in cryptography. While most vendors open their own transform mechanism and the position of the reference value in their public documents, some keep it in secret. If failing to find the documents, reverse engineering of the target program would be only the choice.

According to [8], Microsoft Office 2007 derives the encryption key from PKCS#5[17], which specifies the secure password generation mechanism, as described in the algorithm 1. As of now, H( ) denotes SHA-1, E( ) does AES, and '+' does concatenation.

**Algorithm 1** *Password-transform*

```
Salt <= 16 bytes random value
Input password(unicode) load
$H_0 = H(\text{salt} + \text{password})$
For integer $i$ from 0 to 49999
    $H_n = H(i + H_{n-1})$
$H_{50000} = H(H_{49999} + \text{0x00000000})$
DataBlock = $H_{50000}$ XOR 64 bytes data stream
  consisted of 0x36
EncryptionKey = H(DataBlock)
```

Table 1 summarizes cryptographic algorithms used for transforming user password in famous programs. Fortunately, the programs on the table give the detailed document which gives the algorithms and transform mechanism.

Table 1. Cryptographic algorithms used for some applications' password transform

| Programs | Cryptographic algorithm(s) |
|---|---|
| WinRar (v3.62) | SHA1 |
| PDF(v7.0 - v9.0) | MD5, RC4 |
| MS-Office 2007/2010 | SHA1, AES |

In Unix system, each registered user's login password is hashed and the result is stored. When a user wants to log in the system, he would be asked to put the password. The way to prove he is an authorized user is to show the hash value of the password exists in the system. In this case, the hash value of user's password is the reference value for future verifications. If readers consider the hashing process of passwords in Unix system's user authentication as transform process shown in figure 1, it can be said that forensic password verification procedure is basically the same as Unix's.

Microsoft Office Word 2007 also stores the reference value in the file as metadata[8]. For authentication of user passwords, Microsoft Office 2007 derives two values named EncryptedVerifier and EncryptedVerifierHash through the way describes in algorithm 2. The Salt value used in algorithm 1 and this two values are stored in the encrypted file's header field.

**Algorithm 2** *Reference value generation for verification*

```
Verifier <= 16 bytes random value
EncryptedVerifier = E(EncryptionKey, Verifier)
EncryptedVerifierHash = E(H(Verifier))
```

The investigator who found an encrypted Microsoft Office WORD 2007 file was required to extract the three values of (Salt, EncryptedVerifier, EncryptedVerifierHash) from the file for password recovery. After obtaining the EncryptionKey according to the word through algorithm 1, he must decrypt EncryptedVerifier and EncryptedVerifierHash, respectively. Finally, if the hashed value of decrypted EncryptedVerifier is the same as the value of decrypted EncryptedVerifierHash, the word might be the correct password.

Now, imagine again the examiner knows how to mix SHA-1 and AES algorithms to transform user password in Microsoft Office Word 2007 and where to find reference value for the verification. He can try verification with any word. If he gets the same as reference value after transforming a word he chooses, the word must be a password. The problem is how fast he can process the verification. Under the blind situation about the password, his available choice would be a bruteforce verification. Figure 2 shows a flow chart for password verification of encrypted files.

## 2.2 How fast to do

Because most investigators are assigned to multiple cases and have a finite amount of time to allocate to each investigation, they are often unable to devote adequate time to each examination[7]. However, just one encrypted file in which the critical evidence may be stored can make all investigators involved in a case must to be stuck. The reason is the password recovery is very time consuming process in

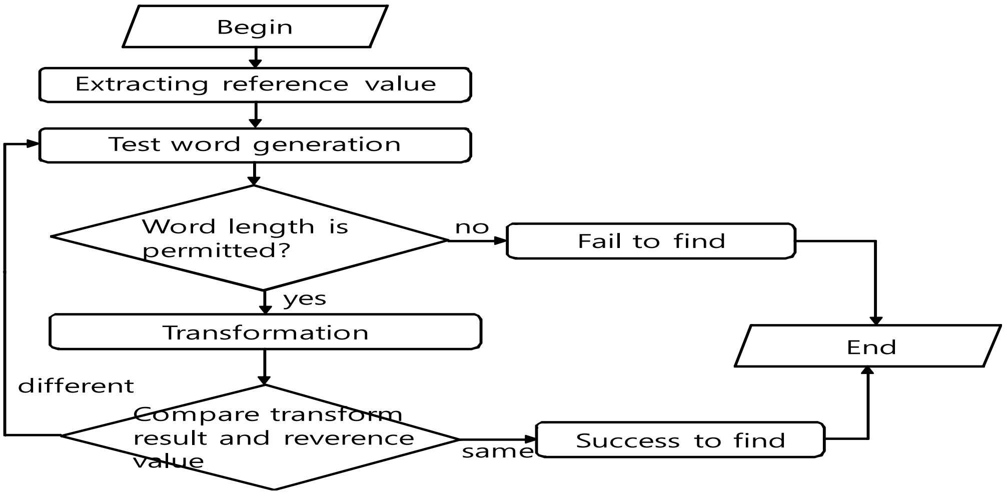

Begin
Extracting reference value
Test word generation
Word length is permitted?
no
Fail to find
yes
Transformation
End
different
Compare transform result and reverence value
same
Success to find

Figure 2. The flow chart for password verification of encrypted files

general. As mentioned before, vendors use safety-proof cryptoalgorithms for password transform. Since the algorithms were designed to survive the bruteforce attack, transform mechanisms also have the resistance to the same attack. Thus, the time consumed for password recovery relies on the performance of verification systems.

Many research groups have studied various methods to increase the speed of cryptographic algorithm processing for long time. In the early 1990s, the researchers paid their attentions to the software implementation techniques for the purpose. Due to the great advances in semiconductor technology, the cost for hardware solutions kept down. By the late 1990s, many researchers focused on specific hardware solution like FPGAs. It gave researchers not only high feasibility like software implementation but also high speed data processing like ASIC. Especially, they had great interested in built-in characteristics of parallel processing. By the middle of 2000s, many researchers and engineers have studied on applying multicore processors for parallel processing of cryptoalgorithms.

In Black Hat Europe 2008, CrackStation[9], an excellent accelerator for MD5 processing, was introduced. Taking advantage of the Cell's vector architecture[9-10], it showed up to 1.4 billion MD5 calculations per second. It was 90-100 times faster than performance of Intel-based architecture at that time. Nick Breese, the developer of CrackStation, implemented it on Sony’s Play Station 3. An alternative to CELL for the cryptographic accelerator might be graphic processing units(GPUs). Actually, GPU and Cell are close cousins from a hardware architecture point of view. They are flexible and easy to program using high level languages and APIs. Two major differences are the number of multicore processors and major usage. While CELL has 8 internal core processors, The NVIDIA GeForce 8800 GTX GPU is comprised of 16 streaming multiprocessors (SMs). Each SM has 8 streaming processors (SPs), with each group of 8 SPs sharing 16 kB of per-block shared memory[12-13]. Thus, GPUs can be regarded as massively parallel processors with 10 times faster computation and 10 times higher memory bandwidth than CPUs[11]. And, while CELL is general purpose microprocessor, GPUs are graphic processing accelerators mainly. Recently, NVIDIA introduced CUDA[13-14] to support developers who want to use GPUs for general purposes. Recently, most commercial products for password recovery of encrypted files use GPUs for fast processing. Figure 3 shows password recovery speed for Microsoft Office 2007/2010 supported by ElcomSoft AOPR. As one can see, NVIDIA GeForce 480 gives about 10 times faster recovery speed than Intel Core i7 processor.

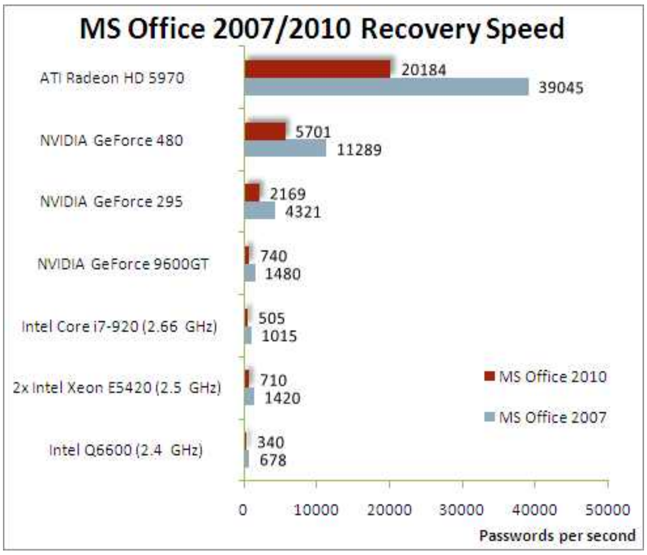


Figure 3. MS Office 2007/2010 password recovery speed by ElcomSoft AOPR[15]

## 2.3 How to reduce the number of words

In computer systems, the number of all printable characters in English language is 95. One can easily understand the number would be increased in exponential manner according to increase of word length. Thus, possible 5-length combinations with all printable characters are 955 = 7,737,809,375. Let's look over the figure 3, again. In the beginning of section 2, we assume an examiner found an encrypted Microsoft Office Word 2007 file. If he uses Elcomsoft AOPR program run on a computer system with a NVIDIA GeForce 480 board, he must have more than 190 hours to test all 5-length words. An worse thing is there in no guarantee the correct password is in the words. Who knows the password is 6-, 7-, or more than 7- length word. The time to be required for testing all combination of characters is increased in exponential manner, too.

As you can see in figure 3, password recovery speed for Microsoft Office Word 2010 is almost 0.5 times slower than the speed for Microsoft Office Word 2007. Thus, one can predict Microsoft would make transform mechanism more complicate to come up with increase of password recovery speed. In this point of view, password recovery for forensic investigation must not absolutely depend on the performance of any solution.

If one can reduce the number of words to be tested, then estimated time of verification will be reduced, too. For example, people tend to use words combined with lower case English characters and digit numbers as their passwords. That means the possibility to find the password would be very high from the verification with those characters. In this case, possible 5 length combinations with those characters is drastically reduced to (26 + 10)5 = 60,466,176, and estimated time of verification for the combinations about Microsoft Office Word 2007 would be less than 1.5 hour.

Another way to reduce the words to be tested is using a dictionary. In contrast with a bruteforce attack, where a large proportion key space is searched systematically, a dictionary attack tries only those possibilities, which are most likely to succeed, typically derived from a list of words for example a dictionary or a holy bible etc. Dictionary attacks succeed because many people have a tendency to choose passwords which are short (7 characters or fewer), single words found in dictionaries or simple, easily-predicted variations on words, such as appending a digit. John the Ripper, the famous password cracking tool, supports the dictionary attack. It takes text string samples (usually from a file, called a wordlist, containing words found in a dictionary), encrypting it in the same format as the password being examined (including both the encryption algorithm and key), and comparing the output to the encrypted string[16].

There is no doubt in that the success possibility of dictionary attacks depends on the dictionary used for the attacks. Thus, a dictionary file for the attacks doesn't need to be targeted for unspecific people. For the forensic investigation, the dictionary had better contain selected words derived from information about a suspect. For example, his/her name concatenated his/her birth year can be a choice. This dictionary has much smaller size than the general dictionary file included in John the Ripper tool package.left.

# 3 Future Issues

While finding vulnerability in transform mechanism might be the ideal case for password verification, it is actually impossible. As described before, the encryption key in Microsoft Office Word 2007 is derived from PKCS#5 through the transform process. The current version of PKCS#5 is announced in 1999, but the weakness or vulnerability of it has not been reported, yet.

Rather, to scale up the computing resources can be practical. In the example, the estimated time of Microsoft Office Word 2007 5-length password verification with 95 characters 190 hours under one NVIDIA GeForece 480 board. If a hundred of boards are applied for verification, the estimated time would be reduced to 1.9 hour. Famous password recovery solutions of Elcomsoft, Passware, and Passcovery provide distributed password recovery in client-server model. In this case, parallel processing by GPU in PC inside and parallel processing in distributed nodes are combined. However, the cost to maintain the massive system is a problem. And new version of commercial programs can easily mix the transform mechanism up so that the verification time is increased drastically. Thus, the computer systems of each police division or law enforcement organization are required to interoperate. For this purpose, integrated framework to guarantee the reliable and secure operation of linked systems must be designed.

And, to reduce the number of words for password verification, it is required to share the information about the crimes and criminals with each law enforcement organization. The profiling pattern from one crime case can be helpful to make a profile for the other cases.

# 4 Conclusions

In computer forensics, One of the greatest challenges faced by a forensic examiner must be finding the key used for encrypt a file or data. To handle this effectively, a great computing power and techniques to reduce the number of words to be verified for password recovery are required.

For the former, the specific hardware system like GPU, CELL, or FPGA, which inherits the parallel processing characteristics, is used. In addition, distributed computing via system aggregation or networking gives more powerful performance. For the later, dictionary based verification is in common.

Unfortunately, the word length people use as their passwords gets longer. And, program vendors make the password transform mechanism more complicate. To come up with the situation, a reliable and secure framework for interoperation of systems belonging to each law enforcement organization must be proposed to aggregate the computing resources.

And, information reported from investigation of crimes must be shared between law enforcements. For example, the password pattern found from investigation of crime case A might be helpful to make password dictionary for investigating a similar crime.

## 5 References

[1] M Reith, C Carr, G Gunsch, "An examination of digital forensic models," International Journal of Digital Evidence, 2002.

[2] Carrier, B, "Defining digital forensic examination and analysis tools," Digital Research Workshop II, 2001.

[3] Casey, Eoghan, Digital Evidence and Computer Crime, Second Edition, Elsevier, ISBN 0-12-163104-4. 2004.

[4] "Electronic Crime Scene Investigation Guide: A Guide for First Responders," National Institute of Justice, 2001.

[5] Vincent Liu and Francis Brown, "Bleeding-Edge Anti-Forensics," Infosec World Conference & Expo, April 3, 2006.

[6] Simson Garfinkel, "Anti-Forensics: Techniques, Detection and Countermeasures," 2nd International, Conference in i-Warefare and Security, pp 77, 2007.

[7] Mark Whitteker, "Anti-Forensics : Breaking the Forensic Process," ISSA Journal, pp 10, 2008..

[8] MS-OFFICE, "[MS-OFFCRYPTO]:Office Document Cryptography Structure Specification", available from http://msdn.microsoft.com/en-us/library/cc313071 (v=office.12).aspx

[9] CrackStation, available from http://www.security-assessment.com/files/presentations/crackstation-njb-bheu08-v2.pdf

[10] CELL, http://en.wikipedia.org/wiki/Cell_(microprocessor)

[11] A. Ailamaki, N. K. Govindaraju, S. Harizopoulos, and D. Manocha, "Query co-processing on commodity processors," Proceedings of the 32nd international conference on Very large data bases, pages 1267–1267. VLDB Endowment, 2006.

[12] Shuai Che, Jie Li, Jeremy W. Sheaffer, Kevin Skadron, John Lach, "Accelerating Compute-Intensive Applications with GPUs and FPGAs,", 2008 Symposium on Application Specific Processors, pp.101-107, 2008.

[13] E. Lindholm, J. Nickolls, S. Oberman, and J. Montrym, "NVIDIA Tesla: A unified graphics and computing architecture," IEEE Micro, 28(2), pp39-55, 2008.

[14] J. Nickolls, I. Buck, M. Garland, and K. Skadron, "Scalable parallel programming with CUDA," ACM Queue, 6(2):pp40-53, 2008.

[15] ElcomSoft, http://www.elcomsoft.com/aopr.html

[16] John the ripper, http://en.wikipedia.org/wiki/John_the_Ripper

[17] ] PKCS #5: Password-Based Cryptography Specification, available from http://www.ietf.org/rfc/rfc2898.txt

# A study of the Graphical User Interfaces for Biometric Authentication System

**Hiroshi Dozono**[1], **Takayuki Inoue**[1], **Masanori Nakakun**[2]**i**
[1] Faculty of Science and Engineering, Saga University, 1-Honjyo Saga, 840-8502 JAPAN
[2] Information Technology Center, Fukuoka University, 8-19-1, Nanakuma, Jonan-ku, Fukuoka 814-0180 JAPAN

**Abstract**— *Recently, many mobile devices are equipped with touch panel, like smart phones, tablet devices, MP3 players and portable gaming consoles. These devices are also equipped with network devices for connecting internet, and they are accessible to personal information or pecuniary system such that net shopping or banking. For accessing such system, user authentication is required. In this paper, we propose an biometric authentication system using touch panel as input device. As the biometrics, behavior biometrics of moving images on the touch panel is used, and the user interface for acquiring the features of behavior effectively is studied. The effectiveness of the proposed method is examined by authentication experiments.*

**Keywords:** Biometrics, Authentication, Touch panel, Mobile device

## 1. Introduction

Recently, many consumer devices, such as game consoles, tablet devices and mobile phones, are connected to the internet. These devices are used for web browsing, watching videos, and also for net shopping, trading and banking which require highly secure systems. As the authentication system, password authentication method, which is used on the conventional computers, is often used on these devices. However, many of these consumer devices are not equipped with keyboard, so on screen keyboard is used for typing password. During typing the screen keyboard, the typed characters are displayed on the display, thus typed password is easily peeped.

For this problem, biometric authentication is considered to be effective. Biometric authentication is classified to biological biometrics and behavioral biometrics. Biological biometrics uses biological characteristics, such as fingerprint, iris pattern, face and vein pattern for authentication. Biological biometrics can be applied to the authentication system with high accuracy. However, special hardwares for detecting biological characteristics are required. On the other hand, behavior biometrics uses the features of the behaviors, such as hand written patters[1], key stroke timings[2], signatures[3] and walking patters, for authentication. Some of these features can be obtained from the sensors conventionally equipped to the devices. However, accuracy of behavior biometrics is less than that of biological biometrics. In this paper, the features obtained from the touch panel are used for authentication. Recently, many devices are equipped with touch panels, thus this method can be widely applied to those devices.

As the authentication system using touch panel, the hand written signatures are often used. However, the devices, which are used in home, are also used by children who cannot write signature, and it is difficult to write uniform signature on slippery screen or capacitive type touch screen which are recently used as multi-tap panel for many devices. We proposed an authentication method using the hand written symbol for the touch panel of resistance film type[1]. As the features for authentication, the pen speed and pen pressure are used as multi-modal biometrics which can improve the accuracy of authentication[4]. However, this method can not be applied to recent capacitive touch panel because it can not detect pen pressure. For this problem, we proposed an authentication method using the movement pattern of some images on the touch panel. The combination of the movement of images as the password and the speed of movement as behavior biometrics. Using the combination of password and behavior biometrics, the authentication system becomes more secure.

In this paper, we propose some improvements for this method. At first, the path for moving images are specified on the screen as curvilineal lines, because we reported that the pen speed tracing on the curvilineal line is more adequate as behavior biometrics. Secondly, the selection of the images are extended to increase the patterns of combination used for password. Finally, the novel user interface for authentication is proposed to improve the usability and security. For each system, the authentication experiments are conducted. For evaluating the relative accuracy among the system, the simple authentication algorithm based on the thresholds are used.

## 2. Touch Panel Devices

Recently many mobile devices are equipped with touch panel. There are some types of touch panel devices. In past days, the touch panel devices using resistance films as sensors were mainly used. For this type of touch devices, special pen is used for inputting information or operating devices. Using resistance film touch panel, exact coordinate on the panel can be obtained, and it is possible to obtain pen pressure on the panel. The pen pressure contains rich features as behavior biometrics, and can be used for improving the accuracy of biometric authentication. However, resistance film touch panel is not sensitive enough to realize comfortable user interface, and it is impossible to detect multi-point on the touch panel to realize multi touch operation which is common for recent mobile devices.

Recently, most of the mobile devices use capacitive touch panel. Especially, the touch panel device which uses projected capacitance are mainly used because of the multi touch capability of this type of touch panel. Capacitive touch panel is operated by nude fingers, and can not be operated by solid pens which are used for resistance film touch panel. Because of this limitation, the accuracy of the coordinate detected by capacitive touch panel is worse than that of resistance film touch panel, and it is impossible to detect pen pressures. However, capacitive touch panel is sensitive enough for realizing confirmable user interfaces for operating mobile devices, and it can realize the multi-touch user interfaces.

In this study, the tablet PC equipped with capacitive touch panel is used in experiments. Fig.1 shows the device. This tablet PC is operated in Windows 7. We developed the authentication system directly on this computer using Microsoft Visual Studio to avoid inefficient cross development environment which is required for developing Android or iOS software.

Figure 1: Tablet PC(FMV LIFEBOOK TH40/D) with capacitive touch panel

## 3. Authentication System with Moving Images on Touch Screen(System-0)

In this section, the authentication method with moving images on touch screen which is reported in [5] is mentioned. Fig.2 shows the display image of the system. This system is called as System-0. In each box on the upper side, a image is set. The user is authenticated with moving the images to the box on the lower side in the pre-registered order for each user. The authentication is performed with matching the order of moved images to pre-registered order and with matching behavior characteristics during moving the images on touch screen to that of registration process. At each authentication, the images are arranged in random order in the upper boxes. The x-y coordinates of the moving images can be obtained from touch panel during moving the images. As the feature vectors for authentication, the sequence of velocity of the pen at each sampling time is used because the coordinate data of moving on the same traces did not be suitable for authentication and capacitive touch panel can not detect the absolute coordinates exactly.

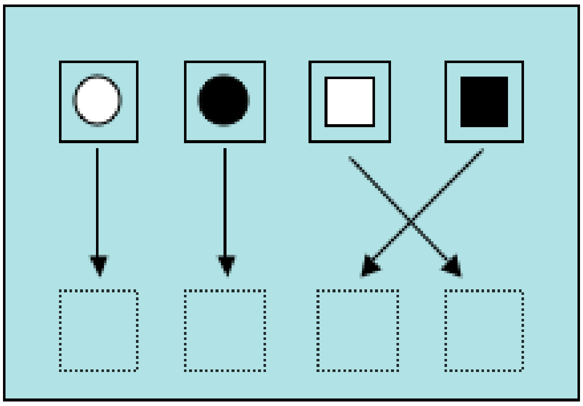


Figure 2: The authentication system with moving images(Sytem-0)

To examine the accuracy of behavior biometrics, the authentication experiment is conducted without changing the order of the images for all users. All users moves the images in same order, and the authentication is performed using velocity data. For evaluating the relative accuracy among the system, the simple authentication algorithm based on the thresholds are used. For each i-th-feature used for authentication, average $m_i$ and standard deviation $\delta_i$ is calculated from the data registered for authentication. For each test data $x_,$ it is authenticated as the specified user if difference of each feature $|x_i - m_i|$ is smaller than $\alpha\delta$ where $\alpha$ is the parameter for tuning authentication accuracy. In the following experiments, $\alpha$ is tuned at the value of cross points of FAR(False Acceptance Rate) and FRR(False Reject Rate) as shown in Fig.3.

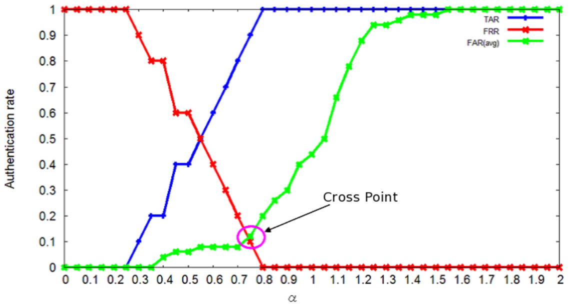


Figure 3: Cross point

FAR denotes the rate of accepting false users, FRR denotes the rate of rejecting the true user, and TAR denotes the rate of accepting true user. Table 1 shows the result.

Table 1: Result of authentication experiment of System-0

| User | TAR | FRR,FAR | $\alpha$ |
|---|---|---|---|
| A | 0.9 | 0.1 | 0.5 |
| B | 0.6 | 0.4 | 0.95 |
| C | 0.6 | 0.4 | 0.45 |
| D | 0.65 | 0.35 | 0.6 |
| E | 0.7 | 0.3 | 0.85 |
| F | 0.85 | 0.15 | 0.55 |
| G | 0.8 | 0.2 | 0.6 |
| H | 0.5 | 0.5 | 0.95 |
| I | 0.8 | 0.2 | 0.6 |
| J | 0.5 | 0.5 | 0.65 |
| avg | 0.69 | 0.31 | 0.67 |

For some user, good authentication accuracy over 70% is marked, however it was not good for more than half users. Almost all user move the images linearly from upper box to lower box, and the paths are not identical, thus the accuracy is degraded.

## 4. Authentication System using Curvilineal Path for Moving Images(System-1)

System-0 cannot mark good accuracy of authentication because of the fluctuations of the paths of moving images. For this problem, the path is specified as line in the display, and the user trace the line with finger for authentication. To improve the accuracy again, the curvilineal paths are specified because we reported that the pen velocity during tracing curvilineal path was more adequate as biometrics than that of linearly path[1]. Fig.4 shows the display image of this system. The authentication experiment is conducted using the behavior biometrics of pen velocity, without using the order of the images as password. Table 2 shows the result. The authentication accuracy is obviously better than that of System-0, and all users mark over 70% accuracy.

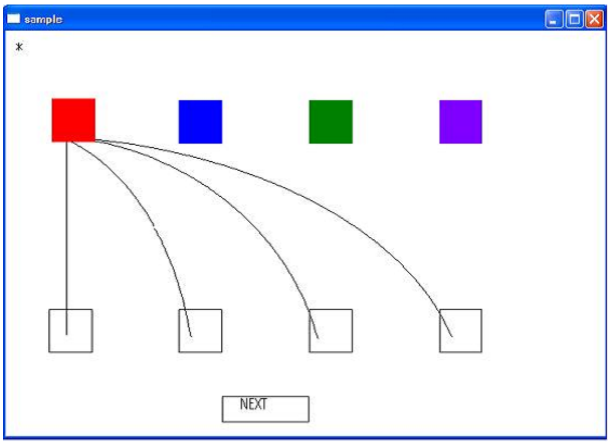


Figure 4: The authentication system using curvilineal path (System-1)

Table 2: Result of authentication experiment of System-1

| User | TAR | FRR,FAR | $\alpha$ |
|---|---|---|---|
| A | 0.7 | 0.3 | 0.7 |
| B | 0.8 | 0.2 | 0.8 |
| C | 0.8 | 0.2 | 0.45 |
| D | 0.9 | 0.1 | 0.9 |
| E | 0.85 | 0.15 | 0.6 |
| F | 0.8 | 0.2 | 0.4 |
| G | 0.78 | 0.22 | 0.4 |
| H | 0.8 | 0.2 | 0.85 |
| I | 0.92 | 0.08 | 0.9 |
| J | 0.75 | 0.25 | 0.5 |
| avg | 0.81 | 0.19 | 0.65 |

## 5. Authentication System using Separate Image Selecting Window(System-2)

System-1 shows good accuracy of authentication considering that it only uses behavior biometrics of pen velocity. However, the number of combination of the order of the images is only $4! = 24$, and this is too small as the variation of password. To increase the combinations, the extension of the number of images is considered to be simple method, however too long sequences causes forgetting the password. For this problem, the separate window which can display more images is introduced. After selecting a image, the window flipped to the window for moving the image.

As the images, Landolt Cs, which are usually used for the visual acuity expression are used. The availability for the user and robustness to the peeping can be tuned with changing the size of circles and gaps. For increasing variation of images, the colored Landolt Cs are used.

Fig.5 and FIg.5 show the display image of the system.

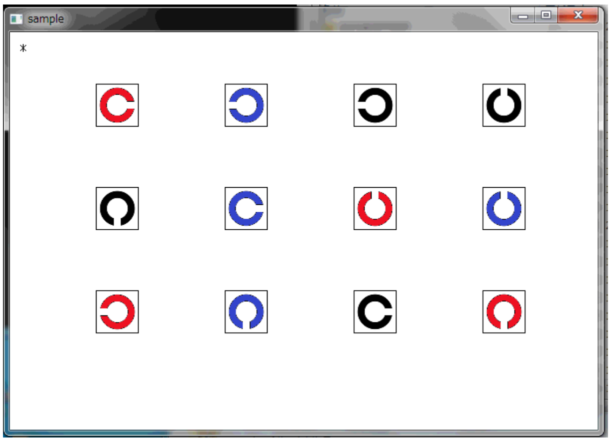


Figure 5: Image selection window of System-2

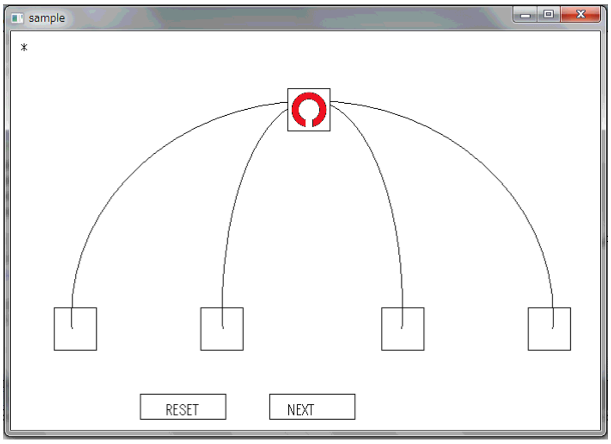


Figure 6: Image moving window of System-2

The user selects the image in the selection window. Then the window is flipped to image moving window, and the user move the image to the registered box tracing the lines. The authentication experiment is conducted using the behavior biometrics of pen velocity, without using the order of the images as password. Table 3 shows the result.

The authentication accuracy becomes worth than that of System-1. The reason is considered that the length and curvature of the path becomes smaller for System-2 than that of System-1. The longer and the adequate curvature are considered to be important to extract the feature of pen velocity as biometrics.

Table 3: Result of authentication experiment of System-2

| User | TAR | FRR,FAR | $\alpha$ |
|---|---|---|---|
| A | 0.6 | 0.4 | 0.85 |
| B | 0.8 | 0.2 | 2.0 |
| C | 0.9 | 0.1 | 1.45 |
| D | 0.72 | 0.28 | 0.8 |
| E | 0.8 | 0.2 | 0.85 |
| F | 0.6 | 0.4 | 1.1 |
| G | 0.8 | 0.2 | 1.35 |
| H | 0.58 | 0.42 | 0.8 |
| I | 0.6 | 0.4 | 0.7 |
| J | 0.8 | 0.2 | 0.7 |
| avg | 0.72 | 0.28 | 1.06 |

As for the usage of the direction of Landolt Cs and colors used as passwords, 7 users used identical color for all images, and 5 users used Landolt Cs sequentially rotated in identical direction. It may be difficult for the user to memorize the combinations of arbitrary directions of Landolt Cs and colora. From this result,, it is better for the user to use familiar images as password.

## 6. The Authentication System using Circular Path (System-3)

For the problem found in System-2, the authentication system using familiar images and Circular Path is proposed. FIg.7 shows the window for selecting images. In this experiment, the images of animals are used as familiar images.

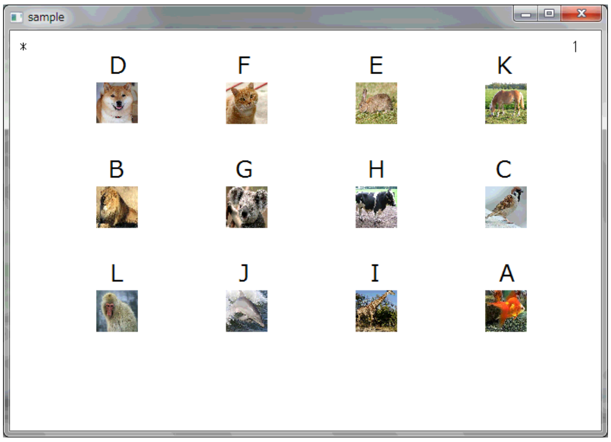


Figure 7: Image selection window of System-3

Fig.8 shows the window for moving mages. The user select a image on selection window in FIg.7. Then, the window is flipped to the image moving window shown in

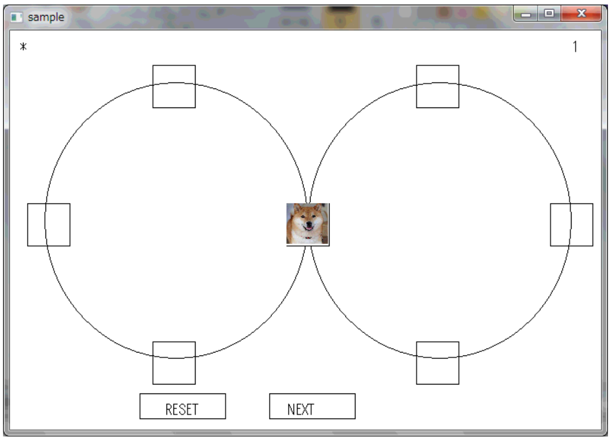


Figure 8: The authentication system using circular path(System-3)

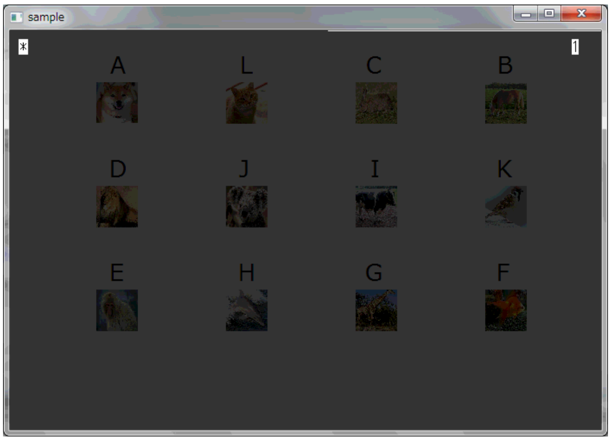


Figure 9: Image selection window with changing contrast level

Fig.8, and the user move the image to the registered box tracing the registered direction of the circle. Compared with System-2, the variations of destination places and moving directions are increased. If the system evaluates the length of total path is too short, the system can suggest inverse direction to the user during registration. The authentication experiment is conducted using the behavior biometrics of pen velocity, without using the destination places of the images as password. Table 4 shows the result. The accuracy of authentication of System-3 using pen velocity as biometrics is almost as same as that of System-1. Considering that System-3 can provide more variations of password, System-3 is much better than System-1 as authentication System.

Table 4: Result of authentication experiment of System-3

| User | TAR | FRR,FAR | $\alpha$ |
|---|---|---|---|
| A | 0.92 | 0.08 | 0.5 |
| B | 0.72 | 0.28 | 0.7 |
| C | 0.8 | 0.2 | 1 |
| D | 0.8 | 0.2 | 0.8 |
| E | 0.9 | 0.1 | 0.5 |
| F | 0.7 | 0.3 | 0.65 |
| G | 0.83 | 0.17 | 0.7 |
| H | 0.8 | 0.2 | 0.85 |
| I | 0.83 | 0.17 | 0.4 |
| J | 0.77 | 0.23 | 0.8 |
| avg | 0.807 | 0.193 | 0.69 |

## 6.1 Control of the visibility

As mentioned in the previous section, System2- uses Landolt Cs as images for controlling the visibility of the images. The familiar images may be easily found out when it is peeped by another person. To control visibility, the contrast of image the screen is modified depending on the security level. Fig.9 shows the window for selecting images with changing contrast level. In this printed proceedings, it becomes much difficult to identify the images, however it is not so much difficult to identity on computer screen. We made experiment of measuring the angle of view of 5 users with changing the 5 level of contrast. Table 5 shows the result. For level 1 of contrast, the contrast of the window is not changes. For all users, the view of angle becomes narrower with changing the contrast to higher level. The visibility of the another user who is peeping the screen can be controlled with changing the contrast to appropriate level.

Table 5: Angle of view for each contrast level

| User | Level1 | Level2 | Level3 | Level 4 | Level 5 |
|---|---|---|---|---|---|
| A | 130 | 110 | 100 | 90 | 80 |
| B | 130 | 110 | 90 | 90 | 90 |
| C | 130 | 100 | 100 | 90 | 80 |
| D | 130 | 110 | 90 | 80 | 60 |
| E | 150 | 120 | 90 | 80 | 60 |

# 7. Conclusion

In this paper, we propose the user interface of the authentication system using the touch panel as input device. With moving the images on the touch panel, the proposed system uses the password which is represented by the destination places and behavior biometrics of pen velocity. Proposed system marks good accuracy considering the simpleness of authentication method with using only pen velocity as biometrics and simple authentication algorithm. Security of this system will be guaranteed with using the destination places of images as password.

As the future work, the authentication accuracy should be more strengthened. In this paper, simple algorithm using threshold is applied for the fair evaluation. The more smart algorithm should be examined as authentication algorithm. And, in this paper, multi-touch capability is not applied to authentication system. It may possible to develop more effective authentication method using multi-touch input.

## References

[1] Hiroshi Dozono and Masanori Nakakuni, et.al: The Analysis of Pen Pressures of Handwritten Symbols on PDA Touch Panel using Self Organizing Maps, Proceedings of the International Conference on Security and Management 2005, pp.440-445(2005)

[2] F. Monrose and A.D. Rubin: Keystroke Dynamics as a Biometric for Authentication, Future Generation Computer Systems, March(2000).

[3] J. J. Brault and R. Plamondon: A Complexity Measure of Handwritten Curves: Modelling of Dynamic Signature Forgery, IEEE Trans. Systems, Man and Cybernetics, 23:pp.400-413(1993)

[4] Hiroshi Dozono and Masanori Nakakuni,,et.al: An Integration Method of Multi-Modal Biometrics Using Supervised Pareto Learning Self Organizing Maps., Proc. of the Internal Joint Conference of Neural Network 2008,(2008)

[5] Hiroshi Dozono, Takayuki Inoue and Masanori Nakakuni,et.al, Study of Biometric Authentication Method using Behavior Characteristics on Game Consoles, Proc. of SAM 2009, (2009)

# Securing Sensitive Data Stored on Smartphones

## Using Face Recognition to Unlock Mobile Devices

**Mark Wilson**
Department of Computer Science
Sam Houston State University
Huntsville, TX 77341
mrw004@shsu.edu

**Lei Chen**
Department of Computer Science
Sam Houston State University
Huntsville, TX 77341
lxc008@shsu.edu

**Abstract – Smartphones and tablets have become the new staple of the business and commercial industries, allowing employees to complete necessary tasks without requiring them to be confined to an office. However, with this mobility it is required that employees using these devices remain cautious to secure the sensitive data that is stored on their smartphones. To better aid those employees that use smartphones to complete their daily tasks at work, it is imperative to develop the best security features possible in order to safeguard the data they carry. This paper presents security features that are already being employed to lock smartphones, and why they are weak in comparison to biometric security features. Biometric features will be outlined, specifically face recognition, and how it could be used to secure sensitive data stored on a smartphone.**

**Keywords**: Smartphone, Biometric, Face Recognition, Security

# 1 Introduction

Due to the overwhelming increase in popularity of smartphones and other mobile devices being used in commercial business, industry, and the classroom alike, strong security measures must be in place to protect the sensitive data stored on those devices [3][10][12]. The smartphone, in particular, has boomed in the business world because it allows for employers and employees to complete more complex tasks in a more efficient manner. These devices empower each user with the ability to immediately contact their employer via a company website, email, or text message and increase productivity by improving communication by sending and receiving data in real-time. Both the options to retrieve available company data and communicate instantly with other employees allow for more informed decisions to be made and for productions to be streamlined [3].

Smartphones are imperative to business management and for employees to properly maintain their designated workloads while away from the office [1]. The primary method of securing all of the data stored on one's smartphone is employing a strong password. A password is a secret word or string of characters that is used as authentication to prove identity or to gain access to a particular resource [10]. Passwords are able to serve as an entry point to aspects of the user's everyday personal life, along with his professional life, including: logging into a computer, checking emails from a server, transferring funds, shopping online, accessing programs and databases, and even composing or reading work-related documents. Due to the potential negative impact of a compromised password, it is crucial that a new technique to unlock smartphones be developed in order to keep the contents of that smartphone protected from unauthorized users. As even more smartphones are introduced into both the workforce and personal lives of users, smartphones will need to be personalized through biometric methods in order to safeguard against physical intrusion.

This paper will be structured as follows. In Section 2, passwords, screen locks, and conventional methods to secure smartphones will be discussed along with a brief introduction of biometric security measures. Section 3 will outline face recognition and two of the most commonly cited issues that make face recognition unreliable. Section 4 discusses how face recognition could be used to best protect a smartphone using hardware that is already available in most modern mobile devices. Section 5 draws conclusions and Section 6 proposes future research in this field.

# 2 Background

## 2.1 Passwords and Screen Lock

The most popular method of securing data employed on modern smartphones is the screen lock. Locking the screen essentially blocks all access to the smartphone unless some type of word, code, or pattern is entered to unlock the smartphone's feature and data. The use of passwords and pass codes has remained the most popular among smartphones, but are likely the weakest form of security. Password strength can be measured by the complexity of the password itself, using such measures as: length of the word or string of characters, use of upper and lower case, inclusion of numbers and symbols, and use of dictionary words. Personal information is also often used as a security measure, using a spouse's name or birthday as a password, for example [10].

Employing a password or pass code to unlock a smartphone is a weak method of security for a number of reasons. When setting a password on many websites, the site will prompt the user for a password and gauge the strength of that password, often requiring a certain number of letters, numbers, special characters ($, %, &, #), etc. [10][13]. Unfortunately, this type of password strength gauge is not a common feature among smartphones; those smartphones with supported apps to measure the strength of a password are generally third party applications and cost an additional fee to download. This may leave a novice user with a weak password and unsecured device. Furthermore, the smartphones that use a password or pass code to unlock the screen lock generally only require a four-digit number, providing a scant ten thousand possibilities. Though many smartphone pass codes allow for sixteen-digit numbers, the majority of users do not employ that increased security due to lack of convenience. Convenience is one of the top reasons that an actual password is not used to unlock a smartphone. Having to input a word using either an onscreen keyboard or a QWERTY keyboard each time the user wishes to use their device may prove frustrating and unrealistic depending upon the length of the word. Finally, a strong method of unlocking a smartphone, introduced by Android, requires the user only to swipe a pattern across a three by three square of dots. This method provides strong security because it offers a multitude of different combinations, and remains simple and convenient all while looking aesthetically pleasing.

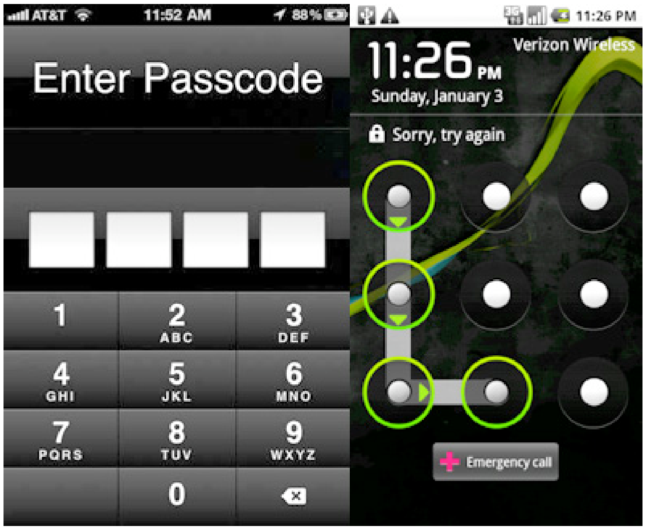


Figure 1. Examples of pass code and pattern security methods. Apple iPhone pass code (left) and Android security pattern (right)

Each of these methods protects a user's smartphone to varying degrees, but it is important to evaluate not only the technical merit of password combinations, but the physical security as well. Due to the majority of today's smartphones having touch screens that require user interaction to access data, regardless of the length of a pass code or the complexity of a pattern, the screen itself can give an intruder the solution to unlock the device. The residues and oils from a user's fingertips are naturally transferred to the smartphone's touch screen, especially those areas of the screen that are touched frequently (password, pass codes, and pattern swipes). It has been discovered that photographing a smartphone's touch screen and adjusting the contrast can reveal the user's pass code or pattern [7].

## 2.2 Different Types of Biometric Methods

Due to passwords being the weakest component of any important security system, it is necessary to attempt to fashion security in such a way that only authorized users are able to access restricted data [13]. Biometrics is the science and technology used to uniquely identify individuals based on their personal traits. Biometric security methods personalize access codes to ensure that only recognized users are able to view or modify sensitive data. This allows for a much stronger security scheme because access to information is based upon who a user is, rather than what that user has or remembers [9].

A particular biometric security scheme that would work well for securing smartphones is keystroke biometrics on an onscreen number pad. This method would not only require that the pass code be correct, but could also identify the user by typing rhythms that are compiled from the user's own style of typing [6]. Over time the smartphone itself would evaluate how the user inputs a pass code by compiling timestamps of touch and release of each digit on the screen. This method would be able to detect minor differences in habits such as using the index finger to enter a pass code versus the thumb. Upon detecting such a difference, this biometric method may determine that the person inputting the pass code is not the authorized user of the smartphone and additional security procedures could be deployed.

Voice recognition is another biometric security method that would readily apply to smartphones. This is a very plausible alternative to the standard pass code because the smartphone itself is able to offer all necessary hardware without having to attach external devices to properly capture the user's voice. Upon setting up the smartphone account the user could record a pass phrase that the smartphone would then compare and match to gain access to data in the future. This is a secure alternative to today's popular methods because a pass phrase could be known to others without risk of the smartphone being compromised because though the pass phrase could be duplicated, the voice could not [9].

# 3 Security through Face Recognition

Obviously the most secure way to protect any important data is to employ a security measure with millions of possible combinations. Depending upon the length of the password being typed, or the complexity of the security pattern being swiped, those millions of possibilities may be achieved. However, in order to truly personalize the smartphone and secure it, the biometric method of face recognition should be employed. Face recognition security measures have improved exponentially over the past decade, but two particular factors are generally cited as a hardship: alignment and illumination [2][8][11]. Though these two factors present difficulty for many face recognition applications, both would be easily solved when applied to smartphone screen lock security.

## 3.1 Face Alignment

Face alignment is an extremely important factor when employing any face recognition security measure because human faces have so many variations and viewing angles. This issue is most notable in unmonitored video surveillance in which face recognition software must analyze a dynamic image and match that image to a known face. Due to constant motion of a person's body, poor resolution, side-views and profiles, and changes in appearance such as facial hair, sunglasses, etc., the recognition device may provide false negatives or false positives when matching an identity with a face.

Face recognition is dramatically improved when faces are analyzed with known parameters, including: fixed distance, centered in front of the camera or scanner, known expression, eyes open, and non-moving [5][11]. Naturally, images such as mug shots or passport photos in which these parameters are strictly controlled offer the highest accuracy of face recognition with the actual identity of a particular person.

Facial alignment is an easily solved problem on the modern smartphone. When entering the pass code or security pattern, the user generally looks at the screen and touches the appropriate places to unlock the smartphone. Many of today's smartphones are installed with forward-facing cameras to allow for video conferencing; this same camera could be used to capture an image of the user's face for recognition purposes. By using the forward-facing camera, the user would be able to see himself on the screen and guarantee the image was centered for proper recognition. The image would also be taken from a fixed distance (arm's length, or shorter) that would be approximately the same each time the user unlocked his smartphone.

## 3.2 Face Illumination

Similar to the alignment of the user's face to ensure proper recognition, illuminating the face is also necessary for an accurate result [11]. Depending upon the location and the model used, a sufficient amount of light may not be available to collect an image that can be used for face recognition. Often used in video, night vision or infrared settings may allow for a suitable image to be collected to accurately identify a person's face. However, due to the high contrast of both methods, they are not considered reliable.

Obtaining proper and consistent exposure ensures that details of the user's face are captured and can be analyzed accurately. Research shows that multiple lighting sources on a flat, lighted background, similar to the standard studio photography lighting format, produce the best results that can be most accurately analyzed [11]. This setup is unrealistic in reality when dealing with face recognition as a security measure, but would apply if attempting to identify a person from an archived digital image taken in a studio environment.

Few smartphones and mobile devices on the market today feature a forward-facing flash to accompany the forward-facing camera, but by either installing a low-level diffused flash or a swiveling flash to the smartphone, a consistent amount of illumination could be provided to accurately analyze the user's face. This provides the smartphone itself a means of illumination so users would not have to depend on ambient light sources that may not provide the amount of light necessary to confirm their face.

# 4 Face Recognition on Smartphones

Because business professionals and personal users alike store so many increasingly sensitive items on their mobile devices, it is important to secure those devices with the strongest method possible. By deploying a face recognition biometric security method to smartphones and other mobile devices users can be confident that data remain secure. The face recognition software would be used to unlock a screen-locked mobile device to prevent any unauthorized users from gaining access to that device.

## 4.1 Unlocking Smartphones With Only the User's Face

To employ the biometric face recognition software to secure smartphones and mobile devices, the opening screen that typically requests the user's password or security pattern would be replaced with a blank screen with a large crosshair. The user would activate the front-facing camera by touching an onscreen button titled, "push to unlock." Once the camera is activated, the user would hold his smartphone in front of his face at a distance to properly align his face within the oval of the crosshair. The face should be positioned in such a way that the vertical line of the crosshair runs down the bridge of the user's nose and the horizontal line of the crosshair would run through the user's eyes.

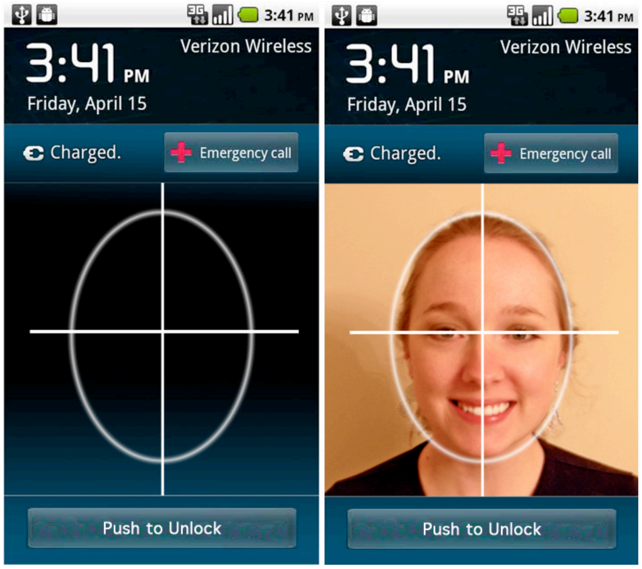


Figure 2. Example of the face capture crosshair (left) and proper face positioning in the crosshair (right)

By properly positioning the user's face in the crosshair, the common problem of face alignment would be corrected because a consistent image of the face would be taken each time. In order to center the user's face within the oval and the eyes along the horizontal line, the smartphone would capture an image that had extremely similar proportions each time. This would solve the obvious problems of side-views and profiles, and the distance from where the image was taken in relation to the camera.

For the business professional, who uses his smartphone most during daylight hours and inside of office buildings, etc., ambient light would not likely create an illumination problem. However, for the common user that does use his smartphone at various times of the day and direct ambient light was not available, a flash from the smartphone itself would be extremely helpful in order to capture an image that was bright enough to analyze. It is recommended that any flash used for this security technique either be diffused or set to a lesser power than the usual flash for a camera phone. This would both conserve the smartphone's battery life and allow for the image to capture facial detail and not overexpose the image.

After capturing the user's image, the smartphone would analyze it and compare it to a previous image taken to verify that it is the same authorized user. The comparison method could be performed using two different methods; the captured image could be compared to an image of the designated user upon initially setting up the security measure on his smartphone, or images of successful attempts to unlock his smartphone would be saved and used for comparison. The latter option would compare each subsequent captured image to a previously authorized image allowing the user's images to somewhat evolve and change slightly over time. This option would allow for minor facial changes, such as: five o'clock shadow, healing injuries, and possibly conservative haircuts. This would also allow for the smartphone to only save one captured image at a time, reducing the amount of hard drive space necessary to run the software.

In the event a dramatic facial change, such as: plastic surgery, new injury, shaving facial hair, etc., along with possible error in the software, it is important that a secondary security measure be available until a new image of the authorized user be captured for comparison. However, it is stressed that the secondary security measure be as strong as possible and not to simply rely on the biometric technology. Users often create a weaker password to secure their data when they believe they are being protected by two or more security methods [13].

Touching the "push to unlock" button a second time would capture an image. Upon comparing and verifying the captured image to the saved image of the authorized user, the smartphone would be unlocked and access would be granted as usual. This level of security is neither needed nor intended for the common user, but would benefit business professionals, college professors, government employees, and tech enthusiasts alike. With the proper speed and resources, taking a picture of one's self to unlock a smartphone may prove no more time consuming than entering a password or swiping a pattern.

## 5 Conclusion

This paper proposed a new method of safeguarding sensitive data stored on a smartphone or mobile device using face recognition software. The popularity of smartphone and other mobile devices have increased tremendously over the past decade, and society is finding a growing number of people using those devices. Depending upon the occupation of the user, be it corporate businessman, government employee, or even student, each store their own types and amounts of sensitive data on their devices. The screen lock is the first line of defense against unauthorized users accessing other's devices, but unfortunately, the average user has a weak or predictable password or pass code that would fail to safeguard one's smartphone. Presented in this paper is not a new method of face recognition or biometrics, but a new way to employ face recognition to secure technology that is becoming exponentially more prevalent in modern society. By using a biometric technology to secure smartphones, the smartphone itself is trained to react strictly to the user and may be considered impervious to break-in.

## 6 Future Work

Future work for this security model will include the development of the interface to both iOS and Android platforms. An extensive amount of time will be budgeted to integrate a face recognition software into the smartphone OS that will provide both speed and user-friendliness while performing accurately and securely. Smartphone hardware will be evaluated to determine the maximum efficiency to power a camera flash, store captured facial images, analyze and compare images, and how quickly these operations can be completed. Most importantly, once properly developed, surveys must be conducted to determine if the general public would use this security model. As quickly as smartphones are improving, if the technology is not able to perform at a rate similar to entering a pass code or pattern users may opt to employ a less secure method. Finally, research must be conducted to determine any major faults or holes in the security of a smartphone face recognition system, and how to repair those faults.

## 7 References

[1] F. Douglis, “As I Emerge From the Mobile Phone Dark Ages, I Look Around in Fear and Wonder,” *IEEE Internet Computing*, vol. 14, pp. 4-6, Jul./Aug. 2010.

[2] M. Hanmandlu, et al., “An Experimental Study of Different Features for Face Recognition,” in International Conference on Communication Systems and Network Technologies, 2011 IEEE. doi: 10.1109/CSNT.2011.121

[3] I. Henri and L. Aurelie, “Give Me a Mobile Phone, and I Will Work Harder! - Assessing the Value of Mobile Technologies in Organizations: An Exploratory Research,” in International Conference on Mobile Business (ICMB), 2006 IEEE.

[4] G. Hua et al., “Introduction to the Special Section on Real-World Face Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 1921-1924, Oct. 2011.

[5] P. Li, et al., “Automatic Recognition of Smiling and Neutral Facial Expressions,” in Digital Image Computing: Techniques and Applications, 2010 IEEE. doi: 10.1109/DICTA.2010.103

[6] R.A. Maxion and K.S. Killourhy, “Keystroke Biometrics with Number-Pad Input,” in IEEE/IFIP International Conference on Dependable Systems & Networks (DSN), 2010 IEEE.

[7] C. Moor, “Your Finger Smudge Could Be Your Downfall” available at: http://www.talkandroid.com/10653-your-finger-smudge-could-be-your-downfall/#TtaFGkZ3yHI

[8] P.J. Phillips, “Improving Face Recognition Technology,” in Computer, 2011 IEEE. available at: http://gala.cs.iastate.edu/coms510/references/IEEEComputer_FaceRecognition_March2011.pdf

[9] A. Pocovnicu, “Biometric Security for Cell Phones,” *Informatica Economica*, vol. 13, no. 1, pp. 57-63, 2009.

[10] M.S. Vijaya et al., “Password Strength Prediction Using Supervised Machine Learning Techniques,” in International Conference on Advances in Computing, Control, and Telecommunication Technologies, 2009 IEEE. doi: 10.1109/ACT.2009.105

[11] A. Wagner, et al., “Towards a Practical Face Recognition System: Robust Alignment and Illumination by Sparse Representation,” *Trans. Pattern Anal. Mach. Intell.,* unpublished. doi: 10.1109/TPAMI.2011.112

[12] J. White and H. Turner, “Smartphone Computing in the Classroom,” in Pervasive Computing, 2011 IEEE. available at: http://www.computer.org/pervasive

[13] H. Wimberly and L.M. Liebrock, “Using Fingerprint Authentication to Reduce System Security: An Empirical Study,” in IEEE Symposium on Security and Privacy, 2011 IEEE. doi: 10.1109/SP.2011.35

[14] R. Xia et al., “Business Models in the Mobile Ecosystem,” in Ninth International Conference on Mobile Business / Ninth Global Mobility Roundtable, 2010 IEEE. doi: 10.1109/ICMB-GMR.2010.30

# An Efficient Acceleration of Digital Fonensics Search Using GPGPU

Chung-han Chen and Fan Wu
*Computer Science Department*
*Tuskegee University*
*Tuskegee, AL 36088*
*E-mail: jchen@mytu.tuskege.edu and wuf@mytu.tuskegee.edu*

***Abstract*—Graphics Processing Units (GPU) have been the extensive research topic in recent years and have been successfully applied to general purpose applications other than computer graphical area. The nVidia CUDA programming model provides a straightforward means of describing inherently parallel computations. In this paper, we present a study of the efficiency of emerging technology in applying General Purpose Graphics Processing Units (GPGPU) in high performance digital forensics search solutions. We implemented forensics search algorithm using the novel CUDA platform on nVidia Geforce 280 GTX and compared its performance with an optimized CPU implementation on a high-end AMD Opteron Dual Core CPU. Our experimental results show that GPGPU can perform as an efficient digital forensics search accelerator and the developed GPU based implementation achieve a significant performance improvement over CPU based implementation and the maximum observed speedups are about 100 times.**

***Keywords - Digital Forensics; NTFS; General Purpose Graphics Processing Unit; CUDA***

## I. INTRODUCTION

Graphics Processing Units (GPU) have been the extensive research topic in recent years and have been successfully applied to general purpose applications other than graphical area [1]. This trend toward general-purpose computation on GPUs (GPGPU) is spurred by a large number of arithmetic units and a high memory bandwidth available in today's GPUs. In certain applications, where there is a high compute to memory bandwidth ratio the GPU has the potential to be orders of magnitude faster than conventional CPUs due to the parallel nature of GPUs versus CPUs, which are inherently optimized for sequential code. In addition, the computational power of GPUs is growing at a faster rate than what Moore's Law predicts for CPUs (Figure 1).

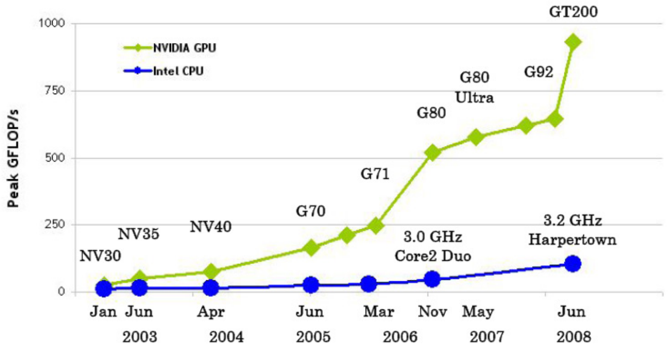


Figure 1: GPU vs. CPU GFLOPS performance over time [2]

Traditional single-core microprocessors are having difficulty achieving higher clock frequencies. The limitations imposed by deep pipelining, transistor scaling, and power and thermal constraints have forced CPU vendors to find other ways to meet high performance computing needs.

One solution is that of multi-core architectures, which integrate multiple cores onto a single chip. Examples are off-the-shelf Intel Duo-core and Quad-core products, the Sony/Toshiba/IBM alliance's Cell Broadband Engine [3], MIT RAW [4], and Sun's Niagra [5]. Another powerful solution is the GPU. GPUs represent highly specialized architectures designed for graphics rendering, their development driven by the computer gaming industry. Recently, there has been a trend to accelerate computationally intensive applications, including scientific applications, on graphic processors. This trend introduced the new term GPGPU or General-Purpose computation on the GPU.

The GPU has several key advantages over CPU architectures for highly parallel, compute intensive workloads, including higher memory bandwidth, significantly higher floating-point throughput, and thousands of hardware thread contexts with hundreds of parallel compute pipelines executing programs in a SIMD fashion. The GPU provides parallel processing of large quantities of data relative to what can be provided by a general CPU. The GPU can be an attractive alternative to CPU clusters in high performance computing environments.

The demand of efficient digital forensics solutions has been continuously growing in the last decades as a consequence of using the internet in critical areas like government and law enforcement agencies. As a result, many hardware based SSL acceleration solutions have been studied and proposed both in research and industrial fields. While the performance improvement that can be derived from accelerations is significant, only a relatively small number of systems employ such dedicated hardware.

Recent announcement of technologies like CUDA [2] by nVidia proved their effort to extend both the programming and memory models. CUDA (for Compute Unified Device Architecture) is a new data-parallel, C-language programming API that bypasses the rendering interface and avoids the difficulties of classic GPGPU. Parallel computations are instead expressed as general-purpose, C-language kernels operating in parallel over all the points in a domain.

This paper introduces an emerging technology in porting digital forensics search to CUDA and analyzes its performance. With the introduction of native integer and binary operations in the latest generation of GPUs, we believe that digital forensics search and its related applications are ideally suited to the GPGPU programming model. Digital forensics search has a great deal of data-parallelism and benefit from the GPU's parallel computing resources. We present an effort in developing an efficient novel approach in implementation of digital forensics search on the GPU based on CUDA. We compare results on an NVIDIA Geforce 280 GTX against an AMD Opteron Processor both under Solaris. All of our experimental results show satisfactory speedups. The maximum observed speedups are about 100 times.

The rest of the paper is organized as follow: Section II introduces some previous work; Section III describes the background on GPU and CUDA briefly; Section IV presents the digital forensics search algorithm and our CPU and GPU implementation; Our experimental results are presented in Section V; Finally, Section VI concludes this paper with our future direction.

## II. AN OVERVIEW OF CUDA ARCHITECTURE

The Workstation that we have used in our implementations is nVidia's GeForce GTX 280, part of the G80 series, which is DirectX 10 compliant. It is one of Nvidia's fastest processors that support the CUDA API and as such all implementations using this API are forward compatible with newer CUDA compliant devices. All CUDA compatible devices support 32-bit integer processing. The 280 GTX consists of 30 SIMD processors, called Streaming Multiprocessors (SM). Each SM contains 8 ALUs which operate in lockstep controlled by a single instruction unit. The simplest instructions are executed by the SMs in 4 clock cycles, which creates an effective SIMD width of 32. Each SM contains a small amount of fast local storage which consists of a register file, a block of shared memory, a constant memory cache and a texture memory cache. The main bulk of the GPU's storage is off-chip global memory, and is thus considerably slower than the on-chip storage. A lot of the programming effort concerning a GPU is the careful usage of the different types of available storage. This is due to the scarcity of fast storage and so performance can degrade dramatically with naive implementations. The code which runs on the GPU is referred to as a kernel. A kernel call is a single invocation of the code which runs until completion. The GPU follows a Single Process, Multiple Data (SPMD, or recently SIMT) threading model. All threads must run from the same static code and must all finish before a kernel can finish. Via the CUDA API, the programmer can specify the number of threads that are required for execution during a kernel call. Threads are grouped into a programmer defined number of CUDA blocks, where each block of threads is guaranteed to run on a single SM. Threads within a block can communicate using a synchronization barrier which ensures that all threads within the block fully commit to the same instruction before proceeding. The number of threads per block is also programmer defined. They should be allocated in groups of 32, called a CUDA warp, to match the effective SIMD width mentioned above. If the thread execution path within a warp diverges, all paths must be executed serially on the SM. An important consideration for GPU performance is its level of occupancy. Occupancy refers to the number of threads available for execution at any one time. It is normally desirable to have a high level of occupancy as it facilitates the hiding of memory latency.

### A. *C Language Extensions*

One of the CUDA's characteristics is it provides a minimal set of extensions to the C language that allow the developer to target portions of the source code for execution on the device. The runtime library of the architecture is split into three major components:

1) **Host component:** Runs on the host and provides functions to control and access one or more compute devices from the host;

2) **Device component:** Runs on the device and provides device-specific functions;

3) **Common component:** provides built-in vector types and a subset of the C standard library that are supported in both host and device code. It should be emphasized that the only functions from the C standard library that are supported to run on the device are the functions provided by the common runtime component.

### B. *Threads, Kernels, Blocks and Grids*

Using CUDA as an extension allows the developer to create special C functions, called kernels. Each kernel executes on N different CUDA threads. The special part of kernels is that it is executed in parallel on each thread. Regular C functions can only execute one.

The dimension of the grid is specified by the first parameter of the <<<…>>> syntax. Each block within the grid can be identified by a one-dimensional or two-dimensional index accessible within the kernel through the built-in blockIdx variable. The dimension of the thread block is accessible within the kernel through the built-in blockDim variable.

Thread blocks are required to execute independently: It must be possible to execute them in any order, in parallel or in series. This independence requirement allows thread blocks to be scheduled in any order across any number of cores, enabling programmers to write scalable code.

The number of thread blocks in a grid is typically dictated by the size of the data being processed rather than by the number of processors in the system, which it can greatly exceed.

### C. Memory

The GPU memory architecture is shown in Figure 2.

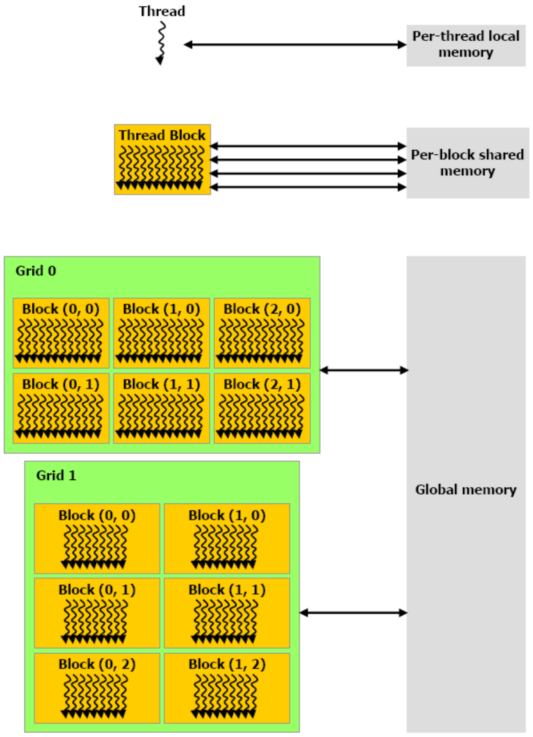


Figure 2: GPU Memory Architecture [2]

Each thread has a private local memory. Each thread block has a shared memory visible to all threads of the block and with the same lifetime as the block. Finally, all threads have access to the same global memory. There are also two additional read-only memory spaces accessible by all threads: the constant and texture memory spaces.

## III. DIGITAL FORENSICS SEARCH

In the currency research, we focus on the keyword(s) search on a disk image using GPGPU. The idea is simple and straight: the disk image is divided into a number of smaller memory spaces, roughly the same number of the threads. Each thread then conducts the keyword search and returns the location (the file name which contains the key word. We will extend the GPGPU application to other digital forensics functions.

We chose to use disk images to perform the digital forensics search experiment instead of using a physical hard drive because a disk image can exactly replicate the disk partitions. Hash function has been used to ensure that the image is authentic and validated to the original disk. The image used in our experiment is NTFS formatted, starting form volume id (at byte offset 0x03, eight bytes in length), 512 bytes per sector (at byte offset 0x0B, two bytes in length), the sectors per cluster (at byte offset 0x0D, one byte in length), and that the logical cluster number for the Master File Table (MFT) (at byte offset 0x30, eight bytes in length). [4] The master file table consist the links to all the system files, program and data files, and directories. Both CPU-based and GPU-based (CUDA) programs were written to locate and read the boot sector and Master file table (MFT), and to perform the digital forensics (Keyword) search function.

### A. CPU-Based Algorithm

The CPU-Based algorithm pseudo code is shown in Algorithm 1:

**Algorithm 1: CPU-based Algorithm Pseudo Code**

```
1. Locate and read Master File Table (MFT)
2. Input keyword
3. Start timer
4. Begin keyword search
5.      Do
6.        Next file
7.        Search keyword in file
8.          If keyword found
9.              Output filename
10.       Until (all file searched)
11. end timer
```

### B. GPU-Based Algorithm

Let $T^*(n)$ be the time complexity of a sequential algorithm to solve a problem P of input size n. Let $T_p(n)$ be the time complexity of a parallel algorithm to solve P on a parallel computer with p processors. The total speedup is calculated as equation 1:

$$S_p(n) = T^*(n) / T_p(n) \tag{1}$$

And $S_p(n) \le p$ . The best scenario is $S_p(n) = p$ , when $T_p(n) = T^*(n) / p$

The efficiency is calculated as equation 2:

$$E_p(n) = T_1(n)/(pT_p(n)) \tag{2}$$

where $T_1(n)$ is when the parallel algorithm run in 1 processor. The best scenario is $E_p(n) = 1$. The GPU-Based algorithm pseudo code is shown in Algorithm 2:

**Algorithm 2: GPU-based Algorithm Pseudo Code**

```
Determine the device configuration
Initialize variables
Allocate device memory for constants
Copy constants from host memory to device constant memory
Setup Execution Parameters
Locate and read Master File Table (MFT)
Input Keyword
Create and start timer
Execute kernel
    Synchronize threads
    Do
        Next file
        Search keyword in file
        If keyword found
            Update file information in shared memory
        Synchronize threads
    Until (all file searched)
Copy file information from device memory to host memory
Stop timer
Check if kernel execution generated an error
Clean up memory
End program
```

Our CUDA kernel performs the main computation of the digital forensics (keyword) search algorithm. It is executed by 256-thread blocks in GPU, which means we can process 256 data in parallel. It improves overall parallel work efficiency and provides more opportunities to hide latency.

## IV. Experimental Results

### A. Description

Our experiment attempts to measure the runtime of a GPU digital forensics search algorithm to return the file information which contain a given keyword. The experiment will also attempt to emulate what the GPU does using a similar CPU digital forensics search algorithm and will use the comparison to show the pros and cons of such implementations.

### B. Setup and Device Configuration

The setup for this experiment requires the CUDA Runtime Library, GeForce GTX 280 graphics card, AMD Opteron, and the CUDA Programming Guide 3.0. We created a program we entitled Brue Force it to generate our metrics and figures. The CPU component will hold a menu driven component to allow the user to decide which approach they would like to make (CPU or GPU). The program then allocates memory for the CPU and GPU accordingly to facilitate the programs keyword search function. The two programs differ on how many slaves are performing keyword search at a time. The average of these runtimes is then computed and used to determine the performance of both approaches.

### C. Experiments for GPU-Based algorithm

The first step for the GPU based algorithm is to setup device configuration. We determine the size of the memory needed to be allocated and save it to a variable. After determining the amount of memory needed we can allocate space for the keyword and MFT to be saved to constant memory. Once copied to constant memory they can be accessed by all threads on the device. Next we setup the execution parameters of the device. To ensure peak performance we have to consider the available memory on the device and the resources needed by each active thread. The number of threads is a multiple of the warp size to reduce wasted resources. The block size is set to have at least two blocks per multi-processor. This is done the parameters of the kernel invocation <<<…>>>. After setting up the grids of the threads that will execute the kernels the keyword search done by the kernels can begin on the device in parallel. Once the kernel begins it uses the threadIdx variable provided by the runtime library to access the individual threads. The threads are synchronized to ensure reads to memory are up to date. Next we read the keyword and MFT from constant memory. Then we begin a loop to conduct the search function. If key word is found in a file, the file information (i.e, path and name) is copied to the local memory. We synchronized the threads again here to control threads making bad reads. The loop ends when all files are searched. Finally we destroy timer and clean up memory. The results of the GPU approach shows that the GPU approach is about 100 times speedup, compared to the CPU approach.

## V. Conclusion and Future Work

We have presented an emerging technology in efficiently performing digital forensics search on GPGPUs. We implemented high performance digital forensics (keyword) search utilizing highly parallel computations capability of GPGPU on nVidia CUDA. We have demonstrated GPU can perform significantly faster than CPU in the field of digital forensics. Our experimental results indicate that our GPU-based implementation shows a significant performance improvement over CPU-based implementation and the maximum observed speedups are about 100 times.

There are several avenues for future work. We would like to test our algorithm on different GPUs and explore the new performance opportunities offered by newer generations of GPUs. It would also be interesting to explore efficient implementation of other digital forensics functions. Finally, another interesting direction is exploring digital forensics on multiple GPGPUs.

## Acknowledgment

This work has been supported by U.S. NSF grant# 0753351.

## References

[1] J.D. Owens, D. Luebke, N. Govindaraju, M. Harris, J. Krger, A.E. Lefohn, and T.J. Purcell,: A survey of general-purpose computation on graphics hardware. *Computer Graphics Forum 26(1) 80-113*, 2007

[2] NVIDIA Corporation. NVIDIA Programming Guide 3.0. Retrieved February, 2010. www.nvidia.com..

[3] J. Kahle, M. Day, H. Hofstee, C. Johns, T. Maeurer, and D. Shippy,: Introduction to the cell multiprocessor. *IBM Journal of Research and Development.,49(4/5):589C604*, 2005.

[4] M. B. Taylor, W. Lee, J. Miller, D. Wentzlaff, I. Bratt, B. Greenwald, H. Hoffmann, P. Johnson, J. Kim, J. Psota, A. Saraf, N. Shnidman, V. Strumpen, M. Frank, S. Amarasinghe, and A. Agarwal,: Evaluation of the raw microprocessor: An exposed-wire-delay architecture for ilp and streams. In *Proceedings of the 31st annual international symposium on Computer architecture*, Washington, DC, USA, 2004. IEEE Computer Society.

[5] P. Kongetira, K. Aingaran, and K. Olukotun. Niagara,: A 32-way multithreaded sparc processor. *IEEE Micro*, 2005.

# SESSION

# HARDWARE SECURITY

## Chair(s)

**TBA**

# New Error Detecting Codes for the Design of Hardware Resistant to Strong Fault Injection Attacks

Zhen Wang and Mark Karpovsky
Reliable Computing Laboratory, Boston University, Boston , USA
wang.zhen.mtk@gmail.com, markkar@bu.edu

*Abstract*—**Cryptographic devices suffer from fault injection attacks. The security of crypto-systems protected by traditional error detecting codes rely on the assumption that the information bits and the error patterns are not both controllable by the attacker. For applications where the assumption is not valid, the security of systems protected by traditional error detecting codes can be easily compromised. In this paper, we present constructions for algebraic manipulation detection (AMD) codes based on the nonlinear encoding functions. For a $(k, m, r)$ AMD code, a message contains three parts: $k$-bit information data $y$, $m$-bit random data $x$ and $r$-bit redundancy $f(y, x)$. For any error $e$ and information $y$, the fraction of $x$ that masks the error $e$ is less than $1$. In this paper we describe lower and upper bounds on AMD codes and show that the presented constructions can generate optimal or close to optimal AMD codes in many cases. We presented efficient encoding and decoding methods for AMD codes minimizing the number of multipliers using the multivariate Horner scheme. The proposed codes can provide a guaranteed high error detecting probability even if both the information bits of the code and the non-zero error patterns are controllable by an attacker. These codes can be used for design of secure multipliers, secure memories or secure hardware implementing cryptography algorithms resistant to fault injection attacks.**

***Keywords*-Error Detecting Codes, Nonlinear Codes, Secure Hardware, Fault Injection Attacks.**

## I. INTRODUCTION

Error detecting codes are widely used for communication channels and for computation channels to protect reliable and secure devices against soft errors, hard errors and malicious attacks in applications like Internet, data storage, cryptosystems and wireless communications.

Most of the existing reliable and secure architectures [1], [2], [3], [4], [5], [6] are based on linear codes such as 1-d parity codes, duplication codes, Hamming codes, BCH codes, Reed-Solomon codes, etc. The error detecting capabilities these architectures largely depend on the accuracy of the error model and may not be sufficient if an attacker can control errors distorting the received messages for communication channels or errors distorting outputs of a device protected by an error detecting code for computation channels.

Robust codes based on nonlinear encoding functions were proposed in [7], [8], [9], [10], [11]. A code $C \in GF(2^n)$ is robust if $\{e|c \oplus e \in C, \forall c \in C\} = \{\mathbf{0} \in GF(2^n)\}$. These codes can provide nearly equal protection against all error patterns [7], [8]. The error masking probabilities $Q_C(e) = |C|^{-1}|\{c \in C, c \oplus e \in C\}|$ ($|C|$ is the size of the code) for robust codes are upper-bounded by a number less than 1 for all non-zero errors. Compared to the systems based on linear codes, systems based on robust codes can provide a guaranteed protection regardless of the accuracy of the error model. Variants of robust codes – partially robust and minimum distance robust codes – were proposed in [10], [11], which allow tradeoffs in terms of the robustness and the hardware overhead.

One limitation of robust codes is that these codes assume the information bits of messages or outputs of the device-to-be-protected are uniformly distributed and are not controllable by external forces, e.g. by an attacker during error injection attacks on devices. The reliability and the security of the communication or computation channels protected by robust codes will be largely compromised if both information bits of the messages and the non-zero error patterns can be controlled by the attacker.

*Example 1.1:* Suppose the 32-bit device is protected by a robust duplication code $C = \{y, f(y)\}$, where $y, f(y) \in GF(2^{32})$, $f(y) = y^3$ and all operations are in $GF(2^{32})$. It is easy to prove that any non-zero error $e$ will be masked by at most two codewords [7], i.e. for any non-zero error $e = (e_y, e_f)$ there exist at most two vectors $y_1, y_2 \in GF(2^{32})$ such that $(y_1 \oplus e_y)^3 = y_1^3 \oplus e_f$ and $(y_2 \oplus e_y)^3 = y_2^3 \oplus e_f$. Assume that an attacker cannot control the fault-free outputs $y$ during attacks and the outputs of the original device are uniformly distributed, then the probability that the attacker conducts a successful attack ($e = (e_y, e_f)$ is not detected) is at most $2^{-31}$. If an attacker has the ability to control the inputs of the device (hence the fault-free outputs) and can inject arbitrary error patterns at the output, let $(v, y)$ be an input-output pair, i.e. $y$ is the output of the device when the input to the device is $v$. Then the attacker can easily derive an error pattern $e^* = (e_y^*, e_f^*)$, $e_y^*, e_f^* \in GF(2^{32}), e_y^* \neq \mathbf{0}$ that will be masked by $y$, i.e. $(y \oplus e_y^*)^3 \oplus y^3 \oplus e_y^* = \mathbf{0}$. During the attack, the attacker can simply input $v$ to the device and inject the corresponding $e^* = (e_y^*, e_f^*)$ at the output of the device. In this case, the attack will always be successful.

For the situation shown in the above example, all previous protection technologies based on traditional error detecting codes will not be sufficient. A coding technique based on adding to $k$ information bits $m$ random bits and $r$ redundant bits, which can still provide guaranteed reliability and security under the above circumstance, is called **algebraic manipulation detection (AMD) code** . (The formal definition of AMD codes will be given in the next Section, see Definition 2.2). A simple AMD code was first presented in [12]. A much more versatile strong AMD code was introduced in

The first author now works for Mediatek Wireless, Inc. The work of the second author is sponsored by the NSF grant CNR 1012910.

[13], where the construction of optimal AMD codes was presented for $k = br$ information digits and $m = r$ random digits ($r$ is the number of redundant digits). In [14], the authors introduced the concept of AMD codes and put all previous constructions in a unified framework. Compared to the widely used Message Authentication Codes, AMD codes do not require a secret key and have simpler encoding and decoding. Codes combining AMD codes and list-decoding are described in [15]. Applications of AMD codes for the design of non-malleable codes are presented in [16].

The main contributions of this paper are as follows. We present lower bounds for the probability of error masking for systematic AMD codes (Section II) and present several new constructions of systematic AMD codes (Section III), which are generalizations of the construction shown in [14]. Some of the presented codes are optimal or close to be optimal. We showed the relationship between AMD codes and classical codes such as the Generalized Reed-Muller codes and the Reed-Solomon codes (Section II and III). We also describe in Section IV an efficient encoding and decoding algorithm for the presented codes based on the multivariate Horner scheme.

The proposed codes can be used for many different applications such as robust secret sharing schemes, robust fuzzy extractors [14] and secure cryptographic devices resistant to fault injection attacks. All the codes described in this paper are binary. Generalization to a nonbinary case is straightforward.

## II. Definitions and Bounds for Algebraic Manipulation Detection Codes

Throughout the paper we denote by $\oplus$ the addition in $GF(q), q = 2^r$. All the results presented in the paper can be easily generalized to the case where $q = p^r$ ($p$ is a prime). Due to the lack of space, proofs for corollaries are omitted.

A code $V$ with codewords $(y, x, f(y, x))$, where $y \in GF(2^k), x \in GF(2^m)$ and $f(y, x) \in GF(2^r)$, will be referred to as a $(k, m, r)$ code. We will assume that $y$ is a $k$-bit information, $x$ is an $m$-bit uniformly distributed random vector (generated by a random number generator) and $f(y, x)$ is an $r$-bit redundant portion of the message $(y, x, f(y, x))$. The general architecture of using AMD codes for the protection of computation channels is shown in Figure 1.

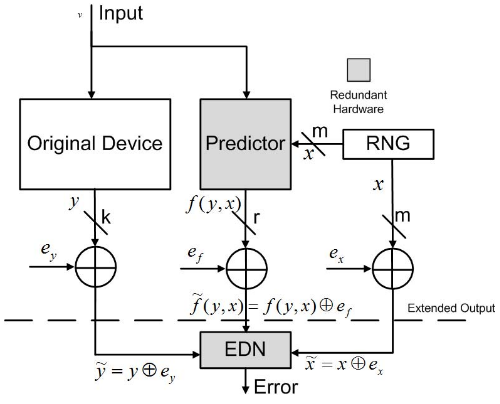


Fig. 1. Computation channel protected by a systematic $(k, m, r)$ AMD code.

*Definition 2.1:* (**Security Kernel**) For any $(k, m, r)$ error detecting code $V$ with the encoding function $f(y, x)$, where $y \in GF(2^k), x \in GF(2^m)$ and $f(y, x) \in GF(2^r)$, the **security kernel** $K_S$ is the set of errors $e = (e_y, e_x, e_f), e_y \in GF(2^k), e_x \in GF(2^m), e_f \in GF(2^r)$, for which there exists $y$ such that $f(y \oplus e_y, x \oplus e_x) \oplus f(y, x) = e_f$ is satisfied for all $x$.

$$K_S = \{e | \exists y, f(y \oplus e_y, x \oplus e_x) \oplus f(y, x) \oplus e_f = \mathbf{0}, \forall x\}. \quad (1)$$

We note that in many applications $e_y \neq 0$ is a necessary condition for an attacker to conduct a successful fault injection attack. However, for secure architectures such as the one shown in [9], [17], the integrity of not only the information bits but also the redundant bits of the codes can be critical. Thereby, to conduct a more general analysis, we do not impose $e_y \neq 0$ in the above definition of the security kernel.

Non-zero errors $e$ in the security kernel can be used by an advanced attacker to bypass the protection based on the error detecting code. For the case of communication channels we assume that an attacker can select any $k$-bit vector $y$ as the information bits of a message $(y, x, f(y, x))$ and any error $e = (e_y, e_x, e_f)$ that distorts the message. For the case of computation channels (Figure 1), we assume the attacker can inject faults that manifest as $e \in K_S$ at the output of the device and select $y$ for which $e$ is always masked. Under the above attacker model for communication or computation channels, the attacker can always mount a successful attack. Thereby an AMD code that can provide a guaranteed error detecting probability under the above strong attacker model should have no errors in the security kernel except for the all zero vector in $GF(2^n)$, where $n = k + m + r$ is the length of the code.

*Definition 2.2:* A $(k, m, r)$ error detecting code is called Algebraic Manipulation Detection (AMD) code iff $K_S = \{\mathbf{0}\}$, where $\mathbf{0}$ is the all zero vector in $GF(2^n)$, $n = k + m + r$.

*Remark 2.1:* The original definition of AMD codes in [14] is for both systematic and nonsystematic codes defined in any group. In this paper we consider binary systematic AMD codes, which is the most practical for hardware implementation. The above definition and all other results in this paper can be easily generalized for non-binary cases.

AMD codes $V = \{(y, x, f(y, x))\}$ have no undetectable errors no matter how the attacker select $e = (e_y, e_x, e_f)$ and $y$. AMD codes for the case $m = r$ and $k = br$ were introduced in [13] and were used in [14] for robust secret sharing schemes and for robust fuzzy extractors.

For a $(k, m, r)$ code $V$, denote by $Q_V(y, e)$ the probability of missing an error $e$ once $y$ is fixed. Then $Q_V(y, e)$ can be computed as the fraction of random vectors $x$ such that $e$ is masked (see (2)) and $K_S = \{e | \exists y : Q_V(y, e) = 1\}$. The code $V$ is an AMD code if and only if $Q_V(y, e) < 1$ for any $y$ and any $e \neq 0$.

$$\begin{aligned} Q_V(y, e) &= 2^{-m} |\{x \,|\, (y, x, f(y, x)) \in V, \\ &\quad (y \oplus e_y, x \oplus e_x, f(y, x) \oplus e_f) \in V\}|. \quad (2) \end{aligned}$$

For a $(k, m, r)$ AMD code $V = \{(y, x, f(y, x)), y \in GF(2^k), x \in GF(2^m), f(y, x) \in GF(2^r)\}$, for any given $y^* \in GF(2^k)$ and $e^* = (e_y^*, e_x^*, e_f^*)$, $e_y^* \in GF(2^k), e_x^* \in GF(2^m)$,

$e_f^* \in GF(2^r)$, $f(y^* \oplus e_y^*, x \oplus e_x^*) \oplus e_f^*$ considered as functions of $x \in GF(2^m)$ should all be different.

*Example 2.1:* Let $k = m = tr$, $y = (y_0, y_1, \cdots, y_{t-1})$, $y_i \in GF(2^r)$ be the information digits and $x = (x_0, x_1, \cdots, x_{t-1}), x_i \in GF(2^r)$ be the random digits. Let $f(y, x) = x_0 \cdot y_0 \oplus x_1 \cdot y_1 \oplus \cdots \oplus x_{t-1} \cdot y_{t-1}$ be the encoding function, where all the operations are in $GF(2^r)$.

It is easy to verify that when $e_y = \mathbf{0}$, for any $e_x$ and $e_f$ ($e_x, e_f$ are not both $\mathbf{0}$), there always exist $y$ such that $e = (\mathbf{0}, e_x, e_f), e \neq \mathbf{0}$ will be masked for all $x$. Thereby, this code is not a AMD code. In this case, $K_S$ contains all vectors $e = (\mathbf{0}, e_x, e_f)$.

Suppose $e_y = (e_{y_0}, e_{y_1}, \cdots, e_{y_{t-1}}), e_{y_i} \in GF(2^r)$ is always non-zero. Without loss of generality, let us assume $e_{y_0} \neq \mathbf{0}$. Then the monomial $e_{y_0} \cdot x_0$ will appear in the error masking equation $f(x \oplus e_x, y \oplus e_y) \oplus f(y, x) \oplus e_f = \mathbf{0}$. Since $e_{y_0} \neq \mathbf{0}$, for every $e, y$ and $x_1, x_2, \cdots, x_{t-1}$, there is a unique solution for $x_0$. Thereby the error is masked with probability $2^{-r}$.

Let $C$ be a $q$-ary code ($q = 2^r$) of length $2^m$ with an encoding function $f : GF(2^m) \rightarrow GF(2^r)$. Let us define the **orbit of** $f$ by (3). We note that for any $f \in C$, $1 \leq |Orb(f)| \leq q2^m = 2^{m+r}$. If $|Orb(f)| = 2^{m+r}$, then for any $e_x$ and $e_f$ there exists $x$ such that $f(x) \neq f(x \oplus e_x) \oplus e_f$. Moreover, if $\varphi \notin Orb(f)$, then $Orb(\varphi) \bigcap Orb(f) = \emptyset$.

$$Orb(f) = \{\varphi | \varphi(x) = f(x \oplus e_x) \oplus e_f\}, \tag{3}$$

where $e_x \in GF(2^m), e_f \in GF(2^r)$.

*Definition 2.3:* We will say that a $q$-ary ($q = 2^r$) code $C$ of length $2^m$ is a **code with full orbit** if for any $f \in C$, $|Orb(f)| = 2^{m+r}$ and $Orb(f) \subseteq C$.

The notion of codes with full orbit will be used in the lower bound for the probability of error masking (see Theorem 2.1).

Any $q$-ary code $C$ of length $2^m$ with full orbit is a union of disjoint orbits of size $q2^m$. The size of $C$ is a multiple of $q2^m$. We note that codes with full orbit are nonlinear and for any code $C$ with full orbit, $\mathbf{0} \in GF(2^m)$ is not a codeword of $C$.

*Example 2.2:* Let $C$ be a binary code of length 8 and Hamming distance 2 containing all vectors with an odd number of $1's$. Let $y = (y_0, y_1, y_2), y_i \in GF(2)$ and $f_y(x) = y_0 \cdot x_0 \oplus y_1 \cdot x_1 \oplus y_2 \cdot x_2 \oplus x_0 \cdot x_1 \cdot x_2$. It is easy to verify that for any $y \in GF(2^3)$, $|Orb(f_y)| = 16$. Thus $C$ is a code with full orbit and $|C| = |\cup_{y \in GF(2^3)} Orb(f_y)| = 128$.

The optimal AMD code should minimize $\max_{y, e \neq 0} Q_V(y, e)$ among all codes with the same parameters. Thus, the criterion we use to construct good AMD codes is

$$\min_{V \in V_{k,m,r}} \max_{y, e \neq 0} Q_V(y, e), \tag{4}$$

where $V_{k,m,r}$ is the set of all $(k, m, r)$ error detecting codes.

We note that the optimization criterion selected in this paper is different from the one shown in [14]. The computational complexity of the encoding function for AMD codes is determined by both $m$ and $r$. In cryptographic applications, the $m$ random digits can be generated by a random number generator (RNG), which is already integrated in most of the modern cryptographic devices. Since the RNG is also used for other purposes such as generating the random mask for countermeasures against power analysis attacks, the number of random digits available for AMD codes in every clock cycle may be limited. The above criterion was selected to maximize the security level of the cryptographic device given the number of available random digits in every clock cycle and the amount of hardware redundancy we can bear.

Let $Q_V = \max_{y, e \neq 0} Q_V(y, e)$ and $Q(k, m, r) = \min_{V \in V_{k,m,r}} Q_V$. Denote by $\hat{d}_q(2^m, M)$ the maximum Hamming distance of a $q$-ary ($q = 2^r$) code of length $2^m$ with full orbit containing $M$ codewords. Obviously,

$$\hat{d}_q(2^m, M) \leq d_q(2^m, M), \tag{5}$$

where $d_q(2^m, M)$ is the maximum possible Hamming distance of a $q$-ary code with length $2^m$ and $M$ codewords.

We next present a lower bound for $Q(k, m, r)$. The constructions of codes providing tight upper bounds for $Q(k, m, r)$ can be found in Section III.

*Theorem 2.1:* For any $(k, m, r)$ AMD code, where $k$ is the number of information bits, $m$ is the number of random bits and $r$ is the number of redundant bits,

$$\begin{aligned} Q(k, m, r) &= \min_{V \in V_{k,m,r}} \max_{y, e \neq 0} Q_V(y, e) \\ &\geq 1 - 2^{-m} d_q(2^m, M), \end{aligned} \tag{6}$$

where $d_q(2^m, M)$ is the maximum possible Hamming distance of a (not necessarily systematic) $q$-ary code $C$ ($q = 2^r$) with length $2^m$ and $M = |C| = 2^{k+m+r}$ codewords.

*Proof:* Let $V$ be a $(k, m, r)$ AMD code composed of vectors $(y, x, f(y, x))$, where $y \in GF(2^k), x \in GF(2^m)$ and $f(y, x) \in GF(2^r)$. When $y$ is fixed, $f$ is a function of $x$. Let us denote this function by $f_y$. Since $V$ is an AMD code, $f_y(x \oplus e_x) \oplus e_f$ is not the same as $f_{y'}(x \oplus e'_x) \oplus e'_f$ for any $y, y', e_x, e'_x, e_f, e'_f$, assuming that elements of at least one of the pairs $(y, y')$, $(e_x, e'_x)$ and $(e_f, e'_f)$ are not equal. Thereby, for different $y$, $e_x$ and $e_f$, $f_y(x \oplus e_x) \oplus e_f$ corresponds to $2^{k+m+r}$ different functions.

Let $C_V = \cup_{y \in GF(2^k)} Orb(f_y)$ be a $q$-ary ($q = 2^r$) code of length $2^m$ with full orbit. Then $|Orb(f_y)| = 2^{m+r}$, $|C| = 2^{k+m+r}$ and $Q_V = max_{y, e \neq 0} Q(y, e) = 1 - 2^{-m} d(C_V)$, where $d(C_V)$ is the Hamming distance of $C_V$. By (5) and (6) we have

$$\begin{aligned} Q(k, m, r) &= 1 - 2^{-m} \max_{V \in V_{k,m,r}} d(C_V) \\ &\geq 1 - 2^{-m} \hat{d}_q(2^m, M) \\ &\geq 1 - 2^{-m} d_q(2^m, M). \end{aligned} \tag{7}$$

■

The following Corollary follows directly from Theorem 2.1.

*Corollary 2.1:* There is no AMD codes $V$ with $k > r2^m - m - r$. ($Q(k, m, r) = 1$ if $k > r2^m - m - r$.)

*Remark 2.2:* We note that the bound in Theorem 2.1 is much stronger than the trivial bound $Q(k, m, r) \geq 2^{-r}$. In fact, $Q(k, m, r) \geq 2^{-r}$ is equivalent to $d_q(2^m, 2^k + m + r) \geq 2^m - 2^{m-r}$, which is a sub-case of Theorem 2.1.

Theorem 2.1 shows the relationship between the worst case error masking probability $Q_V$ for an AMD code $V$ and the Hamming distance of the corresponding code $C_V$ with full orbit. The exact value of $\hat{d}_q(2^m, M)$ is hard to derive. However, the Hamming distance of $C_V$ should not exceed the maximum possible distance for a $q$-ary code with length $2^m$ and $2^{k+m+r}$ codewords, $q = 2^r$. We note that $d_q(2^m, M)$ can be estimated by classical bounds from coding theory such as the Hamming bound, the Johnson bound, the Singleton bound, the Plotkin bound, etc [18].

When $d_q(2^m, M)$ is estimated by the Singleton bound, the lower bound for $Q(k, m, r)$ can be written in a compact form as it is shown in the following Corollary.

*Corollary 2.2:* For any $(k, m, r)$ AMD code,

$$Q(k,m,r) \geq \lceil \frac{k+m}{r} \rceil 2^{-m}. \tag{8}$$

*Example 2.3:* Let $k = m = 3$ and $r = 1$. According to (8), $Q(3, 3, 1) \geq \frac{6}{8}$. Let $V$ be the code composed of all vectors $(y, x, f(y, x))$, where $y, x \in GF(2^3)$ and

$$f(y,x) = x_0{\cdot}x_1{\cdot}x_2 \oplus x_0{\cdot}y_0 \oplus x_1{\cdot}y_1 \oplus x_2{\cdot}y_2, f(y,x) \in GF(2). \tag{9}$$

The error masking equation is $f(x \oplus e_x, y \oplus e_y) \oplus f(y, x) = e_f$, which is a polynomial of $x$ with degree 2. The function on the left hand side of the error masking equation corresponds to a codeword of the second order binary Reed-Muller code $RM_2(2, 3)$ with 3 variables [18]. Any codeword of $RM_2(2, 3)$ has a Hamming weight of at least 2. Thus the number of solutions for the error masking equation is upper bounded by 6. $V$ is a AMD code with $Q_V = \frac{6}{8}$. It follows from (8) that this code is optimal and $Q(3, 3, 1) = 0.75$.

Optimal $(k, m, r)$ AMD codes attain the equality in (6) and minimize the worst case error masking probability among all codes with the same parameters.

In the next section, we will present several general constructions of AMD codes. Some of the generated codes are optimal with respect to the lower bounds (6) or (8).

## III. Constructions of AMD Codes

The codewords of a $(k, m, r)$ AMD code $V$ are in the format $(y, x, f(y, x))$, where $y \in GF(2^k), x \in GF(2^m)$ and $f(y, x) \in GF(2^r)$. When $y$ is fixed, $f_y$ is a function of $x$. In the proof of Theorem 2.1, we have shown that the necessary condition for $V$ to be an AMD code is that $f_y(x \oplus e_x) \oplus e_f$ cannot be the same function as $f_{y'}(x \oplus e'_x) \oplus e'_f$ for any $y, y', e_x, e'_x, e_f, e'_f$, assuming elements in at least one of the pairs $(y, y')$, $(e_x, e'_x)$ and $(e_f, e'_f)$ are not equal.

To compute $Q_V(y, e)$ for the code $V$, the error masking equation $f(x \oplus e_x, y \oplus e_y) \oplus f(y, x) \oplus e_f = \mathbf{0}$ should be evaluated for all $2^m$ possible $x \in GF(2^m)$.

We will say that an AMD code $V = \{(y, x, f(y, x))\}$ is based on code $C$ if the error masking polynomial $f(y \oplus e_y, x \oplus e_x) \oplus f(y, x) \oplus e_f$ is a codeword of $C$ for all $y, e_x, e_y$ and $e_f$. Let us re-write $f(y, x)$ as $f(y, x) = A(x) \oplus B(y, x)$, where $A(x)$ is independent of $y$. We next show that by selecting $A(x)$ and $B(y, x)$ based on different error detecting codes such as the Generalized Reed-Muller codes and the Reed-Solomon codes, we can construct good (and in many cases optimal) AMD codes for different $k$ and different $Q_V = \max_{y,e \neq 0} Q_V(y, e)$ for given $m$ and $r$.

### *A. Constructions of AMD codes Based on the Generalized Reed-Muller Codes*

Let $x = (x_0, x_1, \cdots, x_{t-1}), x_i \in GF(q)$, $q = 2^r$. A $b^{th}$ order $q$-ary Generalized Reed-Muller code $GRM_q(b, t)$ [19] with $t$ variables $(1 \leq b \leq t(q-1))$ consists of all codewords $(f(\mathbf{0}), f(\gamma^0), \cdots, f(\gamma^{q^t-2}))$, where $f(x)$ is a polynomial of $t$ variables $x_0, x_1, \cdots x_{t-1}$ and $\gamma$ is a primitive element of $GF(q^t)$. The degree of $f(x)$ is less or equal to $b$.

It is shown in [19] that the dimension of $GRM_q(b, t)$ is

$$k_{GRM_q(b,t)} = \sum_{j=0}^{t} (-1)^j \binom{t}{j} \binom{t+b-jq}{b-jq}, \tag{10}$$

where $q = 2^r$, $\binom{i}{j} = 0$ when $j < 0$.. If $b = u(q-1) + v, 0 \leq v \leq q-2$. Then the distance of $GRM_q(b, t)$ is $d_{GRM_q(b,t)} = (q-v)q^{t-u-1}$ [19].

Suppose $b + 2 = \alpha(q-1) + \beta \leq t(q-1), 0 \leq \alpha \leq t, 0 \leq \beta \leq q-2$. Assume that $b$ is odd when $t = 1$. Let

$$A(x) = \begin{cases} \bigoplus_{i=0}^{t-1} x_i^{b+2} & \text{if } \alpha = 0, b \text{ is odd;} \\ \bigoplus_{i=1}^{t-1} x_0 x_i^{b+1}, t > 1 & \text{if } \alpha = 0, b \text{ is even;} \\ \bigoplus_{i=0}^{t-1} x_i^{\beta} \prod_{j=1}^{\alpha} x_{|i+j|_t}^{q-1} & \text{if } \alpha \neq 0, \alpha \neq t; \\ \prod_{i=0}^{\alpha-1} x_i^{q-1} & \text{if } \alpha = t; \end{cases} \tag{11}$$

where $x_i \in GF(2^r)$, $|i+j|_t$ is the modulo $t$ addition, $\bigoplus$ is the sum in $GF(2^r)$.

Let

$$B(y,x) = \bigoplus_{1 \leq j_0+j_1+\cdots+j_{t-1} \leq b+1} y_{j_0,j_1,\cdots,j_{t-1}} \prod_{i=0}^{t-1} x_i^{j_i}, \tag{12}$$

where $0 \leq j_i \leq q-1$, $y_{j_0,j_1,\cdots,j_{t-1}} \in GF(2^r), x_i \in GF(2^r)$, $\prod_{i=0}^{t-1} x_i^{j_i}$ is a monomial of $x_0, x_1, \cdots, x_{t-1}$ of a degree between 1 and $b+1$ and $\prod_{i=0}^{t-1} x_i^{j_i} \notin \Delta B(y, x)$, where $\Delta B(y, x)$ is defined by (13).

We note it follows from (13) that when $\alpha = t$, $\Delta B(y, x) = \{x_i^{q-2} \prod_{j \neq i} x_j^{q-1}, 0 \leq i \leq t-1\}$.

*Example 3.1:* Let $r = 3, q = 8, t = 2$ and $b = 10$. Since $b + 2 = 12 = \alpha(q-1) + \beta$, we have $\alpha = 1$ and $\beta = 5$. By (11) and (13), we have $A(x) = x_0^5 x_1^7 \oplus x_0^7 x_1^5$ and $\Delta B(y, x) = \{x_0^5 x_1^6, x_0^6 x_1^5\}$. It is easy to verify that $A(x \oplus e_x) \oplus A(x) \oplus B(y \oplus e_y, x \oplus e_x) \oplus B(y, x)$ is always a non-zero polynomial corresponding to a codeword in $GRM_8(11, 2)$.

AMD codes can be constructed based on $A(x), B(y, x)$ and the Generalized Reed-Muller codes as shown in the next Theorem.

*Theorem 3.1:* Let $f(y, x) = A(x) \oplus B(y, x)$ be a $q$-ary polynomial with $y_{j_0,j_1,\cdots,j_{t-1}} \in GF(q)$ as coefficients and $x \in GF(q^t)$ as variables, where $1 \leq b \leq t(q-1) - 2, q = 2^r$ and $A(x), B(y, x)$ are as shown above. Suppose $b+2 = \alpha(q-1)+\beta$ and $b+1 = u(q-1)+v$, $0 \leq \alpha, u \leq t, 0 \leq \beta, v \leq q-2$. Assume $b + 2 \neq t(q-1) - 1$ and $b$ is odd when $t = 1$. Then the code $V$ composed of all vectors $(y, x, f(y, x))$ is an AMD

$$\Delta B(y,x) = \begin{cases} \{x_0^{b+1}, x_1^{b+1}, \cdots, x_{t-1}^{b+1}\} & \text{if } \alpha = 0, b \text{ is odd;} \\ \{x_1^{b+1}, x_0x_1^b, x_0x_2^b, \cdots, x_0x_{t-1}^b, t > 1\} & \text{if } \alpha = 0, b \text{ is even;} \\ \{x_i^{\beta} x_{|i+1|_t}^{q-2} \prod_{j=2}^{\alpha} x_{|i+j|_t}^{q-1}, 0 \le i \le t-1\} & \text{if } \alpha \neq 0. \end{cases} \quad (13)$$

code with $m = tr$,

$$\begin{aligned} k &= (k_{GRM_q(b+1,t)} - t - 1)r \\ &= (\sum_{i=0}^{t}(-1)^i \binom{t}{i}\binom{t+b+1-iq}{b+1-iq} - 1 - t)r, \quad (14) \end{aligned}$$

and

$$\begin{aligned} Q_V &= 1 - d_{GRM_q(b+1,t)}2^{-m} \\ &= 1 - (2^r - v)2^{-(u+1)r}. \quad (15) \end{aligned}$$

Thus

$$\begin{aligned} &Q((\sum_{i=0}^{t}(-1)^i \binom{t}{i}\binom{t+b+1-iq}{b+1-iq} - 1 - t)r, tr, r) \\ &\le 1 - (2^r - v)2^{-(u+1)r}. \quad (16) \end{aligned}$$

*Proof:* An error $e$ is masked by $V$ if and only if for all $x$

$$A(x \oplus e_x) \oplus A(x) \oplus B(y \oplus e_y, x \oplus e_x) \oplus B(y,x) \oplus e_f = \mathbf{0}. \quad (17)$$

1) If $e_x = \mathbf{0}$ and $e_y = \mathbf{0}$, the error is always detected unless $e_f$ is also $\mathbf{0}$. If $e_x = \mathbf{0}$ and $e_y \neq \mathbf{0}$, the left hand side of the error masking equation (17) is a polynomial of degree from 1 to $b+1$, which corresponds to a codeword of a $(b+1)^{th}$ order $q$-ary Generalized Reed-Muller code. Since $d_{GRM_q(b+1,t)} = (q-v)q^{t-u-1}$, there are at most $q^t - (q-v)q^{t-u-1}$ solutions for the error masking equation.
2) If $e_x \neq \mathbf{0}$, the left hand side of (17) does not contain any monomials of degree $b+2$ due to the fact that $A(x)$ and $A(x \oplus e_x)$ have exactly the same monomials of degree $b+2$. Moreover,
    a) If $\alpha = 0$ and $b$ is odd, $x_i^{b+1}$ appears in (17) iff $x_i$ is distorted, $0 \le i \le t-1$;
    b) If $\alpha = 0$ and $b$ is even, $x_1^{b+1}$ appears in (17) iff $x_0$ is distorted, $x_0x_i^b$ appears in (17) iff $x_i$ is distorted $1 \le i \le t-1$;
    c) If $\alpha \neq 0$, since $b + 2 \neq t(q-1) - 1$, $x_i^{\beta} x_{|i+1|_t}^{q-2} \prod_{j=2}^{\alpha} x_{|i+j|_t}^{q-1}$ appears in (17) iff $x_{|i+1|_t}$ is distorted, $0 \le i \le t-1$. (When $\alpha = t$, ${x_i}^{q-2} \prod_{j \neq i} x_j^{q-1}$ appears in (17) if $x_i$ is distorted.)

    Thereby, (17) always contains monomials of degree $b+1$, the left hand side of the error masking equation again is a codeword in $GRM_q(b+1,t)$. Thus the number of solutions for the error masking equation is still upper bounded by $q^t - (q-v)q^{t-u-1}$.

Thus for any fixed $y$ and $e$, the probability $Q_V$ of error masking is upper bounded by

$$(q^t - (q-v)q^{t-u-1})q^{-t} = 1 - (2^r - v)2^{-(u+1)r}.$$

The left hand side of (17) contains monomials of a degree from 1 to $b+1$ except for the $t$ monomials from $\Delta B(y,x)$. Hence the number of different monomials in $B(y,x)$ is

$$k_{GRM_q(b+1,t)} - 1 - t = \sum_{i=0}^{t}(-1)^i \binom{t}{i}\binom{t+b+1-iq}{b+1-iq} - 1 - t. \quad (18)$$

The number, $k$, of bits in $y$ is equal to the number of monomials in $B(y,x)$ multiplied by $r$, which is

$$(\sum_{i=0}^{t}(-1)^i \binom{t}{i}\binom{t+b+1-iq}{b+1-iq} - 1 - t)r. \quad (19)$$

■

*Example 3.1 (Continued)* For the code shown in Example 3.1, $k = 55 \times 3 = 165$. Since $b = 10 = u(q-1) + v, q = 8$, we have $u = 1$ and $v = 3$. The worst case error masking probability is $Q_V = 1 - 5 \times 2^{-6}$. Thus by (8), $1 - 7 \times 2^{-6} \le Q_V(165, 6, 3) \le 1 - 5 \times 2^{-6}$.

*Corollary 3.1:* When $b = t(q-1) - 2, q = 2^r$, codes generated by Theorem 3.1 are optimal. We have

$$Q(2^{tr}r - tr - 2r, tr, r) = 1 - 2^{-tr+1}. \quad (20)$$

*1) Special Case: $r = 1$:* For this case the dimension of a $(b+1)^{th}$ order binary Reed-Muller code of $t$ variables is $k_{RM_2(b+1,t)} = \sum_{i=0}^{b+1}\binom{t}{i}$ $(t = m)$ [18]. The distance of $RM_2(b+1,t)$ is $d_{RM_2(b+1,t)} = 2^{t-b-1}$. As a result, the dimension of the resulting AMD code $V$ constructed by Theorem 3.1 is $k = \sum_{i=0}^{b+1}\binom{t}{i} - t - 1$. The worst case error masking probability of the code is $Q_V = 1 - 2^{-(b+1)}$.

*Corollary 3.2:* When $q = 2$, the code $V$ generated by Theorem 3.1 is a $(\sum_{i=0}^{b+1}\binom{t}{i} - t - 1, t, 1)$ AMD code with $Q_V = 1 - 2^{-(b+1)}$. The code is optimal when $b = 1$ or $b = t - 2$.

*Example 3.2:* Suppose $m = 7$ and $r = 1$. Let $b = 1$ and

$$f(y,x) = x_0 \cdot x_1 \cdot x_2 \oplus x_3 \cdot x_4 \cdot x_5 \oplus x_0 \cdot x_3 \cdot x_6 \oplus \sum_{i=0}^{6} x_i \cdot y_i. \quad (21)$$

It is easy to verify that $f(y \oplus e_y, x \oplus e_x) \oplus f(y,x) \oplus e_f$ is a polynomial of degree 2, which is a codeword of $RM_2(2,7)$. The distance of $RM_2(2,7)$ is 32. The worst case error masking probability of the resulting AMD code is $Q_V = 0.75$. This code is optimal.

*Remark 3.1:* When $q = 2$ and $b = 1$, the code $V$ generated by Theorem 3.1 is an optimal AMD code with $k = \binom{t}{2}$ and $Q = 0.75$. Obviously, for all $k < \binom{t}{2}$, optimal AMD code with the same $m$ and $r = 1$ can be constructed by deleting some codewords from $V$. The worst case error masking probability for the new code is still 0.75.

*2) Special Case: $b \le q - 3$:* Another special case of Theorem 3.1 is the case $b \le q-3$. In this case $k_{GRM_q(b+1,t)} = \binom{t+b+1}{t}$ and $d_{GRM_q(b+1,t)} = (q-b-1)q^{t-1}$ [19]. The dimension of the resulting AMD code is $(\binom{t+b+1}{t} - 1 - t)r$.

The worst case error masking probability is $(b+1)2^{-r}$.

*Corollary 3.3:* Assume $b$ is odd when $t = 1$. When $b \leq q-3$, the code $V$ generated by Theorem 3.1 is a $(((\binom{t+b+1}{t}) - 1 - t)r, tr, r)$ AMD code with $Q_V = (b+1)2^{-r}$.

When $b = 1$ $B(y,x)$ is the quadratic form $x_0 \cdot y_0 \oplus x_1 \cdot y_1 \oplus \cdots \oplus x_{t-1} \cdot y_{t-1}$, where all the operations are in $GF(2^r)$. If $e_y \neq 0$, it is easy to verify that the number of codewords masking the error is upper bounded by $q^{t-1}$.

*3) Special Case:* $t = 1$ *[13], [14]:* When $t = 1$ and $b$ is odd, $A(x) = x^{b+2}$ and $B(y,x) = x \cdot y_0 \oplus x^2 \cdot y_1 \oplus \cdots \oplus x^b \cdot y_{b-1}$. The code generated by Theorem 3.1 coincides with the construction shown in [13], [14]. For this code, $k \leq r(q-3) = r(2^r - 3)$.

*Corollary 3.4:* [13], [14] When $b \leq q-3$ is an odd number, the code $V$ composed of all vectors $(y, x, f(y,x))$, where $y \in GF(q^{bt}), x \in GF(q), q = 2^r$ and $f(y,x) = x^{b+2} \oplus x \cdot y_0 \oplus x^2 \cdot y_1 \oplus \cdots x^b \cdot y_{b-1}, f(y,x) \in GF(q)$, is an optimal $(br, r, r)$ AMD code with $Q_V = \max_{y,e \neq 0} Q_V(y,e) = (b+1)2^{-r}$. Thereby, $Q(br, r, r) = (b+1)2^{-r}$.

*Remark 3.2:* One limitation of Corollary 3.4 is that $b$ can only be an odd number when the characteristic of the field $GF(q)$ is 2. Otherwise, $A(x \oplus e_x)$ for $A(x) = x^{b+2}$ and $e_x \neq \mathbf{0}$ does not contain any monomial of degree $b+1$. The resulting code is not a secure AMD code as pointed out in [14]. When $b$ is even, $A(x)$ can be chosen as $x^{b+3}$. In this case, $Q_V = (b+2)2^{-r}$.

*Remark 3.3:* When $t = 1$, the left hand side of the error masking equation $f(y \oplus e_y, x \oplus e_x) \oplus f(y,x) \oplus e_f = \mathbf{0}$ is a codeword of an extended $q$-ary $(q, b+2, q-b-1)$ Reed-Solomon code, $q = 2^r$ [18].

When $t > 1$, codes $V$ generated by Theorem 3.1 may have larger number of codewords than codes generated by Corollary 3.4 ($t = 1$), assuming the two codes have the same $Q_V$ and the same $r$.

*Example 3.3:* Suppose $r = 16$, $Q_V = 2^{-14}$. Then for $t = 1$ and $b = 3$, for codes generated by Corollary 3.4, the maximum number of codewords is $2^{br} = 2^{48}$. When $t > 1$, the maximum number of codewords for codes generated by Theorem 3.1 depends not only on $b$ but also on $t$. When $t = 2$, for example, the number of codewords of codes generated by Theorem 3.1 can be $2^{((\binom{t+b+1}{t})-1-t)r} = 2^{192}$.

To end the section, we summarize cases the when codes constructed by Theorem 3.1 are optimal in the Table I.

## IV. Encoding and Decoding Complexity for AMD Codes

In this section, we estimate the hardware complexity for the encoders and decoders for AMD codes based on $q$-ary Generalized Reed-Muller codes (Theorem 3.1). The hardware complexity for the encoders and decoders for AMD codes based on the product of GRM codes can be estimated in a similar way.

It is well known that a multivariate polynomial of $t$ variables $x_i, 0 \leq i \leq t-1, x_i \in GF(2^r)$ can be efficiently computed using the multivariate Horner scheme [20]. When $t = 1$, any polynomial of degree $b+1$ defined over $GF(2^r)$ can be represented as

$$f(x) = a_0 \oplus x(a_1 \oplus x(\cdots(a_b + a_{b+1}x))),$$

where $a_i \in GF(2^r), x \in GF(2^r)$. The computation of the polynomial requires $b+1$ multipliers and $b+1$ adders in $GF(2^r)$.

When $t > 1$, we can first apply Horner scheme as if $x_0$ is the variable and $x_1, x_2, \cdots, x_{t-1}$ are coefficients. In this case coefficients will be polynomials of $t-1$ variables $x_1, x_2, \cdots, x_{t-1}$. To compute these polynomials, we can select one of the remaining $x_i, 1 \leq i \leq t-1$ as variable and apply the Horner scheme again. We repeat the procedure until all $x_i, 0 \leq i \leq t-1$ are factored out.

*Example 4.1:* In Theorem 3.1, let $t = b = 2$ and assume $r$ is large enough. Then the resulting code is a $(7r, 2r, r)$ AMD code. At most 8 multipliers and 7 adders in $GF(2^r)$ are required for the encoding or the decoding. The corresponding encoding network is shown in Figure 2. The critical path of the encoder contains 4 multipliers and 4 adders in $GF(2^r)$.

It is easy to verify that for the encoder of a $(k, m, r)$ AMD code generated by Theorem 3.1 using the multivariate Horner scheme shown above, the number of multipliers and adders is upper bounded by $\lceil \frac{k+m}{r} \rceil$. The latency of the encoder will be $(b+1)(T_M + T_A)$, where $T_M$ and $T_A$ are the latency for a multiplier and an adder in $GF(2^r)$. We note that the actual number of multipliers in the encoder may be smaller than $\lceil \frac{k+m}{r} \rceil$ due to the fact that the power operation can be simplified. For example, in the normal base Galois field, the square operation can be implemented by cyclic shifting [21]. In this case, the multiplier marked in Figure 2, which is used to compute $x_1^2$, is not needed and the total number of multipliers in the encoder becomes 7.

An estimation of the overhead for secure Galois field multipliers based on AMD codes in $GF(2^{239})$ and $GF(2^{409})$ for the elliptic curve cryptographic devices can be found in [22]. We showed in [22] that the area overhead for the protection architectures based on AMD codes is between 110% and 160%. Moreover, when the encoder is pipelined, the protected multiplier has no latency penalty and can achieve the same performance as the original device. (The work in [22] only considered the special case of AMD codes with $b \leq q-3$ (see Section III-A2).)

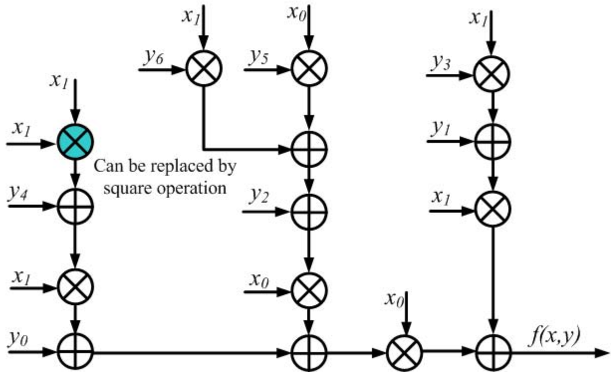


Fig. 2. Encoder Architecture for the $(7r, 2r, r)$ AMD Code Based on $GRM_q(3,2)$ code

TABLE I
OPTIMALITY OF $(k, m, r)$ AMD CODES CONSTRUCTED BY THEOREM 3.1

| $k$ | $m$ | $r$ | $Q_V$ | Optimality |
|---|---|---|---|---|
| $2^{tr}r - tr - 2r$ | $tr$ | $r$ | $1 - 2^{-tr+1}$ | Optimal (Corollary 3.1) |
| $\sum_{i=0}^{b+1} \binom{t}{i} - t - 1$ | $t$ | 1 | $1 - 2^{-(b+1)}$ | Optimal when $b = 1$ or $b = t - 2$ (Corollary 3.2) |
| $(\binom{t+b+1}{t} - t - 1)r$ | $tr$ | $r$ | $(b+1)2^{-r}$ | Optimal when $t = 1$ (Corollary 3.4) |

## V. CONCLUSIONS

In this paper, we presented bounds, general constructions and encoding/decoding procedures for algebraic manipulation detection (AMD) codes based on $q$-ary Generalized Reed-Muller codes and their products. Some of the presented codes are optimal. These codes can provide a guaranteed level of reliability and security even if both the information bits and the non-zero error patterns are controllable by external forces. The same characteristic cannot be achieved by any previously known reliable and secure architectures based on error detecting codes. These codes can be applied for many different applications such as robust secret sharing scheme, robust fuzzy extractors and secure cryptographic devices resistant to fault injection attacks. An efficient encoding and decoding method minimizing the number of required multipliers are given for the presented AMD codes.

## ACKNOWLEDGEMENT

We would like to thank Prof. Leonid Reyzin from the Computer Science Department of Boston University in the United States and Prof. Keren Osnat from the Bar Ilan university in Israel for their important comments and suggestions for the paper.

## REFERENCES

[1] T. Malkin, F.-X. Standaert, and M. Yung, "A comparative cost/security analysis of fault attack countermeasures," in *Fault Diagnosis and Tolerance in Cryptography*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2006, vol. 4236, pp. 159–172.

[2] G. Bertoni, L. Breveglieri, I. Koren, P. Maistri, and V. Piuri, "Error analysis and detection procedures for a hardware implementation of the advanced encryption standard," *IEEE Transactions on Computers*, vol. 52, no. 4, pp. 492–505, 2003.

[3] G. Bertoni, L. Breveglieri, I. Koren, and V. Piuri, "Fault detection in the advanced encryption standard," in *Proc. of the International Conference on Massively Parallel Computing Systems (MPCS 2002)*, Ischia, Italy, Apr. 2002, pp. 92–97.

[4] N. Joshi, K. Wu, and R. Karri, "Concurrent error detection schemes for involution ciphers," in *Cryptographic Hardware and Embedded Systems - CHES 2004*, ser. Lecture Notes in Computer Science, vol. 3156. Springer Berlin / Heidelberg, 2004, pp. 153–160.

[5] R. Karri, K. Wu, P. Mishra, and Y. Kim, "Concurrent error detection of fault-based side-channel cryptanalysis of 128-bit symmetric block ciphers," in *38th Design Automation Conference (DAC 2001)*. ACM Press, 2001, pp. 579–585.

[6] ——, "Concurrent error detection schemes for fault-based side-channel cryptanalysis of symmetric block ciphers," *IEEE Transactions on CAD of Integrated Circuits and Systems*, vol. 21, no. 12, pp. 1509–1517, 2002.

[7] M. G. Karpovsky and A. Taubin, "New class of nonlinear systematic error detecting codes," *IEEE Transactions on Information Theory*, vol. 50, no. 8, pp. 1818–1820, 2004.

[8] M. Karpovsky., K. Kulikowski, and A.Taubin, "Robust protection against fault-injection attacks on smart cards implementing the advanced encryption standard," ser. Proc. Int. Conference on Dependable Systems and Networks (DNS 2004), July 2004.

[9] G. Gaubatz, B. Sunar, and M. G. Karpovsky, "Non-linear residue codes for robust public-key arithmetic," in *Workshop on Fault Diagnosis and Tolerance in Cryptography (FDTC '06)*, 2006.

[10] K.Kulikowski, Z.Wang, and M.G.Karpovsky, "Comparative analysis of fault attack resistant architectures for private and public key cryptosystems," in *Proc of Int. Workshop on Fault-tolerant Cryptographic Devices*, 2008.

[11] Z. Wang, M. Karpovsky, and K. Kulikowski, "Design of memories with concurrent error detection and correction by nonlinear SEC-DED codes," *Journal of Electronic Testing*, pp. 1–22, 2010, 10.1007/s10836-010-5168-5. [Online]. Available: http://dx.doi.org/10.1007/s10836-010-5168-5

[12] S. Cabello, C. Padr, and G. Sez, "Secret sharing schemes with detection of cheaters for a general access structure," *Designs, Codes and Cryptography*, vol. 25, pp. 175–188, 2002, 10.1023/A:1013856431727. [Online]. Available: http://dx.doi.org/10.1023/A:1013856431727

[13] Y. Dodis, B. Kanukurthi, J. Katz, L. Reyzin, and A. Smith, "Robust fuzzy extractors and authenticated key agreement from close secrets," in *In Advances in Cryptology CRYPTO 6*. Springer, 2006, pp. 232–250.

[14] R. Cramer, Y. Dodis, S. Fehr, C. Padró, and D. Wichs, "Detection of algebraic manipulation with applications to robust secret sharing and fuzzy extractors," in *Proceedings of the theory and applications of cryptographic techniques, 27th annual international conference on Advances in cryptology*, ser. EUROCRYPT'08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 471–488. [Online]. Available: http://portal.acm.org/citation.cfm?id=1788414.1788441

[15] V. Guruswami and A. Smith, "Codes for computationally simple channels: Explicit constructions with optimal rate," in *51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 10 2010.

[16] S. Dziembowski, K. Pietrzak, and D. Wichs, "Non-malleable codes," in *Innovation in Computer Science, Cryptology ePrint Archive: Report 2009/608*.

[17] Z. Wang, M. Karpovsky, B. Sunar, and A. Joshi, "Design of reliable and secure multipliers by multilinear arithmetic codes," in *Information and Communications Security*, ser. Lecture Notes in Computer Science, 2009, vol. 5927, pp. 47–62.

[18] F. MacWilliams and N. Sloane, *The Theory of Error-Correcting Codes*. North-Holland, 1998.

[19] E. F. Assmus, Jr, and J. D. Key, "Polynomial codes and finite geometries, Manuscript," 1995.

[20] M. Ceberio and V. Kreinovich, "Greedy algorithms for optimizing multivariate horner schemes," *SIGSAM Bull.*, vol. 38, pp. 8–15, March 2004. [Online]. Available: http://doi.acm.org/10.1145/980175.980179

[21] S. Gao, "Normal bases over finite fields," Ph.D. dissertation, University of Waterloo, 1993.

[22] Z. Wang and M. Karpovsky, "Algebraic manipulation detection codes and their applications for design of secure cryptographic devices," in *IEEE 17th International On-Line Testing Symposium (IOLTS)*, 2011, pp. 234–239.

# Security Best Practices and Risk Assessment of SCADA and Industrial Control Systems

**Guillermo A. Francia, III, David Thornton, and Joshua Dawson**
Jacksonville State University
Jacksonville, AL 36265 USA

**Abstract -** *The nation's critical infrastructures, such as those found in Supervisory Control and Data Acquisition (SCADA) and industrial control systems (ICS), are increasingly at risk and vulnerable to internal and external threats. Security best practices on these systems come at a very opportune time. Further, the value of risk assessment of these systems is something that cannot just be relegated as irrelevant. In this paper, we present a review of security best practices and risk assessment of SCADA and ICS and report our research findings on an on-going risk modeling of a prototypical industrial control system using the CORAS framework tool.*

**Keywords:** Security Best Practices, Risk Assessment, SCADA, Industrial Control Systems.

## 1 Introduction

The nation's critical infrastructures, such as those found in SCADA and industrial control systems (ICS), are increasingly at risk and vulnerable to internal and external threats. These are ushered by insecure connectivity to traditional network systems for the purposes of convenience and also by the vulnerabilities typically found in control system devices and applications. Simply stated, these devices are not ready to be publicly exposed on the Internet.

Security best practices on these systems come at a very opportune time. Further, the value of risk assessment of these systems is something that cannot just be relegated as irrelevant. In this paper, we present a review of security best practices and risk assessment of SCADA and ICS and report our research findings on an on-going risk modeling of a prototypical industrial control system using the CORAS framework tool.

In recent years, the recognition of critical infrastructure vulnerabilities and the consequences of successful attacks have garnered increasing attention. Fortunately, this has led to an ever-growing corpus of best practices and security guides published by governmental and industrial entities.

The rest of the paper is organized into four parts. First, we present a concise overview of security guidelines and best practices for protecting critical infrastructures. Second, we cover risk assessment tools and models. In the third section, we describe the CORAS framework and our motivation in using it. Finally, we present our work in developing a risk assessment model for SCADA and industrial control systems.

## 2 Security Guidelines and Best Practices

Pointers to the set of guidelines, best practices, security tools and new technologies developed by governmental agencies and industrial associations are provided by Ralston, et al. [20].

NIST has published a guideline for security best practices for Information Technology [17]. The NIST has established the Industrial Control System Security Project to research, among its other objectives, the applicability of the NIST SP800-53 recommendations to ICS [12]. The report concluded that an organization, conforming with the baseline sets of security controls in SP 800-53, will also comply with the NERC CIP requirements with regard to the management, operational and technical controls. However, the report pointed out that the so called "business risk reduction" requirements of NERC CIP are not being met because of SP 800-53 is solely focused on information security controls.

Like many information security documents, NIST SP800-53 divides security controls into three categories – technical, management, and operational. These are further subdivided into eighteen families, as shown in Table 1 below.

This set of controls does not represent an exhaustive list, but rather a fundamental set for most practitioners. An additional, much larger list can be found in the appendices and then customized for an individual information security program.

NIST has also released a more targeted security guide [18] which focuses on industrial control systems (ICS) security. This includes the subcategories of SCADA, distributed control systems (DCS), and other control systems like the programmable logic controller (PLC). It contrasts ICS with the more common IT system, underlining threats that are particular to ICS. The broad categories of these threats include policy and procedure vulnerabilities, (such as lack of personnel training and awareness), platform vulnerabilities (such as delayed patching and lack of configuration), and network vulnerabilities (such as weak encryption and lack of redundant hardware). It also explains some of the major risk factors relevant to ICS, including standardized protocols, increased network connectivity, rogue connections, and (somewhat ironically) public information. A cautionary section on documented incidents of attack and hypothetical attack scenarios provides concrete examples of the aforementioned security vulnerabilities.

The President's Critical Infrastructure Protection Board released a concise and approachable overview of key SCADA security concerns in their document, "21 Steps to Improve Cyber Security of SCADA Networks" [19]. Though it was first published in 2002, its content is still relevant today. It represents a good entry point for industry professionals who are taking initial steps toward addressing information security. For more in-depth coverage, the Center for Protection of National Infrastructure's 2010 guide "Configuring and Managing Remote Access for Industrial Control Systems" pairs practical security adjustments with solid justifications for implementing them [4]. It outlines the host of stakeholders related to ICS and the ramifications of attacks on each of them.

In "Best Practices for Government to Enhance the Security of National Critical Infrastructures" [16], the National Infrastructure Advisory Council addresses the need for government to intervene in some markets where the risk of attack and the concordant damage are high. They provide useful advice for managing security in specific industry sectors, and they advocate maintaining high security standards through peer pressure or market competition whenever possible.

Another excellent source for industrial control systems managers was released in 2009 by the Department of Homeland Security [8]. In "Recommended Practice: Improving Industrial Control Systems Cybersecurity with Defense-In-Depth Strategies", the authors provide a comprehensive explanation of technical security vulnerabilities along with strategies to mitigate or eliminate them. It also includes a brief coverage of management techniques and operational security controls.

| **Technical** | **Operational** |
|---|---|
| Access Control | Awareness and Training |
| Audit and Accountability | Configuration Management |
| Identification and Authentication | Contingency Planning |
| System and Communications Protection | Incident Response |
| **Management** | Maintenance |
| Security Assessment and Authorization | Media Protection |
| Planning | Physical and Environmental Protection |
| Risk Assessment | Personnel Security |
| System and Services Acquisition | System and Information Integrity |
| Program Management | |

**Table 1. The NIST Security Controls**

In January of 2009, the Department of Homeland Security also released the latest version of its National Infrastructure Protection Plan (NIPP), which aims to unify critical infrastructure and key resources (CIKR) security concepts [8]. The NIPP is written for a wide audience, including government personnel, CIKR owners, and academia. As such, it provides a broad understanding of the vulnerabilities and repercussions of successful attacks against CIKR.

## 3 Risk Assessment

Risk management is the process of finding the best among many alternatives in order to minimize the impact of uncertain events [5]. It may also be considered as an assessment process used to determine the controls that are needed to adequately and cost effectively protect critical assets. The five main factors involved in the process are:

- value of assets to be protected;
- threats to these assets;
- vulnerabilities of these assets;
- types of losses that these threats would inflict; and
- controls that will mitigate these threats.

For general risk assessment, the ASIS International Guidelines Commission recommended the following general security risk assessment steps [1]:

1. Understanding of the organization, its people and assets at risk;
2. Specifying risks and vulnerabilities;
3. Establishing the probability of risks and frequency of events;
4. Determining impacts;
5. Developing mitigation;
6. Considering the options; and
7. Performing cost and benefit analysis.

Advances in probabilistic risk assessment that can be applied to estimate the risk from SCADA and DCS installations are described in [20]. Also, in the same paper, the authors provided a comparison of approaches to quantifying the risk, threat impact and cyber-security on SCADA and DCS networks.

Previous work on risk assessment studies specifically for the SCADA systems are found in [6], and [10].

## 3.1 Risk Assessment Tools

The Carnegie Mellon University's CERT Coordination Center developed the Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE) [2], which is a suite of tools, techniques, and methods for risk-based information security strategic assessment and planning. It uses the event/fault tree model to analyze threats to critical assets.

A freely available model-driven risk analysis tool, Cost of Risk Analysis System (CORAS), is available for download at http://coras.sourceforge.net/. A guided tour of the CORAS method can be located at the website and at the CORAS book [13]. The CORAS tool is a computerized implementation tool which is designed to support documenting, maintaining and reporting the analysis resulting from the CORAS risk modeling.

Risk Watch for Critical Infrastructure **(**Nuclear Power compliant) is a commercial product that provides compliance and risk assessments for critical infrastructure entities, more specifically the nuclear power sector. It is based on the new Nuclear Energy Institute guidelines contained in the NEI 04-04 Revision 1: "Cyber Security Program for Nuclear Power Reactors". Both the Nuclear Regulatory Commission and the Nuclear Energy Institute participated in the development of this software, which was funded by the U.S. Department of Defense through the Technical Support Working Group [21].

## 3.2 Risk Models and Methodologies

Perhaps, one of the earliest quantitative risk assessment methodologies uses the Annualized Loss Expectancy (ALE) model. The ALE is calculated by multiplying the Single Loss Expectancy (SLE) by the Annualized Rate of Occurrence (ARO), the expected frequency of the event.

The eight-stage security risk assessment model proposed by Drake and Morse [9] includes the following stages: 1) threat obstruction; 2) threat occurrence; 3) detection threat occurrence; 4) recovery from threat occurrence; 5) security breach; 6) detection of breach; 7) damage elimination; and 8) identifying external losses. The external losses include mission failure, personnel loss, loss of resources, revenue loss, and time loss.

The Central Computer and Telecommunications Agency (CCTA) Risk Analysis and Management Method (CRAMM) [11], developed by the UK Security Service, is based on the matrix method to assess risk based on data gathered from questionnaires. It consists of three stages: 1) Asset identification; 2) Vulnerability identification; and 3) Counter-measure installation.

The Quantitative Threat-Risk Index Model (QTRIM) [3] is used to predict the risk of a terrorist attack against a national infrastructure. It is built and tested at the Idaho National Engineering and Environmental Laboratory (INEEL) and calculates risk using terrorist specific constraints, objectives, value systems, logistics, and opportunities on a balance scorecard framework.

The Fault Tree Analysis (FTA) [25] method uses a deductive, failure-based approach. While the leaf node represents the triggering event, the root node represents an unwanted event, or failure, and the different events that may lead to the top event are modeled as branches of nodes.

Attack trees [22] provide a formal way of describing the security of a system by using the FTA model and replacing the fault as the attack goal and event probabilities for failure rates. The capability of attack trees to represent a highly comprehensive attack sequences are highly dependent on the experience of the security analysts that build them [14].

McQueen, et al, [15] developed quantitative techniques to calculate risk reduction estimates for a small SCADA control system. Directed graph structures, wherein nodes represent stages of a potential attack and edges represent the expected time-to-compromise for different attacker skill levels, are used as framework model.

# 4 The CORAS Framework

The CORAS framework is made up of a methodology for model-based risk assessment, a Unified Modeling Language (UML) based specification language, a library of reusable packages, an integrated platform for data repository, and a risk assessment reporting system.

The following objectives of the CORAS project [13] are:

- To develop a practical framework for risk analysis;
- To assess the applicability, usability, and efficiency of the framework; and
- To investigate its commercial viability.

The CORAS platform is an open-source and data-portable tool for risk assessment. It was developed in 2002 by a consortium of partners from four European countries.

The CORAS method is divided into two major groups of procedures. The first group establishes a common understanding of the target for analysis as well the documentation of assumptions and constraints needed for the subsequent risk analysis. The second group is focused on the actual risk analysis [13].

We opted to use this tool primarily because it is open-source, system independent and very user friendly for rapid risk model development.

# 5 Risk Modeling of an Industrial Control System

In this paper, we assume that the initial steps of initial discussions on the target, preparatory analysis, and documentations of the CORAS method have been completed. We start with asset identification.

## 5.1 Assets

We identify and classify each asset with a corresponding degree of importance (1=most important and 5=least important). Table 2 depicts a partial list of assets using the management view of business objectives.

| *Asset* | *Importance* | *Type* |
|---|---|---|
| ***Safe Operation*** | 2 | Direct |
| ***Regulatory Compliant*** | 3 | Direct |
| ***Company Reputation*** | 2 | Indirect |
| ***Customer Service*** | 4 | Direct |
| ***Company Information*** | 5 | Direct |
| ***Profitability*** | 1 | Indirect |

**Table 2. Asset Table**

## 5.2 Threats

Next, in Table 3 we list the threats and possible scenarios that may affect the identified assets.

| *Threat* | *Scenario/Incident* |
|---|---|
| ***Employee*** | Intended/unintended service disruption |
| ***System Failure*** | Power outage |
| ***Network Failure*** | Denial of Service |
| ***Hacker*** | Intrusion or Service Disruption |
| ***Malware*** | Disruption |
| ***Eavesdropper*** | Listening on communication channels |
| ***Natural disaster*** | Tornado, flood, earthquake |

**Table 3. Threat Table**

## 5.3 Vulnerabilities

Vulnerabilities are security weaknesses that could be exploited or inadvertently triggered. Table 4 depicts some of the control system vulnerabilities.

| *Area* | *Vulnerability* |
|---|---|
| ***System*** | Misconfiguration, Missing anti-virus software, Outdate patch, Weak authentication, Web server flaws, Database system flaws. |
| ***Network*** | Firewall misconfiguration, rogue access points, unrestricted personal device access, weak authentication. |
| ***Physical*** | Unlocked facilities, weak entry authentication, Record access unrestricted |
| ***Employee*** | No training, undisciplined web access, unrestricted system access, social engineering |
| ***Information*** | Unrestricted access, unencrypted transmission, lack of data duplication policy, improper media disposal. |

**Table 4. Vulnerability Table**

## 5.4 CORAS Snapshots

The symbols used in the CORAS framework include an accidental human threat (e.g. an untrained technical staff), a deliberate human threat (e.g. a hacker), a vulnerability (e.g. an unpatched system), a direct asset (e.g. company information), an indirect asset (e.g. the company reputation), a non-human threat (e.g. natural phenomenon), an unwanted incident (e.g. disclosure of company secrets), and a threat scenario (e.g. a rogue access point connected to the company network). The symbols are shown in Figure 1.

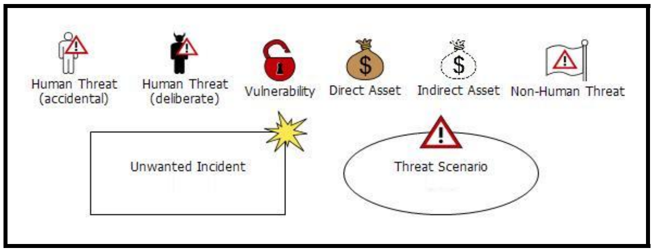


**Figure 1. CORAS Symbols [13]**

Initial threat diagrams are developed during a brainstorming session participated by stakeholders, security professionals, and risk modelers. Preliminary threat diagrams are developed during the initial stage of the brain storming sessions. A sample initial threat diagram is shown in Figure 2. In this preliminary threat diagram two non-human threats are modeled with respective vulnerabilities. This causes a threat scenario in which the system becomes inaccessible and thereby causing an unwanted incident (service disruption). Finally, the service disruption event is triggered and thereby affecting safe operation, a direct asset of the industrial control system.

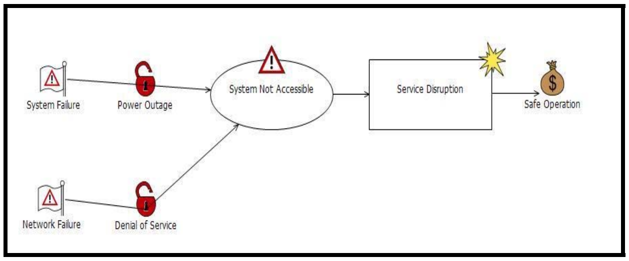


**Figure 2. Non-Human Threat Diagram**

An example of one of our final threat diagrams is depicted in Figure 3. In this final threat diagram we looked at three different actors, ICS Technician, Hacker and an Eavesdropper.

The ICS Technician introduced various vulnerabilities into the ICS system. These vulnerabilities are then exploited by the hacker or eavesdropper to gain access to the ICS system or compromise the confidentiality of the system. In this scenario, the ICS technician receives no training which leads to a misconfiguration of the ICS system and also the introduction of a rogue access point in the ICS network. Another vulnerability associated with the ICS technician is giving the technician unrestricted system access. In this scenario a hacker can attack the ICS system by means of social engineering, where the hacker tricks the ICS technician into releasing information, or discovering the rogue access point and using that to gain access to the network.

The hacker can compromise access to the ICS system which leads to an ICS system disruption. The diagram shows that a disruption in service affects two direct assets: customer service and safe operation as well as two indirect assets: profitability and company reputation. Another route the hacker can take is to compromise the commands issued to the ICS system which leads to compromise the ICS's data integrity. This affects the direct assets: safe operation, company information, and regulatory compliance as well as the indirect assets: profitability and company reputation. The hacker may also pose a threat to the ICS system by using the rogue access point introduced by the ICS technician. The hacker bypasses the network authentication and is able to access the network which, in effect, affecting the direct assets: safe operation and customer service as well as the indirect assets: profitability and company reputation.

The eavesdropper will use the rogue access point to access the network. The ICS system has the vulnerability of unencrypted transmissions so the eavesdropper can compromise the confidentiality of the ICS system resulting in the loss of company information. This diagram shows the various vulnerabilities in the ICS system, how these vulnerabilities can lead to threats, and finally how deliberate actions by humans can lead to these vulnerabilities being exploited resulting in damage to the direct and indirect assets of the company.

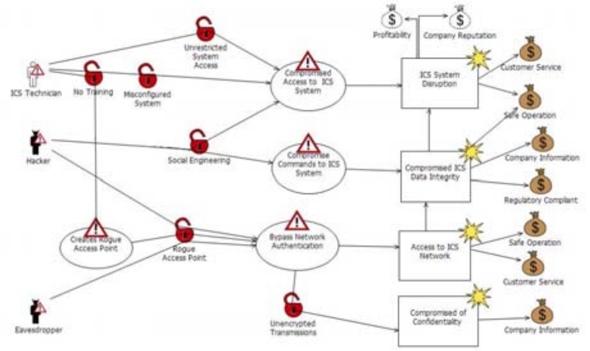


**Figure 3. Final Threat Diagram**

## 6 Conclusions and Future Plans

This paper presented a review of security best practices and risk assessment of SCADA and ICS systems. We also presented our research findings on an on-going risk modeling of a prototypical industrial control system using the CORAS framework tool

The challenge for the authors will be in the continual development of the SCADA and ICS system risk model. Future plans include:

- Development of a risk simulation model that mimics the actual risks endemic to SCADA and ICS systems; and
- Expansion of the current risk model to include parameters representing additional assets, risks, and vulnerabilities.

## 7 Acknowledgements

This paper is based upon a project partly supported by the National Science Foundation under grant awards DUE-0726486 and OCI-0959687. Opinions expressed are those of the authors and not necessarily of the Foundation.

## 8 References

[1] ASIS International. 2004. General Security Risk Assessment Guidelines. www.tisp.org/index.cfm?pk=download&id=10948&pid=10261

[2] Alberts, Christopher and Dorofee, Audrey. 2003. Managing Information Security Risks: The OCTAVE (SM) Approach. Addison-Wesley Professional Publishing. 2003.

[3] Beitel, G.A., Gertman, D. I., and Plum, M.M. 2004. Balanced Scorecard Method for Predicting the Probability of a Terrorist Attack. Risk Analysis IV:581-592, WIT Press, Brebbia, C.A., ed..

[4] Center for the Protection of National Infrastructure. 2010. Configuring and Managing Remote Access for Industrial Control Systems. http://www.us-cert.gov/control_systems/pdf/Recommended_Practice-Remote_Access_1-6-2011.pdf

[5] Cardenas, A., Amin, S., Lin, Z., Huang, Y., Huang, C., and Sastry, S. Attacks Against Process Control Systems: Risk Assessment, Detection, and Response. In Proceeding of the ASIACCS'11 Conference, (March, 2011). ACM, DOI 978-1-4503-0564-8-8/11/03.

[6] Craig,P., Mortensen, J., and Dagle. J.E. 2008. Metrics for the National SCADA Test Bed Program. Technical Report. PNNL-18031, Pacific Northwest National Laboratory (PNNL), Richland, WA.

[7] Department of Homeland Security. 2009. National Infrastructure Protection Plan http://www.dhs.gov/xlibrary/assets/NIPP_Plan.pdf.

[8] Department of Homeland Security. 2009 Recommended Practice: Improving Industrial Control Systems Cybersecurity with Defense-In-Depth Strategies. http://www.us-cert.gov/control_systems/practices/documents/Defense_in_Depth_Oct09.pdf.

[9] Drake, D.L. and Morse, K.L. 1994. The Security-specific Eight Stage Risk Assessment Methodology. In Proceedings of the 17th National Computer Security Conference (San Diego, CA).

[10] Hamoud, G., Chen, R.L., and Bradley, I. 2003. Risk assessment of power systems SCADA. In IEEE Power Engineering Society General Meeting, 2003, volume 2.

[11] Jones, Andy, and Ashenden, Debi. 2005. Risk Management for Computer Security: Protecting Your Network and Information Assets. Butterworth-Heinemann, UK.

[12] Katze, S., Stouffer, K., Abrams, M., Norton, D., and Weiss, J. 2006. Applying NIST SP 800-53 to Industrial Control Systems, NIST 2006. http://csrc.nist.gov/groups/SMA/fisma/ics/documents/papers/Apply-SP-800-53-ICS-final-22Aug06.pdf.

[13] Lund, M. S., Solhaug, B. and Stolen K. Model-Driven Risk Analysis. 2011. The CORAS Approach. Springer-Verlag Berlin Heidelberg .

[14] McQueen, M., Boyer, W., Flynn, M., Alessi, S. Quantitative Risk Reduction Estimation Tool for Control Systems, Suggested Approach and Research Needs. Idaho National Laboratory. International Workshop On Complex Network and Infrastructure Protection (2006A).

[15] McQueen, M., Boyer, W., Flynn, M., Beitek, G. Quantitative Cyber Risk Reduction Estimation Methodology for a Small SCADA Control System, In Proceedings of the 39th Hawaii International Conference on System Science, Kauai, Hawaii. (2006B).

[16] National Infrastructure Advisory Council. 2004. Best Practices for Government to Enhance the Security of National Critical Infrastructures. http://www.dhs.gov/xlibrary/assets/niac/NIAC_BestPracticesSecurityInfrastructures_0404.pdf.

[17] National Institute of Standards and Technology (NIST), SP 800-53, "Guide to Industrial Control Systems (ICS) Security," Website: http://csrc.nist.gov/publications/nistpubs/800-53-

Rev3/sp800-53-rev3-final_updated-errata_05-01-2010.pdf.

[18] National Institute of Standards and Technology (NIST), SP 800-82, “Guide to Industrial Control Systems (ICS) Security,” Website: http://csrc.nist.gov/publications/nistpubs/800-82/SP800-82-final.pdf, September 2008.

[19] President’s Critical Infrastructure Protection Board. 2002. 21 Steps to Improve Cyber Security of SCADA Networks. http://www.oe.netl.doe.gov/docs/prepare/21stepsbooklet.pdf.

[20] Ralston, P.,Graham, J., and Hieb, J. 2007. Cyber security risk assessment for SCADA and DCS networks. ISA Transactions, 46(4), 583–594.

[21] Risk Watch for NEI. Website: http://www.riskwatch.com/index.php/nei-compliance. Last access: March 06, 2012.

[22] Schneier, Bruce. 1999. Attack trees: Modeling security threats. Dr. Dobb's Journal of Software Tools, 24(12), 21-29.

[23] Stølen, Ketin. 2001. CORAS-A Framework for Risk Analysis of Security Critical Systems. In supplement of the 2001 International Conference on Dependable Systems and Networks, pages D4 - D11, July 2-4, 2001, Gothenburg, Sweden.

[24] Stouffer, K., Falco, J., Scarfone, K. 2006. Guide to Supervisory Control and Data Acquisition (SCADA) and Industrial Control Systems Security: Recommendations of the National Institute of Standards and Technology, National Institute of Standards and Technology (2006). http://www.cyber.st.dhs.gov/docs/NIST%20Guide%20to%20Supervisory%20and%20Data%20Acquisition-SCADA%20and%20Industrial%20Control%20Systems%20Security%20(2007).pdf.

[25] Vesely W. Fault Tree Analysis (FTA): Concepts and Applications. Website: http://www.hq.nasa.gov/office/codeq/risk/docs/ftacourse.pdf. Access date: March 05, 2012.

# SCARF: profile-based Side Channel Analysis Resistant Framework

**Juhan Kim[1,2], Kyunghee Oh[1], Dohoo Choi[1], and Howon Kim[2]**
[1]Cyber Security-Convergence Research Department, ETRI, Daejeon, Korea
[2]School of Computer Science & Engineering, Pusan National University, Pusan, Korea

**Abstract** - *SCARF, Side Channel Analysis Resistant Framework which we are developing, is a side channel analysis system which provides profile-based analysis steps that consist of a trace collection process, preprocessing processes and analysis processes in a profile. In this paper, we will introduce SCARF profile function which is very useful to share know-how among organizations who want to evaluate the security level of any cryptography module against side channel attacks. The profile includes a flow diagram about all processes from a trace collection process to analysis processes. It also manages the result trace set, parameters configuration and status of each process. Therefore, if one makes a profile about an electronic device and shares the profile with any other organization, the organization can become aware immediately what they should do to analyze the device. Because the profile has all the information which is required to evaluate the secure level of the device against side channel attacks.*

**Keywords:** Side Channel Analysis System, SCA

## 1 Introduction

Side Channel Analysis (SCA) system [1] is to verify secure level of software or hardware cryptography implementations in electronic devices against side channel attacks. The SCA system collects side channel information like power consumption [2] or electromagnetic leakage [3] during a cryptographic computation in an electronic device such as IC card, smart phone, and any device which includes a cryptography implementation in it. It can perform some preprocessing steps like alignment, compression and filtering with the collected traces. Then it analyzes the preprocessed traces to determine the key of the cryptography module which has a cryptography algorithm such as AES, DES, SEED and RSA with analysis methods like SPA [4], DPA [5] and CPA [6].

There exist several side channel analysis systems such as DPA Workstation of Cryptography Research [7], Inspector of Riscure [8], and Sideway of Brightsight [9]. These systems allow analyzing the vulnerability of cryptographic devices against side channel attacks in a reliable approach. However, they are trace-centric systems which load a trace set, do a preprocessing job and make another trace set as a result of the job. That is, the systems produce more and more trace sets whenever the systems do a preprocessing job such as alignment, filtering, compression and so on. Even one who made the trace sets may not confirm that which preprocessing jobs and which parameters were used to make the trace sets as the number of trace sets increases and as time goes on. Because they neither provide any relation view about the jobs that they did nor save any property which was applied to each process. Therefore, only with trace sets it is not easy to propagate one's know-how what he did to analyze a device.

In this paper, we will introduce SCARF profile function which is very useful to share know-how among organizations who want to evaluate the security level of any cryptography module against side channel attacks. The profile includes a flow diagram about all processes from a trace collection process to analysis processes. It also manages the result trace set, parameters configuration and status of each process. Therefore, if one makes a profile about a device and shares the profile with any other organization, the organization can become aware immediately what they should do to analyze the device because the profile has all the required information.

## 2 Architecture of SCARF profile

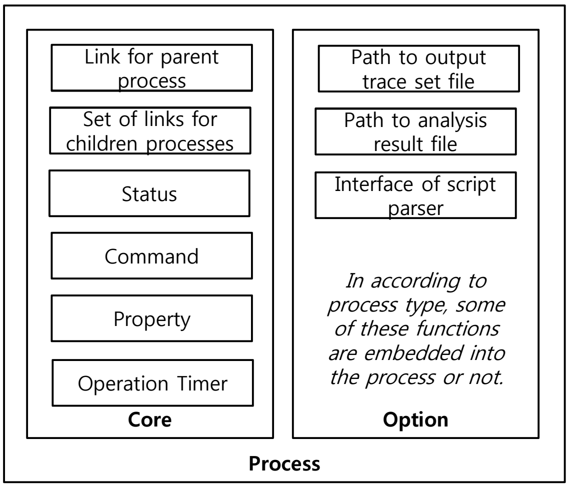

Link for parent process
Set of links for children processes
Status
Command
Property
Operation Timer
Core
Path to output trace set file
Path to analysis result file
Interface of script parser
In according to process type, some of these functions are embedded into the process or not.
Option
Process

Figure 1. Process architecture

It is necessary for a side channel analysis system to design process-centric processing architecture like figure 1 for sharing an analysis method conveniently.

A process in SCARF profile includes the following:

- Link to parent process: It is used to provide a result trace set of parent process to the process.
- Set of links for children processes: It is required to manage process chain of a profile.
- Status: This presents current status of the process. Ex) ready, waiting, running, paused, finished, stopped
- Command: It is for user interface to run, stop, pause and initialize the process.
- Property: Every process has its own property which consists of parameters for it.
- Path to output trace set file: If one process is finished, the process outputs its result trace file and manages the file with the path.
- Path to analysis result file: In case of an analysis process, it has a path to analysis result file instead of the path to output trace file which includes list of candidate key, list of correlation values of the key and a trace set related with time peak of the key.
- Interface of script parser: In case of a collection process, it is connected to a script parser which can control oscilloscopes and SCA boards as contents of a script.
- Operation timer: it checks the time to finish the process job.

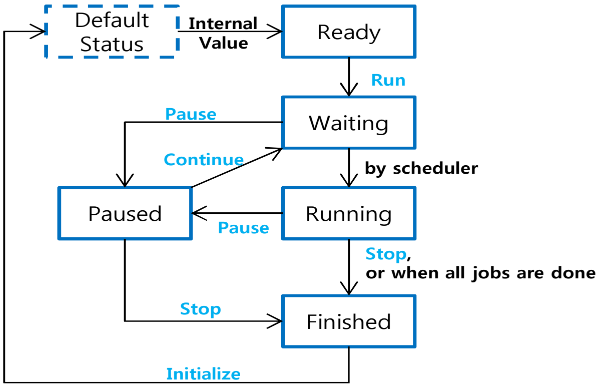


Figure 2. Status Diagram

Figure 2 shows status diagram of a process. The status of a process can consist like the following:

- Ready: When a user creates a new process or initializes an existed process, the status of the process goes into ready. It means that the process is ready to run.
- Waiting: When a user run a process, the status of the process goes into waiting. The process is waiting until its parent finishes at least designated number of jobs. If its parent finished the jobs, a scheduler changes its status to running. Default designated number is 1.
- Running: A process is running.
- Paused: A process is paused by a user.
- Finished: When a user stops a process or where all jobs in the process are done, the status of it can be finished.

SCARF can run all of processes in a profile at once using profile control commands.

- ActivateProfile: All processes which are in ready status start to work.
- StopProfile: All processes which are in progress are halted. Processes in the running or the paused status go into the finished status. Other processes become to have the ready status.
- PauseProfile: It changes status of all processes from waiting/running status to paused status.
- InitializeProfile: It initializes all process and changes status to ready. Then result files of all processes are deleted.

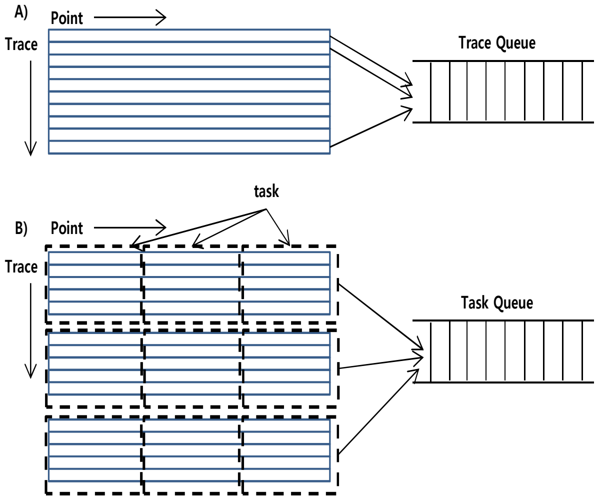


Figure 3. Basic processing unit of SCARF process, A) Trace - input and output data of collection process & preprocessing process, B) Task - input data of analysis process data

Every process in SCARF has input data and output data like figure 3. If the process is kind of collection and preprocessing process in table 1 on next page, it has a trace as input and output data. However, if it is kind of analysis process, it has a task. Queues in Figure 3 are for distributed processing function which will be not described in this paper.

Figure 4 shows the structure of SCARF profile. There can be many profiles in a project. A profile can have three types of process such as collection process, preprocessing process and analysis process like figure 4. It can include lots of and duplicated processes except processes in collection process group. It should have only one collection process as a root process and no process can be positioned after an analysis process.

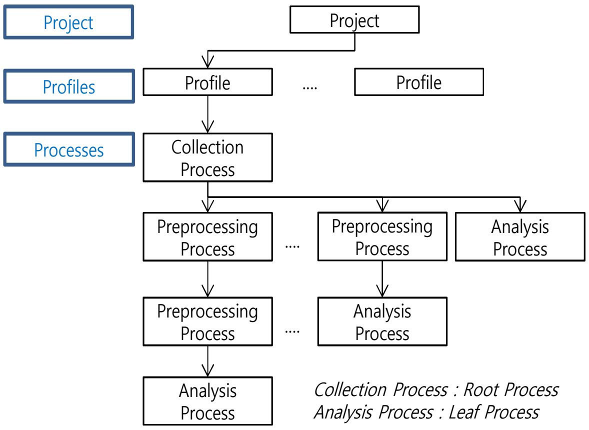


Figure 4. SCARF profile structure

In collection process group, there are 'Oscilloscope' process which gathers traces from an oscilloscope device and 'Trace File' process which reads traces from a file that was collected previously.

The preprocessing group consists of 'Range' process which outputs selected points of traces, 'Trace' for selected traces in a traces set, 'Alignment' for aligning traces, 'Frequency' for transforming time domain traces to frequency domain traces, 'ALaw', 'Filter' and 'Compression'.

SCARF can evaluate cryptography algorithms such as DES, AES, SEED, ARIA, RSA and ECDSA with analysis methods like DPA, CPA, IDPA, and HODPA. These are listed in table 1.

Table 1. SCARF process

| Process Group Name | Process Name |
|---|---|
| Collection | Oscilloscope, TraceFile |
| Preprocessing | Range, Trace, Alignment, Filter, Frequency, A Law, Compression, … |
| Analysis | DES-DPA, DES-CPA, DES-IDPA, DES-FDPA , DES-HODPA, DES-DCA |
| | AES-DPA, AES-CPA, AES-IDPA, AES-FDPA , AES-HODPA, AES-DCA |
| | SEED-DPA, SEED-CPA, SEED-IDPA, SEED-FDPA, SEED-HODPA, SEED -DCA |
| | ARIA-DPA, ARIA-CPA, ARIA-IDPA, ARIA - FDPA, ARIA-HODPA, ARIA-DCA |
| | RSA-CRT-Recom, RSA-CRT-MRED, RSA-L2 R, RSA-R2L |
| | ECDSA-R2L |

Figure 5 show design specification of SCARF process which describes relation among processes and related modules with them.

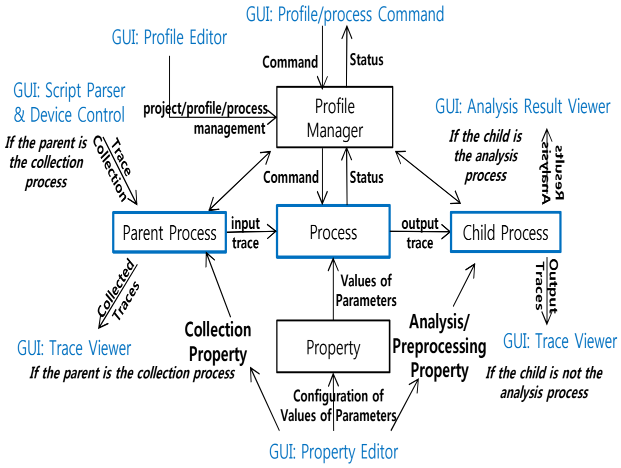
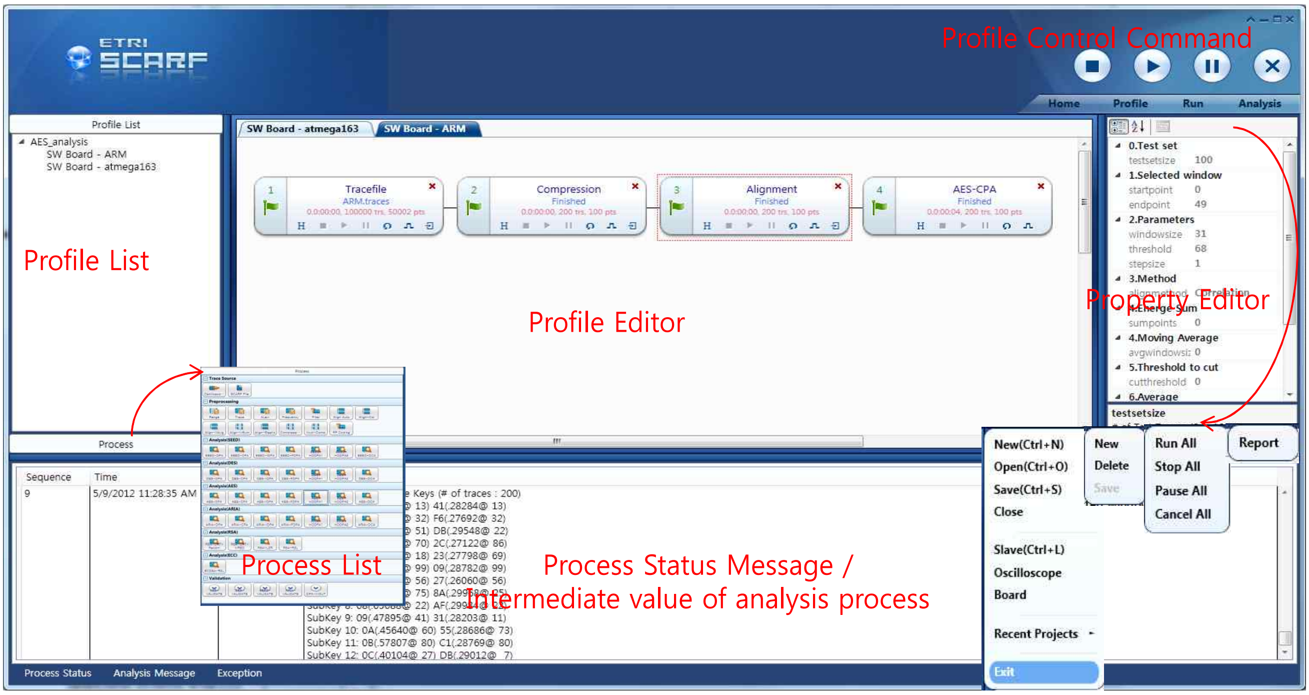

GUI: Profile/process Command
GUI: Profile Editor
Command
Status
GUI: Script Parser & Device Control
project/profile/process management
Profile Manager
GUI: Analysis Result Viewer
If the parent is the collection process
Trace Collection
Command
Status
If the child is the analysis process
Analysis Results
Parent Process
input trace
Process
output trace
Child Process
Collected Traces
Values of Parameters
Output Traces
GUI: Trace Viewer
Collection Property
Property
Analysis/ Preprocessing Property
GUI: Trace Viewer
If the parent is the collection process
Configuration of Values of Parameters
If the child is not the analysis process
GUI: Property Editor

Figure 5. Design specification of SCARF process

There are other components for parallel processing function and distributed processing function to manage jobs of processes. However, we will not present them in this paper to focus on profile-based architecture.

The profile manager in figure 5 can activate or run a process with functions such as activate and run function in the process. Detailed roles of the profile manager are like the following:

- The profile manager calls 'activate' function in a process when the process in ready or paused status and a user operates the process to run.
- When a process is in running status, the profile manager calls 'run' function of the process. The function has a trace or a task as an input data and outputs a processed trace or fractional analysis result. If the process is finished, its final output will be a trace set file or full analysis result file. And the process includes information about a link of the file.
- If a user runs a process and its child process is in running status, the profile manager calls each 'run' function of the child processes with a parameter which is output trace of the process.
- The profile manager saves all result traces or analysis result to files for each process at the current profile directory in the project. And it maintains the location information unless the process is deleted or initialized.
- It provides information about profile count, profile names, and process chain of each profile to the profile editor when a user opens a project. It also supplies status of all processes to the editor.
- It provides the property related with the process to the property editor when a user clicks a process icon on the profile editor. Then the user can edit values of parameters in the property editor when the status of the process is ready.



AES_analysis
SW Board - ARM
SW Board - atmega163

Tracefile ARM.traces 0.0:00:00, 100000 trs, 50002 pts
Compression Finished 0.0:00:00, 200 trs, 100 pts
Alignment Finished 0.0:00:00, 200 trs, 100 pts
AES-CPA Finished 0.0:00:04, 200 trs, 100 pts

0.Test set
testsetsize 100
1.Selected window
startpoint 0
endpoint 49
2.Parameters
windowsize 31
threshold 68
stepsize 1
3.Method
alignmethod Correlation
4.Energe Sum
sumpoints 0
4.Moving Average
avgwindowsiz 0
5.Threshold to cut
cutthreshold 0
6.Average
testsetsize

| Sequence | Time |
|---|---|
| 9 | 5/9/2012 11:28:35 AM |

```
e Keys (# of traces : 200)
@ 13) 41(.28284@ 13)
@ 32) F6(.27692@ 32)
@ 51) DB(.29548@ 22)
@ 70) 2C(.27122@ 86)
@ 18) 23(.27798@ 69)
@ 99) 09(.28782@ 99)
@ 56) 27(.26060@ 56)
@ 75) 8A(.29958@ 25)
@ 22) AF(.29934@ 22)
SubKey 9: 09(.47895@ 41) 31(.28203@ 11)
SubKey 10: 0A(.45640@ 60) 55(.28686@ 73)
SubKey 11: 0B(.57807@ 80) C1(.28769@ 80)
SubKey 12: 0C(.40104@ 27) DB(.29012@  7)
```

Figure 6. SCARF GUI Configuration

- It also stops, pauses, initializes, and runs the process when a user clicks one of command icons on the process icon.
- Through the profile manager, a user can see the result trace set or analysis result of any process only to click one of commands icon of the process.

# 3 Implementation

The SCARF system which we have developed can be operated on Windows 64-bit or 32-bit machine. However, it is optimized on Windows 64-bit environment. We used .Net Framework and C# language to implement it.

Figure 6 shows implemented SCARF system which is configured by profile list, process list, process chain in process editor, property value of the align process in property editor and message box area for printing process status or intermediate analysis value.

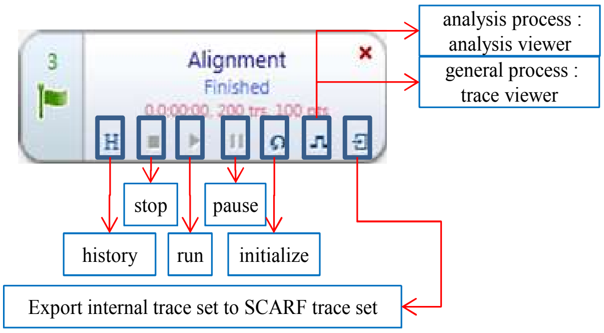


Figure 7. Process control commands

Figure 7 and figure 8 describe control commands for a process and a profile, respectively. Figure 7 also presents that the status is finished, its processing time is below 1 second, its result traces are 200 and a trace has 100 points.

The history information in figure 7 presents how the previous processes and the current process worked. If a process is not on finished status, the history command will be deactivated.

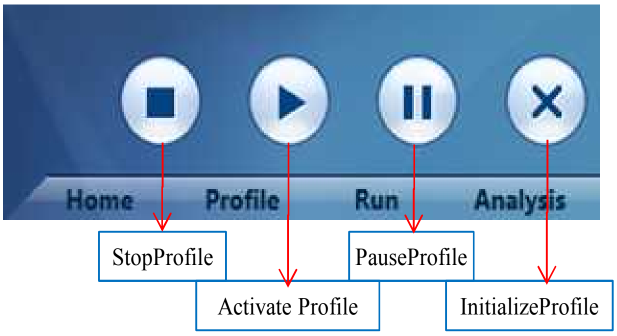


Figure 8. Profile control commands

The example profile on figure 6 describes a side channel analysis test about an AES S/W implementation with ETRI S/W analysis board that has ARM processor. In this profile, we used a trace file which had been collected previously with an oscilloscope when the implementation had been operating on the board. We compressed traces in the file with 500 times and aligned the compressed traces. Then AES-CPA, an analysis process, was operated to analyze the vulnerability of the AES S/W implementation against side channel attacks.

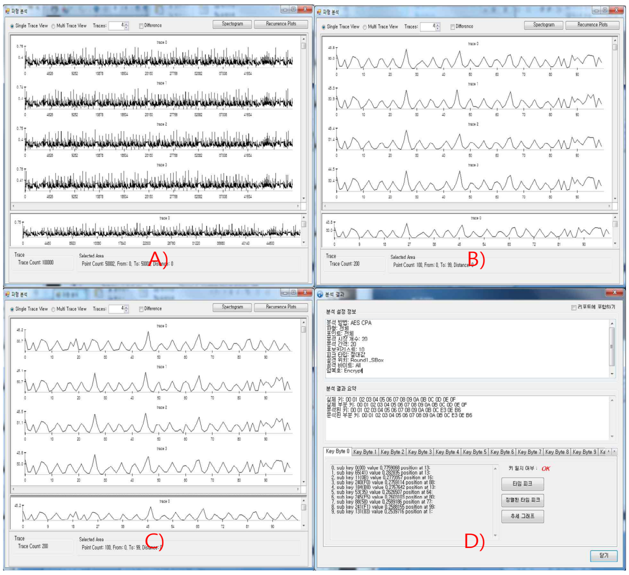


Figure 9. Result Data – A) Collected traces in a file but it had been gathered from an oscilloscope, B) Compressed traces, C) Aligned traces, D) Analysis result

The analysis result, D) in figure 9 has time peak graph of candidate keys. It also includes tendency graph of intermediate results of an analysis process like B) in figure 10. It has interrelationship between A) and B).

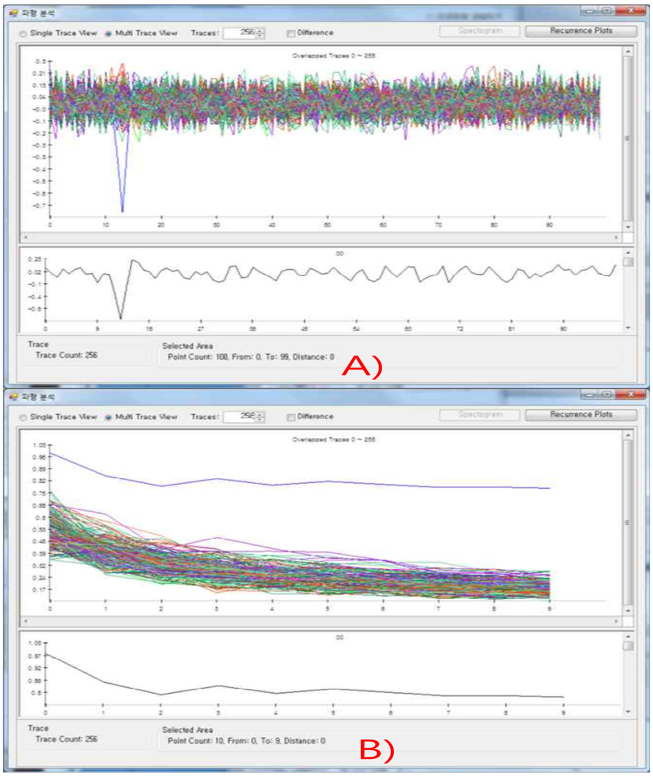


Figure 10. A) Time peak graph about candidate keys, B) Tendency graph of the candidate keys

Time peak graph is used to decide which candidate key is most similar to the real key. The A) graph in figure 10 shows correlation values of all candidate keys. We can think that the real key is 0x00 because the time peak of 0x00 is extremely different from them of any others. However, this is restricted to one byte in a real key which consists of 16 bytes because full key length of the AES algorithm is generally 16 bytes. Therefore, there are 16 time peak graphs and tendency graphs in the analysis result like figure 9 D). However, in according to cryptography algorithm, it has different key byte length.

The tendency graph shows when the best candidate key is extracted. In figure 10 B), we can see that the best is overwhelming from first intermediate result.

Below steps describes how SCARF works to analyze to evaluate the security level of any cryptography implementation against side channel attacks.

- Create SCARF project.
- Make a new profile.
- Drag & drop processes in process list to the profile editor.
- Click a process in the profile editor and set values of parameters in the property editor.
- Run a process or use the profile control commands to run all processes in the profile at once.
- Check the results of analysis processes.

If an organization takes a profile from another for evaluating the security level of any cryptography module against side channel attacks, the organization just open the profile with SCARF and can easily get the know-how of another organization that sent the profile only by checking a flow of processes and properties of them. A profile is propagated easily like the following:

- Copy a project directory like figure 11 and send it.
- Paste the directory and open 'project.config' file in the directory with SCARF.

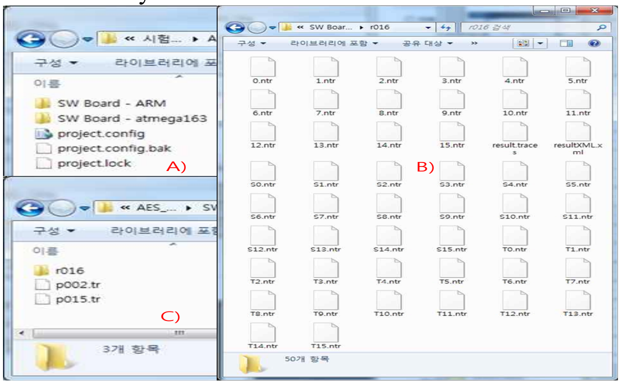


Figure 11. A) Project folder, B) Analysis process folder in figure 6, C) Profile folder

# 4 Conclusion

In this paper, we introduced the profile-based side channel analysis system, SCARF. This approach provides the following advantages:

- A user can reuse the profile which was made by another user without any modification.
- The profile can be initialized by a user and then results of all processes are removed from the profile. However, process chain in the profile and property of each process are remained. Therefore one can restart to evaluate the security level of any cryptography implementation against side channel attacks from trace collection to analysis with the profile which includes another user's know-how.
- It is easy to copy and move a profile. The only one thing has to do is "copy & paste the project folder". That is, if an evaluation target device and its cryptography implementation are similar to them of the previous profile, one can reuse the profile by making a copy.

SCARF is very useful to propagate easily one's evaluation know-how to others. With a profile one can become immediately aware of an oscilloscope used to collect traces, configuration of the oscilloscope, necessary preprocessing processes and properties of the processes, the results of analysis processes and so on.

During an analysis experiment a user may add lots of processes and duplicated processes with different values of parameters in a profile. However the user is always able to check easily which processes are done or not, how the results of the processes are and which processes are necessary by comparing the results of analysis processes. Therefore the user easily recognizes optimized process chain and properties of those processes in the chain. Such information is the know-how to analyze the security level against side channel attacks.

Thus, SCARF can contribute to efficiency improvement of analysis and easy propagation of one's know-how to others by providing a well-organized view of processes and properties to users.

## Acknowledgement

This work was supported by SCARF project which is the R&D program of KCA. [Development of the Technology of Side Channel Attack Countermeasure Primitives and Security Validation]

# 5 References

[1] P. Kocher, Timing attack on implementation of Diffe-Hellman, RSA, DSS and other systems, Proc. Advanced Cryptology, 104-113, 1996.

[2] P.Kocher, J.Jaffe, and B.Jun, Differential power analysis, Proc. CRYPTO, pp. 388-397, 1999.

[3] J.Quisquater and D.Samyde, Electromagnetic Analysis (EMA): Measures and countermeasures for smart cards, Proc. e-Smart, pp.200-210, 2001.

[4] P.Kocher, J.Jaffe, and B. Jun, "Introduction to differential power analysis and related attacks," 1998, White Paper, Cryptography Research, http://www.cryptography.com/dpa/technical, 1998.

[5] P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," Advances In Cryptology - CRYPTO' 99, LNCS 1666 Springer-Verlag, pp.388-397, Santa Barbara, USA, August 1999.

[6] E. Brier, C. Clavier, and F. Olivier, "Correlation power analysis with a leakage model," Cryptographic Hardware and Embedded Systems 2004. LNCS 3156 Springer-Verlag, pp. 16-29, 2004.

[7] Cryptography Research. DPA Workstation. Available online at http://www.cryptography.com/technology/dpa-workstation.html

[8] Riscure. Inspector - The Side-Channel Test Tool. Available online at http://www.riscure.com/archive/Inspector_brochure .pdf

[9] Brightsight. Unique Tools from the Security Lab. Available online at http://www.brightsight.com/documents/marcom-materials/Brightsight_Tools.pdf

# Design and Low Power VLSI Implementation of Triple -DES Algorithm

Alexandra Camacho, Isaac Sanchez, Eugene B. John and Ram Krishnan
Department of Electrical and Computer Engineering
The University of Texas at San Antonio
One UTSA Circle, San Antonio, TX 78249-0669

(oci482@my.utsa.edu; isanchez87@gmail.com; eugene.john@utsa.edu; ram.krishnan@utsa.edu)

***Abstract*— Triple DES (Data Encryption Standard) is a widely used encryption algorithm known to achieve good performance and high security. In this paper, we describe the design and low power VLSI implementation of the well-known triple DES algorithm. The implementation includes two main parts: key generation and the encryption/decryption process. In the DES module, the key generation part takes the given key and produces 16 distinct keys to be used during the encryption stage. The DES process is then repeated three times for added security. The chip was implemented using TSMC 180nm process. The designed chip has an area of 766,359 µm$^2$ and the power dissipation is 32.38mW for a Vdd of 1.8V.**

***Keywords-Encryption, DES, 3DES, Low Power, Cryptosystem.***

## I. INTRODUCTION

With today's technology, there is a large amount of data constantly transferred electronically at any given time. It is typical that at some point in the day many people in the U.S. send emails, pay bills, or communicate wirelessly. Needless to say, much of this information is sensitive and the users have a desire for secure transactions. Ideally when data is transferred it is desired that the original information is transformed into apparent nonsense, so that in the event of third parties intercepting the data, it will appear illegible or as complete nonsense. Only the sender and receiver are able to decipher the data using special knowledge of the key used to transform the data. This transformation is called encryption and the algorithm used to map the data to its transformed state is known as a cipher [1]. There are many of such algorithms in existence however the proposed algorithm of interest is referred to as the Data Encryption Standard (DES). It was selected by the National Bureau of Standards as an official Federal Information Processing Standard (FIPS) for the United States in 1976 and which has subsequently enjoyed widespread use internationally [2]. Along with the growth of electronic data transmission was the development of the devices used to transmit the data. These include many processors used to drive data through landlines and wireless mediums. Many of these devices use encryption implemented on hardware for speed and use less computational memory of computers to run the algorithm [3]. As such, on this paper we describe the design and low power VLSI implementation of a DES algorithm for sending plain text information. In addition we seek to improve encryption by building on the DES chip and developing a Triple-DES algorithm that is more secure and commonly used by financial institutions for secure transactions.

## II. BACKGROUND

### A. Data Encryption Standard

The Data Encryption Standard (DES) is a block cipher that uses shared secret encryption. DES is the archetypal block cipher which is an algorithm that takes a fixed-length string of plaintext bits and transforms it through a series of complicated operations into another ciphertext bitstring of the same length. The block size for DES is 64 bits and the cipher text produced appears as nonsense to third parties not intended to intercept the data. DES uses a customizable key for the transformation, so that decryption can supposedly only be performed by those who know the custom key used to encrypt. The key consists of 64 bits; however, only 56 of these are actually used by the algorithm. Eight bits are used solely for checking parity, and are thereafter discarded. Hence the effective key length is 56 bits, and it is never quoted as such. Every 8th bit of the selected key is discarded, that is, positions 8, 16, 24, 32, 40, 48, 56, 64 are removed from the 64 bit key leaving behind only the 56 bit key [2]. The availability of improved computational power eventually made the original DES algorithm less secure as it became more susceptible to brute-force attacks. To counter this, Triple DES was developed as a relatively simple method of increasing the key size of DES to protect against such attacks, without the need to design a completely new block cipher algorithm. Triple-DES has a number of advantages. First, it builds upon the concept of multiple encryption where a single block of plaintext is encrypted 3 times, typically using 2 different keys K1 and K2. Specifically, given a plaintext block, the DES algorithm is run in encryption mode using K1, then run in decryption mode using K2 and finally run in encryption mode using K1. Thus every plaintext block goes through encrypt (K1), decrypt (K2) and encrypt (K1) steps. Second, Triple-DES in encrypt(K1)-decrypt(K2)-encrypt(K1) is backward compatible with single DES since if K1=K2 in this mode, the first two steps cancel each other resulting in single encryption. This allows legacy applications developed to work with single DES to continue to inter-operate with Triple-DES implementations. Finally, Triple-DES when run in this mode with two keys still maintains an effective key size of 112 bits

[4][5]. This is because a meet-in-the-middle known plaintext attack on multiple encryptions still requires an attacker to brute-force through two steps involving two different keys.

### *B. Method and Design*

The hardware implementation was derived from the algorithm described in [7] and [8]. DES uses a 64-bit key to encrypt 64-bit blocks of data through 16 rounds of permutations, xors, and table look-ups. The key is also shifted and permuted at each stage to increase security. DES is a symmetric algorithm, which means that ciphertext created using a particular key can be decrypted into plaintext using the same hardware and key. Figure 1 shows a data flow graph of how DES works.

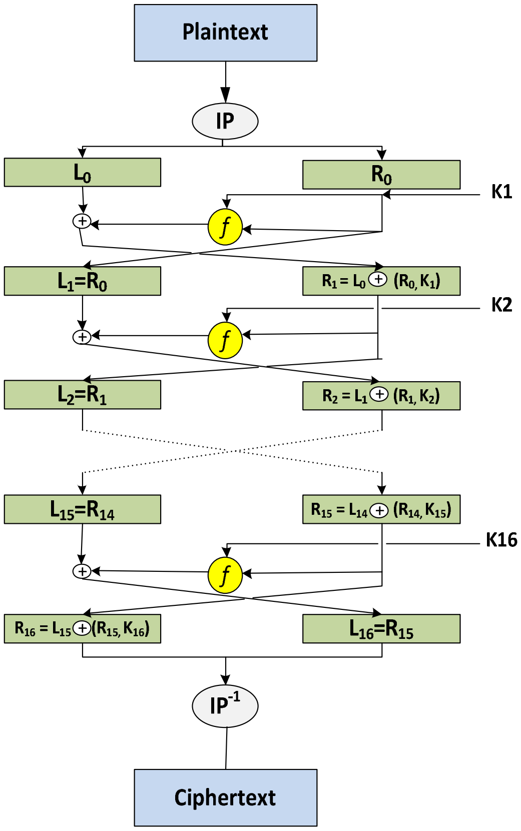
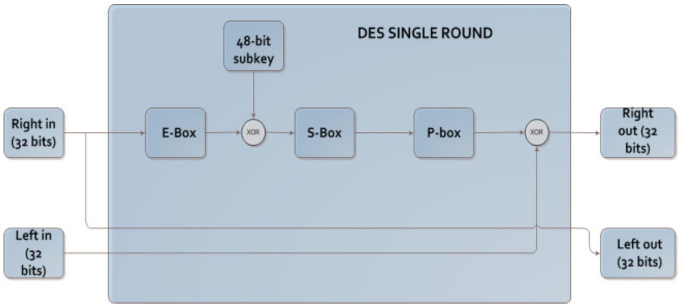

Plaintext
IP
$L_0$
$R_0$
K1
$f$
$L_1$=$R_0$
$R_1$ = $L_0$ ⊕ ($R_0$, $K_1$)
K2
$f$
$L_2$=$R_1$
$R_2$ = $L_1$ ⊕ ($R_1$, $K_2$)
$L_{15}$=$R_{14}$
$R_{15}$ = $L_{14}$ ⊕ ($R_{14}$, $K_{15}$)
K16
$f$
$R_{16}$ = $L_{15}$ ⊕ ($R_{15}$, $K_{16}$)
$L_{16}$=$R_{15}$
$IP^{-1}$
Ciphertext

**Figure 1: DES Data Flow Graph**

During each of the 16 rounds , the data is separated into 32-bit halves. The right half is passed through to become the left half of the next round and is expanded through a permutation to 48-bits before being XORed with the key. The XORed value then becomes the addresses for the 8 sboxes. These sboxes are basically small Look up tables. The output of the sboxes is passed through another permutation before being xored with the left half. This criss-crossing is known as the Feistel scheme. The Feistel structure ensures that decryption and encryption are very similar processes and the only difference is that the subkeys are applied in the reverse order when decrypting. One round of the Feistel Scheme is shown below in Figure 2:

**Figure 2: Feistel Scheme**

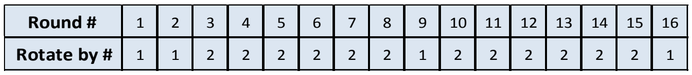


The first step is to pass the 64-bit key through a permutation called Permuted Choice 1, or PC-1 for short. The table for this is given in Figure 4. Note that in all subsequent descriptions of bit numbers, 1 is the left-most bit in the number, and n is the rightmost bit.

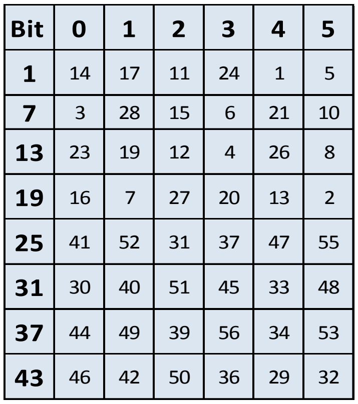
| Bit | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| **1** | 57 | 49 | 41 | 33 | 25 | 17 | 9 |
| **8** | 1 | 58 | 50 | 42 | 34 | 26 | 18 |
| **15** | 10 | 2 | 59 | 51 | 43 | 35 | 27 |
| **22** | 19 | 11 | 3 | 60 | 52 | 44 | 36 |
| **29** | 63 | 55 | 47 | 39 | 31 | 23 | 15 |
| **36** | 7 | 62 | 54 | 46 | 38 | 30 | 22 |
| **43** | 14 | 6 | 61 | 53 | 45 | 37 | 29 |
| **50** | 21 | 13 | 5 | 28 | 20 | 12 | 4 |

**Figure 4: PC-1 Subkey Mapping**

For example, we can use the PC-1 table to figure out how bit 30 of the original 64-bit key transforms to a bit in the new 56-bit key. Find the number 30 in the table, and notice that it belongs to the column labeled 5 and the row labeled 36. Add up the value of the row and column to find the new position of the bit within the key. For bit 30, 36 + 5 = 41, so bit 30 becomes bit 41 of the new 56-bit key. Note that bits 8, 16, 24, 32, 40, 48, 56 and 64 of the original key are not in the table. These are the unused parity bits that are discarded when the final 56-bit key is created. Now that we have the 56-bit key, the next step is to use this key to generate 16 48-bit subkeys, called K[1]-K[16], which are used in the 16 rounds of DES for encryption and decryption. These are generated by splitting the current 56-bit key, K, into two 28-bit blocks, L and R. Using the subkey rotation table in Figure 5 below, L and R are shifted by the number of bits indicated by the subkey index. L and R are then rejoined and permuted choice 2 (PC-2) is applied as PC-1 was before to get the final resulting subkey. Sixteen 48-bit subkeys are gereated for each round of the DES algorithm.



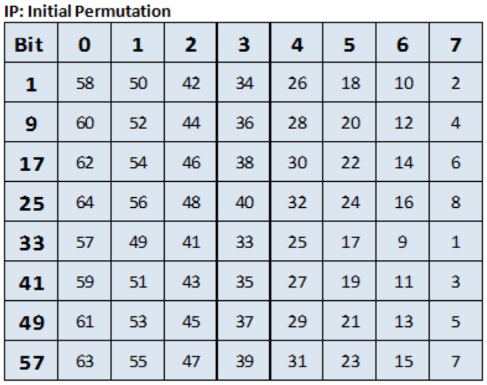
| Round # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Rotate by # | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 |



| Bit | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 1 | 14 | 17 | 11 | 24 | 1 | 5 |
| 7 | 3 | 28 | 15 | 6 | 21 | 10 |
| 13 | 23 | 19 | 12 | 4 | 26 | 8 |
| 19 | 16 | 7 | 27 | 20 | 13 | 2 |
| 25 | 41 | 52 | 31 | 37 | 47 | 55 |
| 31 | 30 | 40 | 51 | 45 | 33 | 48 |
| 37 | 44 | 49 | 39 | 56 | 34 | 53 |
| 43 | 46 | 42 | 50 | 36 | 29 | 32 |

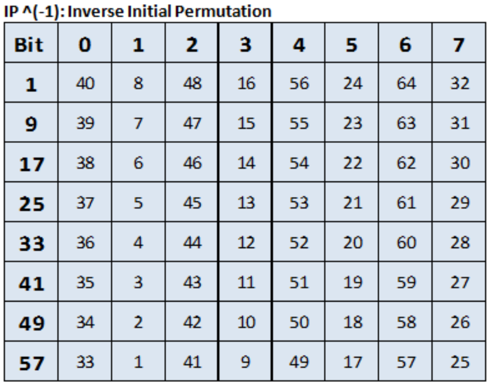
**Figure 5: Rotation table and PC-2**

As stated earlier the original 64-bit block of data is split into two 32-bit halves. Before that, the data undergoes an initial permutation similar to PC-1 and PC-2. The inverse of this permuation is performed at the end of the 16 rounds of DES algorithm. The initial permutation and its inverse are shown in Figure 6.



IP: Initial Permutation

| Bit | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 1 | 58 | 50 | 42 | 34 | 26 | 18 | 10 | 2 |
| 9 | 60 | 52 | 44 | 36 | 28 | 20 | 12 | 4 |
| 17 | 62 | 54 | 46 | 38 | 30 | 22 | 14 | 6 |
| 25 | 64 | 56 | 48 | 40 | 32 | 24 | 16 | 8 |
| 33 | 57 | 49 | 41 | 33 | 25 | 17 | 9 | 1 |
| 41 | 59 | 51 | 43 | 35 | 27 | 19 | 11 | 3 |
| 49 | 61 | 53 | 45 | 37 | 29 | 21 | 13 | 5 |
| 57 | 63 | 55 | 47 | 39 | 31 | 23 | 15 | 7 |

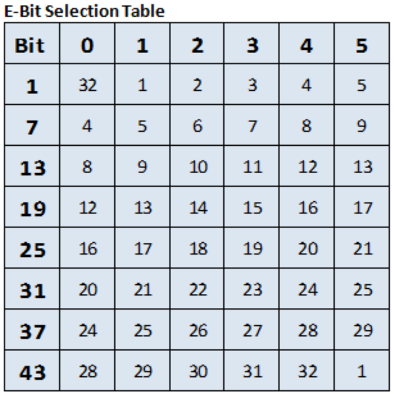


IP ^(-1): Inverse Initial Permutation

| Bit | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 1 | 40 | 8 | 48 | 16 | 56 | 24 | 64 | 32 |
| 9 | 39 | 7 | 47 | 15 | 55 | 23 | 63 | 31 |
| 17 | 38 | 6 | 46 | 14 | 54 | 22 | 62 | 30 |
| 25 | 37 | 5 | 45 | 13 | 53 | 21 | 61 | 29 |
| 33 | 36 | 4 | 44 | 12 | 52 | 20 | 60 | 28 |
| 41 | 35 | 3 | 43 | 11 | 51 | 19 | 59 | 27 |
| 49 | 34 | 2 | 42 | 10 | 50 | 18 | 58 | 26 |
| 57 | 33 | 1 | 41 | 9 | 49 | 17 | 57 | 25 |

**Figure 6: Initial Permutation and its inverse.**

Whenever the permuted data is split, the right half is expanded from 32 bits to 48 bits using the expansion table in Figure 7. The original right half is stored to become the left half for the next round in the DES algorithm later on. Notice that the expansion table includes repeat of bits.

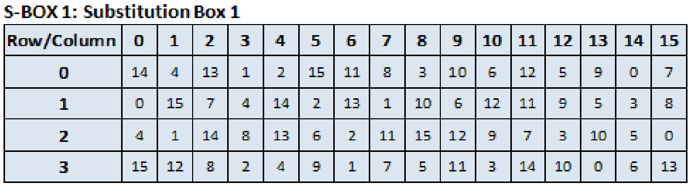


E-Bit Selection Table

| Bit | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 1 | 32 | 1 | 2 | 3 | 4 | 5 |
| 7 | 4 | 5 | 6 | 7 | 8 | 9 |
| 13 | 8 | 9 | 10 | 11 | 12 | 13 |
| 19 | 12 | 13 | 14 | 15 | 16 | 17 |
| 25 | 16 | 17 | 18 | 19 | 20 | 21 |
| 31 | 20 | 21 | 22 | 23 | 24 | 25 |
| 37 | 24 | 25 | 26 | 27 | 28 | 29 |
| 43 | 28 | 29 | 30 | 31 | 32 | 1 |

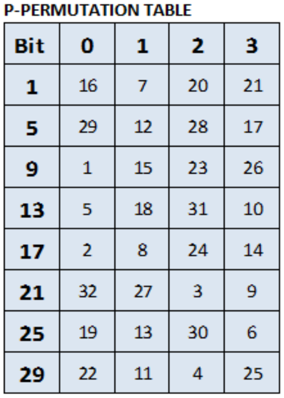
**Figure 7: Expansion Table**

The output of the expansion block is xored with the 48-bit subkey previously generated for that round. This result is split into eight 6-bit blocks. Each block is used as an address for its corresponding s-box look up table, therefore there are eight s-box tables. Below s-box 1 is shown. The first and last bit of the 6-bit blocks are used to index the row of the table. The middle 4 bits index the columns. The value at those indices are concatenated with remaining s-box results. The eight addresses produces 4 bit values each, so the resulting data is 32-bits. At this point another permuation is performed using the p-box table in Figure 8. This result is xored with the original left half during the round and the result becomes the right half for the next round and as stated earlier the next rounds left half is the current rounds initial right half.



S-BOX 1: Substitution Box 1

| Row/Column | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 14 | 4 | 13 | 1 | 2 | 15 | 11 | 8 | 3 | 10 | 6 | 12 | 5 | 9 | 0 | 7 |
| 1 | 0 | 15 | 7 | 4 | 14 | 2 | 13 | 1 | 10 | 6 | 12 | 11 | 9 | 5 | 3 | 8 |
| 2 | 4 | 1 | 14 | 8 | 13 | 6 | 2 | 11 | 15 | 12 | 9 | 7 | 3 | 10 | 5 | 0 |
| 3 | 15 | 12 | 8 | 2 | 4 | 9 | 1 | 7 | 5 | 11 | 3 | 14 | 10 | 0 | 6 | 13 |



P-PERMUTATION TABLE

| Bit | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 1 | 16 | 7 | 20 | 21 |
| 5 | 29 | 12 | 28 | 17 |
| 9 | 1 | 15 | 23 | 26 |
| 13 | 5 | 18 | 31 | 10 |
| 17 | 2 | 8 | 24 | 14 |
| 21 | 32 | 27 | 3 | 9 |
| 25 | 19 | 13 | 30 | 6 |
| 29 | 22 | 11 | 4 | 25 |

**Figure 8: S-Box and P-Box**

The previous steps are repeated for the duration of the 16 rounds of the DES algorithm. The ciphertext is produced when the inverse initial permutation block is applied. We are easily able to implement a triple DES by applying DES 3 times to 64 bit plaintext data with 3 different keys and sets of permutations each time the DES is applied.

## III. Experimental Methods and Results

In this section we describe the experimental methods and the design results. Figure 9 depicts the synthesized top level DES module. The inputs include the 64-bit key and plain text data, a clock, enable, encrypt/decrypt and reset. By toggling the encrypt/decrypt input from 0 to 1, we can set the DES to encrypt the plain text or decrypt the cipher text. Decrypting works exactly in the same manner; however, the subkeys are applied in reverse order. The image on the right shows the next level schematic which is a more detailed view of all necessary connections and modules instantiated for the full 16 rounds. The rounds and constituting sub modules are explained in more detail in the following paragraphs.

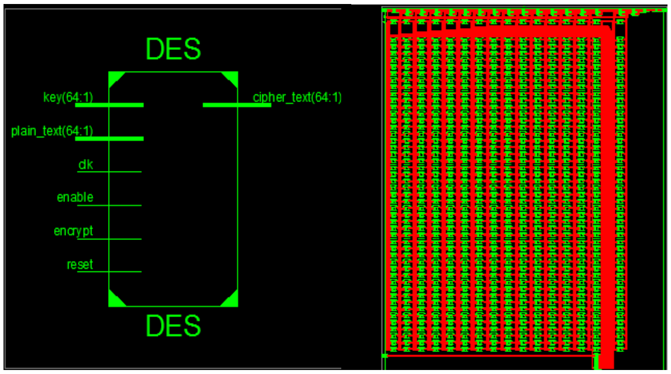


**Figure 9: DES Module (Left): Top Module (Right): Next Level**

Figure 10 shows the schematic of 1 of the 16 rounds in the DES algorithm. Notice the 8 S-boxes. The original algorithm flow chart is also shown to show the translation.

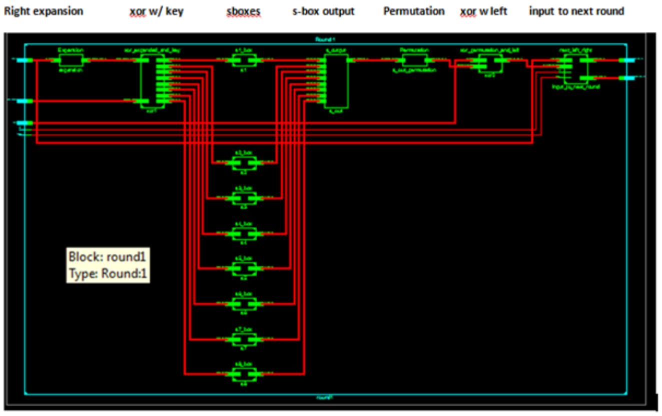


**Figure 10: Single Round of DES Algorithm**

Figure 11 shows the top level schematic of the Triple DES. It works exactly the same, but requires 3 keys to be input into the top module.

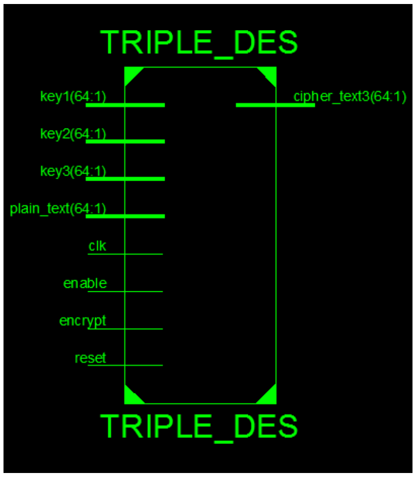
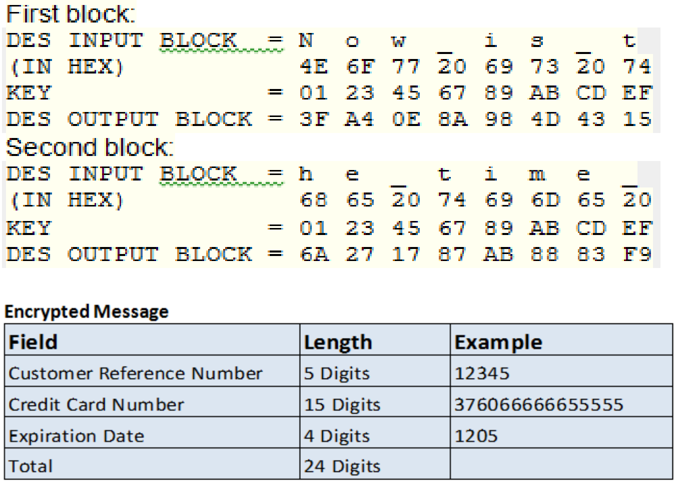


**Figure 11: Triple DES Top Module**

**Behavioral Test:**

In order to test the functionality of our design we used the data shown in Figure 12. The top portion of the figure shows 2 blocks of plain text data to test the DES algorithm. Together the blocks form the string "Now is the time". As shown, the string is broken up into 8 byte blocks, since the DES algorithm is designed for 64bit data. Each character is represented in Hex as well. The bottom portion of Figure 12 shows an example of credit information that is generally encrypted using the triple DES algorithm. We sought to successfully encrypt and decrypt this data.



First block:

```
DES INPUT BLOCK  = N  o  w  _  i  s  _  t
(IN HEX)           4E 6F 77 20 69 73 20 74
KEY              = 01 23 45 67 89 AB CD EF
DES OUTPUT BLOCK = 3F A4 0E 8A 98 4D 43 15
```

Second block:

```
DES INPUT BLOCK  = h  e  _  t  i  m  e  _
(IN HEX)           68 65 20 74 69 6D 65 20
KEY              = 01 23 45 67 89 AB CD EF
DES OUTPUT BLOCK = 6A 27 17 87 AB 88 83 F9
```

Encrypted Message

| Field | Length | Example |
|---|---|---|
| Customer Reference Number | 5 Digits | 12345 |
| Credit Card Number | 15 Digits | 376066666655555 |
| Expiration Date | 4 Digits | 1205 |
| Total | 24 Digits | |

**Figure 12: Behavioral Data**

The correct operation of this circuit was verified by running simulations for both the single DES and Triple-DES designs. The encryption input was switched between 1 and 0 to encrypt/decrypt the information seen in Figure 12.

### *Final Layout*

Figure 13 shows the final layout of the Triple DES chip that was synthesized using Cadence Encounter and RTL compiler for TSMC 180nm process. Through the power and timing analysis we were able to distinguish the specs for the Triple DES and Single DES by analyzing the single DES modules.

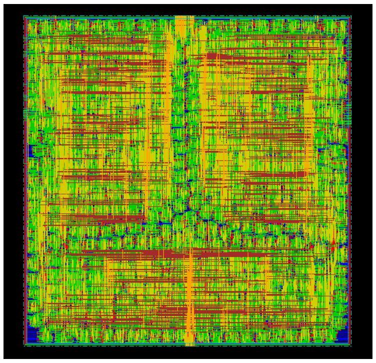


**Figure 13: GDS2 Layout of Triple DES Chip**

The specs for the designed Triple-DES chip are as follows: (1) Designed using TSMC 180 nm process (2) Required VDD: 1.8V (3) Area: 766,359 µm$^2$ (4)Total number of pins on chip: 326 (Three 64bit Keys, One 64bit plaintext input, One 64bit ciphertext output and 6 additional pins for clk, enable, encrypt, vdd, and gnd) (5) Leakage power dissipation of the chip is 20.72 µW and the switching power dissipation is 32.36 mW. The total power consumption of the Triple-DES is 32.38 mW. A summary of the specifications is given in Table 1.

| Chip Specifications | |
|---|---|
| Technology | TSMC 180 nm |
| VDD | 1.8 V |
| Area | 766,359 µm2 |
| Number of pins | 326 |
| Leakage Power | 20.72 µW |
| Switching Power | 32.36 mW |

**Table 1: Triple-DES Chip Specifications**

## IV. Conclusion

In this paper we present the design and low power VLSI implementation of Single DES and Triple-DES algorithm. Since the original DES algorithm was less secure as it became more susceptible to brute-force attacks, triple DES was developed as a relatively simple alternative without the need to design a completely new block cipher algorithm. In our design we used 3 instances of our Single DES algorithm to improve the encryption strength. It should be noted that the Triple DES implementation is considerably slower, but more secure. The designed chip can be used for financial data transfer such as credit card information where a high level of security is needed. The chip was designed using TSMC 180nm process for 1.8 V Vdd. The area of the designed chip is 766,359 µm$^2$, the number of pins on the chip is 326, and the power dissipation is 32.38mW.

## References

[1] Bellare, Mihir; Rogaway, Phillip (21 September 2005). "Introduction". Introduction to Modern Cryptography. p. 10

[2] NBS FIPS PUB, ªData Encryption Standard,ºNat'l Bureau of Standards, US Dept. of Commerce, Jan. 1977.

[3] http://www.via.com.tw/en/initiatives/padlock/whyhardwareisbetter.jsp

[4] Ralph Merkle, Martin Hellman: On the Security of Multiple Encryption , Communications of the ACM, Vol 24, No 7, pp 465–467, July 1981.

[5] Paul van Oorschot, Michael J. Wiener, A known-plaintext attack on two-key triple encryption , EUROCRYPT'90, LNCS 473, 1990, pp 318–325.

[6] NIST Special Publication 800-57 Recommendation for Key Management Part 1: General (Revised), March, 2007

[7] www.jhdl.org/documentation/latestdocs/code/JHDL_examples1.html

[8] http://www.tropsoft.com/strongenc/des.html

# SESSION

# SECURITY APPLICATIONS I

## Chair(s)

**Dr. Kathy Liszka**

# Security in the Cloud: A stake holder's perspective

Vivek Shandilya, *Student Member, IEEE*, and Sajjan Shiva, *Fellow, IEEE*

*Abstract*—**The cloud computing is a paradigm involving many disparate stake holders. Any system built upon this paradigm and a business run on such system architecture would have similarly disparate scope of access, activities and responsibilities for the stake holders. Depending on the scope of the activities and responsibilities, a stakeholder has to device the security measures to secure his activity space. We formulate the problem formally with this premise. We model and survey, depending upon the generic cloud architecture abstracted from the prominent ones, the actvity space of each stake holder and consequently security measures each stake holder has to take, to ensure the over all security of the system. We survey the recent works related to each facet of the formal model and classify them to present a contemporary perspective.**

*Index Terms*—**Coud computing, Formal representation, Security, Activity Space, Best Practices, Stake holders.**

## I. Introduction

Cloud computing is becoming popular as the infrastructure solution for many medium and small scale business companies all over the world. This trend is a game changing paradigm shift. Even though the underlying principle of operation of individual constituents is same, since the cloud architectures differ from each other in their construction details, the exact access and action spaces also vary for the stake holders operating on systems built on different clouds.

But considering a generic cloud architecture which absorbs the essential features of most cloud computing infrastructures, we can identify the prominent stake holders in a business system built over such a cloud. Then we can determine the scope of access, activities and controls each of them have. On that basis we can determine who is responsible for what security aspects in the whole operation, and what action each of the stake holder needs to take to ensure security in his own space and eventually the whole system. There are many cloud service providers with varied features. Many of them focus on a set of niche target customers and specialize on the features that are important to such clientle. Though such differences make each cloud service provided by different service providers different, we can form a generic set of features that constitutes a cloud by taking the commonalities of the major cloud services. The major stake holders are the end user of the business, the business organization, the network (Internet) provider facilitating the communication to and from the cloud with the organization, the cloud service provider, the provider of the operating system and other software used to construct the cloud, the hardware manufacturers and the governments who administer and monitor the legalities as applicable to the whole operation. From the systemic point of view, the key issues of security and privacy are identified and is argued to be a problem of risk management in [1]. It is not only important to study the security from the point of view of the systems that functionally constitute the whole operation built on a cloud infrastructure but also from the perspective of the above mentioned stakeholders so that a framework for creating best practices, policies and laws would be facilitated in an informed way. In this paper we study the scope of security measures for each of the stake holders based on the scope of their activity in the cloud based business systems.

The main contributions of this paper are as follows.

1. We formally represent the stake holders and their profiles in a business system based on a generic cloud computing infrastructure.
2. We delineate the security concerns and responsibilities of each stake holder to ensure the overall security of the business, based on the survey of current literature.

The Section II describes a formal representation of the security-activity profile for a generic cloud based system with the perspective of the stake holders and their activity space.The section III describes in detail, the generic architecture of cloud computing services used by the businesses and their security concerns. Section IV describes the activity spaces and profile of the stake holders in the cloud based system. Section V describes the security activity spaces and profile of the cloud based systems. The section VI discusses the related works. The section VII presents the conclusion.

## II. Formal representation of the security profile of a cloud

In this section, we formulate the structure of the security problem in a cloud-based system formally.
The cloud based system involves $n$, where $0 < n < \infty$, different stake holders. The $i^{th}, 1 < i \leq n$ stake holder $h_i$ has access to an unambiguously defined part of the system. This is referred to as the **access space** $s_{i_A}$ of the stake holder $h_i$.

This access space $s_{i_A}$, is considered as the set of $m$ areas. $s_{i_A} = \{s_{i_{A_1}}, s_{i_{A_2}}, s_{i_{A_3}}, \ldots, s_{i_{A_j}}, \ldots, s_{i_{A_m}}\}, 1 < j \leq m, 1 < m \leq \infty$.

In each of these areas, the stake holder $h_i$ can choose to do one action, at a time, out of the clearly defined *set of actions*, $s_{i_{a_j}}$ corresponding to $j$. The ordered pair of this *access area* and the corresponding *set of action* is referred to as the activity area $s_{i_{\alpha_j}} = (s_{i_{A_j}}, s_{i_{a_j}})$.
The union of all such sets of actions available at, that is, corresponding to, each of the areas in the access space is

Vivek Shandilya is with the Department of Computer Science, University of Memphis, Memphis, TN 38152–3240. Phone: +1 901 848–1763, e-mail: vhmsndly@memphis.edu

Sajjan Shiva is with the Department of Computer Science, University of Memphis, Memphis, TN 38152–3240. Phone: + 90 1 678–5465, fax: + 90 1 678–1506, e-mail: sshiva@memphis.edu

referred to as the stake holder's action space $s_{i_a}$ .

$s_{i_a} = \bigcup_{j=1}^{m} s_{i_{a_j}}$

The union of all sets of access areas is the access space of the stake holder $h_i$.

$s_{i_A} = \bigcup_{j=1}^{m} s_{i_{A_j}}$

The union of all the activity areas of the stake holder $h_i$, is called the activity space of $h_i$.

$s_{i_\alpha} = \bigcup_{j=1}^{m} s_{i_{\alpha_j}}$

The sequence of all the activity spaces of the stake holders is referred to as the activity profile $S_\alpha$ of the cloud.

$S_\alpha = (s_{1_\alpha}, s_{2_\alpha}, \dots, s_{n_\alpha})$

The security measures each stake holder has to take must be accomplish-able with some combination of the legal actions that are available to him. It may involve one or more actions in tandem to be taken in each access area. The security measures in the access area $s_{i_{A_j}}$ the stake holder $h_i$ has to take be the set $r_{i_{A_j}}$,

such that $r_{i_{A_j}} \subseteq s_{i_{a_j}}$.

Thus, the security activity a stake holder does in the access area $s_{i_{A_j}}$ is given by the tuple $r_{i_\alpha}$,

such that $r_{i_{\alpha_j}} = (r_{i_{A_j}}, r_{i_{a_j}})$.

Any practical security measure generally involves a sequence of actions across many activity areas to be effective. And the total security activity of the stake holder $h_i$ is given by $r_{i_A}$,

such that $r_{i_\alpha} = \bigcup_{j=1}^{m} r_{i_{\alpha_j}}$

The sequence of all the security activity spaces of the stake holders is referred to as the security activity profile $S_\alpha$ of the cloud.

$R_\alpha = (r_{1_\alpha}, r_{2_\alpha}, \dots, r_{n_\alpha})$

A security measure by a stake holder can be cognized as an a priori plan of security actions at each access area. We distinguish security measure from the defense measure which happens, that is the series of executions of actions, *while* defending a system against a malfunction either due to an internal system fault or a malicious attack. Thus, this gives the context to make the plan considering the combined general risks and threats depending on the given system characteristics. Since there are different stake holders with having their own access spaces, either intersecting or not with that of other's, each must take their own security measures. This is described below. A security measure of the stakeholder $h_i$, is given by a sequences of actions taken at each instant of operation, that is, it has a time stamp in the operations, chosen from the security profile $r_{i_\alpha}$. This security activity profile gives a formal way to represent quantitatively who can do what and where to defends which assets and how.The above frame work would be useful for the policy makers to confirm what security actions each stake holder would have to take. In the next section, we shall present a survey of the current literature how the activity space and security activity spaces are being used by stake holders in different scenarios.

## III. Cloud based systems: architecture and security concerns

The architectures and the security concerns related to them go hand in hand. The more sophisticated the architecture is the more involved will be the security concerns.

### A. Architectures

The general security and privacy concerns were identified and discussed along with the direction of research addressing them in [2]. They categorize the concerns as traditional security, availability and thried-party data control concerns. There is a new class of problems that are identified which are *cheap data and data analysis, cost effective defense of availability, increased aunthetication demands and mash-up authorization*. The *cheap data and data analysis* of unheard proportions enable even scantly equipped attacker huge information-advantage enabling a sophistication in the attack.*Cost-effective defense of availability* is a concern dealing with the counter measures against an attacker with the sole motive to sabotage activities. Since any disruption yeilds a positive payoff for the attacker, cloud crystallizes the problem to be that of a single point of failure. *Increased authentication demands* encourage the use of thin client at the client side. This emphasizes increased authentication demands on the cloud side.The cloud model encourages users to mash-up their data. *Mash-up authorization* will lead to problems of data-leaks and in terms of the number of sources of data a user may have to pull data from. To address these concerns *information-centric security, high-assurance remote server attestation and privacy-enhanced business intelligence* are identified as the fertile fields for further research.

The security concerns of each of the components constituting the cloud based system not only are pertinent but also new concerns are emerging. One of them is detectability of the hardware infrastructures hosting virtual machines (VM) delivering the payload of a client, leading to cross-VM side-channel attacks to extract information from another target VM on the same machine.This was explored practically by devising an attack launched over Amazon's EC2 to establish the vulnerability by [3].Its a potential breach to be careful about.

The main cloud computing security issues are fundamentally not new, and are tractable to the concerns in the previous time sharing era. But the complexities of multi party trust concerns, and ensuring the need for mutual audit-ability are distinct to cloud computing. These ideological analyses are done by [4].

An analyses of the major cloud provider for performance for the cost and the types of services offered was done in [5].

The vulnerability of using the standard TCP's congestion control is analyzed and showed to provide opportunities of DOS attacks and as an alternative, a network bandwidth allocation scheme called Seawall is presented in [6].

The configuration of the complex cloud infrastructure is important to be such that, it does not provide any security holes to be exploited to gain undue information by malicious activity. An automated auditing process is provided to check the configuration in [7]

### B. Security concerns

The main security risks concerning operations over cloud were topically pointed out in [8]. The security concerns for an enterprise wanting to move their infrastructure over a cloud, and the resulting risk management are discussed in [9]. The technicalities of the security threat are for most part a reincarnation of the previously known security issues in the older classical computing paradigms. An extension of such previous issues into modern cloud computing security issue is suggested to be mitigated by measures which also are inspired from the extending the previously done counter measures. The distinction of novelty of particular issues is discussed and presented in [10].

## IV. Activity spaces

In this section we survey the current literature to get an idea of what the generic activity spaces are in the clouds. It is true that most of the cloud service providers tune up their system architecture to cater some specialized or emphasized services leading to disparate system setups. But we shall take a general consideration and pick the common factors to consider a generic cloud, for our analysis, which can be later customized diligently to any specific architecture with exact details filled in.

The role of the distributed data locations in the cloud architecture is provided and the need for data location compliance is studied in the thesis in [11]. In the Amazon EC2 in [12], Microsoft Azure and other prominent clouds used by many small and medium sized business systems, the cloud service provider takes care of all the hardware, either distributed or not, and the basic operating system that boots those hardware. Some times these hardware are provided by a third party. In that case, both the hardware and the software that boots on those hardware are provided by third parties. Over this operating system a hyper-wiser is used and controlled by the cloud service provider. This is the case in many public and (voluntary and non-profit) community based clouds. Then the cloud service provider provides a actualization through the hyper-wiser and a running operating system over it. This is going to be maintained by the cloud service provider, by installing regular updates and so on. Some of the software applications running on this operating system, as requested by and provided to, the business clients is given as their access space and activity space. Everything below this would be the access space of the cloud service provider if he has not outsourced the hardware and its maintenance to a third party. The business companies will install, run, and configure their applications and that whole space becomes their access area. The end users and clients of the business will be dealing with the companies, by accessing some instances of the software processes they are given, and the data structures they are entitled to access. This forms their access space and activity space.

### A. Activity Profile

Activity profile is illustrated in the following works.

The analysis of mutual dependencies and the trustiness resource legitimacy in cloud computing is provided in [13]. Functionally, the activity profile imply the trustiness in an implicit manner. That is, each stake holder has agreed to do his bit and believes the others shall do theirs. This is reflected in many works that discuss how this trust is established, sustained, verified, maintained and actuated. The more complex constructions would be in inevitable but shall provide new challenges. Such moves will alter the activity profiles enormously. One such case is presented here. At the data link layer and network layer of the cloud, the trends of hybrid electrical/optical data-center networks pose many new challenges. An analyses of these with suggestions and directions towards plausible solutions is in [14]. Trusted block as a service in the context of cloud is discussed in [15]. A new trust model for file sharing in the cloud is discussed in [16]. Virtual machines need accounting and monitoring for ensuring authenticity and integrity. Eventually this should lead to bringing reliability, transparency and security in client model for client satisfaction. To do this a mobile agent based architecture is proposed which can dynamically move in the network to accomplish this task. The trust between the stake holders is an important issue. To dynamically assess it and implement decisions based on it, a mechanism is proposed based on mobile agents to collect the information in [17].

As a main concern of security, data integrity in cloud is important. For that, using third party data integrity Management Service (IMS) has its draw backs. To avoid it, a different cloud storage architecture was proposed with services and protocols and implemented on Amazon S3, with favorable results as presented in [18]. To address the problem of secure data transfer a trust-based file sharing is used. There are many open issues related to this approach which are discussed a new model is proposed in [16] Trustworthy clouds underpinning the future Internet from an overarching perspective is discussed in[19]. The data of the end user is kept on the resources of the cloud provider. to make is safe the user could encrypt his data. If the user wants to do any computation using that data on cloud resources itself, then he has to decrypt it and do the computation with the data. This nullifies the privacy of data. To avoid this, a homomorphic encryption scheme is proposed which allows the computations of data in its encrypted form in [20] To ensure the privacy for the data being stored over cloud by the user, the cloud service provider cannot be implicitly trusted. For this a model based upon the principle of dynamic data re-encryption is proposed in [21]. Another work on the business process as service and about remote auditing is presented in [22]. Accountability, Audit-ability and Trust between the stake holders is analyzed in [23]. TrustCloud, a frame work and a system was provided by HP in [23] to establish trust and co-ordinate the stake holders in cloud.

flexibility and low performance overhead.

## V. Security Activity Spaces

A comprehensive picture of the cross organizational, and the disparate stake holders, and their roles in the security activity is discussed in relation to cross-organizational security settings in [24] and their key roles elaborated in [25]. Public survey was made in two countries and an on line survey in more that 150 countries to identify the perception of the end users about the cloud. The results are discussed in [26], as the impression is a key factor in how the public users eventually adopt and turn out to use it. An Analysis of Security and Privacy Issues in Smart Grid Software Architectures on Clouds is discussed in [27] and in [28]. Many parts of the clouds are very convenient to use for attacks leading to breach of privacy and confidentiality of others and unauthorized possession of copyrighted material. An example analysis of Dropbox is done in [29]. [30] reports on legal , privacy, security, access and regulatory issues. This paper raises an awareness of legal, privacy, security, access and regulatory issues that are associated with the advent of cloud computing. An in-depth literature survey is conducted on these and an analysis is drawn from the issues that are identified through the literature survey. Recommendations are then given on how the issues identified in the analysis can be mitigated. The issues of policy interventions, standards, privacy and data protection, traffic and congestion management, business continuity planning, security and regulation are discussed.

Working in various service models ranging from SaaS, PaaS, to IaaS, of cloud computing to mitigate data abuse, encryption is suggested. With data encryption, an issue arises when the data owner who outsourced the data wants to revoke some data-consumers' access privileges, which normally involves key re-distribution and data re-encryption. In this work, a generic scheme was proposed to enable fine-grained data sharing over the cloud, which does not require key-redistribution and data re-encryption whatsoever. The main primitives made use of are attribute-based/predicate encryption and proxy re-encryption, but our construction is not restricted to any specific scheme of its kind. A generic key distribution scheme with a number of advantages over other similar proposals in the literature in given in [31]. The methods of homomorphic encryption and computing on tamper-resistant hardware suffer from high latency. For outsourcing the data and arbitrary computation with lower latency a token based method was proposed in [32].

### A. Security Activity Profile

The data availability issues in cloud computing is discussed and a RAID based model is proposed to address these issues along with error correction using parity encoding of the data in [33]. A rule-based-forwarding network design and an access control mechanism CloudPolice implemented in hypervisor to top DoS attacks is discussed in [34] To mitigate the security concern for running an application over cloud, the program is suggested to be run in two pieces as a user and protected program. It was shown to be computationally secure in [35]. The cloud-based infrastructure has to be eventually used by the business community. Those business which need to be ensured about the security can have to work flow certified by the auditors which is then accessible to the users to be ensured of the security. This end user perspective security measure is discussed in [36]. In massive parallel processing scenarios of cloud computing forensics are a challenge.An over view of the cloud Forensics is given in [37]. Technical issues with forensic in cloud is elaborated in [38]. Isolating instances in cloud for forensic is discussed in [39]. A survey on cloud forensics and critical criteria for cloud forensic capability is presented with a preliminary analysis in [40]. Mutual protection of the stake holders is an important security measure and is discussed in [41]. While many works deal with general issues, a specific attack like DoS is addressed in the following work in [42]. Based on the analysis of several recent attack scenarios, a system that enables periodic and necessity-driven integrity measurements and remote attestations of vital parts of cloud computing infrastructures is provided. It was implemented on top of Xen Cloud Platform and trusted computing technology is used to provide guarantees. The work show how system attests the integrity of a cloud infrastructure even in the presence of DoS attack.

## VI. Related Work

The trust is an important factor between the many stake holders in cloud computing. The policy-making approaches with control mechanisms in place is discussed in [23] The security challenges and the recommended management models to address them are discussed in [43] An analysis of the cloud computing security problem was done in [44], where the perspective from the architecture of the cloud, services delivery models and the stake holders involved were listed. The analysis showed that the problem has a multi-layered and multi-perspective nature. In our work we are building upon what they identified as the stake holder's perspective and expanding the analysis of what factors affect it.

The cloud is used to provide different services, like infrastructure, services, etc and based on its location is classified as public, private and hybrid. Each of these distinctions get their own identity reflected in distinct security issues. The analysis of the security problem in cloud based systems is done based on the issues that have come up and different security models proposed to mitigate them are listed in [45]. A general discussion of the security challenges were presented in [46] where the information security, network security and the process and data security issues were identified.

## VII. Conclusion

The feature of security is resolved in terms of how each stake holder is making secure practices and also how each of them are co-ordinated. An over all regulation is required to channelize the development and thus a need for a regulatory body is discussed in [47]. Thus, just the way the functional scope is now distributed amongst each of the stake holders, the security or the breach of it depends on each of them. This knowledge is important to decide the best practices and policies for each of the stake holders.

## References

[1] W A Jansen, "Cloud hooks: Security and privacy issues in cloud computing," Jan. 2011, pp. 1–10.

[2] *Controlling Data in the Cloud: Outsourcing Computation without Outsourcing Control.* ACM, 2009.

[3] Thomas Ristenpart, Eran Tromer, Hovav Shacham, and Stefan Savage, "Hey, you, get of my cloud: Exploring information leakage in third-party compute clouds," *CCS'09: Proceedings of 16th ACM conference on Computer*, 2009.

[4] Yanpei Chen, Vern Paxson, and Randy H. Katz, "What's new about cloud computing security?," *Technical Report No. UCB/EECS-2010-5*, 2010.

[5] Ang Li, Xiaowei Yang, Srikanth Kandula, and Ming Zhang, "Cloudcmp: Comparing public cloud providers," *IMC'10, Melborne, Australia*, 2010.

[6] Alan Shieh, Srikanth Kandula, Albert Greenberg, Changhoon Kim, and Bikas Saha, "Sharing the data center network," *NSDI*, 2011.

[7] Soren Bleikertz, Matthias Schunter, Christian W. Probst, Dimitrios Pendarakis, and Konard Eriksson, "Security audits of multi-tier virtul infrastructures in public infrastructure clouds," *CCSW'10, Chicago, Illinois, USA.*, Oct2010.

[8] Jon Brodkin, *Gartner: Seven cloud-computing security risks*, Retrived on 15 Feb, 2012. InfoWorld.com.

[9] Anthony Bisong and Syed M. Rahman, "An overview of the security concerns in enterprise cloud computing," *International journal of network security & its application*, vol. 3, no. 1, Jan 2011.

[10] F Maggi and S Zanero, "Rethinking security in couldy world," *Technical Report 2010, Departmento do Electronica e Informazione, Politecnico di Milano*, 2010.

[11] J. Noltes, "Data location compliance in cloud computing," M.S. thesis, University of Twente, Aug. 2011.

[12] Amazon.com, "Amazon web services: Overview of security processes," May2011.

[13] J. P. Yoon and Zhixiong Chen, "Service trustiness and resource legitimacy in cloud computing," Jan. 2011, pp. 250–257.

[14] Hamid Hajabdolali Bazzaz, Malaveeka Tewari, Guohui Wang, George Porter, T. S. Eugene Ng, David G. Andersen, Micheal Kamisky, Micheal A Kozuch, and Amin Vahdat, "Switching the optical divide: Fundamental challenges for hybrid electrical/optimal datacenter networks," *SOCC'11, Cascais, Portugal*, 2011.

[15] Jianan Hao and Wentong Cai, "Trusted block as a service: Towards sensitive applications on the cloud," *IEEE TrustCom/IEEE ICESS/FCST, International Joint Conference of*, vol. 0, pp. 73–82, 2011.

[16] Edna Dias Canedo, Robson de Oliveira Albuquerque, and Rafeal Timoteo de Sousa Junior, "Review of trust-based file sharing in cloud computing," *The fourth International Conference on Advances in Mesh Networks*, 2011.

[17] Priyank Singh Hada, Ranjita Singh, and Mukul Manmohan, "Article: Security agents: A mobile agent based trust model for cloud computing," *International Journal of Computer Applications*, vol. 36, no. 12, pp. 12–15, December 2011, Published by Foundation of Computer Science, New York, USA.

[18] S. Nepal, Shiping Chen, Jinhui Yao, and D. Thilakanathan, "Diaas: Data integrity as a service in the cloud," in *Cloud Computing (CLOUD), 2011 IEEE International Conference on*, july 2011, pp. 308–315.

[19] Rudigger Glott, Elmar Husmann, Ahmad-Reza Sadeghi, and Matthias Schunter, "Trustworthy clouds underpinnning the future internet," *Future Internt Assembly,*, vol. LNCS 6656, pp. 209–221.

[20] Aderemi A Atayero and Oluwaseyi Feyisetan, "Security issues in cloud computing: The potentials of homomorphis encryption," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 2, no. 10, pp. 546–552, October 2011, CSI Journal.

[21] Pitor Tysowski and M. A. Hasan, "Towards secure communication for highly scalable mobile applications in cloud computing systems," *Centre for Applied Cryptographic Research University of WAterloo, Tech Rep*, vol. CACR2011-33, 2011.

[22] R. Accorsi, "Business process as a service: Chances for remote auditing," 2011.

[23] Ryan K. L. Ko, Bu Sung Lee, and Siani Pearson, "Towards achieving accountability, auditability and trust in cloud computing," in *Advances in Computing and Communications*, vol. 193 of *Communications in Computer and Information Science*, pp. 432–444. Springer Berlin Heidelberg, 2011.

[24] Stefan Thalmann, Daniel Bachlechner, Lukas Demetz, and Ronald Maier, "Challenges in cross-organizational security management," *Hawaii International Conference on System Sciences*, vol. 0, pp. 5480–5489, 2012.

[25] S. Thalmann, D. Bachleehner, R. Maier, and M. Manhart, "Key roles in cross-organisational security settings," *European Security Conference*, 2011.

[26] I. Ion, N. Sachdeva, P. Kumaraguru, and S. Capkun, "Home is safer than the cloud! privacy concerns for consumer cloud storage.," *Symposium on Usable Privacy and Security (SOUPS)*, 2001.

[27] Yogesh Simmhan, Alok Gautam Kumbare, Baohua Cao, and Viktor Prasanna, "An analysis of security and privacy issues in smart grid software architectures on clouds," *International Cloud Computing Conference (CLOUD).IEEE*, 2011.

[28] Alok Kumbhare, Yogesh Simmhan, and Viktor Prasanna, "Designing a secure storage repository for sharing scientic datasets using public clouds," Nov. 2011.

[29] Martin Mulazzani, Sebastian Schrittwieser, Manuel Leithner, Markus Huber, and Edgar R. Weippl, "Dark clouds on the horizon: Using cloud storage as attack vector and online slack space," 8 2011.

[30] N Dlodlo, "Legal, privacy, security, access and regulatory issues in clod computing," Apr. 2011, Ted Rogers School of Management, Ryerson University.

[31] Yanjiang Yang and Youcheng Zhang, "A generic scheme for secure data sharing in cloud," in *Parallel Processing Workshops (ICPPW), 2011 40th International Conference on*, sept. 2011, pp. 145 –153.

[32] Ahmad-Reza Sadeghi, Thomas Schneider, and Marcel Winandy, "Token-based cloud computing," vol. 6101, pp. 417–429, 2010, 10.1007/978-3-642-13869-0_30.

[33] Anil Gupta, Parag Pande, Aaftab Qureshi, and Vaibhav Sharma, "A proposed solution: Data availability and error correction in cloud computing," *International Journal of Computer Science and Security (IJCSS)*, vol. 5, no. 4, pp. 405–413, 2011.

[34] Lucian Popa, "Building extensible and secure networks," *Doctoral dissertation in Computer Science, University of California, Berkeley*, 2011.

[35] Kazuhide Fukushima, Shinsaku Kiyomoto, and Yutaka Miyak, "Towards secure cloud computing architecture – a solution based on software protection mechanism," *Journal of Internet Services and Information Securit*, vol. 1, no. 1, pp. 4–17, 2011.

[36] Rafael Accorsi and Yoshinori Sato, "Automated certification for compliant cloud-based business processes," Nov. 2011, vol. 3.

[37] K. Ruan, J. Carthy, T. Kechadi, and M. Crosbie, "Cloud forensics: An overview," *Advances in Digital Frensics*, vol. VII, 2011.

[38] D. Brik and C. Wegener, "Technical issues of forensic investigatinos in cloud computing environments," Tech. Rep., 2011.

[39] Waldo Delport, MS Oliver, and MD Kohn, "Isolating a cloud instance for a digital forensic investigation," in *Information Security for South Africa (ISSA2011) Conference on*, august. 2011, pp. 145 –153.

[40] Keyun Ruan, Ibrahim Baggili, Joe Carthy, and Tahar Kechadi, "Survey on cloud forensics and critical criteria for cloud forensic capability: A preliminary analysis," May 2011, number Annual conference of the ADFSL Conference on Digital Forensics, Security and Law.

[41] Aiiad Albeshri and William Caelli, "Mutual protection in a cloud computing environment," in *IEEE International Conference on High Performance Computing and Communications*, 2010, number 12, pp. 308–315.

[42] R Neisse, D Holling, and A. Pretschner, "Implementing trust in cloud infrastructures," 2011, pp. 524–533.

[43] Kresimir Popovic and Zeljko Hocenski, "Cloud computing security issues and challenges," *The Third International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services, pp. 344-349*, 2010.

[44] *An Analysis of the Cloud Computing Security Problem*, Nov. 2010.

[45] Puneet Jai Kaur and Sakshi Kaushal, "Security concerns in cloud computing," *High Performance Architecture and Grid Computing*, vol. 169, pp. 103–112, 2011.

[46] L Ertual, S Singhal, and G. Saldamli, "Security challenges in cloud computing," *WORLDCOMP 2010*, 2010.

[47] Bikramjit Singh, Rizul Khanna, and Dheeraj Gujral, "Cloud computing: A need for a regulatory body," in *High Performance Architecture and Grid Computing*, vol. 169 of *Communications in Computer and Information Science*, pp. 119–125. Springer Berlin Heidelberg, 2011.

# Design and Analysis of Message Categorization Method
## - Towards Security Support System Preventing Human Errors -

**Shinsaku Kiyomoto**[1], **Kazuhide Fukushima**[1], **and Yutaka Miyake**[1]
[1]KDDI R & D Laboratories Inc.
2-1-15 Ohara Fujimino, Saitama, 356-8502, Japan

**Abstract**—*Incidents involving information leakage by employees are major issues in any enterprise I system. A data importance analysis method is a basic component of any mechanism to resolve the issues; the importance of data is automatically detected by the method and confirms whether the operation suits the security policy for the level of importance of the data. Insider threads are also protected by analyzing data importance and data flows. A design of the automatic analysis mechanism to ascertain data importance is the first step to realize automatic and maintenance-free security systems. he mechanism finds the appropriate category for user sent data in terms of data importance highly important, important, and unclassified. We present a method of the analysis that achieves data categorization with a practical computational cost. he transaction time for categorization is less than msec, the accuracy is around %, and the total data size of signatures is less than Kbytes.*

**Keywords:** Categorization, Security, Human Error, Support System, Data Importance

## 1. Introduction

There has been a trend towards outsourcing data storage to cloud-based services. A feature of cloud computing is distributed architecture based on unfixed nodes. Cloud computing reduces the total cost of a service by sharing all computational resources with other services. However, some security risks have been identified in relation to cloud computing services[1], [2]. In cloud storage services, data are stored on cloud servers with different security levels, and the security level depends on the cost of managing the server. Obviously, highly important data should be stored on a private cloud service with high-level security. On the other hand, the user can store inconsequential data on public cloud services to reduce the total cost of data management. In this situation, users need to check the importance of their sent, received, or processed data and select appropriate servers for the data. Information leakage is another major issue for enterprise IT systems. Employees may mistakenly send important data outside of the company network or store important data without encryption. A data importance analysis method is a basic requirement to resolve this issue; data importance is automatically detected by the method and notified to the user.

An automatic analysis mechanism to ascertain data importance is useful for avoiding human error. The mechanism finds the appropriate category for user-sent data in terms of data importance: highly important, important, and non-important. Generally, two architectures are used to check the importance of user-sent data: gateway architecture and client-based architecture. Gateway architecture uses a server to check the importance of data, and all sent data pass through the server as a gateway. The gateway architecture relies on sufficient computing resources and a database based on all transaction data sent from many users to judge the importance of the data. However, the drawback is that all data traffic is concentrated on the server, making the server a bottleneck for the system. Thus, a high performance server is required for a scalable system. Furthermore, local operations assessing the importance of data cannot be checked.

An analysis program is executed on a client terminal in client-based architecture. The program analyzes the importance of the data sent/received from the client terminal using an optimized database for each client. The client-based architecture is a distributed system and executes the analysis using the computing resources of the client terminal. This architecture is cost-effective; however, a small program that suits the limited computing resources on the client terminals is required. The challenge is to produce a method that is suitable for any terminal, demonstrates high accuracy in importance analysis with a low computing cost and a small-sized database, and for this method to be suitable for any terminal.

In this paper, we present a new method of analysis for a security support system (SSS) that can be applied to the analysis of data importance and that can work independently on each client terminal. Furthermore, we consider an efficient algorithm that performs automatic updates of the database without increasing the size of the database. Our method yielded better error rates than a conventional Robinson-Fisher algorithm, at a feasible computational cost and using a small-sized database.

The rest of the paper is organized as follows: Section 2 presents the concept of the security support system and scope of this paper, and analysis method including a core mechanism for data categorization is proposed in section 3. Evaluation results of a prototype system are presented in section 4. Section 5 introduces related work on this issue. Finally, we conclude this paper in section 6.

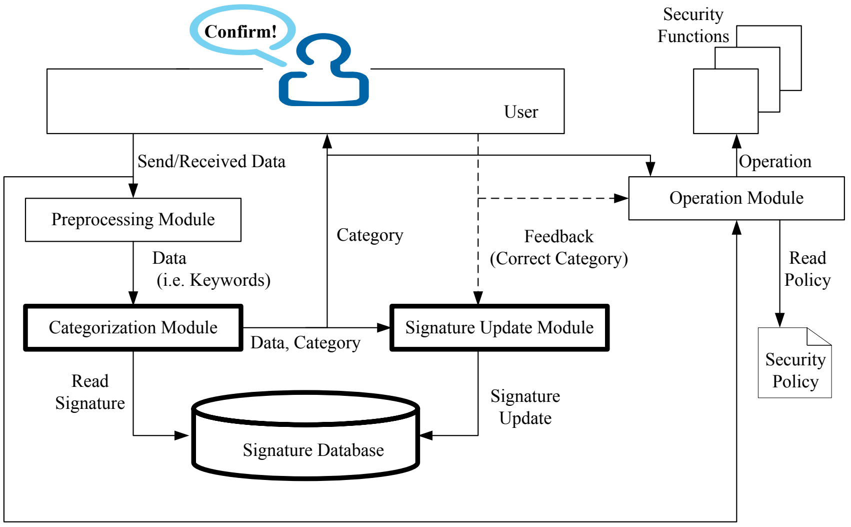


Fig. 1: System Overview

## 2. Concept and Scope

In this section, we introduce a concept of a security support system for enterprise systems, and outline the scope of this paper.

### 2.1 Concept of System

The concept of the security support system is to support user judgment and automatically perform security functions according to a security policy. That is, security functions are automatically applied to data (e.g. emails and business documents) so as to avoid human errors. The system overview is described in figure 1. The system consists of four functional components and a database: preprocessing of data, data categorization, signature update, operation, and a signature database. The system has a security policy for each category, each defined to apply to data that have the same sensitivity or importance. For example, security operations executed on highly important data are different from operations for unclassified data. A signature is prepared for each group, and it becomes a datum defining a property of the category. The preprocessing module for data extracts the properties of data and/or keywords. Then, the module prepares input data for categorization. The categorization module outputs a category appropriate for the data based on the signatures and the signatures in the signature database are updated by the signature update module. The operation module executes security functions for the data according to the security policy and categorization result.

In the initial phase of using the system, a user may have to correct misjudged categorizations, but the system will gain in accuracy with use as the signatures in the signature database are updated. The categorization mechanism will support the user's judgment in operations of security functions according to the security policy, even though the goal of the system is fully automatic operation for security functions.

### 2.2 Scope

In this paper, we focus on three components of the system: the categorization module, the signature update module, and the signature database. In particular, we design concrete algorithms for data categorization and signature updating, and evaluate their feasibility with regard to categorization accuracy and transaction time. The roles of the three components are as follows:

- **Categorization Module.** The module compares user-sent data with the signature database, which has signatures for each category and determines the appropriate category for the data. We assume that the input data is a text transformed by the preprocessing module. The categorization module first extracts words in input data and then the module compares the words with all signatures in the signature database. The extraction includes the removal of common words such as "a" and "the" and modification of words to common roots (e.g., worked is changed to work). The module outputs the category as its result, and both the analyzed data and the result are transferred to the signature update module.
- **Signature Update Module.** The module updates signatures in the signature database according to the results of the analysis performed by the categorization module. The module retains the size of each signature when it is updated. The update operation is a batch operation; the module stores data and results temporarily, and updates the signatures at regular intervals.
- **Signature Database.** The signature database stores signatures for each category. The signatures are computed using training data and updated using previ-

ously analyzed data. The size of the database should be reasonable with regard to implementation on a client terminal and it should not be increased by the update process.

We assume that the system will be executed on a client terminal for each user. Thus, the system has two requirements:

- The computational cost is feasible for a client terminal.
- Signature data sizes for data importance analysis are feasible and are not increased by the update process.

In the next session, we propose a method of analysis that satisfies the above requirements.

# 3. Method

In this section, we propose a categorization algorithm for data importance analysis. The categorization algorithm finds the appropriate category for the data using the signatures for categories. The signatures are preliminarily generated using labeled (already categorized) documents. The JNW algorithm that we propose in this paper evaluates the set of words in the data for comparison with signatures in a signature database. The algorithm is based on two coefficients: the Jaccard coefficient and the new-word-based coefficient. We conducted an experiment to evaluate three major coefficients for categorization algorithms: an overlap coefficient (Simpson's coefficient) [3], the Jaccard coefficient, and a dice coefficient[4]. These coefficients evaluate the similarity of the document and signatures. The Jaccard coefficient is the best coefficient for small training data sets, so we chose to use the Jaccard coefficient. We considered that the accuracy of the algorithm might be improved by also including another coefficient, and concluded as follows; a coefficient that complements the Jaccard coefficient should be added to improve of categorizing accuracy. Our proposal incorporates a new-word-based coefficient as a complement to the Jaccard coefficient; the algorithm evaluates the difference between the document and signatures, and we selected the above two coefficients for the JNW algorithm.

## 3.1 Amended Jaccard Coefficient

The Jaccard coefficient [5] is defined as the size of the intersection divided by the size of the union of the sample sets. The Jaccard coefficient is generally calculated as $\frac{|m \cap m_i|}{|m \cup m_i|}$, where $m$ is a set of words included in a target document and $m_i$ is a set of words included in a document $i$ of group $C_j$. The formula $|x|$ denotes the number of elements in $x$. We obtain the Jaccard coefficient to calculate the ratio of the number of words included in both the target document and a document of group $C_j$. The Jaccard coefficient tends to include errors depending on the size of group $C_j$ and needs to amend a normalized number of elements, $n$. We use $\log(n) \cdot \langle \frac{|m \cap m_i|}{|m \cup m_i|} \rangle_{m_i \in C_j}$ as a new amended Jaccard coefficient, which is obtained by **Theorem 1** in appendix A, where $E(x)$ is the expectation value of $x$. Note that for a calculated values of the amended Jaccard coefficient that were larger than 1, we used 1 as the value of the amended Jaccard coefficient. Furthermore, we decided that no amendment was required and used the original Jaccard coefficient value, where $\frac{|m \cap m_i|}{|m \cup m_i|} > \log(n) \cdot \langle \frac{|m \cap m_i|}{|m \cup m_i|} \rangle_{m_i \in C_j}$. Thus, we calculate the amended Jaccard coefficient $J_n(C_j)$ of the group $C_j$ as follows:

$$J_n(C_j) = max\left[min\left[1, \log(n) \cdot \left\langle \frac{|m \cap m_i|}{|m \cup m_i|} \right\rangle_{m_i \in C_j}\right], \right.$$
$$\left. max\left[\frac{|m \cap m_i|}{|m \cup m_i|}\right]_{m_i \in C_j}\right]$$

## 3.2 New-Word-Based Coefficient

The Jaccard coefficient evaluates the overlaps of words in the target document and the words in documents of the group. In contrast, the new-word-based coefficient calculates the differences between the set of words in the target document and the set of words in a document of the group. Thus, the coefficient carries out a function that complements that of the Jaccard coefficient. The new-word-based coefficient counts the number of words included in the target document and not included in group $C_j$. The algorithm decides that the document is in group $C_j$ by finding a group $C_j$, which has the minimum value of the following coefficient:

$$W_n(C_j) = r_{C_j} \log(|C_j|)$$

where $r_{C_j}$ is the number of words that is included in the target document and not included in group $C_j$.

## 3.3 JNW Algorithm

The algorithm decides whether the target document is categorized into group $C_j$ or not. Thus, the algorithm compares the coefficients for $C_j$ with the coefficients for $\overline{C_j}$, where $\overline{C_j}$ is the complement set of $C_j$ ($\overline{C_j} = \bigcup_{k \neq j} C_k$) and $n$ is $max[|C_j|, |\overline{C_j}|]$. The algorithm calculates $E(C_j)$ and $E(\overline{C_j})$ as follows:

$$E(C_j) = J_n(C_j) W_n(\overline{C_j}), \;\; E(\overline{C_j}) = J_n(\overline{C_j}) W_n(C_j)$$

The algorithm outputs a group $C_j$ that has the maximum value of $E(C_j) - E(\overline{C_j})$.

## 3.4 Training and Updating Signatures

The algorithm requires signatures to determine the appropriate category. The signatures are made using known labeled documents whose categories are given. Thereafter, the signatures should be updated according to current sent/received data. The algorithm preliminarily generates a signature for each category from the labeled documents. The signature for the JNW algorithm consists of data records where each record is defined for each labeled document belonging to the category. Words selected from the labeled document is stored into the record. The signature data for each category consist of at most $v$ records generated from each document belonging to the category. The record includes an $l$ column that consists of a word $m_i$ selected from document $C_i$ and its term frequency $TF(C_i, m_i)$. Each record has data for $l$ words in document $C_i$. The term frequency $TF(C_i, m_i)$ is calculated as $d_i / \sum_i d_i$, where $d_i$ is the number of words $m_i$ in document $C_i$.

In the initial phase, the algorithm uses the signatures generated from labeled documents to categorize the sent/received data, and after categorizing the data, the algorithm adds the data to the signatures according to the results of categorizing and user feedback. With frequent execution, the size of the signatures increases, and the accuracy of the categorizing improves. Thus, a signature update process is needed to keep the size of the signatures constant.

A compression function is used for the signature update. Our method of data compression for the signatures consists of two steps: combining some records and reducing the number of words in the combined record. The compression function is executed for a signature when the number of records in the signature is more than $v$ and consists of the four steps that follow: (1) The function calculates all amended Jaccard coefficients for any pair of two records, and finds a pair of two records with the maximum value of the coefficient. (2) The function creates a combined record from the two records to merge all words in both records. (3) If the combined record has more than $l$ words, the function refers to the TF values of all words in the combined record and removes any words with the lowest TF value until the number of the words in the record is less than $l$. (4) The function executes steps (1) to (3) until the number of records in the signature is less than $v$.

## 4. Evaluation

We implemented a prototype system on a PC (Intel Core i7 CPU-870 2.93 GHz, 4.0 GByte memory) and evaluated the performance and accuracy of the categorization. The 20 newsgroup [6] was used for the evaluation.

Figure 2 shows the error rates of two methods: the Robinson-Fisher algorithm (RF), and the proposed JNW. The RF algorithm is a conventional method for categorization of data. The false rate means the rate at which the categorization algorithm assigned a document to the wrong category. We randomly select three sub-categories in the 20 newsgroup and evaluate the false rates for categorizing items into the three categories. We use 5 patterns of selection of three sub-categories. The JNW algorithm achieved better false rates than the RF algorithm. In cases where we used more than 500 records, JNW achieved around 1.5 - 2.5 % false rates. The difference in the false rates between the no compression case and the compression case of the update algorithm is shown in Figure 3. In the experiment, the compression process was started when over 200 records has been included in the signatures. The compression did not seriously affect the false rate; which remained at about 3 % when we used compressed signatures from 500 records. However, the algorithm with compression cannot achieve better than a 2 % false rate. Figure 4 denotes the data size of all signatures. Our update algorithm kept the size of all signatures less than 140 Kbyte and the false rate was not dramatically increased by the update with signature compression. We also evaluated the accuracy of the categorization algorithm using business documents and confirmed that results were similar.

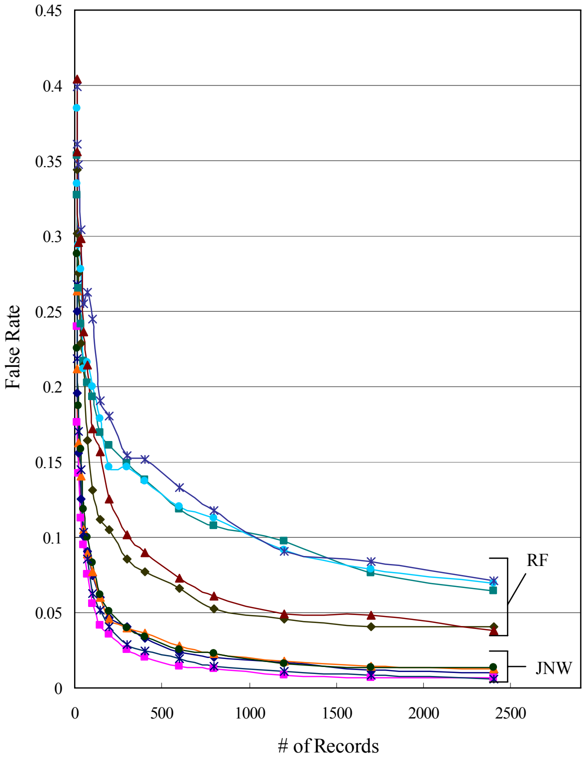


Fig. 2: False Rate

Fugure 5 shows the transaction time for categorization without the signature compression and update of signatures. Each transaction time is an average for randomly selected sample of 1000 data. The row "# of Records" denotes the total number of records in signatures for the JNW algorithm. The transaction time for categorization is about 30 msec when we used 600 records for JNW signatures. A total of 500 - 700 msec was required for updating all signatures, and the transaction time was almost constant as the number of records increased. Thus, the proposed algorithm achieved feasible transaction times for categorization of data and updating the signatures.

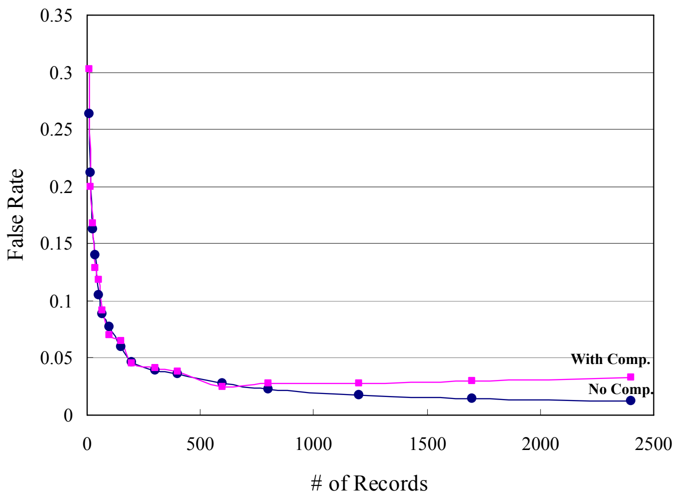


Fig. 3: Comparison of False Rate: With Compression and No Compression

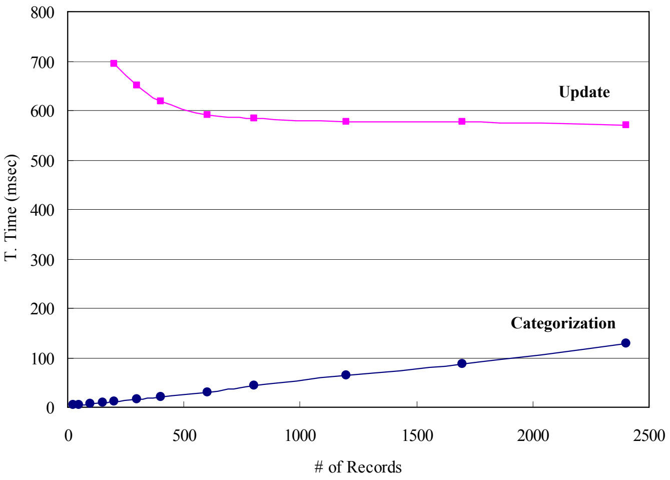


Fig. 5: Transaction Time

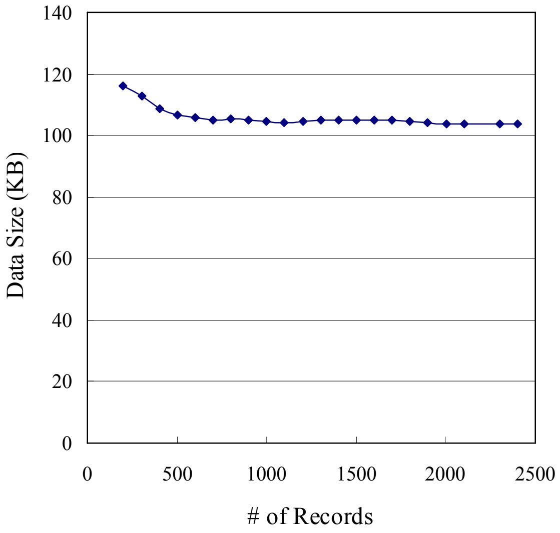


Fig. 4: Data Size of Signatures

Furtheremore, to reduce the transaction time for the update process, we can execute an off-line batch update as an independent process.

## 5. Related Work

There has already been research on data leakage detection. This research has mainly focused on the gateway architecture, and on deliberate misuse of data. Papadimitriou and Garcia-Molina [7], [8] proposed data allocation strategies that inject "realistic but fake" data records into distribution data sets to detect data leakage and identify the responsible party. Bhattacharya *et al.* [9] proposed a privacy violation detection mechanism using data mining techniques. Bottcher and Steinmets [10] proposed an identification algorithm to detect sets of suspicious queries from the logs of an XML database. Ahmed *et al.* [11] presented an audit mechanism for leakage detection and identification of adversaries by counting the frequency of queries. Chow *et al.* [12] described a theoretical framework for inference detection using corpus-based association rules. Kim and Kim [13] proposed a monitoring approach for detecting information leakage using static rules. CutOnce [14] is an information leakage detection program for email systems. There are also several other studies that focus on information leakage detection in email systems [15], [16]. PEEP [16] is a project that aims to develop a privacy compliance system, monitoring outgoing emails in a large organization for potential privacy breaches.

Spam filtering is a categorization technique for text messages[17] and some filtering techniques are good examples for a client-based architecture. Spam filtering techniques categorize email messages into two groups: spam or not spam. Current practical email filtering systems mainly rely on email-specific features such as the email header or the domain names of links embedded in the email [18]. The Robinson-Fisher (RF) algorithm [19] is the most popular algorithm for spam filtering. We also evaluated the transaction time of the categorization for each method to realize practical categorization; for example, the SVM-based approach was dropped due to its long transaction time.

Anomaly-based intrusion detection is a special case of data analysis. Two-group categorization of data is required for anomaly-based intrusion detection systems (IDS); the IDS categorizes data into normal traffic or attack traffic. An anomaly detection algorithm based on the Naive Bayes classification algorithm [20] has been proposed by Maxion and Townsend. Julisch and Dacier presented a new conceptual clustering technique for an intrusion detection system [21]. Lee and Stolfo use data mining for construction and training of classifiers that detect intrusions [22]. SmartSifter [23] is an outlier detection engine addressing fraud detection based on statistical learning theory. For current intrusion detection systems, the trend is to optimize the method of analysis using data mining techniques for particular attacks or applications.

Clustering methods are key techniques for solving the problems addressed in this paper. Several methods have been proposed and compared with other methods in terms of their classification performance [24]. In this paper, we selected an appropriate clustering method and optimized it for data importance analysis. The accuracy of the analysis is generally improved by feedback of the analysis results. Thus, we also considered an efficient algorithm that produced automatic updates of the database without increasing the size of the database or decreasing the categorization accuracy.

## 6. Conclusion

In this paper, we presented a categorization algorithm and an update algorithm for a security support system. By comparing data with signatures generated from old sent/received documents, the categorization algorithm determines whether data are important with 30 msec of computation and with 97 % accuracy. The update algorithm minimizes the size of signatures and the false rates of categorization.

In future research, we will evaluate the false rate to apply our algorithm to a real system and its transaction data.

## Acknowledgment

This work has been supported by the Japanese Ministry of Internal Affairs and Communications funded project, "Study of Security Architecture for Cloud Computing in Disasters."

## References

[1] D. Molnar and S. Schechter, “Self hosting vs. cloud hosting: Accounting for the security impact of hosting in the cloud,” in *Proc. of WEI* , 2010.

[2] K. Hamlen, M. Kantarcioglu, M., L. Khan, and B. Thuraisingham, “Security issues for cloud computing,” *International ournal of Information ecurity and Privacy*, vol. 4, no. 2, pp. 39–51, 2010.

[3] C. D. Manning and H. Schütze, “Foundations of statistical natural language processing,” in *he I Press*, 2002.

[4] L. R. Dice, “Measures of the amount of ecologic association between species,” in *Ecology*, vol. 26, No.3, 1945, pp. 297–302.

[5] P. Jaccard, “Etude comparative de la distribution florale dans une portion des alpes et des jura,” in *Bulletin de la ociete Vaudoise des ciences aturelles*, vol. 37, 1901, pp. 547–579.

[6] J. Rennie, “The 20 newsgroups data set, available at http://people.csail.mit.edu/jrennie/20Newsgroups/,” 2008.

[7] P. Papadimitriou and H. Garcia-Molina, “Data leakage detection,” in *IEEE ransactions on Knowledge and ata Engineering*, vol. 23, 2011, pp. 51–63.

[8] ——, “A model for data leakage detection,” in *Proc. of IC E* , 2009, pp. 1307–1310.

[9] J. Bhattacharya, R. Dass, V. Kapoor, and S. K. Gupta, “Utilizing network features for privacy violation detection,” in *Proc. of First International Conference on Communication ystem oftware and iddleware*, 2006, pp. 1 –10.

[10] S. Bottcher and R. Steinmentz, “Detecting privacy violations in sensitive XML databases,” in *Proc. of ecure ata anagement* , 2005, pp. 143–154.

[11] M. Ahmed, D. Quercia, and S. Hailes, “A statistical matching approach to detect privacy violation for trust-based collaborations,” in *Proc. of the First International IEEE WoW o Workshop on rust, ecurity and Privacy for biquitous Computing*, vol. 3, 2005, pp. 598–602.

[12] R. Chow, P. Golle, and J. Staddon, “Detecting privacy leaks using corpus-based association rules,” in *Proc. of the th AC I K* , 2008, pp. 893–901.

[13] J. Kim and H. J. Kim, “Design of internal information leakage detection system considering the privacy violation,” in *Proc. of International Conference on Information and Communication echnology Convergence (IC C)* , 2010, pp. 480 –481.

[14] R. Balasubramanyan, V. R. Carvalho, and W. Cohen, “Cutonce - recipient recommendation and leak detection in action,” in *Proc. of he AAAI 8 Workshop on Enhanced essaging*, 2008.

[15] N. Boufaden, W. Elazmeh, Y. Ma, S. Matwin, N. El-Kadri, and N. Japkowicz, “Peep - an information extraction based appreach for privacy,” in *Proc. of International Conference on Email and Anti- pam (CEA )*, 2005.

[16] C. Kalyan and K. Chandrasekaran, “Information leak detection in financial e-mails using mail pattern analysis under partial information,” in *Proc. of the th International Conference on Applied Information and Cimmunications (AIC )*, 2007, pp. 104–109.

[17] H. Stern, “A survey of modern spam tools,” in *Proc. of CEA 8*, 2008.

[18] E. Kirda and C. Krügel, “Protecting users against phishing attack,” in *Computer ournal*, vol. 49, No.5, 2006, pp. 554–561.

[19] G. Robinson, “A statistical approach to the spam problem: Using bayesian statistics to detect an e-mail’s spamminess,” in *inux ournal*, 2003.

[20] R. A. Maxion and T. N. Townsend, “Masquerade detection using truncated command lines,” in *Proc. of* , 2002, pp. 219–228.

[21] K. Julisch and M. Dacier, “Mining intrusion detection alarms for actionable knowledge,” in *Proc. of he 8th AC International Conference on Knowledge iscovery and ata ining*, 2002, pp. 366–375.

[22] W. Lee and S. J. Stolfo, “A framework for constructing features and models for intrusion detection systems,” in *AC ransactions on Information and ystem ecurity*, vol. 3, No.4, 2000, pp. 227–261.

[23] K. Yamanishi, J. Takeuchi, G. Williams, and P. Milne, “On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms,” *ata in. Knowl. iscov.*, vol. 8, pp. 275–300, May 2004.

[24] A. McCallum and K. Nigam, “A comparison of event models for naive bayes text classification,” in *Proc. of AAAI- 8 Workshop on earning for ext Categorization*, 1998, pp. 41–48.

[25] B. Eisenberg, “On the expectation of the maximum of IID geometric random variables,” in *tatistics & Probability etters*, vol. 78, 2008, pp. 135–143.

## Appendix

In this appendix, we present **Theorem 1** for calculation of the amended Jaccard coefficient.

**Theorem 1.** If it is assumed that the probability density function $f(x)$ for each $X_i$ is an exponential distribution function, an expectation value of the maximum Jaccard coefficient value $X_{max}$ is estimated as $\langle X\rangle\mathbf{log}(n)$ for a sufficiently large $n$, where $\langle X\rangle$ is the expectation value of Jaccard coefficient values.

*Proof.* We define $F_{max}(x)$ as the probability distribution function of $X_{max}$ and $X_{max} = max[X_1, X_2, ..., X_n]$.

$$\begin{aligned} F_{max}(x) &= Pr[X_{max} \leq x] \\ &= Pr[(X_1 \leq x) \wedge (X_2 \leq x) \wedge ... \wedge (X_n \leq x)] \end{aligned}$$

It can be assumed that each event is independent and it is calculated using $F(x)$, which is the probability distribution function for each $X_i$ as follows:

$$F_{max}(x) = F(x)^n$$

Now, we assume $f(x)$ is the exponential distribution function, that is $f(x) = \lambda e^{-\lambda x}$ $(x > 0)$. $F(x)$ is calculated as $F(x) = 1 - e^{-\lambda x}$ $(x > 0)$. From [25], we obtain

$$X_{max} = \sum_{t=1}^{n} = \frac{1}{\lambda t}$$

Thus,

$$X_{max} = \frac{1}{\lambda k} \approx \frac{1}{\lambda}\mathbf{log}(n) = \langle X \rangle \mathbf{log}(n) \quad \square$$

# Identifying Illegal Digital Images

## An Image Watermark Approach to Preventing the Spread of Pornography

**Mark Wilson**
Department of Computer Science
Sam Houston State University
Huntsville, TX 77341
mrw004@shsu.edu

**Qingzhong Liu**
Department of Computer Science
Sam Houston State University
Huntsville, TX 77341
qxl005@shsu.edu

**Lei Chen**
Department of Computer Science
Sam Houston State University
Huntsville, TX 77341
lxc008@shsu.edu

*Abstract*–**Today, thousands of people are constantly carrying a digital camera in the forms of a smartphone, a key chain, or a tablet computer. Due to the prevalent availability and the simplicity of digital cameras, the classic style 35mm film camera has essentially been replaced. This new technology allows for people to edit and modify their own photos, bypass photo developers for prints, and distribute images via the Internet or email all while keeping their identities and the origin of the image somewhat anonymous. This paper presents different types of illegal images spreading across the Internet, and methods that are currently being employed to identify and stop those images from being distributed. Digitally watermarking the images to include identifying information about the image and the camera will be examined in order to propose a new method of counterstrike against the spread of illegal images.**

**Keywords**: digital watermark, hash value, illegal images, violent pornography, child pornography

# 1 Introduction

The production and distribution of illegal images, typically child pornography, has grown exponentially in the past two decades. Digital Cameras and the Internet provide the anonymity for anyone to take a picture and share it with friends, family, or even strangers. The majority of images emailed across networks or uploaded to servers are completely valid and are not classified as illegal or contraband. However, because of the ambiguous origin of shared images, the ease of distributing illegal images increases dramatically. Currently the United States classifies very few genres of images as illegal, but included in that category are both child pornography and violent pornography.

The term child pornography is defined differently worldwide due to cultural differences, traditions, etc. In the United States an image of a child engaged in real or simulated sexual activities is considered a form of child pornography in the same manner as images depicting parts of a child's body with the intention of stimulating sexual arousal or gratification for the viewer of that image. This definition is based upon the Convention on the Rights of Children, which defines a child as being under the age of 18, and states the government shall protect children from all forms of sexual exploitation and abuse, including: the inducement or coercion of a child to engage in any unlawful sexual activity, the exploitative use of children in prostitution or other unlawful sexual practices, the exploitative use of children in pornographic performances and materials [1]. Violent pornography is defined as explicit material depicting children, adults, or both portraying sexually violent or sexually coercive acts, including rape, genital mutilation, or forms of BDSM (bondage/discipline/sadism/masochism) [2].

To further compound the challenges of the increasing rise of child pornography on the Internet today, law enforcement also has to evaluate virtual images. These images are morphed or blended in order to artificially create images of children depicting sexual activities [1]. United States law prohibits computer generated child pornography when such visual depiction is a computer image or computer-generated image that is, or appears virtually indistinguishable from, a minor engaging in sexually explicit conduct [3].

Next in Section 2, we examine methods to automatically identify digital pornography that are currently being used by law enforcement to combat the problem. Section 3 describes the importance of digital watermarking, how it is currently used, and how it could impact future data. In Section 4 robust digital watermarking will be further examined to identify child and violent pornography by using the hash value of the image itself as well as what security benefits this method could provide. Section 5 draws conclusions from this proposed method of identifying illegal images and Section 6 proposes future research.

# 2 Background

## 2.1 Automatic Nudity Detection

One method currently employed for identifying illegal images is automatic nudity detection. Devices suspected of containing child pornography may contain hundreds of thousands of files. Today, with hard drive space steadily increasing, it is not uncommon to find a terabyte drive filled with digital images or videos, meaning an investigator must sift through the entire contents file by file in order to discover any illegal images. *Paraben Porn Detection Stick* and *Image Analyzer SDK* are two tools that can be used to quickly filter through images, but both tools are relatively new and their nudity detection algorithms and performance results in real forensic cases have not been reported in the literature [4].

The forensic tool, *NuDetective*, analyzes multiple aspects of digital images in order to detect possible nudity. The tool first examines the color space of each digital image. Images are often displayed in RGB (the mixture of red, green, and blue to create a particulasr color) mode that originated from CRT (Cathode Ray Tube) monitors. This method of color identification does not produce optimal results because brightness and color are coupled and thus not suitable for color segmentation in images with unknown lighting conditions [4]. Color space can also be displayed based on hue and saturation (HSV), where hue defines the dominating color and saturation defines the color based on brightness. Both color space methods are used today, and must be evaluated differently in order to produce the desired results.

After processing a particular color space, the image is run through an algorithm that first compares the number of proposed skin pixels to the number of total pixels of the image. Through research a threshold of 15% was defined so the image would continue to be evaluated for possible nudity. If the image included less than 15% skin-colored pixels the detection system determined the image to have no nudity [4].

By processing each color space, particular colors can be identified that are most likely matches to natural skin tones. However, though colors are examined and images including those colors can be filtered so an investigator can visually examine them, results vary. False negatives and false positives occur for a number of reasons, including: variable ambient light, digital camera default color settings, skin color of different persons, and the appearance of different natural objects. Wood, leather, hair, and sand all carry the same colors as possible skin tones, making an automatic nudity detection system less than dependable [4]. These systems are helpful to reduce images possibly containing nudity, but still require visual confirmation to accurately identify illegal images.

## 2.2 Hash Value Comparison

Currently, one of the most popular tactics that law enforcement employs is a hash value comparison to identify and filter illegal images while examining hard drives of people suspect of possessing those images. Many computer forensic suites, including *EnCase* and *Forensic Toolkit (FTK)*, give users the option of calculating hash values of each file residing on the hard drive as the drive is imaged onto a forensic workstation. These forensic suites will also examine each file's header to ensure that the header matches the file extension to highlight any possible discrepancies, and to locate image files that have been assigned a different extension.

By calculating the hash values of files residing on the suspect's hard drive or storage device and comparing them to the hash values of known illegal images, law enforcement is able to quickly and accurately filter images and possibly charge the owner of the computer system with such crimes as manufacturing, possessing, and/or distributing child pornography to others across the Internet.

# 3 Digital Watermarking

Digital watermarking has been an increasingly popular method to copyright and authenticate digital information. This method is able to embed copyright information into the multimedia data through certain algorithms; the information may be author's serial number, company logo, images or text with special significance, etc. The function of this technique is to serve as copyright protection, secret communication, or to authenticate a data file. This watermarked information is generally imperceptible to users unless a particular detector or reader that reveals the possibility of hidden information examines the carrier file. Most notably to this research is that depending upon the type of change made to the carrier file, the hidden data may still be recovered [5].

## 3.1 Features of Digital Watermarks

Digital watermarking has three distinct features that make the process valuable. First, though some watermarks are intentionally embedded into an image file to be visible to the human eye for various reasons, watermarks have the potential to be hidden inside a file and be imperceptible by the viewer. This feature is extremely important when data is embedded into the file to prove authenticity or to merely add data such as file specifics. The feature of imperceptibility is the key function and most popular reason for a digital watermark in the computer industry today. The second feature is the verifiability of the watermark. This provides that the watermark will not change over time as data files are distributed to different destinations. The watermark can be used to determine whether the object is to be protected and monitor the spread of the data being protected, identify the authenticity, and control illegal copying [5]. The watermark's security feature works closely with the verifiability feature in that it allows only authorized users to detect, extract, or modify the watermark for the purposes of copyright protection. Lastly, and arguably the most important feature for this particular research is the robustness of the digital watermark itself. This term refers to the strength of the watermark to withstand any type of processing operation or attack and to remain unchanged and extractable in the event of such an attack. Depending upon the type of watermark and its use, the level of robustness may vary. Robust watermarking is generally employed to sign copyright information of digital works and is able to resist common image processing and lossy compression, ensuring that the watermark is not destroyed after some attack and can still be detected to provide certification [5]. Fragile watermarking is employed for integrity protection and is able to determine if data has been altered.

# 4 Watermarking Images with Hashes

Granted law enforcement has already employed hash value comparison and automatic nudity detection in order to quickly identify contraband images on suspect computer systems, but comparing hash values of a suspect image to a hash value of a known illegal image can prove frustrating if the hash value of the suspect image has changed because the image has been edited. Using robust digital watermarking, the watermark of file information can be preserved within the image itself and compared to known hash values of illegal images regardless if the image has been modified or not.

## 4.1 Employing a New Digital Camera OS

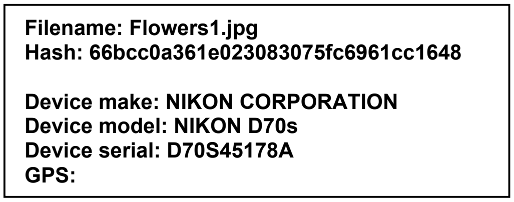
Digital cameras have become the norm for photographers today because of their user-friendliness, instant previews of images captured, and economic advantages. Though digital cameras vary by quality and features, most work exactly the same regardless of brand (Nikon, Fuji, Olympus, Kodak, etc.) thus making a change in the camera operating system relatively easy, aside from proprietary licenses.

When synchronizing the camera to a computer or other device in which photos would be transferred, the camera operating system would examine each of the transferred images and calculate a hash value for each. MD5 or SHA-1 hashes are suggested over a proprietary identification number in order to make the generated values universally recognized and easily comparable. Each of the transferred images would be automatically watermarked with the generated hash value at the time of syncing. Additional information such as camera model, serial number, and GPS coordinates along with data usually included in the file metadata (camera model, aperture, date/time image was captured) could also be watermarked onto each image. The calculation of hash values, the locating of identifying camera information, and the watermarking of each image will slow the syncing process considerably; however the average user would most likely not notice a difference unless transferring a large quantity of images. Once the images have been successfully transferred to the designated device, each image will include a hidden, robust watermark that can uniquely identify it even if the image has been changed by basic photo editors.

## 4.2 Hashing Image Files

For this research a standard JPEG image, measuring 2000 x 3008 pixels was used to act as the original image being synchronized to a computer from a digital camera. A 400 x 400 pixel binary image that includes the generated hash value, along with other identifiers of the image was used to watermark the original JPEG. Figure 1 illustrates the original JPEG image, and figure 2 illustrates the binary watermark. The actual aspect ratio of figure 2 was changed in order to conserve space, but allows for additional data to be included in the future.

Figure 1. Original JPEG image measuring 2000 x 3008 pixels; a default setting for a high-end digital camera.



**Filename: Flowers1.jpg**
**Hash: 66bcc0a361e023083075fc6961cc1648**

**Device make: NIKON CORPORATION**
**Device model: NIKON D70s**
**Device serial: D70S45178A**
**GPS:**

Figure 2. Watermarked image data embedded into the original JPEG during the synchronizing process.

As per [5], highly robust watermarks can be embedded into digital images to withstand attacks such as: compression, cropping, scaling, and median filtering. These changes are considered basic; the majority of people do little more to images aside from these four operations. However, a stronger watermark needs to be employed in order to continue to identify a digital image even after such operations as adding text or painting/drawing the image.

## 4.3 Security Benefits

This technique acts the same as standard hash value comparison (described in section 2.2) that is already widely used in the law enforcement community. However, by employing this method of watermarking original digital images as they are synchronized to another device, law enforcement would be able to identify contraband images in a more timely fashion; time is often of the essence during any type of investigation, especially when children are involved. This method would allow law enforcement to compare a library of hash values from known illegal images to the watermarked hash values found on suspect computers, rather than the hash value of each file.

This comparison is important because, in the case of child pornography, the suspect who initially takes the image may upload it as-is. That original file that was uploaded will carry a particular hash value that may be cataloged by law enforcement at a later date. However, because each child pornographer may have a different type of fetish or interest [6], the original uploaded file may be cropped, lightened, etc., thus changing the hash value each time a new version of the same picture is saved. Today, law enforcement would collect a hash value of each of these edited images when all of them stem from the same original upload. This would be avoided if the original uploaded image contained a robust watermark containing its hash value that could be retrieved after each edit. By matching the watermarked hash value to the known library of illegal images, law enforcement would be able to significantly reduce the amount of time spent searching hard drives of contraband images. Though this technique still requires investigators to visually search images on suspect computers in the event of the hard drive containing original images that are not known to law enforcement, it still reduces the time it would take to find known images and secure an arrest warrant for the individual of the computer being searched.

In extreme cases of child pornography this method of identifying images may simplify prosecution of the suspect by illustrating that multiple computers owned by multiple users all have the same illegal image on the hard drives. Hypothetically, a single illegal image could have infinite hash values associated with it depending on how many times that images had been changed and saved. Due to these infinite possibilities it would be reasonable that images matches on different computers would go unnoticed. By indisputably showing a common hash value between two or more variants of the same image, prosecution could easily prove the distribution of child pornography.

Federal law 18 U.S. Code § 2252A criminalizes certain activities relating to material constituting or containing child pornography. Federal law defines child pornography as "any visual depiction, including any photograph, film, video, picture, or computer-generated image or picture, whether made or produced by electronic, mechanical, or other means, of sexually explicit conduct, where: the production of the visual depiction involves the use of a minor engaging in sexually explicit conduct; or the visual depiction is a digital image, computer image, or computer-generated image that is, or is indistinguishable from, that of a minor engaging in sexually explicit conduct; or the visual depiction has been created, adapted, or modified to appear that an identifiable minor is engaging in sexually explicit conduct." The definition of this law specifically cites the offense that this research method would provide evidence for. Federal law outlines that an offense is committed if "a person knowingly possesses, or knowingly accesses with intent to view any… child pornography."

Not only could this method of quickly identifying illegal images be used as a prosecution tool, but also as a public deterrent. By including not only the image hash value in the watermark but the image's originating camera information (model, serial number, GPS location) as well, investigators would have solid leads to pursue those who are producing illegal images. Federal law also cites "persons who knowingly produces with intent to distribute, or distributes, by any means, including a computer, in or affecting interstate or foreign commerce, child pornography that is an adapted or modified depiction of an identifiable minor."

Many child pornographers are arrested and criminally charged for unrelated crimes prior to any discovery of their involvement with illegal images [7]; this method of tracking may increase the arrests of such persons for crimes relating to pornography.

# 5 Conclusion

In this paper a basic outline of illegal images was discussed along with two different methods that are currently used by law enforcement to identify those images. Due to the increasing popularity of digital cameras worldwide, and the anonymity of the origin of those images captured, illegal images are much easier to produce and distribute. Digital cameras have given persons who produce illegal images an avenue to secretly capture as many as they wish while the Internet provides an outlet to share or sell those images anonymously. In order to successfully combat the future spread of illegal images across the Internet, law enforcement must remain diligent to track each of these images. In this paper, a method of watermarking an image robustly with an original hash value that would not change after the image had been edited was proposed in order to identify a single image, though it may be associated with multiple different hash values. Along with watermarking the original image with a hash value, other camera specifications could be included to give law enforcement investigators solid leads to pursue criminal charges and act as a deterrent to those who produce and distribute illegal images.

## 6 Future Work

Future work will include applying this method of maintaining a digital watermark after the image has been edited. The robustness of the watermark has shown that by using the proper tools the watermark can be recovered after the image has been compressed, cropped, and scaled. More research will have to be conducted in order to make the watermark still more impervious to attacks. Efforts will be focused on JPEG images, but because different cameras are able to capture images in different file types (raw, bitmap, etc.) more research will be conducted to accommodate those file types. In addition, digital camera's operating systems along with their methods of synchronizing to a computer or other digital device must be evaluated in order for the watermark to be properly applied during the synchronization process. Lastly, it must be further researched to discover how crimes involving illegal images identified by a hidden hash value are to be prosecuted in the event of a federal offense.

## 7 References

[1] A. Ibrahim, "Technology For Facilitating and Combating Social Deviations," University of Ontario Institute of Technology: Canada, 2011.

[2] W.A. Fisher, and G. Grenier, "Violent Pornography, Antiwoman Thoughts, and Antiwoman Acts: In Search of Reliable Effects," The Journal of Sex Research, 1994.

[3] S.Ruth, "Pornography and the Internet," IEEE Internet Computing, 2008.

[4] M. de Castro Polastro, and P. M. da Silva Eleuterio, "NuDetective: a Forensic Tool to Help Combat Child Pornography Through Automatic Nudity Detection," 2010 Workshops on Database and Expert Systems Applications. 2010.

[5] J. Xuehua, "Digital Watermarking and its Application in Image Copyright Protection," 2010 International Conference on Intelligent Computation Technology and Automation, 2010.

[6] M. Taylor, and E. Quayle, "Child Pornography: An Internet Crime," Brunner-Routedge: New York, 2003.

[7] P. Jenkins, "Beyond Tolerance: Child Pornography on the Internet," New York University Press: New York, 2001.

[8] J. Fridrich, and M. Goljan, "Robust Hash Functions for Digital Watermarking," Center for Intelligent Systems, 2002.

[9] L. Qiao, and K. Nahrstedt, "Watermarking Schemes and Protocols For Protecting Rightful Ownership and Customer's Rights," Journal of Visual Communication and Image Representation, 2005.

# Cloud computing Security using DOSBAD: Denial of Service Bandwidth Allowance

Biswajit Panja
University of Michigan-Flint,
Flint, MI 48502
bpanja@umich.edu

Dayton Paul
University of Michigan-Flint, Flint, MI 48502
daytonp@umflint.edu

Priyanka Meharia
Eastern Michigan University, Ypsilanti, MI
pmeharia@emich.edu

Alex Henke
University of Michigan-Flint, Flint, MI 48502
ahenke@umflint.edu

*Abstract*— **Over the next decade, cloud computing has a good chance of becoming a widely used technology. However, many challenges face the cloud to be overcome before the average user or business team will trust their vital information with a cloud server. Most of these challenges tie into developing sound security measures for the cloud. One of the largest security obstacles is how to defend against a Denial-of-Service (DOS) or Distributed Denial-of-Service (DDOS) attacks from taking down a cloud server. DOS attacks are nothing new; many strategies have been proposed and tested against DOS attacks on networks. However, none have been able to completely prevent DOS attacks. The search continues for an effective solution to keep data available to legitimate users who need it when the cloud network that stores that data is the target of a DOS attack. The method proposed (DOSBAD) in this paper will explain how effectively detecting the bandwidth limit of a cloud network and the bandwidth currently in use to know when a DOS is beginning.**

***Keywords- Cloud computing; Denial of service; Bandwidth***

## I. INTRODUCTION

It is believed that the world is heading towards a computing resource grid similar to the power grid and being charged based on usage like we are for energy[FZ 08]. There are similarities and differences between grid computing and cloud computing [FZ 08]. They share a lot in common, but the cloud is less secure though it utilizes virtualization. The future of computing may be centralized around cloud computing, and client computing for. People and organization may want to do computing on cloud instead of investing resources locally and implement security in their local computers. That leads the focus on providing security in cloud. Cloud computing needs set standards, has layered architecture (software as a service, platform as a service, infrastructure as a service, and hardware as a service), the different modes of clouds (private, public, and hybrid), types of virtualization used in cloud computing (server, storage, and network virtualization), fault tolerance, security issues, and scalability.

In this paper we propose an approach to avoid Denial of Service (DOS) attacks. In this, an entity integrated into a cloud server can be used to monitor what ratio of available bandwidth is being used. To find the maximum available bandwidth of the server, the entity DOSBAD (Denial-of-Service-Bandwidth-Allowance-Device), will periodically send a series of packets down each possible path within the cloud (router-to-router). Two large packets are first sent to create a queue at the switch between the routers, then two small packets are sent, which will be the ones that have their time of being sent and their time of being received measured. The total time to transfer these packets will be the time at which packet 1 is sent subtracted from the time at which packet 2 is received. Based on the time it takes for the receiver to receive the packets and to acknowledge them, DOSBAD will calculate the bandwidth available between those two routers. DOSBAD will also monitor how much of that bandwidth is in use at each router. The number of incoming packets will be measured, along with the amount of acknowledgement packets that are sent back out. Ideally, the number of packets received should match the number of acknowledgement packets sent back out, indicating that the router is not overwhelmed with the number of incoming packets. When the number of incoming packets starts to outweigh the number of acknowledgement packets sent, that can indicate that the bandwidth limit may be close to being reached. This indicates that either there is an abnormal spike in activity coming into the network (i.e. a flash crowd), or that there is malicious activity being attempted. At this point, DOSBAD may look for common return addresses on incoming packets at the overwhelmed router(s) and then send out a ping to those addresses. If DOSBAD does not receive a response from an address, that may indicate a DOS attack being attempted. DOSBAD signals to the router (or possibly the gateway) to drop all incoming packets from that address. Another feature that the cloud manager may wish to use is to have DOSBAD automatically change the address of the attacked router so that if the attacker tries again from a different attacking address, he is unable to find that router again. All other legitimate traffic will be re-routed to the new address automatically.

DOSBAD keeps a running tab on the addresses of all senders of incoming packets within some time interval. DOSBAD uses this to see from which address the most incoming packets are coming from. Along with this can be stored the signature of each incoming packet. The signatures of packets coming from zombies in a DOS attack sometimes have very specific signatures that can be used to detect that a DOS attack is occurring. If a high ratio of bandwidth is being used, one or more routers are overwhelmed by incoming packets, and a high number of packets are coming in at a router from the same IP address, DOSBAD will proceed to investigate a possible DOS attack by pinging the suspicious address or addresses as mentioned previously.

This paper is organized as follows, section II provides related work. Section III describes the architecture of DOSBAD. Section IV provides the proposed security approach for cloud computing.

## II. Related Work

Dikaiakos et al. [DK 09] talks about the types of things that are expected for the future of cloud computing. Included in this are ideas of infrastructures, platforms, and software being offered as services which are bought by "consumers" or anyone who wants to implement the services. Clouds have some characteristics that help us describe their type, including "internal" or "external/hosted", and "private", "public", or "hybrid". The three layers of a cloud are infrastructure (lowest level), platform (higher abstraction), and application (highest abstraction, provides actual applications that consumers can buy). Also discussed are the challenges that must be solved in order to realize the full potential of the cloud, including the architecture of the cloud versus individual computers, data management and security, cloud interoperability (customers using the cloud applications through different types of machines), and the economics involved with purchasing services.

Armbrust et al [MF 09] starts off by defining what the Cloud is. Their definition is that the Cloud itself is the hardware and software that is needed to provide services of a network to many Cloud users. Cloud providers provide the resources to Cloud users who implement the resources to create applications that Software-as-a-Service users can use for what is called Cloud Computing. Another aspect discussed is the reasons why the Cloud is taking off now and not previously. These reasons include the quick, low-commitment services to users such as PayPal and real-time responsive applications.

The three classes of utility computing, which is what Cloud users purchase from providers, are defined by 3 different abstraction levels for resources provided to Cloud users. For low-level abstraction the user has more flexibility with what kinds of applications they want to program but limit the scalability of the application is very limited (it's hard to change the limits on the application if the demand for it suddenly skyrockets above the set limit). For high-level abstraction the user can make things that are much more scalable but not very flexible for general computing since the user cannot control the low-level hardware. Mid-level abstractions provide some aspects of the previous two classes. General-purpose computing and multiple programming languages are available (low-level) and the libraries help provide limited scalability (high-level). Each of these classes have different models for how they provide computations, storage, and networking to users. For now, none of the three classes have proven to be the most useful out of the three. Each of them is ideal for certain situations.

The next aspect discussed is Cloud economics. The decision of whether to host a service through the Cloud or to continue using datacenter can be answered by looking at several things. If your average utilization and peak utilization very different values, that is a reason for switching to the Cloud because the scalability of resources can help a host not have to pay for unused resources during non-peak times. Another aspect to consider is the cost of transferring all of the user's data from their datacenter to the Cloud: will the time saved by doing so outweigh this cost? If so, the Cloud could be a realistic option. Also, the heating costs saved from using the virtual machines of the Cloud to provide services can be a positive to switching to the Cloud.

The final section of the article goes over ten of the largest obstacles that Cloud Computing must overcome and ten corresponding opportunities that can be used to overcome these obstacles. Obstacle one is the possibility that demand for Cloud Computing could overtake the practical supply of resources that a Cloud provider can meet. The solution discussed is how this encourages a cooperative effort among multiple companies to greatly increase supply. Obstacle two involves service users being locked in to one Cloud user (the one that provides the services that buys resources from the Cloud provider), which can be solved by standardizing services among all Cloud users so that people looking to buy a service can choose who to buy from. Obstacle three involves data on the cloud being secure enough; solutions include things we already implement with networks, such as encryption and firewalls. The fourth through eighth obstacles involve how the Cloud will grow over time, such as data transfer bottlenecks, performance unpredictability, and bugs in the distributed systems. Solutions to these include physically shipping the disks to save money, implementing flash memory to reduce the interrupts and thus increasing performance, and creating a debugger that works with virtual machines (respective to the three mentioned obstacles). The last two obstacles look at the business aspect of the Cloud. The first one is how to prevent reputation fate sharing from a few bad users (spammers, etc.). The concept of trusted email services that already exist could be applied to help guard the reputation of services. Finally, software licensing can cause a problem because a user could purchase a service and not be able to use that service on other computers. The solution is to offer pay-as-you-go options so that the user can pay for what they need as they realize they need it.

In the conclusion of the article, specific implementations of applications, infrastructure, and hardware are mentioned that should be implemented in future systems to be more easily Cloud compatible. Applications should be able to run partly in the client and partly in the Cloud, each part having its own duties. Infrastructures should be designed to run on virtual machines. Hardware systems need to be designed as containers instead of single boxes or racks since users will purchase them in containers. Besides this, questions are posed to the reader as to what the future Cloud systems will be like.

Vouk et al. [VM 04] defines the concept of cloud computing, describes various aspects involved with cloud computing, an example implementation of cloud computing at North Carolina State University, and lastly about research issues involved with the cloud.

Cloud computing can be considered the next step in improving the availability of services and products supplied to users over a network that is in part due to virtualizing the resources. One aspect mentioned that is crucial to clouds is the service-oriented architecture, which means users request services from the cloud provider. Another critical aspect is making services out of components, which can be described by reusability, substitutability, extensibility, scalability, customizability, composability, reliability, availability, and security. A workflow can be used to visually represent services that may be provided by the cloud, usually through a graph. A question posed by the article to the reader is whether or not workflows could be useful in representing the infrastructure of cloud computing. Another aspect discussed related to cloud computing is virtualization of various computing components, such as memory, hardware, and applications. Cloud computing relies heavily on virtualization because it allows computing components to become more portable so they can be provided to users easily as a service. The final and most important aspect of cloud computing are the four types of users this article defines that are involved with the cloud. These are developers, who configure and maintain the Cloud framework, service authors, who develop templates for services from the Cloud framework, service composition experts, who create services for end-users, and end-users, who request services and implement them. The final topic of this article is the research issues of cloud computing, including getting feedback on workflows, collecting, storing, and preventing provenance information, optimization of service components, service portability, cloud computing security, and efficient utilizations of resources.

Mowbray et al. [MP 09] goes into some specific privacy issues with regards to cloud computing and defines one possible solution Privacy Manager that could be used to overcome these issues. The main requirements defined here include minimizing the user's data stored on the cloud to what is necessity, protection to what data is stored on the cloud, limiting the purposes that can use the data and the people who may access the data, user-controlled preferences related to what their data may be used for on the cloud, and feedback given to users about how their information was used afterward.

The solution this article offers to meet the requirements is called Privacy Manager. It uses five main features to both protect user data and give the user a welcoming sense that he or she has control of their own data (customizable features). One of the biggest features is obfuscation of data that goes into the cloud and de-obfuscation of data that is being accessed by the user from the cloud. Obfuscation is similar to encryption, only the user gets to decide on a specific key that is used to modify their data as it goes into the cloud. The key is not provided to the cloud provider so that they cannot de-obfuscate the data themselves. With preference setting, another feature, the user can decide what data gets obfuscated and what data doesn't (sometimes you don't want to obfuscate data for certain applications). The data access feature allows users to see what data they have on the cloud to make sure it is accurate. If the data is obfuscated in the cloud, it gets de-obfuscated by the Privacy manager before being shown to the user. The feedback feature shows the user how their data is being used in the cloud (so that the user will know if their preferences have been violated). Finally, the personae feature can be used so that a user can set up different levels of preferences with different cloud services (obfuscating some information when using certain applications and not obfuscating the same information for other applications). They simply choose the personae that has the preferences they want to apply to the current application.

Cho et al [CB 11] explain with large programs often have a daunting number of lines (usually millions). It is very difficult to track down all the bugs in such a program that could be used by someone malicious to, for instance, initiate a denial-of-service attack or cause a segmentation fault and wreaking havoc. Finding these errors by just using software to send it random input data to check the resulting output does not always find all these errors, as demonstrated by an experiment in this article. A solution this article offers is a new approach to exploring behavior of programs given vast varieties of input data called MACE: Model-inference-Assisted Concolic Exploration.

MACE consists of sending messages to an algorithm called L*. L* infers a state machine based on this input. For every state in this finite state machine (Mealy Machine), L* generates a path to get to that state that is the shortest possible path (i.e. if you want to reach a certain state S, it finds the shortest length of input string that will take you to that state). One input string is used per unique state so that all states can be analyzed using state-space exploration. The output from each of these states plus the input used into L* in the first place are sent through a filter to get rid of redundant inputs and a new list is sent to L* to make a new FSM, and the process keeps going until no new states are found through an iteration.

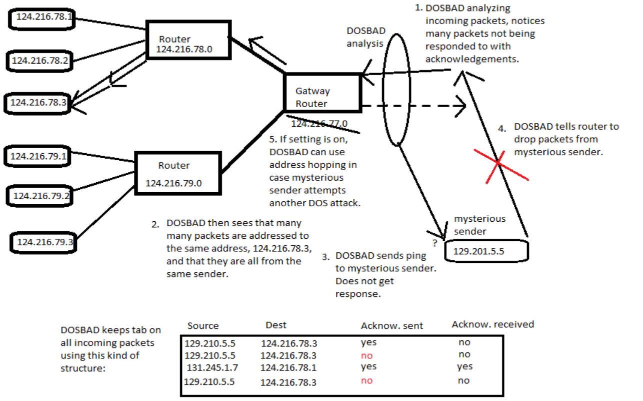
An experiment was ran to test MACE against a baseline method of analyzing programs using the state-space exploration part of MACE without using the component that sends the input back to the L* algorithm. The baseline method uncovered only one vulnerability in the programs tested (Vino and Samba) while MACE found seven between the two, four of which had never been discovered before that on record. MACE also generated a fairly accurate FSM of Vino compared to what it actually was. Other comparisons include the number of detected crashes (30 to 20), unique crashes (9 to 1), and the exploration depth, which showed MACE was more proficient at reaching deeper states than the baseline approach.

Some limitations of MACE are discussed after the experiment. It cannot be guaranteed that MACE found ever possible vulnerability because of how the L* algorithm works. Also, MACE was very good at analyzing user-level programs, but it was not able to go into the kernel level, which limits its effectiveness.

The conclusion poses some questions for the reader to think about for future experiments, including how FSMs can be studied further to find even more effective implementations within the state-space exploration component that MACE uses. Another question is how to find a better way to filter out redundant output messages without eliminating possible new states and thus finding new vulnerabilities. The third question is if there are any other feasible methods besides using FSMs to help generate all possible output sequences for a program.

## III. Architecture of DOSBAD

In this section we discuss the architecture of our proposed model DOSBAD.



1. DOSBAD analyzing incoming packets, notices many packets not being responded to with acknowledgements.

2. DOSBAD then sees that many many packets are addressed to the same address, 124.216.78.3, and that they are all from the same sender.

3. DOSBAD sends ping to mysterious sender. Does not get response.

4. DOSBAD tells router to drop packets from mysterious sender.

5. If setting is on, DOSBAD can use address hopping in case mysterious sender attempts another DOS attack.

DOSBAD keeps tab on all incoming packets using this kind of structure:

| Source | Dest | Acknow. sent | Acknow. received |
|---|---|---|---|
| 129.210.5.5 | 124.216.78.3 | yes | no |
| 129.210.5.5 | 124.216.78.3 | no | no |
| 131.245.1.7 | 124.216.78.1 | yes | yes |
| 129.210.5.5 | 124.216.78.3 | no | no |

Figure 1: Architecture of DOSBAD

DOSBAD periodically measures the available bandwidth along the paths of the network. It does this using a variant of the probe gap model discussed in Huan Liu's article, "A New Form of DOS Attack in a Cloud and Its Avoidance Mechanism. With this variant, more than just a pair of packets is used. A series of 1500 byte packets are sent along the desired measuring path in order to create a queue at the router. Then two 64 byte packets are sent down the path. The time it takes for the second 64 byte packet to reach the receiver is used with the equation:

$$\text{Available bandwidth} = C * \left(1 - \frac{\Delta o - t_p}{\Delta i}\right)$$

Where C is the maximum bandwidth of the path, Δo is the time gap that the receiver measures between receiving each of the 64 byte packets, $t_p$ is the time to transmit the second packet, and Δi is the time gap between sending each of the 64 byte packets.

The packets should usually be sent along the narrowest path in terms of bandwidth in the network, since that is the most vulnerable area in the network. Repeating the packet sending process once every second, a trend can be observed as to if the available bandwidth begins to change drastically. Also, if the available bandwidth falls to a certain amount to inhibit network performance, DOSBAD can detect this and investigate possible causes.

At some ratio of available bandwidth to maximum bandwidth B, the network will become sluggish. If DOSBAD detects a B at or less than this ratio, it begins looking through its traffic list. The way DOSBAD stores traffic information looks something like this:

| Source (32 bits) | Destination (32 bits) | Acknowledgement Sent (0 for no, 1 for yes) (1 bit) | Acknowledgement Received (0 or 1) (1 bit) | Duration in milliseconds (up to 1 second) (10 bits) |
|---|---|---|---|---|
| 129.210.5.5 | 124.216.78.3 | 0 | 0 | 301 |
| 129.210.5.5 | 124.216.78.3 | 0 | 0 | 498 |
| 131.245.1.7 | 124.216.78.1 | 1 | 1 | 543 |
| 129.210.5.5 | 124.216.78.3 | 1 | 0 | 782 |

DOSBAD stores each instance of traffic going through the network, either to or from a host within the network, within the last second. When DOSBAD detects low available bandwidth, DOSBAD can check this dynamic table, checking for many instances of the same source or destination address. In this rather simple example, there are many instances of the IP address 129.210.5.5 sending packets to the destination within the network of 124.216.78.3. We see that an acknowledgement was sent out for the first instance of traffic, but was not received, and that the other instances from this IP address were not even sent an acknowledgement. This means that 129.210.5.5 is the most likely suspect of launching a DOS attack if there is one being launched.

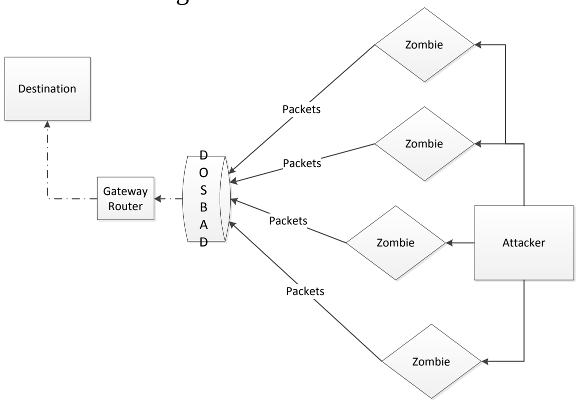


Figure 2: Attack Model

Normally, all incoming packets are going to be encrypted, so DOSBAD cannot check the packets itself to see if the packets look valid from the content. DOSBAD must therefore use packet signature authentication on the packets coming from the suspicious IP address. DOSBAD will store a list of known signatures, much like an antivirus program, and compare the incoming signatures to this list. If they find a match, the means the packet is part of a DOS attack and DOSBAD can have the incoming packets from that IP address dropped.

It may not always be the case that a perpetrator's IP address can be identified. Some attackers spoof their IP address, or use zombie machines to launch a distributed DOS attack. In that case, the only way DOSBAD has to detect the attack is to look at only the destination IP address within the network that has the most packets being sent to it:

| Source (32 bits) | Destination (32 bits) | Acknowledgement Sent (0 for no, 1 for yes) (1 bit) | Acknowledgement Received (0 or 1) (1 bit) | Duration in milliseconds (up to 1 second) (10 bits) |
|---|---|---|---|---|
| 127.215.17.2 | 124.216.78.3 | 0 | 0 | 301 |
| 125.127.18.1 | 124.216.78.3 | 0 | 0 | 498 |
| 131.245.1.7 | 124.216.78.1 | 1 | 1 | 543 |
| 125.117.21.4 | 124.216.78.3 | 1 | 0 | 782 |

Again, DOSBAD notices the many packets being sent to 124.216.78.3. DOSBAD will also still notice the unreturned acknowledgements, even though all the source IP addresses are different. This can indicate a distributed DOS attack against the network. Since it isn't as simple as just dropping the packets from a specific source, DOSBAD will have to check for a 1 on the acknowledgement sent bit with a 0 on the acknowledgement received bit. This ensures that DOSBAD is dealing with one of the zombies since they won't return the acknowledgement. DOSBAD again uses packet signature analysis by comparing the signature of the incoming packet with its list of known attack signatures. Upon finding a match, DOSBAD will again know for sure that this packet is part of a DOS or DDOS. 125.117.21.4 packets will be dropped, then their instances will leave the table. In an updated table, the 1 bit for acknowledgement sent will now move to a different zombie, since the destination has now moved on to trying to verify a different sender. DOSBAD can then check for another 1 0 combination on those two bits and then have the gateway drop that address. This process may continue for a while until network performance returns to normal.

If performance is so bad that no legitimate traffic is getting through at all, it may be beneficial to implement application hopping. The services provided by the destination may be temporarily moved to a different host that isn't receiving nearly as much traffic until the bandwidth to the burdened host can be freed up. This setting is customizable by the cloud service provider.

A situation may occur where DOSBAD's list of known signatures is not up to date with every possible attack signature. In this case, DOSBAD will not be able to detect a signature that is not on its list. With this situation, it may be helpful to log the IP address that DOSBAD does packet signature analysis on, use application hopping, and then using the log to see if there is a new attack signature that can be derived and added to the list of known signatures.

## IV. Implementation

Our implementation was to attempt to simulate DOSBAD. DOSBAD, the software analyzing the packets, is a program developed at our lab for the purpose of taking in readings and converting them to understandable and organized data tables. It then checks them against user generated attacks.

*Example source code for alert detection*

```
private boolean processAlert(Alert alert){
…
rs = server.retrieveNodeHistory(alert.nodeID,
cal.getTimeInMillis() - interval[alert.interval]);
calculator.getData(rs, parameter[k]);
calculator.calcMean();
mean[k] = calculator.getMean();
…
if (mean[k] <= value[k])
        condition[k] = false;
…
alertPanel.displayAlerts(alert);
```

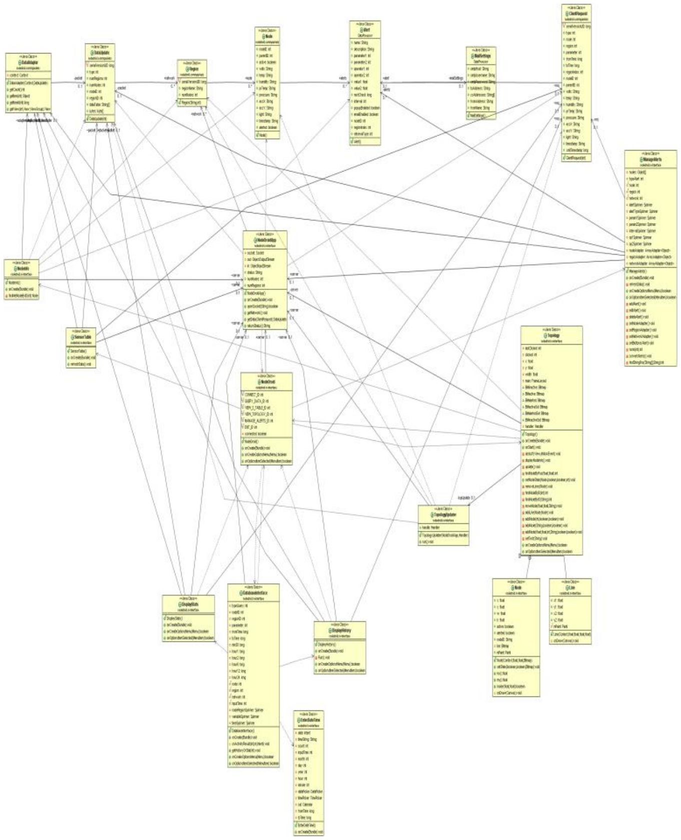


Figure 3: Implementation of DOSBAD

Figure 3 provides the class implementation diagram. It shows the classes and functions used in this implementation. The interaction among the classes shows though lines. Best of our knowledge this implementation is not done before. In order to pass these objects back and forth, both applications needed to utilize Java object input and output streams. When we created these streams and began testing the connection we encountered a problem. We could successfully open both streams, but when we tried to transmit packets, we would get a Java class loader exception. After many days of research and debugging, we located the source of the problem. The packet source code, which should have been identical down to the classpath, was being inadvertently changed.

## V. Conclusion and Future work

In this paper we propose a protocol(DOSBAD) to avoid Denial of Service (DOS) attacks in cloud servers. DOSBAD integrated into a cloud server can be used to monitor what ratio of available bandwidth is being used. To find the maximum available bandwidth of the server, DOSBAD periodically send a series of packets down each possible path within the cloud (router-to-router).

This protocol can be improved by implementing it in actual cloud servers. Different DOS or DDOS attacks can simulated to make sure it can handle multiple attacks at the same time.

## References

[FZ 08] Cloud Computing and Grid Computing 360-Degree Compared Foster, I.; Yong Zhao; Raicu, I.; Lu, S.; Grid Computing Environments Workshop, 2008. GCE '08

[DK 09] Dikaiakos, M.D.; Katsaros, D.; Mehra, P.; Pallis, G.; Vakali, A.; Cloud Computing: Distributed Internet Computing for IT and Scientific Research Internet Computing, IEEE 2009

[MF 09] M Armbrust, A Fox, R Griffith, AD Joseph, RH Katz. Above the Clouds: A Berkeley View of Cloud Computing - 2009, UC Berkley

[VM 04] MA Vouk - Cloud Computing – Issues, Research, and Implementations, Journal of Computing and Information Technology, 2004.

[MP 09 ]Miranda Mowbray, Siani Pearson, A Client-Based Privacy Manager for Cloud Computing, COMSWARE '09 Proceedings of the Fourth International ICST Conference on COMmunication System softWAre and middleware

[CB 11] Chia Yuan Cho, Domagoj Babic, Pongsin Poosankam, Kevin Zhijie Chen, Dawn Song and Edward XueJun Wu, “MACE: Model-inference-Assisted Concolic Exploration for Protocol and Vulnerability Discovery”, To appear in Proceedings of the 20th USENIX Security Symposium, (USENIX Security’11)

[MB 08] Johns, M.; Engelmann, B.; Posegga, J. XSSDS: Server-side Detection of Cross-site Scripting Attacks, Computer Security Applications Conference, 2008. ACSAC 2008.

[WJ 11] Jansen, W.A.; Cloud Hooks: Security and Privacy Issues in Cloud Computing System Sciences (HICSS), 2011 44th Hawaii International Conference on

[XG 09] Jinpeng Wei Xiaolan Zhang Glenn Ammons Vasanth Bala Peng Ning, Managing Security of Virtual Machine Images in a Cloud Environment,CCSW '09 Proceedings of the 2009 ACM workshop on Cloud computing security

[BC 09] Rimal, B.P.; Eunmi Choi; Lumb, I.; A Taxonomy and Survey of Cloud Computing Systems, INC, IMS and IDC, 2009. NCM '09. Fifth International Joint Conference

[FY 08] Foster, I.; Yong Zhao; Raicu, I.; Lu, S.; Cloud Computing and Grid Computing 360-Degree Compared, Grid Computing Environments Workshop, 2008. GCE '08

[BG 09] Andreas Berl, Erol Gelenbe, Marco Di Girolamo, Giovanni Giuliani, Hermann De Meer, Minh Quan Dang, and Kostas Pentikousis, Energy-Efficient Cloud Computing, Incorporating Special Issue: Architecture/OS Support for Embedded Multi-Core Systems, 2009

[WW 09] Cong Wang; Qian Wang; Kui Ren; Wenjing Lou; Ensuring Data Storage Security in Cloud Computing, Quality of Service, 2009. IWQoS. 17th International Workshop

[WW 11] Cong Wang; Qian Wang; Kui Ren; Wenjing Lou , Improved Verifiability Scheme for Data Storage in Cloud Computing, Wuhan University Journal of Natural Sciences 2011

[XB 06] Wei Xu , Eep Bhatkar , R. Sekar, Practical Dynamic Taint Analysis for Countering Input Validation Attacks on Web Applications, 15th USENIX Security Symposium (Vancouver, BC, Canada, August 2006).

[ZS 09] Xinwen Zhang, Joshua Schiffman, Simon Gibbs, Anugeetha Kunjithapatham, Sangoh Jeong, Securing Elastic Applications on Mobile Devices for Cloud Computin, CCSW '09 Proceedings of the 2009 ACM workshop on Cloud computing security

[YR 09] Liang Yan, Chunming Rong and Gansen Zhao, Strengthen Cloud Computing Security with Federal Identity Management Using Hierarchical Identity-Based Cryptography, Cloud Computing Lecture Notes in Computer Science, 2009

# ELECTRONIC IDENTITY MANAGEMENT INFRASTRUCTURE FOR TRUSTWORTHY SERVICES IN E- GOVERNMENT AND E-COMMERCE

**Alberto Polzonetti, Damiano Falcioni, Fausto Marcantoni**
School of Science and Technology
University of Camerino (Italy)
Name.surname@unicam.it

**Abstract** - *A number of potential initiatives are being considered including the creation of a international electronic identity management infrastructure for trustworthy services in e- government and e-commerce. A lot of work has been done in recent years in the field of electronic identity management, including through a series of research programs and pilot projects. While each of these projects contributes new elements to the field of electronic identity management, it is also clear that the results will need to be developed further, refined and integrated. This paper would open the discussion on the need for a "multi-faceted electronic Identification (eID) system for all citizens", as a key enabler for trustworthy interactions between public authorities, businesses, citizens, and within the large spectrum of social networks and communities. This concept, which is also referred to as an ubiquitous eID infrastructure for digital life, is envisaged to offer a wide range of functionalities, including the provision of multiple identity instances, from government- accredited to commercially accepted, and ranging from near-anonymity to strong and unambiguous identification. Furthermore, the system should be user-controlled and privacy- protective, providing the basis for accountability and innovative applications in an open and competitive market.*

***Keywords:*** *Electronic Identity Management, e-government, e-commerce*

## 1 Introduction

International eIDM ambitions are thus high, and it is not yet fully clear how existing initiatives and projects can be integrated into a common vision, or what framework would be needed from a technical, infrastructural, organizational and legal.

There is a need for discussions and consultations to determine exactly what can be expected from a eIDM infrastructure, what the approach and goals should be, and which steps need to be taken next to realize this vision.

The actual debate stresses the need for all states to increase its performance when it comes to the use of innovative ICT solutions, especially in the public sector. In a European context, for example, this emphasis on appropriate public policy is justified, due to the public sector's larger stake in GDP than in other regions of the world. To increase the use of innovative ICT solutions, three interlinked lines of action are proposed :

- Improving the quality and coherence of our investment efforts, as there is currently too much fragmentation which dilutes the efficiency of our investments;
- Raising investment in research and innovation, including through public procurements;
- Stimulating the demand for R&D, by opening up new markets for R&D to respond to real needs and challenges.

A set of measures for the public sector to achieve these goals will be proposed, including large scale actions that go from research to actual procurement and deployment, to ensure that R&D investments have a real impact in practice. This can build on existing building blocks that have already been used in Europe, such as Large Scale Pilots, Public-Private Partnerships and pre-commercial procurement of R&D. One of these areas in which this approach will be applied is the deployment of innovative eID solutions.

## 2 Needs And Objectives

After these introductory remarks and considerations, in this chapter we speak about on the needs and objectives of a eIDM system: what is it that we expect of something termed a "ubiquitous eIDM system for digital life"? What will the expected/desired impact be, and how far do we want to go?

After the development of basic internet services, service paradigms have moved on to web 2.0 services and are now shifting towards a cloud computing model. In this model, eID is often seen as one element of web services that needs to be able to integrate smoothly with other services. If this is to work in practice, a great deal of flexibility will be expected of the underlying eID infrastructure.

One of the first elements of debate in the community was the basic question of what constitutes an eID. This seems to be a very basic question, and a lot of research has been done on this point, but different perspectives can be taken, which will have a very significant impact on how a eIDM infrastructure should be created. Key questions and goals are:

i. The scope and meaning of 'identity' (at least for the purposes of a eIDM infrastructure) needs to be made clear. Intuitively we tend to think of identity in terms of physical people. From an e-government and business perspective, legal entities are equally important however; and it is even possible to consider the broader nothing of an identity of things/objects. The scope and definition of eID changes when we try to outline what we want to identify, and this is particularly important when examining semantics. Currently, exchanging electronic identity information is very complicated simply due to the lack of common semantics (e.g. even the simple notion of a name is interpreted differently from country to country).

ii. Related to this is the question of management of identities: who creates eIDs, and how are these managed? In reality, end users rely on a multitude of "partial identities" to represent or authenticate themselves in specific contexts, and it is unclear how this can be supported in a European eIDM infrastructure, or to what extent it should be. In order to address this question, it needs to be clear who registers and verifies attributes (if at all), and on what conditions these can be exchanged or re-used, or simply confirmed. Relying on market mechanisms to choose an economically optimal solution may not provide desirable results from a data protection perspective.

iii. Thirdly, an advanced identity management system needs to be able to manage links between entities. Simple examples include linking parent A to child B, or linking manager C to company D. Mandate management and role management is the main example of this. There is a lot of work still left to be done on this point: tools need to be created that allow users of a European model system to verify and manage such links.

iv. A fourth crucial element is the reliability of identity information, either in terms of being generally reputable (considered trustworthy) or in terms of real guarantees (accountability in case of problems). The role of the public and private sector was discussed in this regard as an interesting example: 'official identities' or 'formal identities' are often issued or managed by the public sector, but this doesn't necessarily mean that identification services provided by the private sector are less trustworthy or less usable in practice. From the end user's perspective, functionality is more important driver than clear guarantees in relation to the trustworthiness of identity information, as can be seen in the increasing importance of reputation based identification (e.g. in social networks, which are largely based on establishing trustworthiness via peer-to-peer appreciation). From the service provider's perspective, trustworthiness – especially in terms of accountability and liability – is much more important, and reputation as such may not hold sufficient appeal from this perspective. It has already been made clear in the past that future eIDM infrastructure in Europe should be multi-level, i.e. permitting varying levels of security/reliability. This is one of the key gaps that still needs to be filled.

v. Functionally, it would be important to uncouple the provision of electronic identification or authentication services from specific applications. An 'invisible eID infrastructure' is key to creating an open eID model that could be taken up in commercial and public sector applications. In that respect European governance has the benefit of being conceptually based on a roughly "federated" model. A web of services is a model that plugs into this same concept of thinking: application independence (service-independence) of the eID infrastructure is important.

vi. There is also the question on whether a European legal framework, or at least European guidelines for regulations, is needed. This issue was raised in relation to a number of points, including the multi-level reliability issue addressed above: some participants felt that governments needed to set up the rules and regulations to issue/manage tokens/eIDs, preferably based on European guidelines. Currently, national legal barriers impede some approaches that are being explored at the European level; examples include the German ban against the intervention of intermediaries in the relationship with the public sector (including e-government services), which impedes the use of proxy based identification models; and the ban on using permanent unique identifiers for generic purposes in Germany and Hungary, which means that any European approach cannot require the prior existence of such identifiers. Guidance is necessary on what the consequences of European initiatives will be, and how we can operate within the limits of applicable laws, given the lack of direct European regulatory competence to harmonize eID regulations.

vii. Finally, the privacy and security aspect should take a central role. The point was made and discussed that private industry (on-line service providers, financial services, mobile communications, …) does not have much of a problem in getting the identity information that they want and as reliable as they need it to be. But there is a significant problem from the opposite perspective: how do you empower users to enforce their rights and manage their data? This should be addressed in a European eIDM infrastructure as well, and this should be done soon; security and privacy protection cannot be taken up as an afterthought. Innovative systems exist in current research, but the infrastructure must be set up to implement this.

Collectively, the considerations above contain a good summary of what can be expected or should at least be considered as the needs of a eIDM infrastructure (in no particular order):

- Clear definition of scope: what is the concept of identity that we want to address at a European level?
- Management of identity: which entities need to be involved in managing an identity, and what is their function?
- Management of relationships: how do entities whose identities are managed relate?
- Trustworthiness of identity: how can you trust the identity, especially in terms of accountability and liability?
- Identity provisioning in applications and services: how do you use identity in an application?
- Clear legal framework: how to regulate the use and management of identity?
- Privacy protection and secure identity management: how do you integrate users' rights into the infrastructure?

# 3 IMPLEMENTATION

Having discussed the needs in relation to eIDM infrastructure, this chapter examined how an infrastructure meeting these requirements could be implemented, taking into account the diverging and demanding needs in relation to such issues as identity re-use, tiered reliability and trust, private sector support, privacy-by-design and enforcement of applicable rules.

The first aspect extensively discussed in this regard was the strong role that innovative technologies could play in developing this infrastructure. Regardless of the preferred technology, any electronic identity management system is inherently dependent on the use of a secret in some form over another. There are already advanced identity management models in place that allow you to spread a secret robustly over many locations, and that allow you to limit the disclosure of identity information (such as e.g. IBM's Idemix or Microsoft's U- Prove). This allows you to increase security and reliability and improve data protection enforcement. Such PETs need to be developed and deployed further, and it needs to be examined in particular how take-up of such advanced models can be encouraged. The development of a business case around such models is crucial in this regard, as will be further discussed below. Finally, any approach taken at the European level needs to be sufficiently flexible to take up newer approaches to identity management that might emerge or increase in popularity, including e.g. identification based on biometric encryption (through local verification of biometric information) or mobile identification.

As a complement to the technological tools deployed, the architecture as a whole also needs to be designed to meet the objectives above. The role of validation services and proxy services was mentioned in this respect, as solutions that were currently being tested in STORK and PEPPOL, and that were also being examined by private sector partners. These approaches are appealing, as the main issue to be resolved here is to determine reliability/authentication levels; other issues could then be handled by federating (i.e. managing them at the national level). However, other participants in the meeting rightly indicated that such solutions would need to implement strict safeguards to address privacy and security issues: it should be ensured that such solutions cannot become a single point of failure, and that they do not retain information on identity use; otherwise, they constitute a significant privacy threat. Other approaches should therefore also be considered.

Both with regard to technology and infrastructure, the importance of working with industry partners was generally recognized to be crucial. Public-private partnerships and systematic coordination with industry was seen as a key way of ensuring that any model adopted at the European level would also see substantial take-up in reality. It is necessary however to consider the different stakeholders, and particularly the different interests between eIDM users and eIDM vendors. Without a proper link to industry however, European initiatives risk remaining at the theoretical or pilot level, or seeing limited practical use. The integration of harmonized eID middleware implementations in existing operating systems distributed by major vendors was given as an example to be looked at. By harmonizing protocols, the integration and use of existing and new eID solutions could be facilitated to a significant extent.

However, measures to achieve the desired outcomes should not be focused exclusively on the technological and infrastructural aspects, but also on legal issues. There was some doubt whether European regulation was a useful (or even possible) route forward, given the fact that identity management is generally regarded as a national competence, but it was considered that guidance and support could be provided once an appropriate paradigm for an eIDM infrastructure was established.

In addition to the technical, infrastructural and legal challenges, perhaps one of the most challenging issues is creating a model that has sufficient appeal to end users and service providers, i.e. ensuring that the eIDM platform has real business appeal. To do so, we need to make sure that our own goals and expectations as described above match those of the stakeholders. For instance, while data protection issues and user control are societal needs that must be protected to safeguard our European values, end users' perceptions seem be driven more by short term convenience. There may be a need to reflect on future needs and values in the discussion between experts and end users in this respect.

Naturally, we need to make sure that there is a real business model that makes sense to stakeholders. The example of banks was discussed on this point: even banks that could use a generic eID token (like a government issued eID card) are generally reluctant to do so, even if it would be more secure than their own existing solutions. At least part of the reason is that having their own solutions gives them full and exclusive control over the business model, and that their own tokens act as an advertising medium in a way that generic eIDM tokens would likely not be able to offer. Can this be addressed appropriately? This concern however would be completely different for small innovative service companies.

Globally, while there was a strong consensus on the importance of each of the aforementioned issues (technology, infrastructure, legal framework, business case), it was also felt that some additional research would be required to offer satisfactory answers that would allow the creation of a coherent and suitable European eIDM framework. The question was raised on whether an 'eIDM research roadmap' was needed, and if so, what it would look like. This is a complicated issue, due to the need to continuously take into account the changing eID landscape in each of the countries involved and in the eID industry. A flexible approach would thus be needed, with a strong emphasis on maintaining open communications with industry representatives.

Despite this complexity, if we want to go from research to implementation as envisaged by the planned Communication, we need to make sure that our knowledge of the eIDM landscape is complete, and research on a number of key issues still seems needed. Principally, the conceptual model behind the functionality that we are looking for is not clear: how can specific roles and responsibilities be defined and organized in a general eIDM framework, and how can the advanced technological options commented above be integrated into this framework? Secondly, the economics behind eIDM are not well understood, or more accurately: it is unclear how the objectives that we have envisaged above can be implemented in a way that is attractive for end users and service providers alike. Broadly painted, many service providers with an extensive customer base and the required infrastructure want cheap access to as much info as they can use, and end users are more interested in convenience than in security; at any rate, it seems unlikely that end users would be willing to pay a premium for security. It would be interesting to see if there are cases currently available that are supported by the market (as opposed to government mandate or subsidies), or what encouragement measures are being applied effectively to improve the economic appeal of electronic identities.

Apart from the concepts and economics, the issue of accountability was presented as an area of discussion. Electronic identity management is needed to support accountability, by giving the service provider a way to reliably link certain actions to certain users. Currently, this operates mostly within closed contexts: service providers can rely on electronic identities either because they issue or manage them themselves, or because they have a clear contractual relationship with the issuer of the credentials. Open eID infrastructures that are not limited to a closed group of service providers see much less uptake, and the issue of accountability plays an important impeding role here. This becomes even more clear when discussing whether private sector issued eIDs should be usable in a public sector context. While there is no objection to this in principle, there is still a substantial lack of trust and a real need for sufficient accountability guarantees.

In addition, as was also noted above, even if accountability from the end user is sufficiently guaranteed to the service provider, the inverse relationship (accountability of the service provider to the end user) is not yet guaranteed in practice; this is an aspect where further research or possibly regulatory guidance might be needed, including in terms of implementing real privacy-by-design solutions, to ensure that our envisaged European eIDM approach is sufficiently focused on the end users' interests as well. In the same respect, the questions of usability and accessibility were raised: solutions need to be inclusive to all users. While a lot of research has already been done in this domain, there is a clear need to link this research to real results.

Globally, there was a consensus that new research would be needed to coordinate existing knowledge and know-how (which is already available to a significant extent in Europe) into a coherent vision. A comprehensive approach would be needed to form a coherent picture of how existing solutions and newer innovative approaches could be integrated into an eIDM infrastructure that supports the needs and objectives defined above. The issues of accountability, economics and inclusiveness were identified as key problems to be addressed in this research. Further efforts could then focus on creating the necessary components in a second stage.

It is thus clear that future research will be instrumental in shaping the approach taken towards creating the envisaged ubiquitous eIDM infrastructure. Specific tools are to steer this research or to bring it to fruition, including through pilot implementations or actual deployment.

A number of interesting possibilities for moving forward were none the less discussed, including:

*Identification and dissemination of best practices in eIDM initiatives, as is currently already being explored (e.g. through the eID Observatory);*

*Collecting and disseminating clear overviews of the art in eIDM solutions, as a way of encouraging take-up of advanced solutions by currently less advanced market players and as a way of permitting frontrunners to explore innovative solutions more easily;*

*Focusing on standardization efforts (e.g. standardization of interfaces) to reduce the complexity of the problems we are facing;*

*Identifying and exploring innovative eIDM approaches, to determine which approaches are already being tested/implemented that could meet some of the requirements above.*

These approaches are appealing, as they would allow progress to be made irrespective of the final outcome to be chosen. However, it is clear that a coherent model for a eIDM infrastructure would need to be determined before the outcomes from these approaches can be leveraged fully, and that the full societal context needs to be considered, including the need for inbuilt privacy protection and security.

## 4 Socio-Economic Impact

In this section we discuss the socio-economic impact of creating a eIDM infrastructure, including in terms of financial gains and general benefits to all stakeholders.

From a macro-economic perspective, one of the first interesting aspects of this debate focused on export possibilities. The European approach to identity is rather particular, and reflects our cultural attitudes towards identity, data protection and privacy. The discussions above (including on technical, infrastructural and legal needs) reflected this: there is a desire to ensure that our eIDM infrastructure matches our cultural perceptions on these issues. While this European approach may not be universally welcome, it does open interesting avenues for exploitation. Some regions (including e.g. in Asia) have shown some interest in European personal data paradigms, and we should thus not overlook the possibility that the eIDM solutions developed in Europe could prove to be valued exports. Thus, from a macro- economic perspective, there appears to be a real potential for validation.

However, the micro-economic perspective must also be considered, and it was clear that on this point the socio-economic impact depends on whose interests you're considering (service providers, end users, or solution vendors). The return on investment therefore also depends on whose perspective you take, and one of the key complexities to be overcome is the need to make sure that there is a fair distribution of benefit; otherwise, the solution will not be taken up. This is linked to the business model question raised earlier: who is profiting from the infrastructure, and who is paying for it? These two aspects need to be sufficiently linked.

You might want to consider an authentication process as an example of a business model. Such a model is not necessarily a best practice (or legally permissible in countries that require CSPs to offer free verification services), but it does illustrate the point: without a real business model that matches cost with benefit, uptake will suffer. The Norwegian and Swedish public sector, for examples the public sector acknowledged that they wanted end users to take up eIDM, and that taking up part of the bill as a government was an acceptable cost of public policy. In contrast, in the UK initiatives relying on the users' willingness to pay for authentication certificates failed. This was acknowledged to be a key question: how do you model pricing and benefits to optimize uptake?

In that respect, it is clear that underlying costs that affect the price tag must also be acknowledged and accounted for. Liability is a key component of cost: during the discussions, Nordic approaches emphasizing trust were contrasted with other European approaches emphasizing accountability. While both approaches can function within their respective markets, interconnecting them will be quite complicated, due to the need to bridge this difference in perception of accountability requirements. Similarly, there is often a price to be paid for simplicity and accessibility: username/password systems may be easy and seem cheap, but when support costs for forgotten passwords are factored in, the picture may change. These elements also play a role if you want to accurately gauge costs and benefits.

## 5 Conclusion

It seems that there was a good consensus on the objectives for a eIDM approached as commented in the first section above, and on the need for additional research on a number of issues, including on accountability, economics and inclusiveness. These should permit the creation of a coherent concept for a ubiquitous European eIDM infrastructure, suitable for adoption by public and private sector service providers, and adjusted to the needs and expectations of the end users. The creation of an appealing business model that links costs to benefits will be crucial to ensure real take-up, keeping into account that both costs and benefits will have clearly visible and less apparent implicit components.

These issues will not be solved in the short term, and further reflection and refining of the positions above will still be needed to arrive to a clearer picture of Europe's post-i2010 objectives and strategies in the field of electronic identity management.

Please use the styles contained in this document for: Title, Abstract, Keywords, Heading 1, Heading 2, Body Text, Equations, References, Figures, and Captions.

*Do not add any page numbers and do not use footers and headers (it is ok to have footnotes).*

# 6 References

[1.] BOTTERMAN, M., GRAUX, H. and MITRAKAS, A., "A Roadmap for a pan-European eIDM Framework by 2010", http://europa.eu.int/information_society/activities/egovernment_research/doc/ eidm_roadmap_paper.pdf (accessed November 2011)

[2.] DUMORTIER, J. and GRAUX, H., Legal Study on Legal and Administrative Practices Regarding the Validity and Mutual Recognition of Electronic Documents. November 2006. Draft Final Report. Tender No. ENTR/04/67. 111 pp., http://ec.europa.eu/enterprise/ict/policy/legal/index.htm (accessed November 2011)

[3.] ENISA "Privacy and Security Risks when Authenticating on the Internet with European eID Cards", November 2009, http://www.enisa.europa.eu/activities/identity-and-trust/eid/eid-online-banking, Accessed November 2011)

[4.] EU research & innovation strategy for digital technologies, http://ec.europa.eu/information_society /newsroom/cf/ document.cfm? action=display&doc_id=597 ( accessed January 2011) ]

[5.] Eurosmart," European Citizen Card: One Pillar of Interoperable eID Success", October 2008, https://www.eid-stork.eu/dmdocuments/public/ecc-position-paper-final.pdf (accessed November 2011)

[6.] Leitold H.and Posch R. and Rannenberg K. and Krontiris J., "eID Interoperability (chapter 12) book title: Handbook of eID Security: Concepts, Practical Experiences, Technologies",{2010},pages = {167 - 186}, publisher = {Walter Fumy, Manfred Paeschke}

[7.] Reinhard Posch,"Trust, Security and Identify Managament: The Austrian Viewpoint",Conference on Trust and Identity Management,year = {2007},pages = {12 - 21}

# An Overview of Laws and Standards for Health Information Security and Privacy

**Francis Akowuah[1], Xiaohong Yuan[1], Jinsheng Xu[1], Hong Wang[2]**
[1]Department of Computer Science, North Carolina A&T State University, Greensboro, North Carolina, USA
[2]Department of Management, North Carolina A&T State University, Greensboro, North Carolina, USA

***Abstract***
*In the complex technological world that healthcare organizations and their business associates operate, there exist security threats and attacks which render individually identifiable health information vulnerable. Laws exist to ensure that healthcare providers take practical measures to address the security and privacy needs of health information. There are also standards that assist healthcare entities to meet the security and privacy requirements of health information. This paper provides a chronological overview of U.S. laws and standards related to health information security and privacy, such as HIPAA, Sarbanes-Oxley Act, HITECH, COBIT, ISO/IEC 27002 2005, and CSF.*

**Keywords**
Health information systems, security and privacy, laws, standards

## 1. Introduction

With the adoption of health information systems, healthcare organizations and their business associates operate in a complex, interconnected, technological world. Individual identifiable health information thus become vulnerable to an entire new set of security threats and attacks such as malicious code, denial of service, and many others. When these threats and attacks are successful, individual privacy becomes woefully invaded. Moreover, it brings about economic loss and reputation damage to healthcare organizations [1].

Fortunately, the United States government has taken keen interest in healthcare provisioning and has enacted laws and regulations over the years to curb the security and privacy problems faced by healthcare organizations. These laws require healthcare entities to take measures to address the security and privacy needs of health information. Notable among these laws are Health Insurance and Portability Act (HIPAA), Sarbanes-Oxley Act, and Health Information Technology for Economic and Clinical Health (HITECH). Standards such as Control Objectives for Information and Related Technology(COBIT), ISO 27002 2005 and Common Security Framework (CSF) also exist to assist these entities to meet the security requirements of the law.

In this paper, we give a chronological overview of the United States laws and standards for health information security and privacy. We consider federal laws such the HIPAA, Sarbanes-Oxley Act, and HITECH. The requirements of these laws and their implications on healthcare providers are discussed. Furthermore, recommendations and best-practices of COBIT, ISO/IEC 27002 2005, and CSF are overviewed. Laws that were enacted before 1996 to protect individual health information are also briefly discussed. Though other laws and standards exist that relate to information systems, only those related to health information security and privacy are discussed in this paper.

The remainder of the paper is structured as follows. Section 2 provides information about pre-HIPAA laws. Sections 3 to 5 overview HIPAA, Sarbanes-Oxley and HITECH respectively. Sections 6, 7 and 8 discuss COBIT, ISO/IEC 27002 2005 and CSF standards. Section 9 concludes the paper.

## 2. Federal Laws Prior To HIPAA

Before HIPPA came into enforcement there had not been a far-reaching federal regulation that ensured or catered for private health information. The Freedom of Information Act passed in 1966 provided the American public the right to acquire information from federal agencies with some exceptions. Among the nine exceptions were access to personnel and medical information, since a disclosure of such information obviously was an invasion of privacy [2]. This exception was however not strong enough to protect patient records and other health information.

As a result, the Privacy Act of 1974 was enacted specifically to protect patient confidentiality. However, only federally operated health care facilities had to comply with this act. It was an important legislation because it explicitly stated that patients had the right to access and amend their medical records. Medical facilities were required under this act to document all disclosures of patients' health information [2].

## 3. Health Insurance Portability And Accountability Act Of 1996 (HIPAA)

HIPAA was signed into law on August 21, 1996. The objectives were to make health care delivery more efficient and to increase the number of Americans with health insurance coverage. Three main provisions were made to achieve the objectives. They are portability, tax and administrative simplification provisions. This paper focuses on the administrative simplification provision.

The Department of Health and Human Services (HHS), as instructed by the administrative simplification provision, issued a number of regulations concerning electronic transmission of health information, which was expanding rapidly in the early 1990s. Although the ultimate aim of the provisions was to standardize the use of electronic health information, Congress realized that advances in electronic technology could compromise the privacy of health information. Consequently, nationwide security standards and safeguards were developed for the use of electronic health care information (referred to as the Security Rule). In addition, privacy standards were created for protected health information (referred to as the Privacy Rule) [3]. In what follows, the HIPAA Privacy Rule and Security Rule are described.

### 3.1 HIPAA Privacy Rule

Standards for Privacy of Individually Identifiable Health Information or the Privacy Rule was first published on December 28, 2000 by the United States Department of Health and Human Services. The rule however became effective in April 2003 [4]. It promulgates detailed regulations concerning the types of uses and disclosures of individually identifiable health information that are permitted by the covered entities [3].

The Privacy Rule defines *"individually identifiable health information"* held or transmitted by a covered entity or its business associate, in any form or media, whether electronic, paper, or oral as Protected Health Information (PHI). “Individually identifiable health information” refers to information, including demographic data that relates to the individual’s past, present or future physical or mental health or condition, the provision of health care to the individual, or the past, present, or future payment for the provision of health care to the individual, and that identifies the individual or for which there is a reasonable basis to believe can be used to identify the individual. Employment data kept by covered entities in their capacity as employer are not considered as PHI [5]. Typical examples of PHI are medical records, billing records, enrollment, payment, claims adjudication, etc. [6].

While allowing the flow of health information to enhance high quality health care and protect public health and well-being, the Privacy Rule ensures that individuals’ health information is properly protected. In other words, while it provides protection of the privacy of the patient’s health information, it also allows important uses. In order to allow the flow of information, covered entities are permitted (not required) to disclose and use health information without individual’s authorization for the following purposes:

- Treatment, payment and healthcare operations
- Public interest and benefit activities
- Limited data set for research, public health

The Privacy Rule also permits informal permission to be obtained from the individual, usually, in situations when he/she is incapacitated. The individual has the opportunity to agree, acquiesce or object. Moreover, individuals can request access to and an accounting of uses and disclosures of health information from covered entities [5]. Patients have the right to access their information and to request for how the information has been disclosed.

### 3.2 HIPAA Security Rule

The Health Insurance Reform: Security Standards or the Security Rule was published in the federal register on February 20, 2003 but compliance began on April 21, 2006. [4]. It establishes national standards to protect individuals’ electronic personal health information that is created, received, used, or maintained by a covered entity. Once again, covered entity refers to organizations that are subject to the rule. The Security Rule requires appropriate administrative, physical and technical safeguards to ensure the confidentiality, integrity, and security of electronic protected health information [5].

Administrative safeguards are policies and procedures that depict how the covered entity complies with HIPAA. These include written privacy policies and the designation of a Privacy Officer. Physical safeguards control physical access to protected health information to avoid unauthorized access to protected data. Technical safeguards control access to computer systems and protect PHI transmitted over open networks from being intercepted [4].

The emphasis must be laid here that while HIPAA Privacy Rule protects the privacy of PHI, HIPAA Security Rule protects only information that a covered entity creates, receives, maintains or transmits in *electronic form*. Hence it does not apply to PHI transmitted orally or in writing. Electronic transmission includes media such as the Internet, extranets, private networks, leased lines, and dial-up lines. It does not include paper faxes, voice mail, telephone calls, or videoconferencing [6].

## 4. Sarbanes-Oxley Act of 2002 (SOX)

The name of the act was coined from Senator Paul Sarbanes and Representative Michael Oxley who drafted the act in 2002. The intent is “to protect investors by improving the accuracy and reliability of corporate disclosures made pursuant to the security laws, and for other purposes”. Section 302 spells new standards for corporate accountability and penalties for acts of wrong-doing. It holds Chief Executive Officers (CEOs) and Chief Financial Officers (CFOs) accountable for the accuracy and completeness of financial statements. According to Section

404 of the act, publicly-traded organizations are supposed to implement internal controls and procedures to communicate, store and protect financial data. These controls must be protected from internal and external threats and unauthorized access, including those that occur through online systems and networks [7]. Organizations must assess the effectiveness of these controls and report to Security and Exchange Committee (SEC).

With regard to healthcare organizations, financial information includes patient direct payments for healthcare, health insurance and other health plan payments. As defined in HIPAA, individually identifiable health information refers to information, including demographic data that relates to "…the past, present, or future payment for the provision of health care to the individual, and that identifies the individual or for which there is a reasonable basis to believe can be used to identify the individual" [5]. It must be noted that the Sarbanes-Oxley Act does not explicitly discuss health information security in the text of the act. However, most modern accounting systems are computer-based and are often incorporated in health information systems. Accurate financial reporting depends on reliable and secure computing environments. Review or assessment of internal controls is not complete without mentioning information assurance or security. As such, information assurance professionals and other Information Technology (IT) professionals need to understand Sarbanes-Oxley Act to develop strategies to assist their organizations to comply with SOX [8] . SOX requirements indirectly compel management to consider information security controls on systems across the organization in order to comply.

Moreover, the act created the Public Company Accounting Oversight Board (PCAOB) to assist in the implementation and oversight of Sarbanes-Oxley Act. The board guides and oversees auditors as they assess a company's compliance with SOX. PCAOB created Proposed Auditing Standard to provide detailed guidance in the assessment process. In a release, PCAOB stated this:

> *Determining which controls should be tested, including controls over all relevant assertions related to all significant accounts and disclosures in the financial statements. Generally, such controls include: …Controls, including information technology general controls, on which other controls are dependent* [9].

In essence, this statement asserts that information technology (IT) general controls form the foundation for many other types of financial reporting controls and therefore, must be assessed for SOX.

The requirements are quite hard to implement. In the first place, security best practices are not well defined in the act. Also, some organizations have budget constraints to enable them to implement the needed security technologies. This also goes in line with obtaining the right security expertise. Hence difficulties in deploying and managing required technology come as a result [7].

# 5. Health Information Technology for Economic and Clinical Health (HITECH)

On February 17, 2009, the Health Information Technology for Economic and Clinical Health (HITECH) Act, enacted as part of the American Recovery and Reinvestment Act of 2009 (ARRA), was signed into law. The aim was to promote the adoption and meaningful use of electronic health record (EHR) technology. Privacy and security concerns associated with the electronic transmission of health information are addressed by Section D of the act [10]. HITECH responded to the increased public awareness and debate over the regulations made by the HIPAA Privacy and Security Rules. HIPAA Security and Privacy provisions and penalties, under HITECH, are applied directly to business associates of covered entities. Hitherto, the provisions and penalties were applied to business associates through contractual provisions with covered entities. Business associates are required to restrict the use and disclosure of protected health information. In default, like covered entities, they shall be subjected directly to the civil and criminal penalties for violating HIPAA regulations [11].

HITECH Act imposes more stringent regulatory requirements than the Security and Privacy Rules of HIPAA. It also increases the severity of penalties for a violation of HIPAA and provides funding and incentives for hospitals and physicians for the adoption of health information technology. The breach notification process was also expanded as new conditions and penalties for noncompliance were also stipulated. For instance, in times of breach, covered entities shall notify victims within sixty days after the discovery of the breach. In the case where covered entity does not have contact information of victims, breach must be posted on their website or make media notifications (local). If more than 500 people are affected by the breach, state media and government notifications are required [10, 11].

In relation to the adoption of Electronic Health Record (EHR) systems, where almost all protected health information (PHI) are digital, patients become more vulnerable to scams and other threats. HITECH thus establishes protocols and certifications for health information products. The protocol refers to the process of notifying breaches. Certification program is handled by the National Institute of Standards and Technology. EHR systems are also required to use some form of encryption technology to render PHI "unusable, unreadable, or indecipherable" to unauthorized individuals. Practitioners must destroy all unencrypted PHI after use [10].

Other new requirements of HITECH include [10,11]:

- Covered entities must honor an individual's request that information be withheld from health plan providers if care is paid for in cash.
- Covered entities must be capable of providing a 3-year audit trail of patient health information disclosures upon request.
- Covered entities may not receive payment for communicating with patients for marketing purposes without the specific authorization of the patient (including fundraising, solicitations, etc.).
- Employees of covered entities or other individuals who knowingly access, use, or disclose PHI for improper purposes will be subject to criminal penalties.
- Civil penalties for violations under HIPAA are increased, depending on the conduct.

## 6. Control Objectives For Information And Related Technology (COBIT)

COBIT is a framework created by Information Systems Audit and Control Association (ISACA) for IT management and IT Governance. The first version was released in 1995 with the latest version being COBIT 5. It is a supporting toolset that allows managers to bridge the gap between control requirements, technical issues and business risks. With COBIT, organizations are able to develop policy and good practice for IT control throughout organizations. It stresses on regulatory compliance, assists organizations to increase the value attained from IT, enables alignment and simplifies implementation of the COBIT framework. Globally, COBIT is used by those with primary responsibilities for business processes and technology, those who depend on technology for relevant and reliable information and control of information technology [12].

ISACA strives to underscore the importance of technology in business processes and the need for management to appreciate it. ISACA also asserts that, in determining the appropriate technology to use and how to control its use, management needs to understand the risks and constraints in order to make good business decisions.

COBIT has four domains which include Plan and Organize, Acquire and Implement, Deliver and Support, Monitor and Evaluate. Thirty-four (34) IT processes are categorized under these four domains. COBIT covers security in addition to all the other risks that can occur with the use of IT. Although one out of the 34 processes is specifically devoted to security, control objectives that address security are scattered throughout the various processes in each domain [13]. COBIT requires deep expert knowledge to implement each application because of its generic nature. Although the guideline of security management is also published, its content is abstract [14].

The process devoted to security is called Ensure Systems Security which is defined under Deliver and Support domain. Ensure Systems Security provides security guidance on the following [13]:

- Manage Security Measures
- Identification, Authentication and Access
- Security of Online Access to Data
- User Account Management
- Management Review of User Accounts
- User Control of User Accounts
- Security Surveillance

IT professionals designing health information systems can follow the guidance provided in COBIT to ensure security.

## 7. ISO/IEC 27002 2005

Previously known as ISO/IEC 17799:2000, ISO/IEC 27002 2005 is a standard that can be applied to general information security management. In other words, it demonstrates what can be done to protect an organization's information assets. When ISO/IEC 17799:2000 was officially published on June 15, 2005, it was known as ISO IEC 17799 2005. On July 1, 2007, the name was formally changed to ISO IEC 27002 2005. However, much of its content is the same with some sections added [15].

As stated in the official title page ISO 27002 is a "code of practice for information Security Management". Any organization seeking to adopt a comprehensive information security management program or improve its existing information security practices can use the standard. Although ISO/IEC recommends a complete consideration of the practices, organizations do not have to implement every recommended security practice stated therein. The important thing is to know what works best for the unique information security risks and requirements. The ISO standard asserts that information can be protected using a wide variety of controls. Such controls include hardware and software functions, procedures, policies, processes and organizational structures. Organizations including healthcare organizations, must develop, implement, monitor, evaluate and improve these types of security controls [15].

## 8. Common Security Framework (CSF)

Released in March 2009, CSF was established by The Health Information Trust Alliance (HITRUST) in collaboration with healthcare, technology and information security leaders. Organizations that create, access, store or exchange personal health and financial information can use

Table 1 Laws and Standards for health Information Security and Privacy

| Law/Standard | Subject | Date | References | Description |
|---|---|---|---|---|
| The Freedom of Information Act | Acquiring information from federal agencies | 1966 | [2] | Provides the American public the right to acquire information from federal agencies with some exceptions, which includes health information. |
| The Privacy Act of 1974 | Patient confidentiality | 1974 | [2] | Required federal operated health facilities to protect the confidentiality of patients' medical records. |
| HIPAA | Health care security and privacy | August 21, 1996 | [5] | To protect individuals' electronic personal health information that is created, received, used, or maintained by a covered entity. |
| Sarbanes-Oxley Act | Financial reporting | 2002 | [7] | Improves the accuracy and reliability of corporate disclosures made pursuant to the securities laws, and for other purposes |
| HITECH | Health information system | February 17, 2009 | [11]<br>[10] | To promote the adoption and meaningful use of health information technology. |
| COBIT | Risk management | 1995 | [12] | Framework for IT management and governance |
| ISO 27002 | Information security management | July 1, 2007 | [15] | Code of practice for information Security Management |
| CSF | Health information security | March 2009 | [17] | Provides the needed structure, detail and clarity that pertains to information security that is tailored to the healthcare industry. |

CSF which is the first IT security framework developed specifically for healthcare information [16]. Organizations benefit in terms of the needed structure, detail, and clarity that pertains to information security that is tailored to the healthcare industry. With the help of prescriptive set of controls and supporting requirements, organizations are able to meet the objectives of the framework. CSF leverages and cross-references existing standards and regulations helping to avoid the introduction of redundancy and ambiguity into the industry. By normalizing the security requirements of healthcare organizations including federal, state and other standards, CSF helps organizations to easily understand their compliance status across a wide range of authoritative sources and standards [17].

CSF is organized into two components:
*Information Security Control Specifications Manual*: It is best-practice-based specification that provides prescriptive implementation guidance. It entails recommended security governance practices and security control practices to ensure the effective and efficient management of information security.
*Standards and regulations mapping:* A reconciliation of the framework to common and different aspects of generally adopted standards and regulations. The CSF includes 42 control objectives and 135 control specifications based on the ISO/IEC 27001:2005 and ISO/IEC 27002:2005 standards [17]

CSF can only be accessed by subscribing to HITRUST Central, the managed online community for healthcare information security professionals. Standard subscriptions at no charge are available to individuals from qualifying organizations as defined by HITRUST. The online, interactive version of the CSF, authoritative sources and the CSF Assurance Kit is available only through a paid subscription [17].

# 9. Conclusion

A number of laws and standards exist to ensure the security and privacy of health information. This paper provides an overview of U.S. laws such as HIPAA, Sarbanes-Oxley Act and HITECH, as well as standards such as COBIT, ISO/IEC 27002 2005 and CSF. These laws and standards can be summarized in Table 1.

Although standards such as COBIT and ISO 27002 2005 are generic in nature, healthcare organizations can implement them to achieve security and privacy of health information as required by federal laws such as HIPAA and HITECH. However, healthcare providers, health plans, business associates and all other covered entities shall only reap the benefits if these standards are implemented properly. Drastic steps must also be taken to comply with all the rules and regulations required by laws. Security is everyone's business, as such, all parties in an organization should be involved in playing their roles in securing health information.

## Acknowledgements

This work is partially supported by NSF under grant HRD-1137516, and by Department of Education under grant P120A090049. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation and Department of Education

## References

[1] Ruoyu Wu, Gail-Joon Ahn, and Hongxin Hu, "Towards HIPAA-Compliant Healthcare Systems," in *Proceedings. of* the *2nd ACM SIGHIT International Health Informatics Symposium*, Miami, 2012, pp. 593-602.

[2] Karen A. Wager, Frances W. Lee, and John P. Glaser, *Health Care Information Systems: A Practical Approach for Health Care Management*, 2nd ed., John Wiley & Sons, Inc., 2009.

[3] Sharyl J. Nass, Laura A. Levit, and Lawrence O. Gostin (Eds.). Institute of Medicine (US) Committee on Health Research and the Privacy of Health Information, *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*, National Academies Press, 2009, Washington DC, US.

[4] Alan R. Heminger and John Chessman, "A Study of U.S. Battlefield Medical Treatment/Evacuation. Compliance with HIPAA Requirements," in *Proceedings of the 42nd Hawaii International Conference on System Sciences*, Hawaii, 2009.

[5] Office of Civic Rights. (2003, May) Department of Human & Health Sciences. Accessed on April 2012, Availabe at: http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/privacysummary.pdf

[6] Charles A. Shoniregun, Kudakwashe Dube, and Fredrick Mtenzi, *Electronic Healthcare Information Security*. New York: Springer Science+Business Media, LLC, 2010.

[7] SOX-Online. (2006) SOX Online. [Online]. HYPERLINK "http://www.sox-online.com" http://www.sox-online.com

[8] Greg Stutts. (2004, May) SANS Institute InfoSec Reading Room. [Online]. HYPERLINK "http://www.cs.jhu.edu/~rubin/courses/sp06/Reading/soxForInfoSec.pdf" http://www.cs.jhu.edu/~rubin/courses/sp06/Reading/soxForInfoSec.pdf

[9] PCAOB. (2004) Public Company Accounting Oversight Board. [Online]. HYPERLINK "http://pcaobus.org/Rules/Rulemaking/Docket008/2004-03-09_Release_2004-001-all.pdf" http://pcaobus.org/Rules/Rulemaking/Docket008/2004-03-09_Release_2004-001-all.pdf

[10] Eric M. Johnson and Nicholas Willey, "Will HITECH Heal Patient Data Hemorrhages?," in *IEEE, System Sciences (HICSS), 2011 44th Hawaii International Conference*, Kauai,Hawaii, 2011, pp. 1-10.

[11] Linn Foster Freedman. (2009, February) The Health Information Technology for Economic and Clinical Health Act (HITECH Act): implications for the adoption of health information technology, HIPAA, and privacy and security issues. [Online]. HYPERLINK "http://www.nixonpeabody.com/publications_detail3.asp?ID=2621" http://www.nixonpeabody.com/publications_detail3.asp?ID=2621

[12] ISACA. (2011) Information Systems Audit and Control Association. [Online]. HYPERLINK "http://www.isaca.org/Knowledge-Center/COBIT/Pages/Overview.aspx" http://www.isaca.org/Knowledge-Center/COBIT/Pages/Overview.aspx

[13] Carol Woodbury. (2004) SkyView Partner. [Online]. HYPERLINK "http://www.skyviewpartners.com/pdf/COBIT_Security.pdf" http://www.skyviewpartners.com/pdf/COBIT_Security.pdf

[14] Shoichi Morimoto, "Application of COBIT to Security Management in Information Systems Development," in *Proceedings of the Fourth International Conference on Frontier of Computer Science and Technology*, Shanghai , 2009, pp. 625-630.

[15] PRGL. (2011, December) Praxiom Research Group Limited. [Online]. HYPERLINK "http://www.praxiom.com/iso-17799-intro.htm" http://www.praxiom.com/iso-17799-intro.htm

[16] Solutionary. (2012, March) Solutionary. [Online]. HYPERLINK "http://www.solutionary.com/index/compliance/security-frameworks.php" http://www.solutionary.com/index/compliance/security-frameworks.php

[17] HITRUST. (2012, March) Health Information Trust Alliance. [Online]. HYPERLINK "http://www.hitrustalliance.net/csf/" http://www.hitrustalliance.net/csf/

# Cyber Semantic Account Management User Behavior Modeling, Visualization and Monitoring

**Keith Shapiro (**kshapiro@securboration.com) **Tony Stirtzinger** (tstirtzinger@securboration.com)
Securboration Inc., Melbourne, FL, USA

**Abstract** - *The research described in this paper presents an automated approach to generating and visualizing user behavior profiles based on activity observed in a networked environment. Derived (initially) from access log data, these profiles can serve as a model of expected behavior. Deviations from expected behavior can be identified by comparing benchmark profiles to profiles based on recent activity. This approach represents a practical solution to the problem of identifying malicious activity associated with compromised log-in credentials or insider attack. In those types of attacks, traditional defensive measures often fall short as the user was granted access to the system via normal channels using valid credentials.*

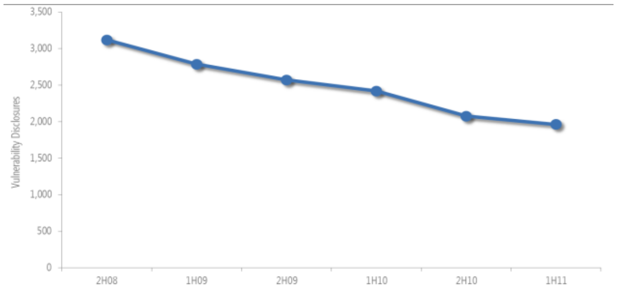
**Keywords:** Cyber Defense, Web Analysis, Behavior Modeling

## 1 Introduction

The field of information technology security has traditionally been focused on protecting against attack vectors based on software or hardware vulnerabilities. In that space, efforts are generally concentrated on firewall configuration, patch management, anti-virus technologies and intrusion detection.

Traditional protection technologies have matured over the years and are becoming an effective defense against 'brute force' types of attacks. Furthermore, as operating systems and software applications evolve, an emphasis on security has resulted in a decline in industry wide vulnerability disclosures. According to data published in the National Vulnerability Database (http://nvd.nist.gov) vulnerability disclosures across the industry in 1H2011 were down 37.1% from 2H2008[1].

Figure 1



Industry Wide Vulnerability Disclosures 2H2008 – 1H2011[1]

In response to an improving defensive landscape, hackers appear to be altering their attack strategies. Recent trends indicate that stealing or forging log-in credentials has become a growing methodology for achieving unauthorized access. This is commonly referred to as the 'masquerade' problem.

In terms of information security, a masquerade is a type of attack where the attacker pretends to be an authorized user of a system in order to gain access or to gain greater privileges than they are authorized for.

A masquerade is usually attempted through the use of stolen or forged logon IDs and passwords. Other methods do exist which include finding security gaps in programs, or through bypassing the authentication mechanism. The attempt may come from an employee within an organization or from an outside user through some connection to the network. Once the attacker has been authorized for entry, they may have full access to the organization's critical data, and (depending on the privilege level they pretend to have) may be able to modify and delete software and data, and make changes to network configuration and routing information.

The research described in this paper describes an automated approach to constructing user behavior profiles. Based on access log data (with planned expansion into other input areas), these profiles can serve as a model of expected behavior. Analytical components can identify deviations from expected behavior and alert analysts to potential indicators of compromised credentials or insider attack.

## 2 Motivation

Most current tools in the log analysis space are designed to roll up information, perform basic statistical analysis and identify trends related to basic usage. Security aspects tend to concentrate on identifying network based attacks. None of these approaches lends itself to identifying the use of compromised/forged credentials or insider attack.

As the masquerade problem grows, the cyber defense mission must be augmented with tools that can identify and alert on this type of malicious behavior.

In fact, the need for this level of analysis and monitoring has been called for at the highest levels within the DOD:

> In August 2009, Robert F. Lentz, Deputy Assistant Secretary of Defense for Cyber, Identity and Information Assurance (CIIA) published a vision, goals, and strategy document with the purpose of influencing the CIIA strategy of all DOD including the Combatant Commands. One of the strategic capabilities identified is Cyber Surveillance and Reconnaissance which is defined in that document as technologies and processes to persistently observe global information grid user, asset, and network behavior and examine their characteristics or configuration in order to detect anomalies, misuse, or unauthorized activity [2].

# 3 Related Work

The research described in this paper builds on initial research performed on the Semantic Policy Broker (SEMPBro) Small Business Innovation Report (SBIR) project. SEMPBro is based upon a semantic model of what system functions and information a user is allowed to access as part of his specific job duties. This research concluded that for the foreseeable future, access control will remain embedded within the implementation of the operational systems, resulting in a technology barrier for effective use of SEMPBro's capabilities.

During this research we identified an alternative application of the core technology within the cyber domain. Cyber threats were identified where valid login credentials were used by unauthorized users. This problem, also known as a 'masquerader cyber-attack,' results in exfiltration of operational data (e.g., mission plans, mission status). It was determined that we could reuse the technology from the original research to observe the patterns of user activity and compare those patterns to what is expected as per the semantic model.

Salem, Hershkop and Stolfo [3] describe the differences and challenges between masquerade detection and insider attack detection. While our initial research focused on masquerade detection, using simple features similar to [4], the ability to quickly and efficiently recalculate benchmarks positions our research to also apply to the insider attack.

Much of the prior cyber-attack research uses the standard masquerade detection dataset provided by Schonlau et al. [5], however this dataset was not applicable to measure our research as attack vectors and commands were not the evidence evaluated for our user profiling. Therefore, our results were measured against analyst generated criteria correlated to the provided dataset.

Finally, [6] used a taxonomy to classify commands in order to remove false positives and increase scalability in their approach to masquerade detection. Even though our approach for detection did not involve UNIX system commands and the sequences of such, this work inspired the use of semantic models (in lieu of taxonomic structure) to classify user actions and relate them to acceptable behavior in the context of operational activities. It is anticipated that extensions of these models will help with detection of false positives.

# 4 Approach

In general terms our approach can be classified as one-class training versus two-class as we use the data from a single user instead of multiple users to build a profile.

Compared to two-class training, one-class requires less source data and eliminates complex profile dependencies. Most important is that research has demonstrated one-class training as effective as two-class for masquerade detection [7].

Our approach to automatically generating user behavior profiles is based initially upon transforming file based access log data into relational database objects. Relational algebra and other analytical capabilities can be leveraged to construct models of user behavior.

Included in our approach is the concept of floating date ranges. This means that behavior models can be constructed for any date range supported by the source data.

With regards to the identification of anomalous behavior we support the concept of benchmark and comparison date ranges. The benchmark range is configurable and is used to construct the model of expected behavior. The comparison date range is also configurable. It can represent a time slice before, after or within the benchmark range.

Benchmark and comparison behavior models are overlaid to identify differences. Behavior in the comparison model that exists in the benchmark model is identified as expected behavior. Behavior in the comparison model that does not exist in the benchmark model is marked as suspect behavior.

We can also infer suspect behavior through statistical analysis. By using heuristic mechanisms to calculate an acceptable bandwidth of daily page activity we can then identify suspect activity as anything exceeding the acceptable bandwidth for the benchmark pattern.

We submit that this approach is both unique and practical. Unique in that compared to solutions based on machine learning algorithms, profiles are stored as dynamic relational objects with relational algebra playing a significant role in anomalous behavior detection and real-time monitoring. Practical in part because the approach does not require specialized computing environments. Relational database products are commonly deployed and managed in the enterprise.

## 4.1 Model Attributes

The behavior attributes have been initially baselined as the primary data fields tracked in access logs. Examples include time stamp, request and encryption information.

These attributes are commonly included in the set of data points used by analysts to manually construct behavior profiles.

### 4.1.1 Derived Patterns

Grouping by time stamp allows the construction of patterns based on daily activity, weekly activity and hourly activity.

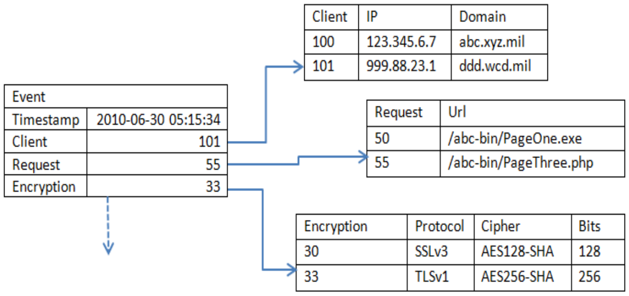
Grouping by Request path allows the construction of patterns based on unique page visits.

Additional patterns can be modeled by breaking down certificate and encryption information into subcomponents.

Each of these derived patterns is included in the user behavior profile.

## 4.2 Relational Design

Due to the potential for enormous amounts of input data and the effect this would have when imported into a database, a high degree of normalization was emphasized when designing relational objects.



| Client | IP | Domain |
|---|---|---|
| 100 | 123.345.6.7 | abc.xyz.mil |
| 101 | 999.88.23.1 | ddd.wcd.mil |

| Event | |
|---|---|
| Timestamp | 2010-06-30 05:15:34 |
| Client | 101 |
| Request | 55 |
| Encryption | 33 |

| Request | Url |
|---|---|
| 50 | /abc-bin/PageOne.exe |
| 55 | /abc-bin/PageThree.php |

| Encryption | Protocol | Cipher | Bits |
|---|---|---|---|
| 30 | SSLv3 | AES128-SHA | 128 |
| 33 | TLSv1 | AES256-SHA | 256 |

Figure 2 - Event Normalization

In addition to physical size considerations, performance was another important factor. Our goal was to design a web based interface that could respond to requests within a 5-30 second time frame.

To support performance goals, user behavior profiles were constructed on demand and placed in special cache tables to allow for extremely fast response times when viewing and analyzing the profiles.

## 4.3 Importing Data

An automated service was developed to import the data from access log files into the relational database. The service is designed to support multiple systems and multiple log file formats.

Log files from systems participating in the research are copied to a location where they could be consumed by the automatic log import service. Copying could be performed manually or by an automated process.

Daily imports are designed to keep the data fresh and allow for monitoring by configuring the most recent day as the comparison range.

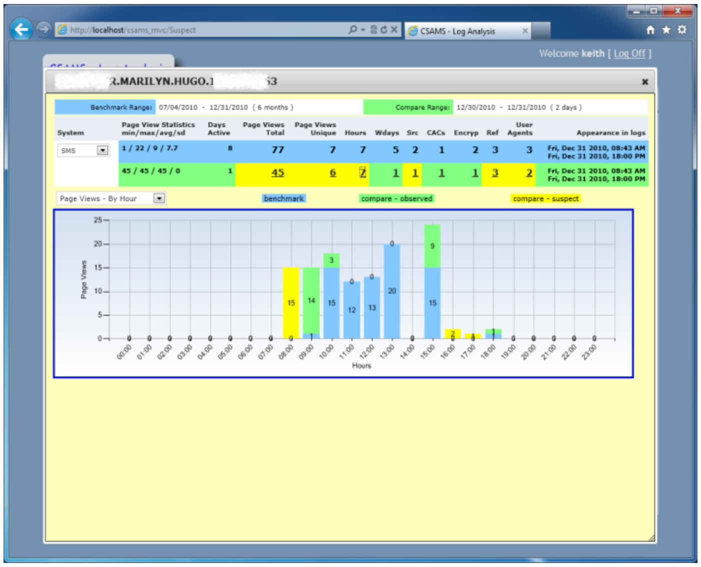
## 4.4 Multiple Systems

The source system is identified and recorded when the input log data is stored in the database. This allows user behavior profiles to link to individual systems. When displaying and analyzing profiles, you can choose to view by individual system or across systems.

## 4.5 Profile Presentation

Presentation options include a grid which allows the analyst to view lists of user behavior profiles corresponding to system and date range selections. When comparing profiles or working with a list of suspect profiles, color coding is used to indicate the users exhibiting suspect activity as well as the specific attributes linked to the suspect activity.

A tabular display of profile attributes and a set of charts visualize expected vs. suspect behavior for individual users. Color coding is used to differentiate between benchmark, expected and suspect behavior.



[redacted]R.MARILYN.HUGO.1[redacted]53

Benchmark Range: 07/04/2010 - 12/31/2010 ( 6 months )

Compare Range: 12/30/2010 - 12/31/2010 ( 2 days )

| System | Page View Statistics min/max/avg/sd | Days Active | Page Views Total | Page Views Unique | Hours | Wdays | Src | CACs | Encryp | Ref | User Agents | Appearance in logs |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SMS | 1 / 22 / 9 / 7.7 | 8 | 77 | 7 | 7 | 5 | 2 | 1 | 2 | 3 | 3 | Fri, Dec 31 2010, 08:43 AM<br>Fri, Dec 31 2010, 18:00 PM |
| | 45 / 45 / 45 / 0 | 1 | 45 | 6 | 7 | 1 | 1 | 1 | 1 | 3 | 2 | Fri, Dec 31 2010, 08:43 AM<br>Fri, Dec 31 2010, 18:00 PM |

Page Views - By Hour

Figure 3 –Sample User Behavior Profile Chart

# 5 Results

To test our approach we imported approximately seven months' worth of log files from a single source system and one month of data from a second source system.

## 5.1 Storage Requirements

The total size of the log file data on disk was approximately 15GB. Once imported, the physical size of the database was approximately 6GB – a 60% decrease in the required storage size. The table below documents properties of the source data.

| | System #1 | System #2 |
|---|---|---|
| Date Range | 7 months | 1 month |
| Days Active | 207 | 32 |
| Total Page Views | 7.2 million | 127,000 |
| Unique Users | 33400 | 475 |

## 5.2 Performance

Performance of the import service was a consideration along with the overall responsiveness of the web based application. The table below documents various performance metrics

Note that tests were performed using a Dell PowerEdge T310 server with 24GB memory and dual Xeon X3440 processors.

| Operation | Elapsed Time (approximate) |
|---|---|
| Import an 80KB log file | 25-35 seconds |
| Generate profiles for a 30 day range (11k users, 33k page views) | 8 seconds |
| Activate Profile View Charts for a single user | 2 seconds |
| Generate benchmark/comparison profiles for 6 month benchmark range (33k users and 6 million page views) and 2 day comparison range (2k users, 53k page views) | 15 seconds |
| Activate Profile Comparison | 4 seconds |
| Generate Suspect profiles for 6 month benchmark range (33k users and 6 million page views) and 2 day comparison range (850 users) | 12 seconds |
| Activate Suspect Profile | 4 seconds |

## 5.3 Identifying Suspect Activity

Using a six month benchmark range (30k users, 6 million page views) and a 2 day comparison range (2k users) the software identified approx. 850 users that exhibited at least one suspect behavior attribute. The table below groups this data by suspect categories (9 total categories exist).

| # Suspect Categories | # of users |
|---:|---:|
| 8 | 2 |
| 7 | 5 |
| 6 | 15 |
| 5 | 42 |
| 4 | 72 |
| 3 | 114 |
| 2 | 194 |
| 1 | 402 |

While there were 850 users identified exhibiting some out of band behavior, the number of users exhibiting behavior that deviated between 5 to 8 dimensions simultaneously was only 64, or .002%. This testing has led to the hypothesis that the more interesting monitoring and analysis should be focused on a combination of: 1) users exhibiting a high number of violated attributes simultaneously; and 2) the combinations of which attributes are consistently in the set of simultaneous violations.

# 6 Conclusions and Future Work

The overall objectives of our research were accomplished; primarily, a.) Develop a mechanism to automatically generate user behavior models based on access log information and b.) Leverage the models to identify suspect behavior.

Furthermore, based on the initial results of this research we believe our storage requirement and performance goals were achieved.

The physical size of the database can become an issue. Therefore, as the data set grows, archive strategies or the creation of multiple instances would be required to ensure performance remains acceptable.

The ability to identify suspect behavior based on expected behavior models was confirmed. For effective monitoring and alerting capabilities we believe the number of suspect users should be reduced.

There are many ways to improve upon this technology to provide an even more effective tool for cyber defense analysts. Significant enhancements and an expansion into enterprise capabilities are planned for subsequent phases.

Enhancements include the use of semantic models and technologies to provide more robust cyber account monitoring capabilities. Semantic technologies can be used to infer anomalous behavior which might not otherwise be noted by statistical analysis.

Refinements in suspect attribute priorities and exception management will reduce the 'noise' of false positives and make for much more effective alerting capabilities.

Future work is also planned to allow a broader input scope. The behavior profiles can be greatly enhanced by incorporating process models including linkages to the system interactions that result from executing those models. Clustering concepts can also provide value in the form of role derivation from observed activity. Inputs from Enterprise Service Bus activity can also add to the strength of the behavior profiles.

## 7 Acknowledgements

This material is based in part upon work supported by the Air Force Research Laboratory contract number FA8750-08-C-0062. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of AFRL or the U.S. Government.

## 8 References

[1] Microsoft. "Microsoft Security Intelligence Report, Volume 11, Worldwide Threat Assessment"; 2011 http://www.microsoft.com/security/sir/default.aspx

[2] http://iase.disa.mil/policy-guidance/dasd_ciia__strategy_aug2009.pdf

[3] Malek Ben Salem, Shlomo Hershkop, and Salvatore J. Stolfo, "A survey of insider attack detection research" in *Insider Attack and Cyber Security: Beyond the Hacker*, Springer, 2008. "Taxonomy-Based Multinomial Modeling Approach in Unix Systems". Columbia University Computer Science Department, Technical Report # cucs-021-08, 2008.

[4] H. Kim, S. Cho, J. Seo, Y. Lee, and S. Cha. Use of support vector machine (svm) in detecting anomalous web usage patterns. In *Symposium on Information and Communications Technology, 2004.*

[5] Schonlau, M., and Theus, M. 2000. Detecting Masquerades in Intrusion Detection Based on Unpopular Commands. Information Processing Letters 76, 33-38.

[6] Malek Ben Salem, Salvatore J. Stolfo. "Masquerade Detection Using a Taxonomy-Based Multinomial Modeling Approach in Unix Systems". Columbia University Computer Science Department, Technical Report # cucs-021-08, 2008

[7] K. Wang and S. J. Stolfo. One-Class Training for Masquerade Detection, In Proceedings of 3rd IEEE Workshop on Data Mining for Computer Security, 2003. http://www.cs.columbia.edu/~kewang/paper/DMSEC-camera.pdf

# SESSION

# SECURITY APPLICATIONS II + ABUSE PREVENTION METHODS + SOFTWARE AND TOOLS

## Chair(s)

**Dr. Nizar Al Holou**

# Towards Security Policy and Architecture for Managing Implantable Medical Devices

Ram Krishnan and Eugene John
Dept. of Electrical and Computer Engineering
University of Texas at San Antonio
Email: Ram.Krishnan@utsa.edu and Eugene.John@utsa.edu

Manoj Panday
School of Medicine
University of Texas Health Science Center at San Antonio
Email:panday@uthscsa.edu

**SAM Track: Security Management**

***Abstract*—Implantable cardiac rhythm management devices (CRMDs) such as permanent pacemakers and internal cardioverter defibrillators (ICDs) utilize embedded computers and radios to monitor chronic disorders and treat patients. Life-saving devices like ICDs, for instance, include pacemaker technology and are designed to communicate wirelessly with a nearby external device programmer (EDP) that can remotely read data and change settings without the need for surgery. An ICD implanted in a patient can sense a rapid heartbeat and administer an electric shock to restore normal heart rhythm. It is has been shown that current ICDs in the market can be reverse engineered and are prone to software radio-based attacks. The ICDs can be remotely disabled or be made to administer an electric shock at random. Existing defense mechanisms include a simple cryptographic approach where a symmetric-key based challenge-response protocol is used between the ICD and an authorized EDP. This approach does not scale. In the real world, large scale deployment and management of shared key material amongst various entities such as CRMDs, EDPs, hospitals, clinics, and ambulances is a major issue. In this paper, we investigate security policy issues applicable to the CRMD ecosystem and issues for architectures that enforce the policy. Given the nature of this domain, these solutions will need to balance security, privacy and risk. For instance, an unauthorized EDP may need to issue a command to the ICD in emergency situations.**

***Index Terms*—Security Policy, Architecture, Key Management, Implantable Medical Devices**

## I. Introduction

This paper is motivated by the need to provide security and privacy for various cardiac rhythm management devices (CRMDs) that are being deployed in the order of millions in the market today. Specifically, the CRMDs that are of interest are those that can perform some computation to provide patient data to medical personnel and/or administer some type of treatment. A well-known example of such a device is the implantable cardioverter defibrillator (ICD). ICDs utilize embedded computers and radios to monitor chronic heart disorders and treat patients. They include pacemaker technology and are capable of wireless communication with a nearby external device programmer (EDP). EDPs can wirelessly read patient data and change settings without the need for surgery. An ICD implanted in a patient can sense a rapid heartbeat and administer an electric shock to restore normal heart rhythm.

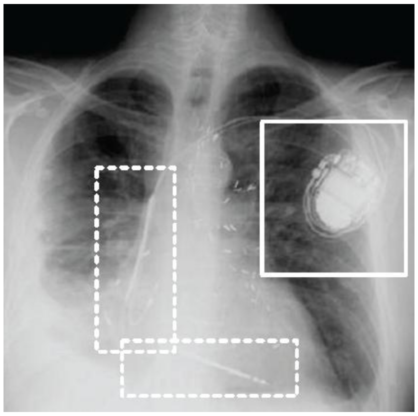


Fig. 1. Chest x-ray of a patient with a dual chamber ICD implanted via the left subclavian vein (source: [1]).

Figure 1 shows a chest x-ray of a patient with a dual chamber ICD implanted via the left subclavian vein. An EDP can be used to configure such an ICD in a patient. The configuration could dictate under what conditions the ICD should activate and regulate heart rhythm by administering corrective electric shocks. An ICD can communicate with the EDP when a magnetic field is generated in its vicinity which closes a switch in the ICD. Subsequently, the EDP can be used to perform diagnostics, read and write private patient data and modify therapy settings.

Figure 2 shows the timeline of a sample communication between an EDP and ICD. As shown, after the application of a magnetic field, the EDP can query the ICD for telemetry and patient data and issue commands to change its configurations. The protocol is not cryptographically protected. In [1], the authors show how such a protocol can be reverse engineered using standard lab equipments. The ICD can be made to respond to unauthorized EDPs, and its configurations can be changed with relative ease. Thus, an attacker could employ an unauthorized EDP to administer an electrical shock.

**Motivation** The current solution involves sharing a secret key (symmetric key) between an ICD and authorized EDPs. This allows one to build a cryptographic strength protocol where the ICD can throw a challenge to any EDP that wants to interact with it. Only if the EDP provides a correct response that is based on the shared secret key, the ICD would accept further commands.

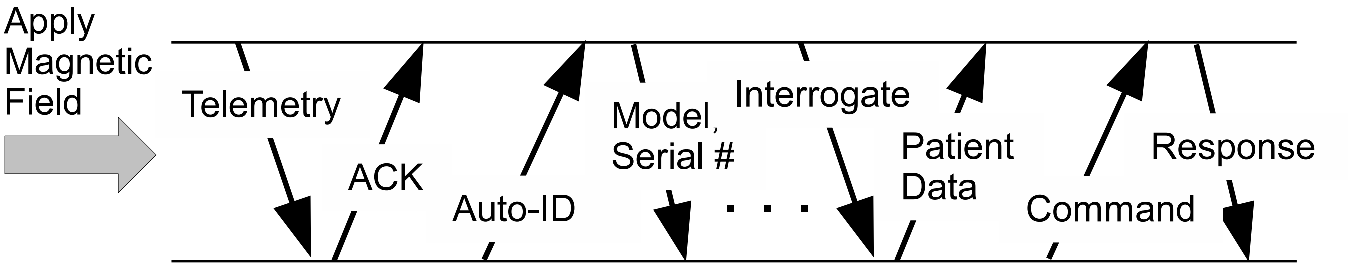


Fig. 2. An example interaction between an ICD and EDP. The interaction is not cryptographically protected. This allows a malicious party to easily issue unauthorized commands to the ICD.

In the real world, this approach does not scale. Large-scale deployment of shared key material amongst various entities such as CRMDs, EDPs, hospitals, clinics, and ambulances poses an unacceptable amount of risk for key compromise At the same time, it would be naive to expect that keys be shared only between the patient's ICD and the corresponding doctor's EDP. Due to nature of this domain, inaccessibility to an ICD in times of emergency could be fatal. Consider a situation where a physician who is new to the patient responds to an emergency situation but is unable to access the ICD since his/her EDP does not have the shared secret key.

**Key Challenges** The key challenge is to develop intuitive, yet effective and scalable models for managing security keys amongst a large number of disparate entities that may belong to different administrative domains. For instance, the patient may consult doctors from different hospitals, various nurses may work with the doctors and patients may be treated by emergency response personnel who may not have prior agreement to access the patient's ICD.

This paper will investigate the requirements of a scalable framework for secure, reliable and risk-aware cryptographic key management for CRMDs and its ecosystem. The techniques developed will need to balance security, privacy and risk due to the sensitive nature of this domain. Security is concerned about integrity of commands exchanged between external control devices and CRMDs, in addition to the confidentiality of data exchanged. Privacy is concerned about appropriate access of collected sensitive data by authorized parties. Risk is concerned about balancing security and privacy in scenarios where the mission is more important (for example, in emergency situations, an unauthorized EDP may need to issue a command to the ICD).

**Relevance to Real-World Threat** Although no known attack has been reported, implantable medical devices are only increasing in sophistication. For instance, CRMDs can now communicate over a much longer range than those developed in the late 90s. Thus, this tremendously increases the threat surface and attack practicality. Furthermore, the knowledge of such vulnerabilities is psychologically troubling to patients that undergo implantation. If not maliciously, there is always room for accidental unauthorized modification of ICD settings and/or the ability to read private patient data. Thus this paper addresses an important problem that has broad impact.

**Prior Work** To the best of our knowledge, prior work in this area does not address this problem either directly or addresses them inadequately. Existing solution to this problem is to simply share a secret key between every ICD and EDP [1]. While conceptually acceptable, this does not scale to real-world scenarios as argued earlier. The challenges involved in manufacturing and providing safe computer-based medical treatments in the presence of unintentional failures have been investigated. However, the proposed work addresses such challenges in the context of intentional failures due to passive and active attacks by a malicious entity. Securing patient data in databases has been studied in the past [7]. Pervasive healthcare security including medical sensor security has been investigated in [10]. There is also ample literature on wireless security in low-power environments. See for example: [2], [8], [11].

## II. Approach for CRMD Management

This section briefly provides an overview of the technical approach of the proposed solution. From a methodology standpoint, security policy issues will be clearly separated from security enforcement issues. The policy model is concerned about "what" needs to achieved while the enforcement model is concerned about "how" the policy can be realized. This allows one to address issues at an appropriate level of abstraction. The goal of this paper is to develop effective key management techniques for CRMDs. (From here on, we will use the more general term CRMD, for implantable *medical devices*, instead for ICD.) Thus various policy models need to be specified so that appropriate key management techniques can be developed at enforcement level.

The policy issues concerning this problem include developing models to specify information sharing policies amongst various entities that are involved in a CRMD ecosystem. For example, it may be the case that a specific CRMD implanted in a patient can communicate with an EDP only in the presence of his/her doctor unless it is an emergency. In another scenario, the policy could be that the CRMD can only communicate with a pre-authorized set of EDPs regardless of who employs them. Yet another policy could specify authorization with respect to the personnel that employ them instead of the identity of the EDPs that are employed. Note that such varying policies require different key management techniques.

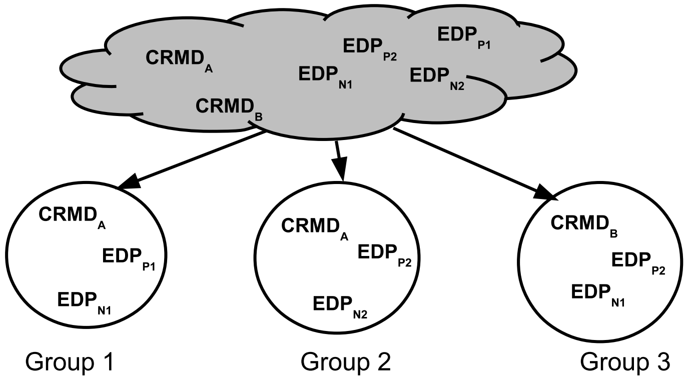


Fig. 3. Various groups can be formed between CRMDs and ICDs from the global set. The groups indicate which EDPs can communicate with which ICDs. Group 1 is formed between a CRMD for patient A and her physician P1 and nurse N1. The same CRMD can belong to another group 2.

The policy models will build upon our preliminary work in abstract policy models for Group-Centric Secure Information Sharing (g-SIS) [3], [9], [4], [6], [5]. In g-SIS, a security policy can be specified for entities that form a group. In our scenario, a group can be formed between a CRMD and the EDPs that may communicate with that CRMD. Furthermore, a CRMD can belong to multiple groups. For example, a CRMD can belong to a group of physician EDPs while also belonging to another group of nurse EDPs. Furthermore, the CRMD can also belong to yet another group of EDPs that will be used by emergency response personnel. A major challenge is in managing such a large number of CRMDs and EDPs and their group memberships. In the prior work in g-SIS, formal models for a single group has been developed. We need to develop models to manage multiple groups that are specific to this domain and those that can specify inter-relationship between groups. For example, we may specify that a CRMD can be a member of group A as long as it is also a member of group B. Thus if membership in group B ceases, the CRMD will lose membership in group A as well.

Figure 3 shows an example scenario. Membership of EDPs in a group indicates that they can communicate with the ICDs in that group. Various permissions can be specified in the group. For example, the physician and nurse in group 1 can read patient data and update ICD's settings while those in group 2 may only read data.

## III. Key Management

Following concrete and intuitive models under which various information sharing policies can be specified using formal logic,[1] various enforcement models that meet the policy specification can be developed. Since the policy models specify information sharing at the group level, group key management techniques can be employed. Specifically, various techniques will be employed for scalable group-centric key management including centralized, decentralized and distributed approaches.

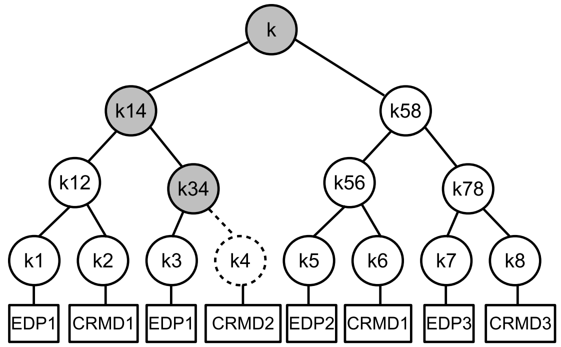


Fig. 4. A Logical Key Hierarchy [12] for group key management. The shaded circles indicate the keys that need to be changed when a new member (indicated by boxes) CRMD2 joins a group in which EDP1 belongs.

In all of these approaches, the critical problem is in handling group creation and membership changes. For example, when an EDP is taken out of a group, the group key for all the remaining members should be updated. In centralized approach, we assume that a group manager exists for each group that will manage the keys shared amongst various entities. In decentralized approach, more than one group manager may exist. In the distributed approach, there is no group manager. Every member can both be a manager and a regular member.

We discuss how logical key hierarchy (figure 4) can be employed to manage keys in this context. In the figure, the boxed nodes indicate entities that need to securely exchange messages. The rest of the figure forms a key graph. Each node contains a symmetric key. Each boxed node stores every key that is in a path from itself to the root. For example, CRMD1 stores keys k2, k12, k14 and k. The key in the leaf node indicates a unique key for each boxed node. Given this setting, data can be securely exchanged between CRMD1 and EDP1 using the key k12. This is because both the entities have knowledge of k12. Similarly, data can be securely exchanged between EDP1 and CRMD2 using k34. Note that data protected using k14 is readable by all the members that fall under that subtree—specifically EDP1, CRMD1 and CRMD2. Finally, data encrypted using k can be decrypted by every member in the system: EDP1, EDP2, CRMD1 and CRMD2.

Such a key graph is typically stored in a server that manages the keys for all participating entities in the CRMD ecosystem. Note that the key graph can be fully customized to accommodate the needs of a specific scenario with respect to group level sharing. For example, by adjusting the degree of each node and the height of the tree, several groups and subgroups can be formed for secure communication.

In practice, under normal circumstances each entity may use its group key to communicate with other entities in the same group. In case of an emergency, if an EDP does not belong to a specific group but still needs to communicate with a CRMD in that group, it can use a key at a higher level. For

[1] For example, First-Order Linear Time Temporal Logic was used in the preliminary work on g-SIS [3].

example, under emergency situations, CRMD3 can be used to communicate with EDP2 using k58.

Membership management is an important issue. For example, a new member, say EDP1, may need to be added to the group that is managed using key k78. Also, a member may leave a group. Different policies may be required in such scenarios. Forward secrecy ensures that after a member leaves a group, it cannot read any new data exchanged the remaining members. Backward secrecy ensures that when a new member joins a group, it is only able to read new data exchanged between the group members and not any data exchanged before the member joined. Thus to ensure forward secrecy, whenever a member leaves, the remaining group members need to change the group key. Similarly, to ensure backward secrecy, whenever a member joins, the group key needs to updated to ensure that the new member is unable to access past data.

Let us consider member management in logical key hierarchy. In figure 4, when CRMD2 joins the group in which EDP1 belongs, a new node with a new unique key (k4) for CRMD2 is created in the key graph. (We assume that the server has prior knowledge of this unique key.) To guarantee backward secrecy, every key in the path from k4 to the root needs to be updated. Thus keys k, k14 and k34 need to be updated. Note that updated versions of these keys (say k', k14' and k34') in the key graph need to share with other devices in the graph. As seen in the figure, this means that EDP1 needs to get k34', k14' and k' and CRMD1 needs to get k14' and k'. This can be achieved by encrypting the new keys with appropriate old keys. For instance, the new keys can be encrypted using k3 to be sent securely to EDP1 and can be encrypted with the old k12 to be sent securely to all members in that subgroup. Similarly, when EDP1 leaves the group with k34, the keys k, k14 and k34 need to be updated and shared with members that are affected by this membership change.

Thus membership management can be efficiently handled in logical key hierarchy. Specifically, when one member leaves, it does not require key updates to every other member.

## IV. Conclusion and Future Work

We investigated group-based security policies and corresponding key management strategies using logical key hierarchy for securely managing large-scale implantable medical devices such as ICDs. We strongly believe that this paper serves as starting point for research in this important area in a number of different avenues. First, a major challenge in implementing this approach for the CRMD ecosystem is ensuring key updates are carried out in a timely manner. This has a number of practical challenges. For instance, there are always periods of time during which a few entities will have outdated keys since they may not be always connected to the server.

Next, a number of other key management strategies can be explored. Logical key hierarchy is a centralized approach where the keys are managed by a single server. We plan to explore decentralized and distributed approaches to key management for the CRMD ecosystem. In the decentralized approach, there are more than one server to manage key updates. In a distributed approach, there is no specific server and client. Every entity is both a server and a client. We plan to investigate a hybrid approach for key management that is both practical and effective in terms of minimizing the time period during which members have outdated keys.

We also plan to explore different policy models and corresponding key management techniques for enforcement. For example, a conditional membership policy may require that a member's membership in a group is contingent upon its membership in another group. Similarly, membership could be hierarchical in which membership in a group automatically guarantees membership in other groups that are dominated by the joining group.

## References

[1] D. Halperin, T.S. Heydt-Benjamin, B. Ransford, S.S. Clark, B. Defend, W. Morgan, K. Fu, T. Kohno, and W.H. Maisel. Pacemakers and implantable cardiac defibrillators: Software radio attacks and zero-power defenses. In *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pages 129 –142, may 2008.

[2] Chris Karlof, Naveen Sastry, and David Wagner. Tinysec: A link layer security architecture for wireless sensor networks. In *ACM SENSYS'04*. ACM, 2004.

[3] Ram Krishnan, Jianwei Niu, Ravi Sandhu, and William H. Winsborough. Group-centric secure information-sharing models for isolated groups. *ACM Trans. Inf. Syst. Secur.*, 14:23:1–23:29, November 2011.

[4] Ram Krishnan, Ravi Sandhu, Jianwei Niu, and William Winsborough. Towards a framework for group-centric secure collaboration. *Proceedings of IEEE International Conference on Collaborative Computing*, 2009.

[5] Ram Krishnan, Ravi Sandhu, Jianwei Niu, and William Winsborough. A conceptual framework for group-centric secure information sharing. *ACM Symposium on Information, Computer and Comm. Security*, March 2009.

[6] Ram Krishnan, Ravi Sandhu, Jianwei Niu, and William H. Winsborough. Foundations for group-centric secure information sharing models. In *Proc. of ACM symposium on access control models and technologies*, 2009.

[7] M. Meingast, T. Roosta, and S. Sastry. Security and privacy issues with health care information technology. In *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, pages 5453 –5458, 30 2006-sept. 3 2006.

[8] Adrian Perrig, Robert Szewczyk, J. D. Tygar, Victor Wen, and David E. Culler. Spins: security protocols for sensor networks. *Wirel. Netw.*, 8:521–534, September 2002.

[9] Ravi Sandhu, Ram Krishnan, Jianwei Niu, and William Winsborough. Group-centric models for secure and agile information sharing. In *Computer Network Security: 5th International Conference, on Mathematical Methods, Models, and Architectures for Computer Network Security (MMM-ACNS 2010)*, Lecture Notes in Computer Science. Springer, 2010.

[10] K.K. Venkatasubramanian and S.K.S. Gupta. Security for pervasive health monitoring sensor applications. In *Intelligent Sensing and Information Processing, 2006. ICISIP 2006. Fourth International Conference on*, pages 197 –202, 15 2006-dec. 18 2006.

[11] S. Warren, J. Lebak, Jianchu Yao, J. Creekmore, A. Milenkovic, and E. Jovanov. Interoperability and security in wireless body area network infrastructures. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 3837 –3840, 2005.

[12] Chung Kei Wong, Mohamed Gouda, and Simon S. Lam. Secure group communications using key graphs. *IEEE/ACM Trans. Netw.*, 8:16–30, February 2000.

# A Secure Communication System Based on Self-organizing Patterns

Paulius Palevicius*, Loreta Saunoriene†, Minvydas Ragulskis‡
Research Group for Mathematical and Numerical Analysis of Dynamical Systems,
Kaunas University of Technology, Studentu 50–222, Kaunas LT–51368, Lithuania,
Email: *Paulius.Palevicius@ktu.lt, †Loreta.Saunoriene@ktu.lt, ‡Minvydas.Ragulskis@ktu.lt

***Abstract*—This paper proposes a secure steganographic communication algorithm based on the evolution of self-organizing patterns. The presented algorithm is a modification of a secure steganographic scheme, presented in our previous work [1]. Algorithm is based on the formation of self-organizing patterns in a Beddington-deAngelis-type predator-prey model with self-diffusion. Computational experiments show that the generation of interpretable target patterns cannot be considered as a safe encoding of secret visual information because the target pattern becomes interpretable only when the cover image (initial distribution of preys) leaks the secret to a naked eye. Therefore, we propose an alternative approach when the cover image represents the self-organizing pattern which has evolved from initial states perturbed using the dot-skeleton representation of the secret image. Such visual communication technique protects both the secret image and communicating parties.**

***Index Terms*—Visual steganography, self-organizing pattern, nonlinear evolution.**

## I. Introduction

The field of research on pattern formation modelled by reaction-diffusion systems, which provides a general theoretical framework for describing pattern formation in systems from variant disciplines including biology, chemistry, physics, etc., seems to be an increasingly interesting area. One of the classical numerical examples illustrating a variety of irregular spatiotemporal patterns comprises a simple reaction-diffusion model with finite amplitude perturbations [2]. The phenomenology of a wide variety of two- and three-dimensional physical-chemical systems displaying prevalent stripe and bubble morphologies of domain patterns in equilibrium is discussed in [3]. Patterns specifying dynamic behavior of chemoresponsive gels undergoing the Belousov-Zhabotinsky reaction are constructed in [4]. Pattern formation mechanisms of a reaction-diffusion-advection system, with one diffusivity, differential advection, and Robin boundary conditions of Danckwerts type, are investigated in [5]. Time-periodic forcing of spatially extended patterns near a Turing-Hopf bifurcation point is studied in [6]. One of the promising applications of the phenomenon of pattern formation could be digital image processing when the evolving pattern would be used to encode the initial image. A digital fingerprint image is used as the initial condition for the evolution of a pattern in a model of reaction-diffusion cellular automata [7], though the possibility to encrypt the initial fingerprint in the evolved pattern is not discussed in [7]. The dynamic behavior of predator-prey model has long been and will also continue to be one of the dominant themes in both ecology and mathematical ecology due to its universal existence and importance. Complex dynamics and spatiotemporal pattern formation in variant predator-prey models are analyzed in [8], [9], [10], [11].

This paper proposes a secure steganographic communication algorithm based on the evolution of self-organizing patterns. The algorithm is a modification of a secure steganographic scheme, presented in our previous work [1].

In general, cryptography is a method of storing and transmitting data in a form that only those it is intended for can read and process [12]. Modern cryptography follows a strongly scientific approach and designs cryptographic algorithms around computational hardness assumptions that are assumed hard to break by an adversary. But cryptography does not always provide safe communication. Steganography is a science of concealing data in a communication in such a way that only the sender and receiver know of its existence [13]. The advantage of steganography, over cryptography alone, is that messages do not attract attention to themselves. Therefore, whereas cryptography protects the contents of a message, steganography can be said to protect both messages and communicating parties [14].

As mentioned previously, we will demonstrate that self-organizing patterns can be effectively exploited as a secure tool for steganographic communication.

## II. The model of the system

We exploit a well known predator-prey model with Beddington-DeAngelis-type functional response with self-diffusion [8]. The system of differential equations describing the dynamics of this model can be written:

$$\begin{aligned}\frac{\partial N}{\partial t} &= r\left(1-\frac{N}{K}\right)N-\frac{\beta N}{B+N+wP}P+d_1\nabla^2 N,\\ \frac{\partial P}{\partial t} &= \frac{\varepsilon\beta N}{B+N+wP}P-\eta P+d_2\nabla^2 P,\end{aligned}\qquad(1)$$

where $t$ denotes time; $N$ and $P$ are densities of preys and predators respectively; $\beta$ is a maximum consumption rate, $B$ is a saturation constant; $w$ is a predator interference parameter; $\eta$ represents a per capita predator death rate; $\varepsilon$ is the conversion efficiency of food into offspring. It can be noted, that the operator

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{2}$$

is the Laplacian operator in the two-dimensional space. Self-diffusion terms $d_1\nabla^2 N$ and $d_2\nabla^2 P$ imply the movements of individuals from a higher to lower concentration region. Self-diffusion coefficients are denoted by $d_1$ and $d_2$, respectively [8].

Non-zero initial conditions

$$N(x, y, 0) > 0; P(x, y, 0) > 0 \tag{3}$$

are set in a rectangular domain $(x, y) \in \Omega = [0, L_x] \times [0, L_y]$, where $L_x$ and $L_y$ is the size of the system in the directions of $x-$ and $y-$ axis. Neumann, or zero-flux, conditions are set on the boundary:

$$\frac{\partial N}{\partial n} = \frac{\partial P}{\partial n} = 0; (x, y) \in \partial\Omega, \tag{4}$$

where $n$ is the outward unit normal vector of the smooth boundary $\partial\Omega$. Zero-flux boundary conditions imply that no external input is imposed from outside.

The first step in analyzing the model is to determine the equilibria (stationary states) of the non-spatial model obtained by setting space derivatives equal to zero, i.e.,

$$\begin{gathered} r\left(1 - \frac{N}{K}\right)N - \frac{\beta N}{B + N + wP}P = 0, \\ \frac{\varepsilon\beta N}{B + N + wP}P - \eta P = 0. \end{gathered} \tag{5}$$

In fact, physically, an equilibrium represents a situation without "life". It may mean no motion of a pendulum, no reaction in a reactor, no nerve activity, no flutter of an airfoil, no laser operation, or no circadian rhythms of biological clocks. And at each equilibrium point, the movement of the population dynamics vanishes.

In the absence of diffusion, the model has three equilibria in the positive quadrant [8]:

1) $(0, 0)$ (total extinct) is a saddle point.
2) $(K, 0)$ (extinct of predators or preys-only) is a stable node if $\varepsilon\beta < \eta$ or $\varepsilon\beta > \eta$ and $K < -\frac{\eta B}{-\varepsilon\beta+\eta}$; a saddle if $\varepsilon\beta < \eta$ and $K > -\frac{\eta B}{-\varepsilon\beta+\eta}$; a saddle-node if $\varepsilon\beta < \eta$ and $K = -\frac{\eta B}{-\varepsilon\beta+\eta}$.
3) a non-trivial stationary state $(N^*, P^*)$ (coexistence of preys and predators), where

$$\begin{aligned} N^* &= \frac{1}{2rw\varepsilon}K(rw\varepsilon - \varepsilon\beta + \eta) \\ &+ \frac{1}{2rw\varepsilon}\sqrt{K^2(rw\varepsilon - \varepsilon\beta + \eta)^2 + 4rKw\varepsilon\eta B}, \\ P^* &= \frac{(\beta\varepsilon - \eta)}{w\eta}N^* - \frac{B}{w}. \end{aligned} \tag{6}$$

The numerical model of predator-prey system is based on standard five-point approximation for 2D Laplacian with the zero-flux boundary conditions. The concentrations $\left(N_{ij}^{n+1}, P_{ij}^{n+1}\right)$ at the moment $(n+1)\tau$ at mesh position $(x_i, y_j)$ are calculated as [8]:

$$\begin{aligned} N_{ij}^{n+1} &= N_{ij}^n + \tau d_1 \Delta_h N_{ij}^n + \tau f\left(N_{ij}^n, P_{ij}^n\right), \\ P_{ij}^{n+1} &= P_{ij}^n + \tau d_2 \Delta_h P_{ij}^n + \tau g\left(N_{ij}^n, P_{ij}^n\right), \end{aligned} \tag{7}$$

where the Laplacian is

$$\Delta_h N_{ij}^n = \frac{N_{i+1,j}^n + N_{i-1,j}^n + N_{i,j+1}^n + N_{i,j-1}^n - 4N_{i,j}^n}{h^2}. \tag{8}$$

Initially, the entire system is placed in the stationary state $(N^*, P^*)$ with a random perturbation. The system evolves either into steady or time-dependent state after a certain number of iterations. Different sets of the model parameters correspond to the special types of final patterns: stripe-like patterns, regular spotted pattern, the mixture of spotted and stripe-like patterns or the spiral wave patterns [8].

## III. A SECURE COMMUNICATION SYSTEM BASED ON SELF-ORGANIZING PATTERNS

We use Beddington-DeAngelis-type predator-prey model with self-diffusion with the following parameter set: $d_1 = 0.01$, $d_2 = 1$, $r = 0.5$, $\varepsilon = 1$, $\beta = 0.6$, $K = 2.6$, $\eta = 0.25$, $\omega = 0.4$, $B = 0.3154$. All our numerical simulations employ the Neumann (zero-flux) boundary conditions with a system size of $200 \times 200$ space units $(L_x = L_y = 50)$. The system in Eq. (1) is solved numerically in two-dimensional space using a finite difference approximation for the spatial derivatives and an explicit Euler method for the time integration (Eq. (7)) with a time step $\tau = 0.01$ and space step $h = 0.25$. The scale of the space and time are average to the Euler method.

The dynamics of the time evolution of preys $N$ is demonstrated in Fig. 1. Fig. 1(a) presents the equilibrium point $(N^* = 0.43058; P^* = 0.718555)$ with small random perturbations.

We use the logistic map

$$x_{i+1} = \mu x_i (1 - x_i) \tag{9}$$

with $\mu = 4$ for the computation of a set of $200 \times 200$ pseudo-random numbers distributed in the interval $[0; 1]$. The dynamics of the logistic map depends on the value of parameter $\mu$. When $\mu = 4$, system in Eq. (9) demonstrates chaotical behavior and therefore is appropriate for the generation of random numbers.

The obtained random set distributed in the interval $[0; 1]$ is linearly transformed into an $\varepsilon$-length interval with zero mean and is added to the initial concentration of preys:

$$[N]|_{t=0} = N^* \cdot [1] + \left[\widetilde{N}\right]; [P]|_{t=0} = P^* \cdot [1], \tag{10}$$

where $[1]$ is a $200 \times 200$ matrix of ones; $\left[\widetilde{N}\right]$ is a $200 \times 200$ matrix of pseudo-random numbers distributed uniformly in the interval $[-\varepsilon/2; \varepsilon/2]$. It is clear that the parameter $\varepsilon$ must be significantly lower than the maximum concentrations in the final $N$ and $P$ patterns; we use $\varepsilon = 10^{-3}$ in computational experiments illustrated in Fig. 1.

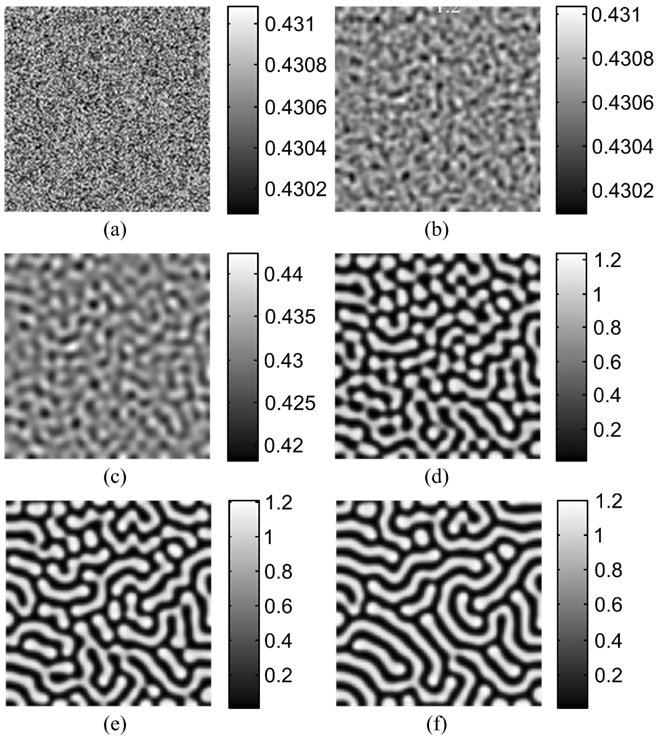


Fig. 1. Dynamics of the time evolution of preys: (a) – the initial distribution ($\varepsilon = 10^{-3}$); (b) – after 2500 iterations; (c) – after 10000 iterations; (d) – after 25000 iterations; (e) – after 50000 iterations; (f) – after 200000 iterations.

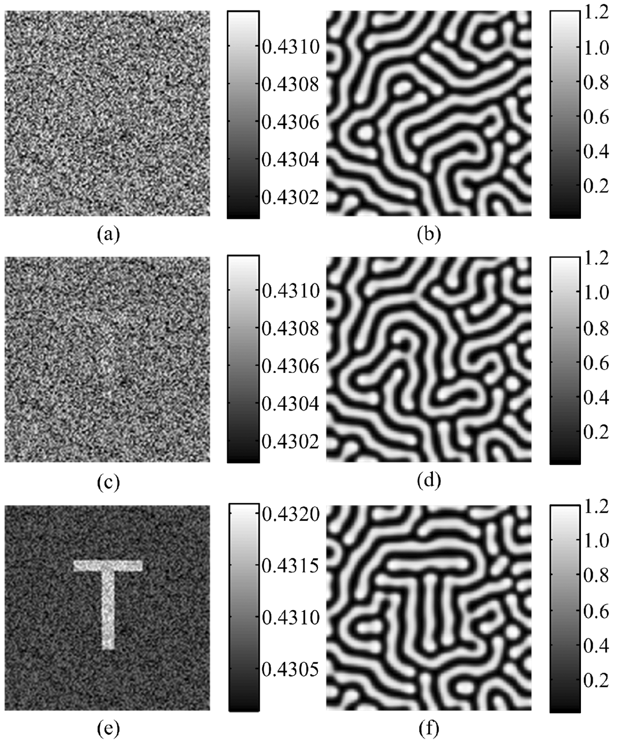


Fig. 2. Time evolution of preys: (a) – the initial density of preys ($\varepsilon = 10^{-3}$; $\delta = 0$); (b) – the pattern of preys after 200000 iterations. (c) and (e) represent initial densities of preys distorted by the T-shaped mask at $\delta/\varepsilon = 0.1$ and $\delta/\varepsilon = 1$ respectively (the same matrix $\left[\widetilde{N}\right]$ is used in all experiments). (d) and (f) illustrate patterns of preys after 200000 iterations.

Fig. 1(b), Fig. 1(c), Fig. 1(d) and Fig. 1(e) show the evolution of the spatial pattern of preys after 2500, 10000, 25000 and 50000 iterations. Time-independent self-organizing pattern of stripes and spots is obtained after 200000 iterations (Fig. 1(f)). It is important to note that the pattern shown in Fig. 1(f) is sensitive to initial conditions. Fig. 2(a) shows the initial distribution of preys and Fig. 2(b) represents the pattern after 200000 iterations (all parameters of the system are kept the same). Different initial perturbations in Eq. (10) (a different set of pseudo-random numbers) evolve into a pattern of the same type as shown in Fig. 1(f) but with a different writing.

### *A. The generation of target patterns*

The evolution of self-organizing patterns is sensitive to initial perturbations. This fact allows construction and manipulation of target patterns by small modifications in the initial distribution of preys. Fig. 1(f) and Fig. 2(b) illustrates that two different realizations of initial concentrations of preys result into apparently similar but locally different patterns of stripes.

Let us assume that the matrix of random perturbations $\left[\widetilde{N}\right]$ is modified by adding a positive constant $\delta$ to numerical values of some pixels in the initial distribution of preys. In general, the initial density of preys then can be described by the following equation:

$$[N]|_{t=0} = N^{*} \cdot [1] + \left[\widetilde{N}\right] + \delta \cdot [M], \qquad (11)$$

where $\delta$ is a fixed constant; $[M]$ is a binary mask matrix holding ones at those pixels where the initial random density of the preys is increased by $\delta$ and zeroes where the random density of preys is kept unchanged.

It is clear that different levels of $\delta$ would lead to the different patterns when the system evolves in time.

Let us assume that the initial random density of preys (shown in Fig. 2(a)) is changed by adding a T-shaped mask. Numerical values of pixels in the zone occupied by the letter T are incremented by $\delta$; all other pixels remain unchanged. Fig. 2(c) and 2(e) represent modifications of the initial distribution of preys for different values of $\delta$. It appears that the striped-spotted pattern of preys mimics the shape of the mask after 200000 iterations if only $\delta$ is sufficiently high. It can be noted that a larger ratio $\delta/\varepsilon$ corresponds to a clearer target image in final patterns (Fig. 2(f)). Unfortunately, the ratio $\delta/\varepsilon = 0.1$ (Fig. 2(c)) does not yield an interpretable pattern (Fig. 2(d)). But even such relatively small modifications in the initial distribution of preys are statistically detectable (the shape of the mask can be seen by a naked eye in Fig. 2(c).

Therefore, such an approach can not be considered as a safe technique for encoding secret information.

### *B. A steganographic communication scheme based on the difference between evolving patterns*

Previous computational experiments show that modifications of the initial random density of preys cannot be consid-

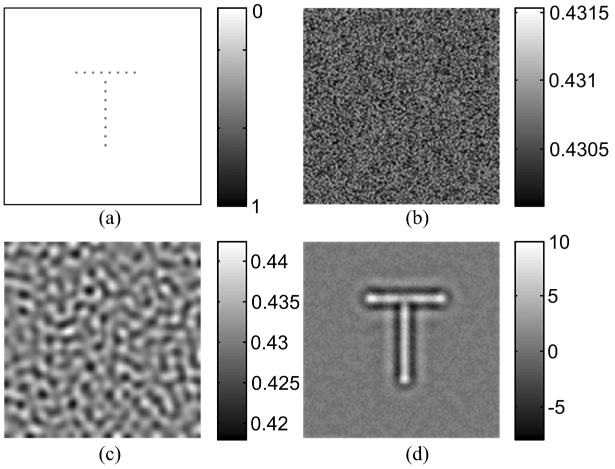


Fig. 3. A steganographic communication scheme based on the difference between evolving patterns. (a) – the dot-skeleton representation of the secret image; (b) – the perturbed initial distribution of preys; $\delta/\varepsilon = 0.3$; (the initial distribution of preys is shown in Fig. 1(a)); (c) – time evolution of (b) after 10000 iterations; (d) – the difference between (c) and Fig. 1(c).

ered as a safe encoding of secret visual information – the target pattern becomes interpretable only when the initial distribution of preys leaks the secret to a naked eye. Therefore, we propose an encoding scheme based not on a target pattern but on the difference between two evolving patterns.

At first we construct the initial random distribution of preys (Eq. (10)) and compute the density of preys after the system evolves $m$ iterations in time (Fig. 1(a) and Fig. 1(c)). In the next step the initial random distribution of preys is perturbed. We use Eq. (11) for the perturbation, but the mask $[M]$ now holds not a target pattern but skeleton dots of the secret image instead (Fig. 3(a)). It can be noted that the matrix $\left[\tilde{N}\right]$ must be kept the same in both computational experiments and that $\delta$ is low enough to prevent statistical identification of the perturbation (we use $\delta/\varepsilon = 0.3$ in Fig. 3(b)). Now, the density of preys is computed after the system evolves $m$ iterations in time (Fig. 3(c)). In fact, differences between Fig. 1(c) and Fig. 3(c) are hardly seen. Anyway, we compute the difference between these two patterns; the resulting image is shown in Fig. 3(d). It can be noted that the colorbar in Fig. 3(d) shows the difference in pixel levels (grayscale levels are measured in the interval $[0; 255]$), while colorbars in Fig. 3(b) and 3(c) show actual concentration of preys.

The secure communication system based on the formation of self-organizing patterns can be described by the schematic diagram in Fig. (4).

The functionality of the proposed technique is demonstrated using a computational example illustrated in Fig. 5. The secret image is shown in Fig. 5(a); its dot-skeleton representation – in Fig. 5(b) (the distance between dots in the direction of the $x-$ and $y-$ axis is 7 pixels). The encrypted image is shown in Fig. 5(c); the evolved pattern from the encrypted image (after 10000 iterations) is shown in Fig. 5(d). The evolved pattern from the random perturbation (without the embedded dot-skeleton representation of the secret image) is shown in Fig. 5(e). The difference between Fig. 5(d) and Fig. 5(e) is shown in Fig. 5(f).

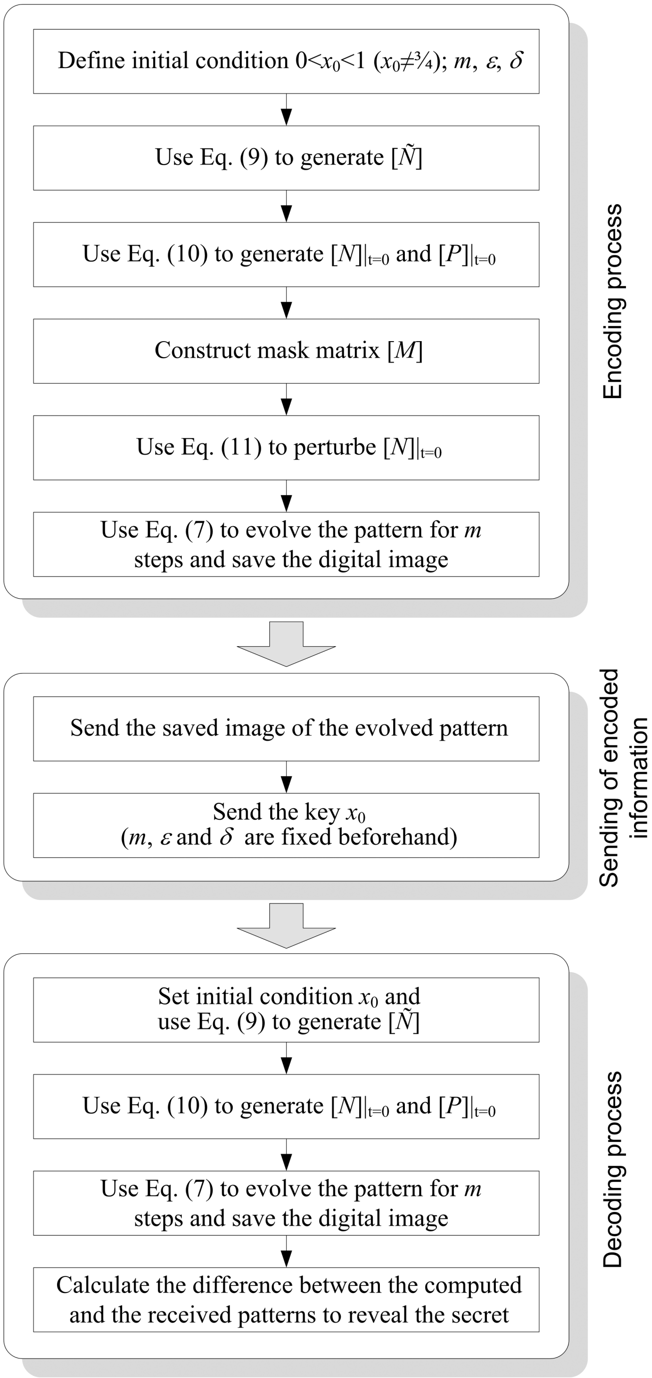

Define initial condition $0<x_0<1$ ($x_0 \neq 3/4$); $m$, $\varepsilon$, $\delta$
Use Eq. (9) to generate $[\tilde{N}]$
Use Eq. (10) to generate $[N]|_{t=0}$ and $[P]|_{t=0}$
Construct mask matrix $[M]$
Use Eq. (11) to perturbe $[N]|_{t=0}$
Use Eq. (7) to evolve the pattern for $m$ steps and save the digital image
Encoding process
Send the saved image of the evolved pattern
Send the key $x_0$ ($m$, $\varepsilon$ and $\delta$ are fixed beforehand)
Sending of encoded information
Set initial condition $x_0$ and use Eq. (9) to generate $[\tilde{N}]$
Use Eq. (10) to generate $[N]|_{t=0}$ and $[P]|_{t=0}$
Use Eq. (7) to evolve the pattern for $m$ steps and save the digital image
Calculate the difference between the computed and the received patterns to reveal the secret
Decoding process

Fig. 4. Schematic diagram of the secure communication system based on the formation of self-organizing patterns.

A naked eye can not see any differences between Fig. 5(d) and Fig. 5(e). But it is important to note that the actual difference between Fig. 5(d) and Fig. 5(e) is a smooth image; the secret information is not hidden at some isolated pixels. Steganalysis procedures [15] would not be able to detect the

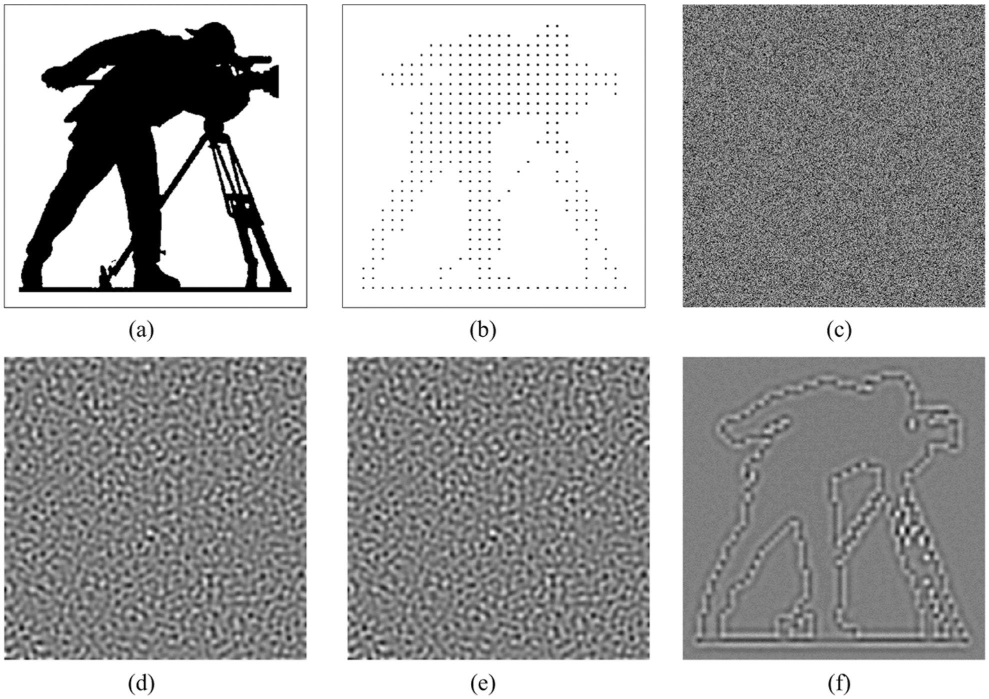


Fig. 5. The illustration of steganographic visual communication system based on self-organizing patterns: (a) – the secret image; (b) – the dot-skeleton representation of the secret image; (c) – the random initial distribution of preys with the embedded dot-skeleton representation of the secret image; (d) – time evolution of (c) after 10000 iterations; (e) – time evolution of the random initial distribution of preys without the embedded dot-skeleton representation of the secret image; (f) – the difference between (d) and (e) reveals the secret image.

fact that some secret information is being transmitted by means of Fig. 5(d).

## IV. Advantages of the proposed communication scheme

As mentioned previously, steganography includes the concealment of information within computer files. Steganographic coding may be present inside of a transport layer, such as a document file, image file, program or protocol. Our approach could be classified as a variant of text steganography inside a cover image. Various algorithms have been proposed to implement steganography in digital images. They can be categorized into three major clusters: algorithms using the spatial domain such as S-Tools [16], algorithms using the transform domain such as F5 [17] and algorithms taking an adaptive approach combined with one of the former two methods, e.g., ABCDE (A block-based complexity data embedding) [18]. Most of the existing steganographic methods rely on two factors: the secret key and the robustness of the algorithm.

A number of different methods exist to utilize the concept of steganography. Least significant bit (LSB) insertion is a common and simple approach to embed secret text information in a cover object. 3 bits in each pixel can be stored by modifying the LSBs of R, G and B array in a 24-bit image as cover. To the human eye, the resulting stego image will look identical to the cover image [19], [20]. The LSB modification concept can be used to hide data in an image [20], [21]. A random LSB insertion method is developed in [22] where the secret data are spread out among the cover image in a seemingly random diffused manner. An LSB insertion steganographic method coupled with high security digital layers is presented in [23]. A heuristic approach to hide data using LSB steganography technique is proposed in [24].

A definitive advantage of the proposed secret communication scheme is determined by the complexity of physical processes exploited in the encoding and decoding of secret visual information. The security of communicating parties is preserved since the transmittance of visual patterns does not attract the attention of eavesdroppers. In that respect our technique outperforms classical steganographic algorithms where some pixels of the cover image are modified in order to conceal a secret message in the cover image [13]. We transmit a smooth pattern which has evolved from perturbed initial conditions. It would be impossible to trace a perturbed pixel in the digital image of the evolved pattern.

## V. Concluding remarks

A new steganographic communication scheme based on evolving patterns is proposed in this paper. We use the perturbed pattern of preys to hide the skeleton of the secret image.

We have exploited the well-known Beddington-DeAngelis-type predator-prey model with self-diffusion for the generation of evolving patterns. The ability to encrypt images in a self-organizing pattern is based on the sensitivity to initial conditions in the evolution of this pattern. In principle any nonlinear physical model of evolving patterns in isotropic systems, which have as equilibrium stripe-like patterns (the reaction-diffusion model, the two-phase flow model, the model of competing species, the disordered plane wave model, etc.) could be used as the algorithm for the computation of evolving Turing's patterns.

The storage capacity of secret information is relatively small and is predetermined by the average width of stripes in the evolving pattern. Nevertheless, the ability of the proposed scheme to hide information and to avoid suspicion outperforms traditional steganographic techniques if the security of communication is considered as a primary objective.

## Acknowledgment

Partial financial support from the Lithuanian Science Council under project No. MIP-041/2011 is acknowledged.

## References

[1] L. Saunoriene and M. Ragulskis, "A secure steganographic communication algorithm based on self-organizing patterns," *Phys. Rev. E*, vol. 84, p. 056213, 2011.

[2] J. E. Pearson, "Complex patterns in a simple system," *Science*, vol. 261(5118), pp. 189–192, 1993.

[3] M. Seul and D. Andelman, "Domain shapes and patterns:the phenomenology of modulated phases," *Science*, vol. 267(5197), pp. 476–483, 1995.

[4] O. Kuksenok, V. V. Yashin, and A. C. Balazs, "Spatial confinement controls self-oscillations in polymer gels undergoing the belousov-zhabotinsky reaction," *Phys. Rev. E*, vol. 80, p. 056208, 2009.

[5] A. Yochelis and M. Sheintuch, "Principal bifurcations and symmetries in the emergence of reaction-diffusion-advection patterns on finite domains," *Phys. Rev. E*, vol. 80, p. 056201, 2009.

[6] C. M. Topaz and A. J. Catllá, "Forced patterns near a turing-hopf bifurcation," *Phys. Rev. E*, vol. 81, p. 026213, 2010.

[7] Y. Suzuki, T. Takayama, I. N. Motoike, and T. Asai, "Striped and spotted pattern generation on reaction-diffusion cellular automata: Theory and lsi implementation," *Int. J. Unconv. Comput.*, vol. 3, pp. 1–13, 2007.

[8] W. Wang, Y. Lin, L. Zhang, F. Rao, and Y. Tan, "Complex patterns in a predator-prey model with self and cross-diffusion," *Commun. Nonlinear Sci. Numer. Simulat.*, vol. 16, pp. 2006–2015, 2011.

[9] W. Wang, L. Zhang, H. Wang, and Z. Li, "Pattern formation of a predator-prey system with ivlev-type functional response," *Ecol. Model.*, vol. 221(2), pp. 131–140, 2010.

[10] M. R. Garvie and C. Trenchea, "Spatiotemporal dynamics of two generic predator-prey models," *J. Biol. Dynamics*, vol. 4(6), pp. 559 – 570, November 2010.

[11] B. Dubey, N. Kumari, and R. K. Upadhyay, "Spatiotemporal pattern formation in a diffusive predator-prey system: an analytical approach," *J. Appl. Math. Comput.*, vol. 31, pp. 413–432, 2009.

[12] V. V. Yaschenko, *Cryptography: An Introduction*. Providence, RI: American Mathematical Society, 2002.

[13] N. Johnson, Z. Duric, and S. Jajodia, *Information hiding: steganography and watermarking: attacks and countermeasures*. Netherlands: Springer, 2001.

[14] F. A. P. Petitcolas and S. Katzenbeisser, *Information Hiding Techniques for Steganography and Digital Watermarking*. Boston: Artech House Publishers, 2000.

[15] H. Wang and S. Wang, "Cyberwarfare: Steganography vs steganalysis," *Communications of the ACM*, vol. 47, p. 76, 2004.

[16] "Online software [s-tools]," accessed 15-August-2011, ftp://ftp.funet.fi/pub/crypt/mirrors/idea.sec.dsi.unimi.it/code/s-tools4.zip.

[17] "Online software [f5]," accessed 15-August-2011, http://wwwrn.inf.tu-dresden.de/∼westfeld/f5.html.

[18] H. Hioki, in *Proceedings of Pacific Rim Workshop on Digital Steganography*, July 2002, p. 30.

[19] N. F. Johnson and S. Jajodia, "Exploring steganography: Seeing the unseen," *Computer*, vol. 31, p. 26, 1998.

[20] S. Bandyopadhyay, D. Bhattacharyya, P. Das, S. Mukherjee, and D. Ganguly, "A tutorial review on steganography," in *IC3*, August 2008, p. 106.

[21] K. M. Singh, S. B. Singh, and L. S. S. Singh, "Hiding encrypted message in the features of images," *IJCSNS*, vol. 7, p. 302, 2007.

[22] M. Sutaone and M. Khandare, "Image based steganography using lsb insertion technique," in *IEEE WMMN*, January 2008, p. 146.

[23] N. N. EL-Emam, "Hiding a large amount of data with high security using steganography algorithm," *J. Comp. Sci.*, vol. 3, p. 223, 2007.

[24] S. K. Bandyopadhyay, D. Bhattacharyya, P. Das, S. Mukherjee, and D. Ganguly, "A secure scheme for image transformation," in *IEEE SNPD*, August 2008, p. 490.

# Software Similarity and Metamorphic Detection

Mausami Mungale
Department of Computer Science
San Jose State University
San Jose, California 95192

Mark Stamp
Department of Computer Science
San Jose State University
San Jose, California 95192
Email: stamp@cs.sjsu.edu

***Abstract*—In this paper, we consider a novel method for measuring the similarity of software. Our technique can be applied to any executable file and no special effort is required when developing the software. In addition, our similarity score can be computed at any point in time—even after the software has been distributed.**

**Our approach was inspired by the success of previous research focused on detecting metamorphic computer viruses. Here, we train a hidden Markov model (HMM) on an opcode sequence extracted from a specific piece of software (the "base" software). This trained model can then be used to score another piece of software, giving a measure of its similarity to the base software. We provide experimental results that show our scheme is robust in the sense that we can extensively modify the base software and still obtain strong scores from the trained HMM. Interestingly, the work presented here has some implications for the metamorphic detection problem that served as the original motivation. We briefly discuss the connections between these two problems.**

## I. Introduction

In this paper we consider a novel technique for measuring the similarity of software. The work presented here is not a watermarking scheme *per se*, but it could be used in a similar way, at least in certain cases. Our approach was inspired by previous studies of metamorphic computer viruses [1], [7], [17], [18]. Metamorphic viruses change their structure (but not their function) each time they replicate. This makes detection difficult, since there is no constant signature available for virus scanning. In this paper, we use elementary morphing techniques to train a hidden Markov model, and we use more advanced metamorphic techniques to test the robustness of our approach. Here, "robust" means that we can properly classify software even in the presence of significant modifications to the code.

Our technique is potentially useful in several contexts. For example, suppose that a company suspects that their copyrighted software has been illegally copied. Then our scoring technique could be used to compare the original software to the suspected copy. A high score would not prove that the suspect software is a modified copy, but it would indicate that further analysis is warranted. In contrast, a low score would imply that the two pieces of software are substantially different.

The remainder of this report is organized as follows. Background information is covered in Section II. Section III contains an overview of our technique, while Section IV discusses our implementation in more detail. Experimental results are given in Section V and Section VI concludes the paper.

## II. Background

In this section, we briefly discuss background topics that are relevant to the discussion in the remainder of the paper. Specifically, we discuss digital watermarking, metamorphism, and hidden Markov models.

### A. Watermarking

Watermarking is a technique for embedding some special mark in an object, so that the object can be identified [2]. Software watermarks could be used, for example, to detect software piracy and provide evidence to prosecute those responsible. Software watermarking can be accomplished in various ways, such as embedding some particular sequence of instructions at the assembly code level [16]. This embedded code—which serves as a watermark enabling us to identify the code—could be read by disassembling the executable. Ideally, such a watermark needs to be robust against attacks involving modification of the code.

Software watermarks can be classified as static or dynamic. For example, embedding a sequence of assembly code instruction (as mentioned in the previous paragraph) is a static technique. A dynamic technique might consist of a particular "secret" value that is output when a special input is provided. A good discussion of software watermarking can be found in [3].

It is worth noting that robust watermarking of digital content has proven difficult to achieve in practice. To illustrate this point, consider the Secure Digital Music Initiative (SDMI) [12]. In September 2000, SDMI issued a public challenge, apparently to show off the supposed strength of four "robust" watermarking technologies [11]. However, in spite of limited information provided with the challenge, all of the watermarking techniques were soundly defeated [4]. The excellent work presented in [4] should serve as a cautionary tale against strong claims of robust digital watermarking.

Various types of attacks against static software watermarks are possible. Additive attacks involve inserting additional watermark information into already watermarked software. The goal of such an attack is to make the original watermark undetectable. Distortion attacks involve using semantic preserving transformations to make a watermark undetectable. Subtractive attacks involve identifying the watermark and removing it without changing the functionality of the program. Although our proposed technique is not a true watermarking scheme, the obvious attacks are similar to those used against static watermarks.

### *B. Metamorphism*

Metamorphism is the process of transforming a piece of code into copies that are functionally equivalent, but structurally different [15]. Metamorphism has been used by virus writers in an attempt to defeat signature based anti-virus software. Metamorphic software also has the potential for positive uses, such as increasing the diversity of software. More diverse software can reduce the impact of many implementation level attacks, such as buffer overflows [14].

*1) Assembly Language Basics:* The full x86 instruction set is large and complex [5]. A typical instruction consists of an "opcode" followed by one or more operands. The operands can be constant values, pointers to a value in memory, or a register. The instructions can broadly be classified as data transfer instructions, arithmetic and logical instructions, and control flow instructions. Table I shows some typical instructions of each of these types.

TABLE I
EXAMPLES OF X86 INSTRUCTIONS

| Data Transfer Instruction | |
|---|---|
| `MOV` | Move byte or word to register or memory |
| `IN`, `OUT` | Input/output byte or word |
| `LEA` | Load Effective Address |
| `PUSH`, `POP` | Push/Pop word on/from stack |
| **Arithmetic and Logical Instructions** | |
| `NOT` | Logical NOT of byte or word |
| `AND` | Logical AND of byte or word |
| `OR` | Logical OR of byte or word |
| `XOR` | Logical XOR of byte or word |
| `ADD`, `SUB` | Add, subtract byte or word |
| `INC`, `DEC` | Increment, decrement byte or word |
| `NEG` | Negate byte or word (two's complement) |
| `MUL`, `DIV` | Multiply, divide byte or word (unsigned) |
| **Control Flow Instructions** | |
| `JMP` | Unconditional jump |
| `JE`, `JNE` | Jump if equal/Jump if not equal |
| `LOOP` | Loop unconditional, count in `CX`, short jump to target address |
| `CALL`, `RET` | Call, return from procedure |

### *C. Metamorphic Code Techniques*

An important component of our system is a metamorphic code generator, which is applied at the assembly code level. There are a large number of semantic preserving transformations that can be applied to assembly code to obtain metamorphic copies [7]. Here, we briefly discuss some elementary metamorphic techniques.

*1) Control Flow Preserving Transformations:* In this type of transformation, we insert instructions that, taken as a whole, do not change the data flow or the control flow of the program. We give a few elementary examples of such transformations and provide examples of each.

1) A `NOP` is a special instruction that has no effect on the execution state—it is simply a "do nothing" instruction. Therefore, we can insert `NOP`s between instructions as shown in Table II.

TABLE II
NOP EXAMPLE

| Original Code | Transformed Code |
|---|---|
| `MOV AL,BL` | `MOV AL,BL` |
| `ADD AL,05H` | `NOP` |
| | `ADD AL,05H` |

2) We can use groups of arithmetic or logical instructions, the net effect of which does not change the value of any registers. Since arithmetic instructions can change the flags, we may need to save and restore the `EFLAG` register when inserting such code. Neglecting effects on the flag bits, Table III gives examples of such instruction groups.

TABLE III
ARITHMETIC EXAMPLE

| |
|---|
| `ADD AX,05H`<br>`SUB AX,05H` |
| `XOR AX,0H` |
| `AND AX,FFFFH` |

3) We can add a label to any instruction and put a `JMP` instruction to that label just before the instruction. This does not change the program behavior—see Table IV for an example.

TABLE IV
JMP EXAMPLE

| Original Code | Transformed Code |
|---|---|
| `MOV AL,BL` | `MOV AL,BL` |
| `ADD AL,05H` | `JMP LOC` |
| | `LOC: ADD AL,05H` |

4) We can `PUSH` the value of a register on the stack and `POP` it immediately to preserve the program semantics. This is illustrated in Table V.

TABLE V
PUSH AND POP EXAMPLE

| Original Code | Transformed Code |
|---|---|
| `MOV AL,BL` | `MOV AL,BL` |
| `ADD AL,05H` | `PUSH AX` |
| | `POP AX` |
| | `ADD AL,05H` |

*2) Dead Code Insertion:* Perhaps the easiest way to morph a program is to insert code that is never executed. Any combination of instructions can be included within a dead code block. Table VI gives an example of dead code insertion.

It is fairly easy to make dead code stealthy, in the sense that it is non-trivial to determine whether or not it is not actually executed. We do not consider this further here, but it

TABLE VI
DEAD CODE EXAMPLE

| Original Code | Transformed Code |
|---|---|
| `MOV AL,BL` | `MOV AL,BL` |
| `ADD AL,05H` | `JMP LOC:` |
| | `PUSH AX` |
| | `POP AX` |
| | `ADD AL,BL` |
| | `LOC: ADD AL,05H` |

is worth noting that we make no effort to remove dead code in our scoring method discussed below. Removing obvious dead code is not difficult and would serve to make our scoring even stronger. In effect, we are treating the inserted dead code as if it were stealthy, although it is not.

*3) Equivalent Code Substitution:* We can transform code by replacing an instruction (or instructions) with an equivalent instruction (or instructions). Examples of equivalent code substitution appear in Table VII.

TABLE VII
EQUIVALENT CODE EXAMPLE

| Original Code | Transformed Code |
|---|---|
| `ADD AL,05H` | `ADD AL,04H`<br>`ADD AL,01H` |
| `MOV AX,BX` | `PUSH AX`<br>`POP BX` |

### *D. Hidden Markov Models*

A Markov chain can be viewed as a statistical model in which there are states and known probabilities for state transitions. In such a Markov model, the states are visible to the observer. In contrast, a hidden Markov model (HMM) has "hidden" states, i.e., the states are not directly observable. Although the states of a hidden Markov model are not visible, there is some observable output, which is probabilistically related to the hidden state [10], [13]. A hidden Markov model consists of state transition probabilities, a probability distribution over all possible output symbols for each state, and initial state probabilities.

We use the following notation to describe an HMM:

$T =$ the length of the observation sequence
$N =$ the number of hidden states in the model
$M =$ the number of distinct observation symbols
$X = \{x_0, x_1, \ldots, x_{N-1}\} =$ the states of the Markov process
$O = \{O_0, O_1, \ldots, O_{M-1}\} =$ set of possible observations
$A =$ state transition probabilities
$B =$ observation probability matrix
$\pi =$ initial state distribution

Figure 1 illustrates a generic HMM, where $X_0, X_1, \ldots, X_{T-1}$ are the hidden states and $O_0, O_1, \ldots, O_{T-1}$ are the observed symbols in each state. Note that the matrices $A$ and $B$ represent the state transition probabilities and observation probabilities, respectively. We represent an HMM compactly as $\lambda = (A, B, \pi)$.

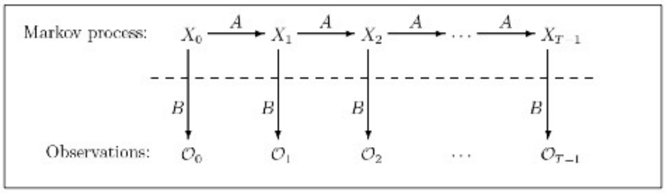


Fig. 1. Generic Hidden Markov Model

HMMs are used in many applications, including speech recognition, sequence alignment, and malware detection. The utility of HMMs derives largely from the fact that there are efficient algorithms to solve each of the following three problems [13]:

(1) Given the model $\lambda = (A, B, \pi)$ and a sequence of observations $O$, find $P(O|\lambda)$. That is, we can score an observed sequence $O$, relative to a given model $\lambda$.
(2) Given the model $\lambda = (A, B, \pi)$ and an observation sequence $O$, find an optimal state sequence for the underlying Markov process. In other words, we can uncover the (most likely) hidden state sequence.
(3) Given an observation sequence $O$, and parameters $N$ and $M$ (i.,e., the number of hidden states and the number of distinct observation symbols, respectively), find the model $\lambda = (A, B, \pi)$ that maximizes the probability of observing $O$. This can be viewed as training the model to best fit the observed data.

Note that when training an HMM, the only free parameter is $N$, the number of hidden states. This is the sense in which an HMM is a machine learning technique—the user only has to specify $N$.

In this paper, we employ the algorithms for problems (1) and (3), above. We use (3) to train a model to match a "base" piece of software. Then we can use (1) to score any piece of code against the model—a high score indicates a high degree of similarity with the base code.

## III. DESIGN OVERVIEW

The goal of this research is to design a robust method for measuring software similarity at a fairly deep level. In this section we give a brief overview of the approach we have followed.

### *A. System Overview*

Our system has two phases. In the first phase, we use a metamorphic generator to create slightly morphed copies of the base software. We extract and append the opcode sequences from these morphed copies. We then use the resulting opcode sequence to train a hidden Markov model. The purpose of using morphed copies of the base software is to avoid having the HMM overfit the training data.

The second phase consists of scoring. In this phase, we extract the opcode sequence from a given piece of software and we score it against the trained HMM obtained in the first phase. A high score indicates that the software in question is closely related to the base software, while a low score shows that the software is "far" from the base code. Below, we show that reasonable scoring thresholds can easily be determined and that the technique described here is highly robust.

*1) Design of Metamorphic Generator:* Our metamorphic generator makes slightly morphed copies of a given base software. To generate these morphed copies, we use some of the techniques discussed in Section II-C. The amount of morphing to be applied is an adjustable parameter, given as a percentage. For example if we select 20% morphing, then after morphing, the program will have expanded by approximately 20%, due primarily to dead code insertion. Figure 2 illustrates the design of the metamorphic generator.

## IV. Implementation

In this section we briefly describe how we have implemented the various components in our system. Additional details can be found in [8].

### A. Metamorphic Generator

In our metamorphic generator, we use dead code insertion and various control preserving transformations (see Section II-C). We make changes directly to assembly code obtained by disassembling the executable file. In addition, we provide a set of "normal" files from which dead code is extracted. After disassembling the base file, the following steps are performed by the metamorphic generator for each morphed copy produced:

(a) Compute the number of lines of code in the base file.
(b) Compute the number of positions where transformed code will be inserted, based on a specified morphing percentage and the number computed in (a).
(c) Select five random locations where dead code blocks will be inserted into the morphed copy.
(d) Insert the transformed code and the dead code at the selected locations.

It is, of course, possible to generate much more highly metamorphic code—see [7] for a discussion of a highly metamorphic generator. However, here we are not trying to make our code extremely metamorphic. Instead, we simply want the code in each morphed copy to differ sufficiently so that we can avoid potential problems caused by overfitting the data when we train the HMM.

### B. Training an HMM

Given the base executable file, we disassemble it and create a collection of morphed versions using the method discussed above. The opcode sequences are then extracted from these morphed files. For training and testing we used a standard five-fold cross-validation approach [18]. Specifically, we do the following:

(1) We partition our opcode sequences into five subsets, each containing opcodes sequences from an equal number of files.
(2) Four of these subsets are selected and all opcode sequences in these subsets are appended to create one long sequence. This long opcode sequence is then used to train an HMM.
(3) To determine threshold values, the sequences in the remaining subset are scored against the model. In addition, a collection of sequences extracted from normal executable files are also scored. All scores are computed as a log likelihood per opcode (LLPO) [18].
(4) Repeat steps (2) and (3), reserving a different subset for testing each time.

The results obtained over the five iterations are averaged. Figure 3 illustrates the training phase.

Once we have trained the HMM and determined a threshold, we can score any executable file. To score a file, we simply extract its opcode sequence and use the trained HMM to compute a score. If the resulting score is higher than our predetermined threshold, we classify the file as being similar to the base file. A score below our threshold implies that the file is more similar to the normal files than to the base program used to generate the HMM. However, we actually obtain more information than a simple "yes/no" answer. In fact, we obtain a score, and the higher the score, the closer the match (in the HMM sense) to the base program.

## V. Results

In this section, we discuss experimental results for one typical test case. Additional experimental results can be found in [8].

The test data that we used for this experiment consists of 30 randomly selected executables from Cygwin version 1.5.19. Cygwin files have been used as representative "normal" files in several previous studies of metamorphic generators [1], [7], [17], [18].

The 30 Cygwin utilities files were named N0.EXE through N29.EXE. Each file was disassembled using IDA Pro, version 4.6.0. For the disassembled files, we added the prefix "IDA" to the respective file names and changed the suffix to "ASM," that is, the disassembled files are denoted IDAN0.ASM through IDAN29.ASM. We randomly selected another Cygwin executable to serve as the base code that we will test against. We named the disassembled version of this particular file IDAW.ASM. For this experiment, we generated 100 morphed copies of IDAW.ASM, which we denote as IDAW0.ASM to IDAW99.ASM.

As discussed in Section IV-B, the files IDAW0.ASM to IDAW99.ASM were used to train the HMM using five-fold cross validation. The files IDAN0.ASM through IDAN29.ASM serve as our normal files, that is they are the test set that will be used in setting the threshold.

In general, we would expect Cygwin utility files to be more alike than randomly selected programs. Consequently, it is inherently more difficult for an HMM—or any other statistical

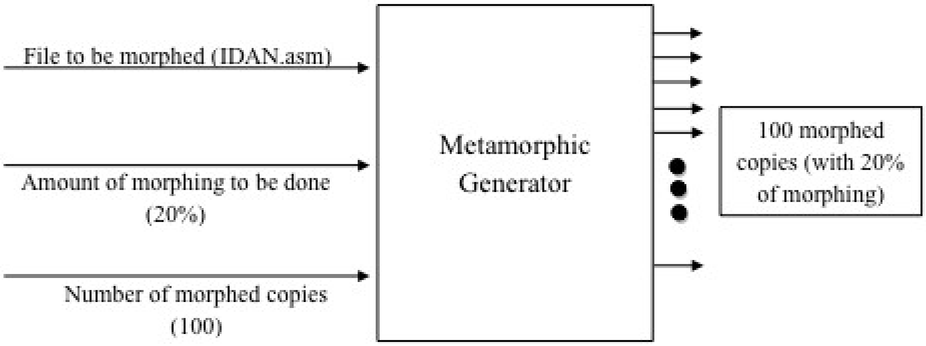

File to be morphed (IDAN.asm)
Amount of morphing to be done (20%)
Number of morphed copies (100)
Metamorphic Generator
100 morphed copies (with 20% of morphing)

Fig. 2. Metamorphic Generator

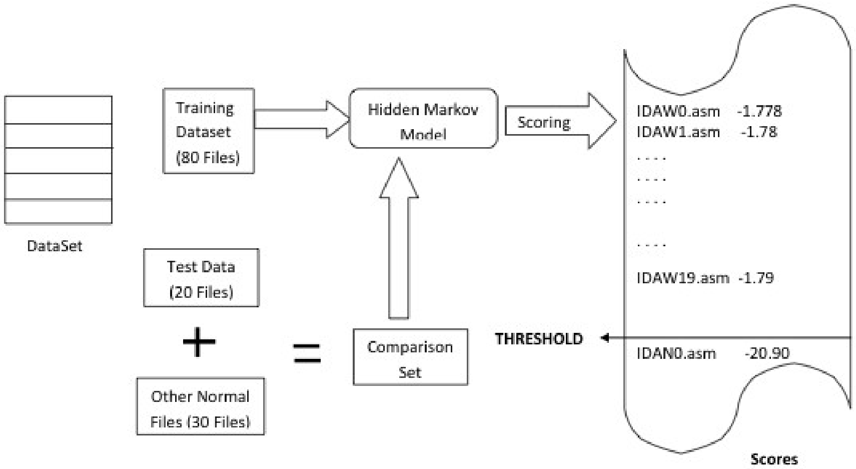


Fig. 3. HMM Training

discrimination technique—to distinguish Cygwin files from each other, as compared to two randomly selected programs. Therefore, by selecting our base program and our normal programs from the collection of Cygwin utility files, we have created a relatively challenging test for our technique.

To evaluate the ability of our scheme to detect software after it has been "attacked" (i.e., modified), we developed a tampering scheme to modify the base file. The tampering is done is a way that makes the tampered file look more like the normal files since, from the attacker's point of view, the goal is to make the HMM unable to properly classify the tampered file.

The tampering scheme used here is essentially the same as that used in [18], where the goal was used to evade HMM-based detection of metamorphic malware. In [7] it was shown that by selecting a relatively small amount of dead code from normal (i.e., non-virus) files, the virus files could evade detection by the HMM technique developed in [18]. It was somewhat surprising that a small amount of dead code would suffice, given that the detection technique in [18] had proven effective against all other metamorphic generators tested, including the strongest hacker-produced generators. We selected the metamorphic generator in [7] since we believe it is the most challenging generator—from the perspective of detection using HMMs—developed to date. This generator should pose a significant challenge for our similarity technique.

To tamper with the base program, we copied instructions from normal files and inserted them as dead code into the base file. We denote these tampered files as IDAT0.ASM, IDAT1.ASM, and so on. Again, each tampered file is a variant of the base file.

As more code from the normal files is inserted into the tampered files, the tampered files become closer (in terms of opcode statistics) to the normal file. From the perspective of the HMM, the tampered files should look more similar to the normal files. In addition, the higher the percentage of "tampered" code, the closer the HMM scores should be to the normal score range. Our goal is to experimentally determine a rough estimate of the amount of tampering that can be tolerated before we are unable to correctly classify the tampered files as matching the base file.

For our experiment, we used 10% morphing, which implies that each training file is a slightly morphed version of the base file. We followed the HMM training and scoring procedure discussed above. For each test, we scored 20 morphed files and 30 normal files to obtain a threshold. The tampering included dead code insertion and equivalent instruction substitution. Note that the dead code was extracted from the normal files, namely, other Cygwin utility files.

Figure 4 shows the misclassification rates for dead code percentages ranging from 10% to 90%. Note that we correctly classify all tampered versions of the base file up to 50% tampering, while at higher tampering rates the classification accuracy drops precipitously. At tampering rates of 70% or greater, we cannot correctly classify any of the tampered files. This shows that our technique is extremely robust, since failure only occurs at very high tampering rates.

Finally, Figure 5 shows the growth in the size of the tampered files, in term of the number of lines of code.

### A. Discussion

In [7], a metamorphic generator was constructed that produced malware variants containing dead code selected from normal files. A similar process was followed in the experiment discussed in this section. However, in [7], only a relatively small amount of dead code insertion was needed to break the HMM detector. That is, a small amount of dead code was sufficient to cause the HMM to fail to properly classify the metamorphic viruses. In contrast, a very high degree of metamorphism is required before the similarity test discussed in this paper fail to properly classify the tampered software. While this may seem contradictory, in fact, it is not.

In the case of metamorphic detection, the HMM is trained on the morphed files, since only the morphed virus files are assumed to be available. This is the most realistic scenario in the case of malware detection. In contrast, for the similarity problem discussed here, we have access to the base file, so we can train our model on a "pure" set of training data, which is independent of the degree of tampering employed by the attacker. For metamorphic malware detection, the model is, in effect, trained on a tainted data set, which implies that the model itself to degraded at higher levels of morphing. When viewed in this light, the large difference between the morphing rates tolerated by these two related techniques is not totally unexpected.

The results in this paper have some interesting implications for the virus detection problem. First, note that if the metamorphic virus generator is available, then it should be possible to train the model on data that does not include dead code from normal files. This should give us results comparable to those presented here, that is, we should then be able to tolerate a much higher degree of "tampering" than when we are forced to train on the actual virus files, as in [18].

The results in this paper also have implications for the virus detection problem, even in case where the generator is not available. Depending on the sophistication of the morphing techniques used, it may be possible to strip some or all of the dead code from the viruses before training [9]. If so, the resulting model would be considerably stronger. In fact, the results here and in [7] indicate that removing dead code from the training data is critical. In contrast, a relatively high degree of tampering can be tolerated in the files that are scored. This is significant in the case of virus detection, since training the model is essentially one-time work, so we can afford to spend considerable effort to obtain pristine training data. In contrast, the scanning phase must be fast. Fortunately, the results here indicate that we can tolerate high levels of tampering at the detection phase (provided the underlying model is strong), which would make scanning much more robust and practical.

## VI. Conclusion

In this paper, we have analyzed a software similarity measure inspired by the success of previous research using hidden Markov models to detect metamorphic computer viruses. Experimental results show that this similarity score is robust, in the sense that it can withstand a very high degree of tampering before the classification fails.

While a high score using our method would not prove that the software in question is a tampered version of the original base file, it would certainly indicate that further investigation is warranted. Conversely, a low score would clearly indicate that the software is "far" away from the base program, in which case further investigation is unlikely to yield incriminating results. Since our approach can be used after the fact, only executable code is required, and the computational cost is minimal, it would be reasonable to compute this similarity score before proceeding to a more detailed—and costly—analysis.

## References

[1] S. Attaluri, S. McGhee, and M. Stamp, Profile hidden Markov models and metamorphic virus detection, *Journal in Computer Virology*, Vol. 5, No. 2, May 2009, pp. 151–169

[2] R. Chandramouli, N. Memon, and M. Rabbani, Digital watermarking, *Encyclopedia of Imaging Science and Technology*, 2002

[3] C. Collberg and C. Thomborson, Software watermarking: Models and dynamic embeddings, 1999

[4] S. A. Craver, et al., Reading between the lines: Lessons from the SDMI challenge, *10th USENIX Security Symposium*, 2001

[5] Intel, IA-32 architectures software developer's manuals, `http://www.intel.com/products/processor/manuals/index.htm`.

[6] K. R. Irvine, *Assembly Language for x86 Processors*, 6th edition, Prentice-Hall, 2010

[7] D. Lin and M. Stamp, Hunting for undetectable metamorphic viruses, to appear in *Journal in Computer Virology*

[8] M. Mungale, Robust watermarking using hidden markov models, Master's Thesis, Department of Computer Science, San Jose State University, Spring 2011

[9] S. Priyadarshi, Metamorphic detection via emulation, Master's Thesis, Department of Computer Science, San Jose State University, Spring 2011

[10] L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, Vol. 77, No. 2, pp. 257–286, 1989

[11] SDMI public challenge, September 2000, `http://www.hacksdmi.org`

[12] Secure Digital Music Initiative, `http://www.sdmi.org`

[13] M. Stamp, A revealing introduction to hidden Markov models, 2004, `http://www.cs.sjsu.edu/faculty/stamp/RUA/HMM.pdf`

[14] M. Stamp, Risks of monoculture, Inside Risks, *Communications of the ACM*, 2004

[15] M. Stamp, *Information Security: Principles and Practice*, 2nd edition, Wiley 2011

[16] S. Thaker, Software watermarking via assembly code transformations, Master's Thesis, San Jose State University, 2004

[17] S. Venkatachalam and M. Stamp Detecting undetectable metamorphic viruses, to appear in *Proceedings of SAM '11*

[18] W. Wong and M. Stamp, Hunting for metamorphic engines, *Journal in Computer Virology*, Vol. 2, No. 3, December 2006, pp. 211–229

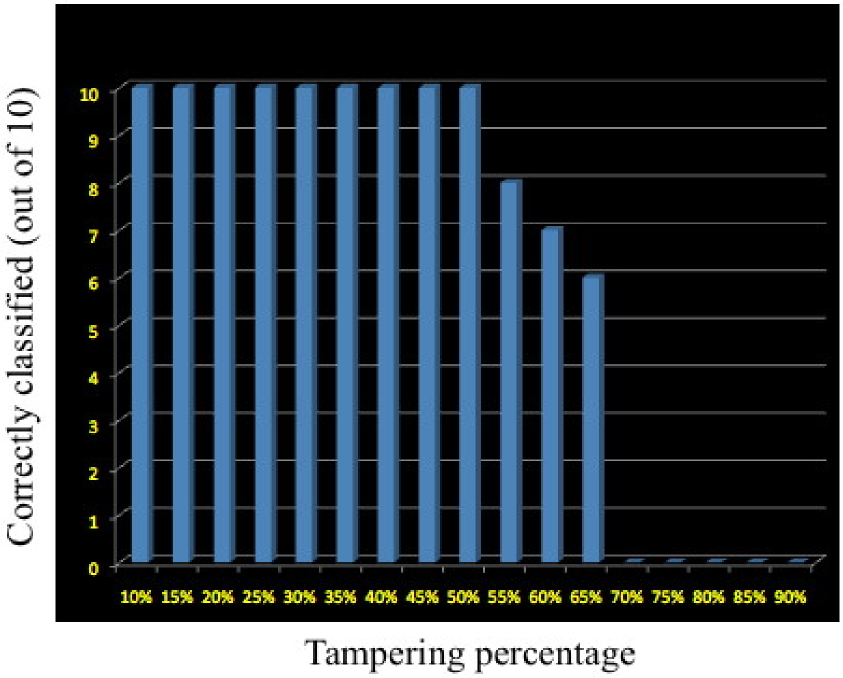


Fig. 4. Detection Results

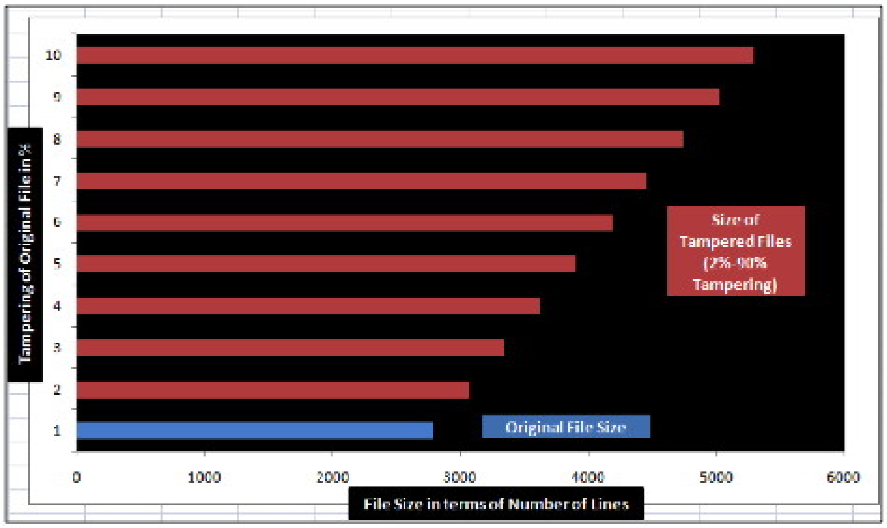


Fig. 5. Size of Tampered and Base Files

# Designing and deploying a secured VO for a  wine geotraceability application

François Barrère, Romain Laborde, Abdelmalek Benzekri
*IRIT Laboratory - Toulouse University, Route de Narbonne 118, 31062 Toulouse, France,*
*(barrere, laborde, benzekri }@irit.fr  ;fbarrere@gmail.com*

**ABSTRACT:**

The concept of the Virtual Organisation (VO) comes from network evolution infrastructure and the growth of collaborative work tools. VO can be deployed to support the cooperation between different partners sharing a same goal. Geowine project aims to bring geo-traceability information in the “wine supply chain” to clients and customers to make them sure they are buying a genuine bottle. Such an innovative service require bottle and content to be protected, but also implies to build and secure the information system used and shared by the partners of the supply chain. In order to match the VO, this paper exposes the expectation of Geowine project, the problems encountered regarding security and the way we have deployed the information system.

**KEYWORDS :**

virtual organisation, security, PKI, Clark and Wilson, integrity.

## I. INTRODUCTION

Virtual Organizations (VOs) offer new challenges and ways of working. It offers a new model of cooperation between enterprises that may be built dynamically. They are using temporary network of independent organizations connected through Information and Communication Technologies (ICTs) [1]. Common projects were primarily economical or educational ones and most often led by the necessity of sharing resources (human, material, etc.), capabilities and competencies. Some domains like organizing e-commerce may benefit from VO implementation. During the last three years we have worked in a project named Geowine which aims to set up a geo-traceability framework for wine lifecycle. This project concerns wine safety and consumer protection.

Within a VO network, organizations are opening their information systems and security concerns are raised to setup collaboration means. VO borders need to be defined: security practices and policies must be implemented to protect the resources from harmful attacks avoiding consequences like business or marketplace losses.

When reasoning with people aware with ICT systems, a global information security policy may be designed and refined to be attached to the VO level. Then it can be built as an extension of the different organizations’ local security policies avoiding to start from scratch. Such a solution brings the opportunity to focus only on security concerns associated with the VO, while respecting those defined to govern exchanges within an organization. The question is how to implement such a system when multiple producers, group of producers, vendors, SME’s, expertise-centers, clients, consumers …  working in different domains (and sometimes far from IT technologies) need to cooperate.

In the Geowine project, we have been faced with such situation. We have been involved in designing and implementing a secured exchange system to guarantee geo-traceability all along wine lifecycle. This paper presents the solution we have deployed to guarantee that trusted information are retrieved by wine consumers. This paper is organized as follow: section 2 describes the Geowine project; section 3 discuss the different architecture that are suitable to meet our objectives ; section 4 present integrity models that can be applied in commerce application. Section 5 presents the main results of our implementation.

## II. GEOWINE PROJECT

### A. Virtual organization setup context

**Geowine project was funded by “Conseil Regional Midi Pyrenées”, by the DGCIS Direction Générale de la Competitivité de l’Industrie et des Services, by the conseil Général of Tarn et Garonne and by Montauban town. Geowine is a research and development project sustain by “AGRIMIP competiveness cluster**” which stand as an association promoting changes in general agriculture context.  The main objective of this cluster is to create synergy among the various partners in the agribusiness chain (corporate groups, innovative SMEs, farmers, training and research establishments. To comply with competitiveness cluster's strategy, the general agricultural context, agribusiness, and nutritional behavior are analyzed in a globalized economic context. Today, the market and the consumers has become the driving force of these sectors and the term "chain" symbolizes this development. In this context, the cluster's scope of action involves to build agri-chains, a concept to guarantee the creation of innovative products by orienting agriculture and agribusiness towards the market and consumer expectations. As a consequence the achievement of a desired result need to build a cooperation called in ICT world a virtual organization where SMEs, R&D departments, farmers become partners and act together into three innovation orientations : a) technologies are used to facilitate the cooperation between partners and the compliance with goal specification like safety, traceability…., b) new processes and control functions are proposed while partners are working together, c) The market and the consumers need to be satisfied with new tools and equipments.

### B. Geowine objective

Geowine is a research project combining geomatics and viticulture. Geomatics stand for geospatial technology (or geomatics engineering) and proposes to gather, store, process, and deliver geographic or spatial referenced information. Some technologies, tools and techniques have been developed and are applied in land surveying, cartography, soil studying, photogrammetry, geography and earth mapping. Viticulture covers the science, production and study of grapes which deals with the series of events that occur in the vineyard. Many vine yards belong to wine producers. Like any authentic product, wine need to be protected. For a long time, many labels have been created and assigned to identify bottles' origin. Some countries are specialized into wine production and are using a labelling system which is considered as too complex. As a simplification is on study, the geowine project aim to use technologies, to be innovative and compliant with a new labelling form and deploy it. Moreover this system is design to guarantee the origin of the content of a bottle.

In fact, as wine commerce is ongoing, some bottles are considered as luxury products and fraud becomes more and more important; It is important to propose a solution that can fight against (as it can be possible) fraud increase. Today wine fraud represents 8-10% of international business and accounts for 20-30% as regards wine in particular countries. Staying in the fields of competiveness cluster the third objective was to initiate a partnership between research and some companies while developping a new innovative product linking "geotraceability authentication".

### C. Using bubble tags to fight against fraud

Wine authenticity is defined as the ability to guarantee that consumers will taste or drink an authentic wine. Whatever it could be designed, the system must ensure that bottles are really filled with the chosen wine. Consequently, all along the chain separating production from the final customer (the consumer), the system must ensure the "integrity" of bottles. Far to be a wine-specific problem, bringing guarantees to a client or a customer affect many domains: who can be sure he or she is wearing an original item, or possessing an authentic watch, an authentic passport? Building a secure and/or safe infrastructure requires to deploy systems depending on each situation, on each product, on each software. With each application, solutions are available to fight against counterfeit (antitheft, RFID…) and regarding each domain the best may be highlighted.

Regarding wine bottle distribution, the use of the Bubble Tag ™ technology was chosen. As chip components can be enclosed in some kinds of passports storing its holder features, unique bubble stickers are made and paste on the bottle.

A Bubble Tag ™ is made of a translucent polymer inside which, as a result of a random phenomenon, self-generated bubbles appear. This constellation of bubbles forms the basis of the identity given to a document or a product. Each Bubble Tag ™ is unique and impossible to replicate, even by "Prooftag" trademark which hold bubble tag patent. Either as a seal or a label, the Bubble Tag ™ is stuck on wine bottles, make them unique and bring them an information that cannot be counterfeit. The established link between the bottle and the tag is stored in a database. Then, it is possible to assign each bottle a certified information (by authenticating or not its bubble tag). The tag is a visible proof of authenticity that can be simply controlled by all partners of the chain: distributors, authorities, consumers... The authenticity of the bubble tag can be first visually controlled and/or electronically verified using a smartphone to capture the bubble tag picture and control it with a matrix (QR code) stored into a tag control server. A request is send to the database to verify the bubble tag value.

.

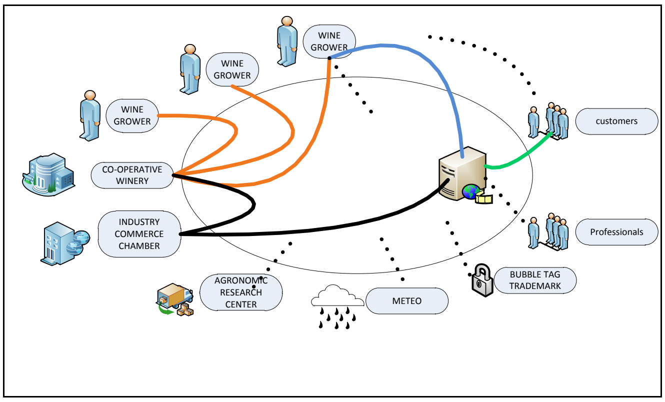
### D. Wine traceability requires information traceability

Traceability represents the ability to trace all what is done regarding either a product or a particular food chain. For the wine production, it means tracing the history of this product from the grapes production soil (beginning of the chain) and not only until the bottle is filled in, but ultimately up to its distribution. Wine-geo-traceability may provide additional information from the production environment (land or location of vineyards and/or their precise position). It is then possible to certify the traditional traceability information through the geographical information and provide relevant details on local environmental conditions. In this context, geo-traceability is adapted to the viniculture/viticulture and it establishes a trace "from the bottle to the vineyard". This is achieved by geo-localizing the land involved and to characterize the wine vintages by identifying the soils for each variety of grapes, their local environmental conditions, and all that has contributed to the particular uniqueness of the wine through pedo-climatic geo-indicators. Wine traceability implies to build and/or share a common information system between all partners who are involved in the wine chain. As the partners are all together responsible for regarding treatments and information dissemination, it is obvious that the system must be secure.

## III. Designing data geo-traceability into a shared information system

### A. VO Partners characteristics

The VO partners are given by private or professional customers, by wine grower which belong to a co-operative winery, industry and commerce chamber, agronomy research centre, and bubble tag producer. As geowine project aims to bring an added value to wine consumers or distributors, the minimal shared information system is implemented by a web portal. Private and professional customers retrieve information from this portal. The other partners (wine grower, winery, weather experts, third tier supplier…) exchange and share VO data and put some information on the web portal site. They keep discretion of their own information system.



!!!B. reflecher

On the other hand, many other domains or applications look like Geowine project architecture. For instance, in the following projects: Value Improvement thought a Virtual Aeronautical Collaborative Enterprise (VIVACE) and Transglobal Secure Collaboration Program (TSCP), we have been faced with different issues when setting up VOs [2]. The same questions are raised and not limited to: how to implement a web portal? Who is responsible for the data accessed by users? Who owns the portal information? Is the portal information in state "update" for the users trying to access it? Is this information qualified as updated and genuine? Answers to these issues are not so easy to be brought.

### B. Centralized access to geowine portal

The web portal may be implemented using a service of a third tier supplier, who collects independently from each "information producer", the ones that must be accessible to the clients of the web server. So, the information system owned by each "information producer" stay independent and separated from the web portal. Such a solution can be understood as an externalized one. The third tier supplier is deploying a centralised web portal. The major risk is that information could not be updated because the supplier needs to get input from the VO partners. From our point of view, this solution is not suitable for Geowine project. As Geowine project refers to the introduction of new technologies and requires proactive users compliant with the web2.0 concept for the success of Geowine application.

### C. Distributed access to Geowine portal

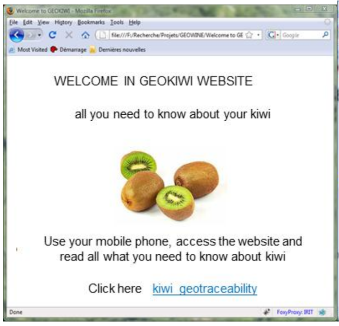
Another solution is to consider that a third tier supplier can join the VO and will be responsible for the website portal management. This could be suitable when all VO partners don't want to lose data control (extranet), and t no one wants to take in charge the website portal management . Whether the web site portal is or not managed by a third tier supplier, the Geowine solution claim to bring geo-traceability information for customers. This implies that the processes executed by the VO partners (from the grapes picking, through the bottle fill- in and wine selling) need to update some kind of information in order to be delivered by the website portal. Each VO partner executes different actions relative to his "know-how" about how to stay free and choose the best time for exporting information on the website portal. Consequently, it is more convenient and compliant with web2.0 concept to set up an IT architecture, where each VO partner realise himself the updates on the website portal. A VO partner can be as qualified as proactive. Moreover, information like geo-indicators are obtained by executing the software, needing input information from different VO partners or bought data from external sources (i.e. weather information, soil information, sunshine information, temperature information,...). Consequently, this distributed solution is suitable for implementing access to the website portal.

### D. VO Partners exchanges

Exchanges between VO partners are realised using the Internet. Consequently, VO partners need to transmit data using a secured solution, due to network risks. Naturally, only users belonging to the VO and who own permissions can put or modify data stored on the website portal. Moreover, all the members in the VO need to be known by other members and an agreement must be written linking the VO partners and defining the policy that must be enforced taking into account information diffusion. If any wrong information is published on the web site portal by a VO partner, it can damage the reputation of all other VO partners. As commercial contract refers to a legally binding agreement between parties, it is out of the scope of our contribution. We have concentrated our efforts on securing exchanges between VO partners.

Even many projects are dealing with the deployment of secure VOs supported by secure IT infrastructures, they are suitable when an information system belonging to different partners need to be open and when many resources must be accessible from one information system to another. Some projects such as CAS [8] and VOMS [9] propose a centralized management solution; for user's authentication and granting rights on resources. Other projects such as TrustCom [10] adopt the identity federation concept and thus, it proposes a decentralized management solution for the VO distributed information system.

To lend weight and credibility to Geowine, it is clear that the information retrievd from the web portal must be relevant and of good quality. While each VO partner is able to define what kind of information is relevant, it is essential to manage the integrity of the information throughout their lifecycle.

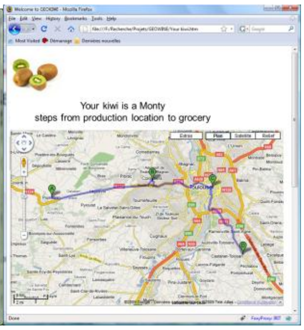


WELCOME IN GEOKIWI WEBSITE

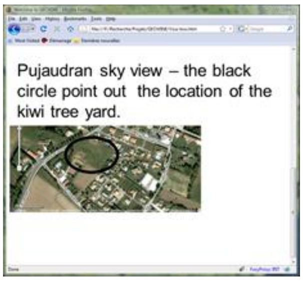
all you need to know about your kiwi

Use your mobile phone, access the website and read all what you need to know about kiwi

Click here kiwi geotraceability



Your kiwi is a Monty
steps from production location to grocery



Pujaudran sky view – the black circle point out the location of the kiwi tree yard.

Let us suppose, for instance, there is a customer that likes Kiwi fruits. This amateur heard that there is a website that gives all information about this fruit. So he/she connects to this web site (Geokiwi). Then, he/she retrieves the following information: the fruit grew in Pujaudran - a little village in the south of France-, then it has been transported by a lorry to a market center, in Toulouse – aeronautical capital -, then transported by a car until Castanet town close to Toulouse, where the customer bought the fruit from a grocery. But something surprises the customer. A sticker is pasted on the fruit, indicating that the fruit is coming from Napoli in Italy.

In fact, all the retrieved information from this web site are wrong. Also, the picture view is wrong: Few months ago there were kiwi trees on it, but the yard was sold and a house will soon appear. An information can be true and later the same information becomes false because an event occurs.

The map indicates that the kiwi was transported from point A to B and next from point B to C. However the web service is called and runs well, while considering the parameters (A,B,C,D) and the map of these parameters. But, the obtained result is false because somebody enters wrong information. Without stringent procedures the customer will evaluate the result as extremely bad and his opinion about the website will be "totally useless". An answer to this issue is that only authorized and trusted users must be able to upload information following their control.

Even if it is obvious that integrity must be considered to get success, its implementation is not easy when multiple VO partners are implied in a joint venture. We propose to give two meaning on integrity criterion. The first one refers to the confidence that is given to the information and the second one to the confidence that is granted to the providers of such information. Security measures must ensure that the VO partner who provides information stored in databases are clearly identified and moreover the system should have a way of ensuring that it is a reliable partner. Prior to the communication of any information, VO partners must verify the accuracy of information and validate it before it can be installed into the Geowine system. This process will ensure the truthfulness of information, and ensures that VO partners who are using or relying information (even if they are not expert in a domain) they will not be able to input involuntary fake information in the system.

Finally, as the integrity of information enforces that geo-traceability information must be retrieved by clients, we have proposed to track and store any critical action. Doing so, VO partners should not be able to deny any executed action. If an error occurs or is signaled by a client, the tracking information will be used to determine when the error occurs and which partner is involved in propagating this error in the system. The identified VO partners will correct the error and will not be able to deny their already executed actions.

## IV. Related works and theory

For many years, formal models have been proposed to deal with security constraints.

### A. Bell LaPadula model

The first well-known is the Bell-LaPadula model [4]. It deals with confidentiality constraints. It aims to give no more grants that are necessary to the one which need to access information. Subjects and objects are assigned values using a rank scale (i.e public, unclassified, confidential, restricted, secret…). The model uses a set of access control rules established on the classification of objects and clearances of subjects. The access to information is granted by comparing the clearance of a subject and the rank of an object. Classification of objects is known to value the risk concerning this object regarding disclosure. With subjects, the scale represents the level of confidence given to a subject in order to execute actions. No "read up" and no "write down" characterize this model. The drawback of this model is the lack of clear distinction between protection and security.

### B. BIBA model

The Biba security model was developed in order to address a weakness in the Bell-LaPadula model. Biba has proposed three models to address integrity considerations Low-Water-Mark, Ring et Strict Integrity. Like Bell-La Padula model, Biba uses objects and subjects. Integrity levels are defined and an order set is assigned to these levels. Biba model is a mean to evaluate integrity when a chain of subjects and objects appear. With BIBA, for example, user “UU” can only read any information whose level is upper or equal to the one that is assigned to user “UU”. A user which is classified as level “LL”, can modify any data whose level is less or equal to level “LL”. This model does not provide any mean to deal with data whose level is not defined. As the trust level is directly linked to the less level that can be highlighted, in that case the confidence level will be null.

If such a model was applied to Geowine, it would imply to define a scale in order to rank VO partners depending on their ability regarding the different operations they could execute. It seems to be hazardous to choose the level that must be assigned to a VO partner when previous experience does not exist. If the assignment is done and it is not the right one, then anyone can access wrong data if they are uploaded by this VO partner on the website. Establishing trust in digital system is not so easy.

### C. Clark and Wilson model

The Clark-Wilson (CW) model is an integrity application level model. It attempts to ensure the integrity properties of commercial data. It provides a framework for evaluating security in commercial application systems. This model respects two principles:

- The principle of separation of duty states no single person should perform a task from beginning to end, but that the task should be divided among two or more people in order to prevent fraud by one person acting alone.
- The principle of well-formed transaction is defined as a transaction where the user is unable to make arbitrarily any change or transfer regarding these data. Any action must be under control. This is the way to preserve or ensure the integrity of the data. A security system in which transactions are well-formed ensures that only legitimate actions can be executed, and also ensures that internal data is accurate and consistent to what it represents in the real world.

Clark and Wilson partitioned all data in a system into CDI (constrained data items) and UDI (unconstrained data items).

- the CDI are objects that the integrity model is applied to,
- the UDI are objects that are not covered by an integrity policy.

When data items are transmitted from one process to another one, two procedures may be applied to them:

- the first one is an Integrity Verification Procedure (IVP), that verifies the valid state or not of the data items (i.e. the verification is done to assert that users or owners are not processing violated data),
- the second procedure is the transformation procedure (TP) or well-formed transaction. A transformation procedure receives input data and transforms it as an output data. If only a transformation procedure is able to change data items, then the integrity of the data is maintained.

Integrity enforcement systems require that all transformation procedures must be logged, to provide an audit trail of data item changes. To ensure that integrity is achieved and preserved, Clark and Wilson are adding certification rules and integrity-preserving rules. The integrity-monitoring rules are enforced by the administrator and they are enforcement rules guaranteed by the system.

**Certification Rules**

- **C1 (IVP Certification)** - The system will have an IVP for validating the integrity of any CDI.
- **C2 (Validity)** - The application of a TP to any CDI must maintain the integrity of that CDI. CDIs must be certified in order to ensure that they result in a valid CDI
- **C3** - A CDI can only be changed by a TP. TPs must be certified to ensure they implement the principles of separation of duties & least privilege
- **C4 (Journal Certification)** - TPs must be certified to ensure that their actions are logged
- **C5** - TPs which act on UDIs must be certified to ensure that they result in a valid CDI

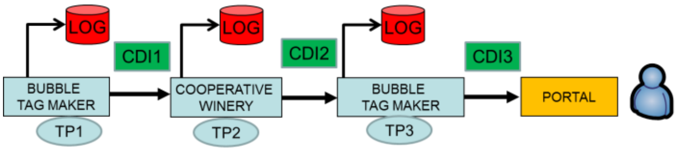
**Enforcement Rules**

- **E1** (Enforcement of Validity) - Only certified TPs can operate on CDIs
- **E2** (Enforcement of Separation of Duty) - Users must only access CDIs through TPs for which they are authorized.
- **E3** (User Identity) - The system must authenticate the identity of each user attempting to execute a TP
- **E4** (Initiation) - Only the administrator can specify TP authorizations

## V. Implementing security into Geowine

### A. Applying CK model

As already defined, the spectrum "know-how" of each VO partner is large and each one is considered as a specialist in its domain. We proposed to apply the CK model into Geowine. Each VO partner is able to clearly indicate what it is able to do, or to make. Also, each VO partner can define the transformation function depending on its own"know-how" spectrum and the constraint data item it will produce or consume during a transformation procedure. For example, the figure below represents the TP and CDI regarding the fabricated "Bubble tag" and its lifecycle.



For example, the Bubble Tag Maker make the seal, take a picture of each seal, and assigns to each seal an ID matrix. It stores this information into a database. Then, it delivers the products to the cooperative winery using postal service ( DHL , UPS…) and sends via network a file containing the associated pictures and the data-matrix . This is called the TP1 transformation procedure. Obviously, the bubble tag maker has to log this completed action into a database, in case any problem occurs it can send such information. The file that is communicated to the cooperative winery needs to be protected. If it is not, integrity cannot be assured and if a theft occurs or if changes are made by any hacker then geo-traceability is no more possible. So this data must be classified as a constraint data item (CDI). Even if the (Bubble Tag, data-matrix) has been sent, the bubble tag maker must consider that seals are not activated. In the contrary, if a seal product is stolen and stuck on bottles belonging to anybody else, the bubble tag maker will produce wrong information.

TP2 is executed by the cooperative winery. When tags are stuck on the bottles, the cooperative winery must update the received file and for control must ask the bubble tag maker to change the status from "not-activate" to "activate". This process must completely be defined by the cooperative wineryand its responsibility is committed. It must control what is to be done before sending the file back to the bubble tag maker. As the file is sent via network, this exchange must be protected. For this reason the file is classified as CDI2 before sending it back.

TP3 is executed by the bubble tag maker. When it receives CDI2, it can update its data base in order to change the tag status. When a user request sent to the web portal and if the bubble code assigned to the bottle is true, the web portal propagates this request to the bubble tag maker. The bubble tag maker must consult its data base and sends a response to the web portal. The transmitted information by the bubble tag maker is classified as CDI3 in order to be sure that integrity is ensured. The customer verifying the authenticity of a bottle will catch this information: the bottle identified by (picture, datamatrix) is authenticated, because the Bubble tag maker certifies the seal is true. It can certify the truthfulness of information because it receives from the cooperative winery a file asking to activate the code. The cooperative winery can certify the content of the bottle, because it fills by itself the bottles and stick the tag on the bottle, and also the tag was received from the bubble tag maker as authentic and registered one. Each VO partner implied in this process must be compliant with the certification rules: for example the bubble tag maker must define the process to be followed internally in order to guarantee its CDI. We propose to make these certification rules to be written either in the security policy, or as an addendum to this security policy. Moreover, the integrity is ensured using CDI and TP all along the chain. That chain constitute a dataflow and an integrity property that can be formally established using the data flow representation described in [11]. The article in [11] deals with verifying the integrity properties when data flows are crossing a network. The idea can be extended to deal while considering CDI belongs to a data flow and TP protects fields of this dataflow.

Many other scenarios describing the relation between VO partners have been designed. This scenario describes only what happens with the constraint data item (CDI). In some scenarios, some unconstraint data item will appear. For example, the cooperative winery is providing some data to the research center responsible to calculate geo-indicators values. This concerns items like soils, vine grower, vine yards. These data are considered as CDI and certification rules are set up to guarantee their use by the research data center. Some other values are provided by weather officers, or administrative officers (rainy or sunny days, vine yard location...). Such data are not CDI but UDI because there is no mean to control their value by the organization giving these information that does not belong to the Geowine consortium. So this is why they are considered as UDI. Nevertheless, if there is some possibility to get some guarantees they could be classified as CDI.

### B. Securing exchanges

Even if transformation procedures are executed under the responsibility of a single VO partner, two kinds of users may be highlighted in that the Geowine system. Some of them are human users, those implied into different actions (fill in, paste stickers and seal on bottle…). The others are given by the software agent or program. As exchanges are realized using Internet, integrity requires to implement secure services:

- First of all as CDI is transmitted from a VO partner to another one, integrity must be ensured while crossing the equipment all along the route,

- Each VO partner must not be able to deny sending CDI or UDI, to the other VO partner,
- Each entity must be able to determine which program, which person sends data. The source authentication must be provided.
- Finally logs associated to the transformation procedure must be implemented. The VO partners could give any information if anything goes wrong.

### C. Implementing a PKI

Internet is an open solution and supports secured connection using different protocols. Tunnels may be set, virtual private networks can be installed. Also, web services can be defined to support exchanges implying only partners, or to download, or to retrieve information from the common website. Whatever the choice would be, PKI is underlying. Basically, PKI implement a cryptography system and use a combination of secret key and public key. PKI enables a number of other security services including data confidentiality, data integrity, and key management. The very foundation or framework for PKI is defined in the ITU-T X.509 Recommendation [X.509]. The Internet Engineering Task Force (IETF) Public Key Infrastructure X.509 (PKIX) working group is the driving force behind setting up a formal and generic model based on X.509. It is suitable for deploying a certificate-based architecture on the Internet and thus secure transaction between parties. Among the services that can be implemented, the PKI digital certificate's cryptographic binding enables a PKI relying party such as the partner receiving a signed or encrypted data (for example CDI), to verify with assurance that this CDI was really signed and/or encrypted by its originator.

There are two ways to deploy a PKI. The first one is to be charged by a professional organization like Thawte, Verizon, Verisign, Globalsign. Such a choice implies additional cost, but seems to be a reasonable solution when doing business. Costs are from 100 euros to 1500 euros per year depending on the number of certificates to manage and the quality of service subscribed. As long as Geowine stays a proof of concept, free solutions are more suitable. We have chosen and implemented EJBCA which is an enterprise class PKI Certificate Authority built on JEE technology. The Geowine virtual organization is composed of different VO partners using heterogeneous platforms and operating systems. EJBCA [7] is a platform independent, a flexible, and a component based CA (Certificate Authority) that can be used as a stand-alone or integrated in other JEE applications.

## VI. Conclusion

VOs allow organizations to realize common projects by sharing resources using a web portal. To support the VO, we need to manage users and resources by implementing a secure IT shared infrastructure. The major difficulty was to design an architecture where the web portal must be used by customers in order to retrieve traceability information. These information need to be pertinent and truth. As they are coming from different VO partners and sometimes are the result of the execution of software programs, data need to be protected. We have applied the Clark and Wilson model i for designing the framework architecture and we have developed a web-service in order to secure data exchange between VO partners. A PKI using EJBCA has been deployed. Moreover, syslog-ng [12] was used to set up auditing even if it is not mentioned in this paper.

## REFERENCES

[1] Bultje, R., van Wick, J (1998). Taxonomy of Virtual Organisations, Based on Definitions, Characteristics and Typology. VOnet Newsletter 2(3), 1998.

[2] Laborde and al. A secure collaborative web-based environment for virtual organisations. International Journal of Web Based Communities Volume 5 Issue 2, March 2009

[3 ] http://www.prooftag.net Prooftag Trademark web site

[4] Bell D., LaPadula L., “Secure Computer Systems Mathematical Foundations”, Rapport Technique MTR-2547, Vol 1, MITRE Corporation, 1973.

[5] Biba K., “Integrity Considerations for Secure Computer Systems”, Rapport Technique MTR-3153, MITRE Corporation, 1977.

[6] Clark D., Wilson D., “A Comparison of Commercial and Military Security Policies”, IEEE Symposium on Security and Privacy, pp. 184-1994, 1987.

[7] http://www.ejbca.org

[8] SALEEM A. et al. : Using the VOM portal to manage policy within Globus Toolkit, Imperial College London, South Kensington Campus, London, 2004

[9] ALFIERI R. et al. : VOMS, an authorization system for Virtual Organizations. Proceedings of the 1st Grids Conference, Feb. 2003, Santiago de Compostela.

[10]TUPTUK N., LUPU E. : State of the art evaluation, the TrustCoM project, June 2004

[11] A formal data flow oriented model for network security policies analysis El Khoury & al ICNS 2012

[12] http://www.balabit.com/network-security/syslog-ng

# A Conceptual Framework for Securing Digital I&C Systems in Nuclear Power Plants

**Jung-Woon Lee, Jae-Gu Song, Cheol-Kwon Lee, and Dong-Young Lee**
I&C and HF Research Division, Korea Atomic Energy Research Institute,
Daejeon, The Republic of Korea

**Abstract -** *Digital technologies have been applied recently to the I&C systems of nuclear power plants. Due to this application of digital technologies, cyber security concerns are increasing in the nuclear industry. In this paper, the characteristics of I&C systems are described in terms of their differences from industrial control systems, and related nuclear regulatory requirements and other guides are introduced. Key features for cyber security including a defensive architecture, possible threats, and vulnerabilities are analyzed. Based on this analysis and an analysis of technical controls presented in the regulatory guide 5.71, a conceptual framework of technical security controls for the I&C systems is proposed, and how to achieve it is discussed.*

**Keywords:** Nuclear Power Plant, I&C system, Cyber Security, Security Controls

## 1 Introduction

The instrumentation and control (I&C) systems in nuclear power plants (NPPs) collect sensor signals of plant parameters, integrate sensor information, monitor plant performance, and generate signals to control plant devices for NPP operation and protection. Recently, digital technologies have been applied to the I&C systems in NPPs. New cyber threats have become more elaborate and are attacking industrial control systems (ICS). This makes cyber security an important issue in the nuclear industry.

Computer systems available in nuclear utilities usually include I&C systems, which consist of safety and non-safety systems, as well as on-site office systems, and off-site corporate business systems, as shown in Fig. 1. Although this paper focuses on I&C systems, on-site office systems are also considered as a boundary connected to the I&C systems. Office systems in general receive data from the plant I&C systems for administration purposes and send plant information to off-site corporate business systems through the Internet. Network isolation has been applied to I&C systems to protect from intrusions by Internet users. However, recent examples of advanced persistent threat (APT) attacks have demonstrated well that network isolation is not enough for securing nuclear power plants.

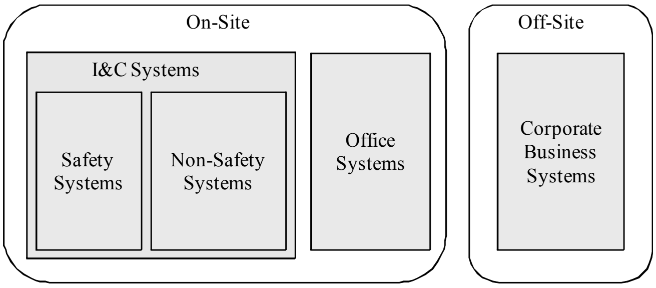


Fig. 1. Typical digital systems of NPPs

The National Institute of Standards and Technology (NIST) has published many guidance documents for the cyber security risk assessments of ICS [1~6]. In the nuclear domain, NRC regulatory guides and regulations [7,8,9], the IEEE Std. 7.4.3.2-2010 [10] and the IAEA technical guidance [11], are available for the cyber security of NPP I&C systems. Based on these documents, a preliminary cyber security assessment was performed for a digital safety system in NPPs in our previous study [12]. Although the guidance documents and nuclear regulatory guides provide the requirements of security controls, a guidance describing which security controls should be applied to which digital assets and how to implement them is still needed. Also, there have been no practical examples available for the application of security controls for NPP I&C systems.

In this paper, the characteristics of NPP I&C systems are described in terms of their differences from ICS and related nuclear regulatory requirements, and other guides are briefly introduced. Then, three key features for the cyber security of NPP I&C systems, including a defensive architecture, possible threats, and vulnerabilities, are analyzed. Based on this analysis and an analysis of technical controls presented in the regulatory guide (RG) 5.71, a conceptual framework of technical security controls for the I&C systems is proposed and how to achieve it is discussed.

## 2 Characteristics of NPP I&C Systems

Fig. 2 shows a typical configuration of an NPP digital I&C system. At the lowest level, sensors and actuators are located to send or receive signals from the devices at higher levels. At the level next to the sensor and actuator level, there are plant protection and control systems that collect sensor signals, evaluate them logically, and send information to information processing systems or a human-machine interface (HMI) located at levels higher than the plant protection and control systems. In some cases, plant control systems send control signals to actuators directly. Monitoring systems located at the level above the plant protection and control system level receive information from the plant protection and control systems, and process the information to send it to the NPP control room operators via HMI in the main control room and remote shutdown facility. Control signals for the actuators of plant equipment and components can be generated by the operators or by the plant protection and control systems. Communication networks are generally used to transfer information among the systems at different levels.

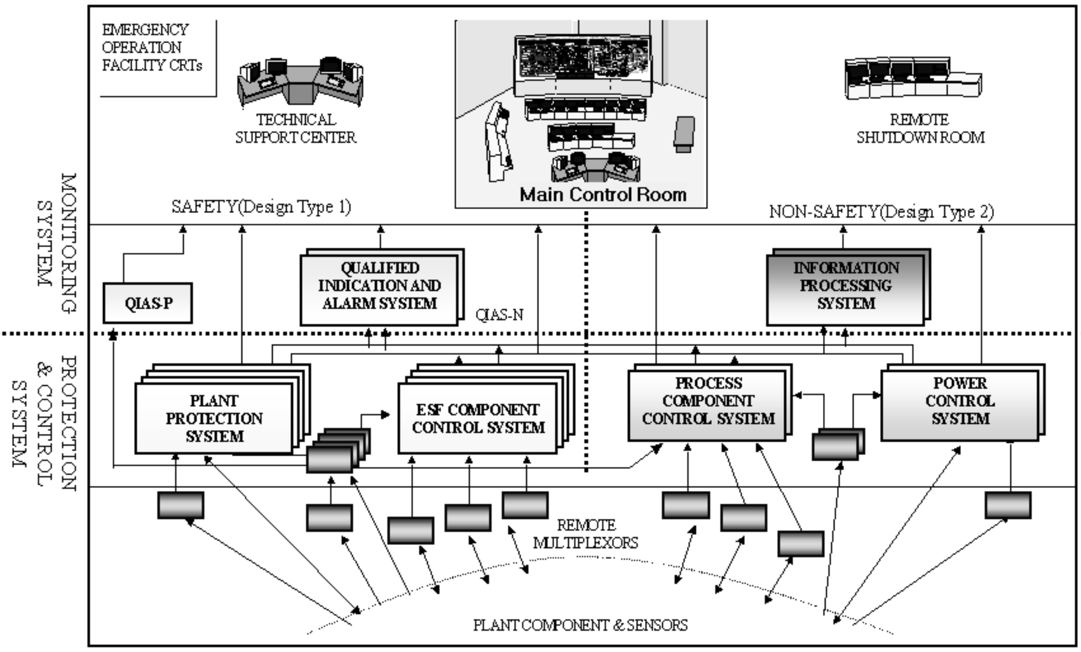

EMERGENCY OPERATION FACILITY CRTs
TECHNICAL SUPPORT CENTER
Main Control Room
REMOTE SHUTDOWN ROOM
MONITORING SYSTEM
SAFETY(Design Type 1)
NON-SAFETY(Design Type 2)
QIAS-P
QUALIFIED INDICATION AND ALARM SYSTEM
QIAS-N
INFORMATION PROCESSING SYSTEM
PROTECTION & CONTROL SYSTEM
PLANT PROTECTION SYSTEM
ESF COMPONENT CONTROL SYSTEM
PROCESS COMPONENT CONTROL SYSTEM
POWER CONTROL SYSTEM
REMOTE MULTIPLEXORS
PLANT COMPONENT & SENSORS

Fig. 2. Typical configuration of I&C system in NPPs

The I&C systems in NPPs can be grouped into two categories: safety systems and non-safety systems. In some regulatory requirements, safety systems can be graded again as either safety-critical or important-to-safety. The safety systems are placed on the left side of Fig. 1, and the non-safety systems on the right side. The safety systems shutdown the reactor safely and maintain it in a safe shutdown condition. The non-safety systems are related to power generation. Except for the safety systems, the NPP I&C systems have a similar structure and constituents to those of the ICS. The safety systems require higher reliability, functionality, and availability than the non-safety systems. Hardware for the safety systems should have redundancy. Failures in the non-safety systems should not cause a loss of safety function, in other words, any signal traffic from the non-safety systems to the safety systems is not allowed. Software for the safety systems should be qualified through rigorous verification and validation processes.

# 3 Nuclear regulatory requirements and other guidance documents

As cyber security has been an emerging concern in the nuclear industry, the U.S. NRC issued the regulatory guide (RG) 1.152 revision 2, "Criteria for Use of Computers in Safety Systems of Nuclear Power Plants," in 2006 [7]. This regulatory guide addresses cyber security for the use of digital computers in the safety systems of NPPs. The IEEE Standard 7-4.3.2-2010 [10] was issued as a revision of the previous version, in which cyber security requirements with a lifecycle approach were newly supplemented. The RG 1.152 revision 2 and the IEEE Std. 7-4.3.2-2010 require that the digital safety system development process should address potential security vulnerabilities in each phase of the digital safety system lifecycle and also system security features should be developed appropriately according to the lifecycle process.

In 2009, 10 CFR 73.54, “Protection of Digital Computer and Communication Systems and Networks,” requires NPP licensees in U. S. to submit a cyber security plan for protecting critical digital assets (CDAs) associated with the following categories of functions from cyber attacks: 1) safety-related and important-to-safety functions, 2) security functions, 3) emergency preparedness functions, including offsite communications, and 4) support systems and equipment which, if compromised, would adversely impact safety, security, or emergency preparedness functions [8]. The RG 5.71 [9] was issued in 2010 for applicants and licensees to comply with the requirements of 10 CFR 73.54. The RG 5.71 provides a framework to aid in the identification of CDAs categorized in 10 CFR 73.54 and the application of defensive architecture and security controls for the protection of CDAs from cyber threats.

The IAEA technical guidance [11] presents guides for the management (Part I) and implementation (Part II) of computer security at nuclear facilities. In Part I, regulatory and management considerations, management systems, and organizational issues are discussed and in Part II implementing computer security, threats/vulnerabilities and risk management, and special considerations for nuclear facilities are described. Notably in Annex III to this document, common human errors during cyber security process are listed.

The NIST has published numerous documents related to cyber security. Among these, NIST Special Publication (SP) SP 800-82 [6] contains valuable information throughout the cyber security program of NPPs. NIST SP 800-30 [1], SP 800-37 [2], and SP 800-39 [3] are helpful for cyber security risk assessments, and SP 800-53 [4] and SP 800-53A [5] provide the detail implementation guides of security controls.

# 4 Key features for cyber security

In our previous study [12], analysis activities and considerations necessary for conducting the cyber security risk assessments of NPP I&C systems are examined for the system design phase and the component design and equipment supply phase in the development of the systems. The assessment process used in the study consists of the following 6 steps:

1) System Identification and Cyber Security Modeling,
2) Asset and Impact Analysis,
3) Threat Analysis,
4) Vulnerability Analysis,
5) Security Control Design, and
6) Penetration test.

This process was applied to our assessment of a sample NPP digital safety system. Based on our experience from the assessment, three key features; defensive architecture, threats, and vulnerabilities are analyzed in this section to establish a basis for devising a framework of security controls.

## 4.1 Defensive architecture

A defense-in-depth strategy should be applied and maintained in I&C systems to effectively protect CDAs from cyber attacks. For this purpose, the security levels should be defined and an appropriate security level should be assigned to each CDA.

NIST SP 800-82 [6] recommends a defense-in-depth strategy including the use of firewalls, the creation of demilitarized zones, intrusion detection capabilities, and other managerial security programs. The IAEA technical guidance [11] recommends a graded approach in which computer systems are grouped into zones and a security level is assigned to each zone. It uses five security levels and defines the graded protective requirements. NEI 04-04 Revision 1 [18] presents a defensive strategy with five levels: level 4, control system network; level 3, data acquisition network; level 2, site local area network; level 1, corporate WAN; and level 0, the Internet. RG 5.71 [9] also requires employing defense-in-depth strategies to protect CDAs from cyber attacks and suggests a defensive architecture configured with five concentric cyber security defensive levels. Systems requiring the greatest degree of security are located within a greater number of boundaries. Fig. 5 shows this defensive architecture in RG 5.71.

The cyber security defensive architecture presented in RG 5.71 was used as a reference in this study. It is assumed that the assets or systems at security levels 1 and 0 may correspond to on-site office systems or external corporate business systems, and security levels 2 through 4 correspond to the I&C systems in Fig. 1. When determining the security levels for digital assets, their direct relationship with safety function and the impacts of a loss of confidentiality, integrity, and availability caused by cyber threats on the plant safety or plant trips are important factors to be analyzed. The security levels for NPP I&C systems are redefined in our study by considering the above factors and are described as follows:

1) Security level 4: This level contains CDAs associated with safety and those important to plant trips. The CDAs at this level should be protected from malfunctions of devices at the lower levels. Only a one-way data flow is allowed from Level 4 to Level 3. Redundant security controls or mitigation measures regarding vulnerabilities should be applied.
2) Security level 3: This level contains the assets or systems that do not impact the safety directly, but may cause the plant trips or are connected to other systems through a network. The assets or systems at this level should not receive any data from the devices at security level 2. Security controls or mitigation measures regarding vulnerabilities should be applied.
3) Security level 2: This level contains independent assets or systems that do not impact plant safety or trips and are not connected to any network. Security controls or mitigation measures regarding vulnerabilities may be applied in consideration of the impact of cyber threats to an asset or system itself.

With this definition, most CDAs of the safety system are at level 4, and some parts of the safety system related to the monitoring function can be assigned to level 3. Non-safety systems can be assigned to levels 3 or 4, and some stand-alone systems may be placed at level 2.

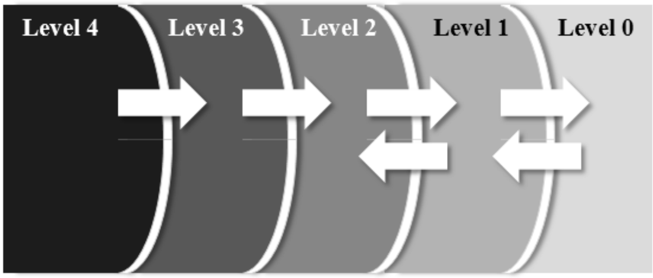


Fig. 3. Simplified cyber security defensive architecture (redrawn from RG 5.71 [9])

## 4.2 Possible threats

.NPP I&C systems generally use closed data and communication networks or air-gaps such that access through the Internet to the systems becomes difficult. However, recent cases of APT attacks demonstrate that NPP I&C systems may also be infected by malware enabling cyber attacks through portable devices such as notebooks and USB drives. Hence, it is important to identify all the connection points between humans with external electronic devices and the I&C systems, and to analyze potential security breaches that can be exploited by cyber threats. These connection points are usually related to the plant maintenance and test tasks.

## 4.3 Vulnerabilities

Use at least 2 centimeters (0.75 inch) for the left and right margins. Leave a 0.6 centimeters (0.25 inch) space between the two columns in the center of the page. Use font size (character size) 10 for text. The text should be prepared with single line spacing. *Do not use bold in the main text. If you want to emphasize specific parts of the main text, use italics.* Leave at least 2.0-2.5 centimeters margin at the page head (top of each page) for placing final page numbers and headers (final page numbers and running heads will be inserted by the publisher). Select a standard size paper such as A4 (210 X 297 mm) or letter (8.5 X 11 in) when preparing your manuscript.

The North American Electric Reliability Council (NERC) listed the top 10 vulnerabilities of control systems and recommended mitigation strategies [4]. The top 10 vulnerabilities are as follows;

1) Inadequate policies, procedures, and culture that govern control system security,
2) Inadequately designed control system networks that lack sufficient defense-in-depth mechanisms,
3) Remote access to the control system without appropriate access control,
4) System administration mechanisms and software used in control systems are not adequately scrutinized or maintained,
5) Use of inadequately secured wireless communication for control,
6) Use of a non-dedicated communications channel for command and control and/or inappropriate use of control system network bandwidth for non-control purposes,
7) Insufficient application of tools to detect and report on anomalous or inappropriate activity,
8) Unauthorized or inappropriate applications or devices on control system networks,
9) Control systems command and control data not authenticated, and
10) Inadequately managed, designed, or implemented critical support infrastructure.

These vulnerabilities contain both managerial and technical vulnerabilities. Among these vulnerabilities, items 1), 2), 7), and 9) may exist in NPP I&C systems, but other items are less related.

In NIST SP 800-82 [6], numerous vulnerabilities are listed in ICS in various categories. These vulnerabilities are evaluated in this study by considering their relevance to NPP I&C systems. Table 1 lists the vulnerabilities selected from this evaluation.

Table 1. Possible vulnerabilities in the I&C systems selected from NIST SP 800-82 [6]

| Category | Vulnerability |
|---|---|
| Policy and Procedure | Inadequate security policy for the ICS |
| | No formal ICS security training and awareness program |
| | No specific or documented security procedures were developed from the security policy for the ICS |
| | Lack of administrative mechanisms for security enforcement |
| | Few or no security audits on the ICS |
| Platform Configuration | OS and application security patches are not maintained |
| | Data unprotected on portable device |
| | Lack of adequate password policy |
| | Inadequate access controls applied |
| Platform Hardware | Unauthorized personnel have physical access to equipment |
| Platform Software | Buffer overflow |
| | Denial of service (DoS) |
| | Use of insecure industry-wide ICS protocols |
| | Use of clear text |
| | Inadequate authentication and access control for configuration and programming software |
| | Intrusion detection/prevention software not installed |
| | Incidents are not detected |
| Platform Malware Protection | Malware protection software not installed |
| Network Configuration | Weak network security architecture |
| | Data flow controls not employed |
| | Inadequate access controls applied |
| Network Hardware | Inadequate physical protection of network equipment |
| | Unsecured physical ports |
| | Non-critical personnel have access to equipment and network connections |
| Network Perimeter | No security perimeter defined |
| | Firewalls nonexistent or improperly configured |
| Network Monitoring and Logging | Inadequate firewall and router logs |
| | No security monitoring on the ICS network |
| Communication | Standard, well-documented communication protocols are used in plain text |
| | Authentication of users, data or devices is substandard or nonexistent |

As in our previous study [12], vulnerabilities in the sample safety system were identified, and the measures for mitigating these vulnerabilities were devised. Table 2 shows the vulnerabilities and mitigation measures.

Table 2. Vulnerabilities and mitigation measures for a sample digital system obtained from our previous study [12]

| Vulnerability | Mitigation Measure |
|---|---|
| DoS attacks and malware execution on other assests communicating with the assest infected during the maintenance works | Establishment of security managing and infection detection systems for PC, USB, and external storage media used for the maintenance works |
| System shut-down by malware infected during the maintenance works | Establishment of device authentication policies |
| Data modification by malware infected during the maintenance works | Monitoring of running services: creating a white list by checking running processes, and detection and blocking of unnecessary services |
| Seizure of system authorities due to vulnerabilities residing in the OS | Network monitoring |
| DoS attacks and malware execution on other systems by vulnerabilities residing in the system | Firewalls/Intrusion Prevention System(IPS)/Intrusion Detection System(IDS) |
| Eavesdropping, data forgery, and attacks by malware | Data encryption |
| Data modification by using known vulnerabilities of standard communication protocols | Vulnerability patches |

## 5 Conceptual framework of technical security controls

Security controls can eliminate or mitigate the vulnerabilities identified for the system or prevent the system from possible cyber threats. Appendix B (Technical Security Controls) and Appendix C (Operational and Management Security Controls) to RG 5.71 [9] provide a comprehensive set of security controls. These controls are developed by incorporating the selected controls from NIST SP 800-53 [4], NIST SP 800-82 [6], and other DHS ICS security guidances. When analyzing the controls in the RG 5.71, it can be found that some security control requirements in Appendix B to RG 5.71 cannot be implemented technically, but should rather be handled as operational and management controls, and in contrast to this, some controls in Appendix C to RG 5.71 contain requirements that should be implemented in the system design.

Based on an analysis of the defensive architecture, possible threats and vulnerabilities in section 4, and the analysis of the security controls described in the RG 5.71, candidate technical security controls that would be implemented for securing the I&C systems are obtained. The candidate controls include data traffic control, access controls for external devices, monitoring and logging, intrusion detection systems (IDS), intrusion prevention systems (IPS), and data encryption. Fig. 4 illustrates the relations of the control elements with the system.

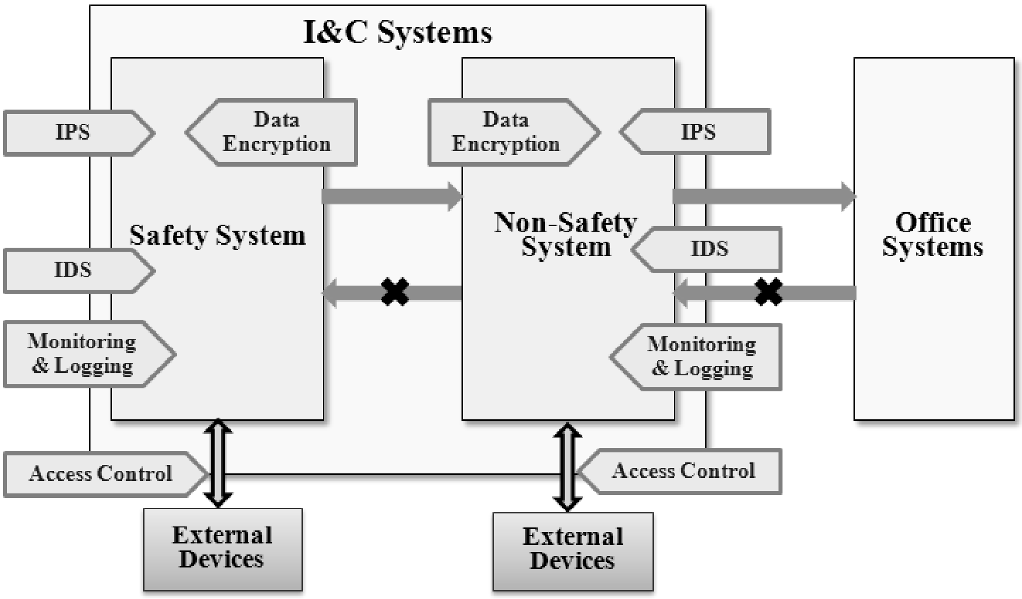


Fig. 4. Conceptual framework of security controls

### 5.1 Data traffic control

In general, security levels become higher in the order of in-site office systems, non-safety systems, and safety systems. Data flow from systems with lower security levels to systems with higher levels should not be allowed. It is required by nuclear regulations that data traffic from non-safety systems to safety system is not allowed. Similarly, data traffic from office systems to non-safety systems should not be allowed either, although this is not required by regulations. Since non-safety systems take charge of plant control, plant trips or device damage may occur if any of the non-safety systems are affected by cyber attacks passing through the office systems. It will be better to apply the same one-way data traffic between office systems and off-site corporate business systems to reduce the chances of cyber attacks through the Internet.

It is possible that different security levels are assigned to the assets inside the safety systems or non-safety systems. In this case, the same data traffic control scheme can be applied between the higher security level assets and lower level assets.

### 5.2 Access controls for external devices

External devices are usually used during the maintenance or tests for I&C systems in NPPs. External devices may include notebooks, USB drives, portable electronic devices, external media, etc. Malware can infiltrate a CDA when using these devices. In this way, these external devices can provide a cyber attack path. It is evidently important to apply access controls to interfaces between the system and external devices. However, deep consideration is required when applying the access controls.

There may be many humans involved in the maintenance and test tasks, such as plant personnel and subcontract workers who perform the tasks and security control administrators. As stated before that humans are the weakest link in cyber security [15,16,17], human errors or violations in the security process may take place during the tasks. For this reason, when selecting or designing the access controls, it will be necessary to analyze carefully the tasks, procedures, possible human errors, the possibilities of attempting shortcuts to bypass inconvenient security controls, etc. Human factors engineering specialists, I&C specialists, and plant maintenance and test personnel, as well as IT security specialists, should form a team to perform this analysis, and then select and apply effective access controls. Education and training on cyber security for plant personnel is also important to maintain security. The following quotation from the I3P report [18] helps us understand this matter well:

> "Information security depends not only on technology, but also on the awareness, knowledge, and intentions of the employees, customers, and others using information-based systems and networks."

### 5.3 Monitoring, logging, IDS, IPS, and data encryption

Logging and monitoring are essential tools for security audits and an analysis of abnormal system behavior induced by malicious activities. IDS aims to detect possible malicious activities inside the system. IPS, in addition to IDS, functions to take actions to prevent or stop activities identified as malicious. Data encryption enhances the secure management of a data flow within a system.

Most of present digital I&C systems do not include any of these functions, even logging and monitoring functions for cyber security purposes. Although cyber security control devices or software may exist on the market, they cannot be applied unless it is verified that the inclusion of these security features shall not cause any adverse impacts on the I&C systems in NPPs. While regulatory requirements specify this matter only for the safety systems, NPP utilities may require the same for the non-safety systems. Another point to keep in mind in the application of security controls in the safety systems is the fact that the implementation of controls will require the same degree of qualification efforts as the safety system itself if the controls are embedded into the safety system or give any signals to the safety system. For this reason, many of security devices or software based on IT security technologies may not be applicable to NPP I&C systems, and even worse, for IPS and data encryption, which may interfere with I&C functions in certain ways.

For all cases of the development of new security controls and the direct application or modifications of existing IT security controls, a test-bed emulating NPP I&C systems should be developed first. This test-bed will be used to verify that the security function works as intended, and that the inclusion of security controls does not cause any adverse impacts on NPP I&C systems. If this test is performed on the real systems installed in NPPs, the test may induce damage to the systems. The development of these security controls, together with establishing a test-bed, requires long term researches. Hence, the development and application of a data traffic control mechanism in data communications and access controls for plant maintenance and tests with external devices can be considered as immediate measures for securing NPP I&C systems.

## 6 Conclusions

Cyber security has become an important issue in the nuclear industry. Based on an analysis of the key features for cyber security, this paper proposes a conceptual framework of technical security controls for securing the I&C systems in NPPs. Data traffic control, access controls for external devices, monitoring and logging, IDS, IPS, and data encryption are suggested as candidates for security controls. Many topics that should be considered with caution were discussed for the development and application of these security controls. Conclusively, the application of security controls that are appropriate for securing NPP I&C systems requires long term researches. It is recommended to develop and apply 1) data traffic control mechanisms among CDAs or systems at different security levels, and 2) access controls during the maintenance and test tasks with external electronic devices as immediate measures for securing NPP I&C systems.

## 7 Acknowledgement

This work was supported by the nuclear Technology Development Program of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea government Ministry of Knowledge Economy (No. 2010161010001E).

## 8 References

[1] NIST Special Publication 800-30, Risk Management Guide for Information Technology Systems, July 2002.

[2] NIST Special Publication 800-37 Revision 1, Guide for Applying the Risk Management Framework to Federal Information Systems, February 2010.

[3] NIST Special Publication 800-39, Managing Information Security Risk, March 2011.

[4] NIST Special Publication 800-53 Revision 3, Recommended Security Controls for Federal Information Systems, August 2009.

[5] NIST Special Publication 800-53A Revision 1, Guide for Assessing the Security Controls in Federal Information Systems, 2010.

[6] NIST Special Publication 800-82, Guide to Industrial Control Systems (ICS) Security, June 2011.

[7] Regulatory Guide 1.152 revision 2, Criteria for Use of Computers in Safety Systems of Nuclear Power Plants, U.S. Nuclear Regulatory Commission, January 2006.

[8] 10 CFR Part 73.54, Protection of Digital Computer and Communication Systems and Networks, U.S. Nuclear Regulatory Commission, Washington, DC., 2009.

[9] Regulatory Guide 5.71, Cyber Security Programs for Nuclear Facilities, U.S. Nuclear Regulatory Commission, January 2010.

[10] IEEE Standard 7-4.3.2-2010, Standard Criteria for Digital Computers in Safety Systems of Nuclear Power Generating Stations, August 2, 2010.

[11] IAEA Nuclear Security Series No. 17, Technical Guidance, Computer Security at Nuclear Facilities, IAEA, Vienna, 2011.

[12] Jae-Gu Song, Jung-Woon Lee, Cheol-Kwon Lee , Kee-Choon Kwon, and Dong-Young Lee, A Cyber Security Risk Assessment for the Design of I&C Systems in Nuclear Power Plants, Nuclear Engineering and Technology, Korean Nuclear Society (accepted for publication on March 13, 2012)

[13] NEI 04-04 Revision 1, Cyber Security Program for Power Reactors, Nuclear Energy Institute, November 18, 2005.

[14] Top 10 Vulnerabilities of Control Systems and Their Associated Mitigations - 2007, North American Electric Reliability Council, December 7, 2006.
http://www.us-cert.gov/control_systems/pdf/2007_Top_10_Formatted_12-07-06.pdf

[15] News Article, "Human error biggest threat to computer security," by Rene Millman, 16 March, 2012.
http://www.itpro.co.uk/115920/human-error-biggest-threat-to-computer-security

[16] News Article, "Human Error Considered Primary Cause of Network Security," By: Marketwire, May. 18, 2011.
http://www.sys-con.com/node/1839156

[17] Sara Kraemer and Pascale Carayon, Human errors and violations in computer and information security: The viewpoint of network administrators and security specialists, pp143–154, Applied Ergonomics, Vol. 38, 2007.

[18] National Cyber Security: Research and Development Challenges: Related to Economics, Physical Infrastructure and Human Behavior, Institute for Information Infrastructure Protection (I3P), 2009.
http://www.thei3p.org/docs/publications/i3pnationalcybersecurity.pdf

# Sterilization of Stego-images through Histogram Normalization

**Goutam Paul**[1] **and Imon Mukherjee**[2]
[1]Dept. of Computer Science & Engineering,
Jadavpur University, Kolkata 700 032, India.
Email: goutam.paul@ieee.org
[2]Dept. of Computer Science & Engineering,
St. Thomas' College of Engineering & Technology, Kolkata 700 023, India.
Email: mukherjee.imon@gmail.com

**Abstract**—*Steganalysis is very popular in order to defeat the steganographic algorithm. But sometimes it is not possible to detect the hidden data without having any prior knowledge of the embedding algorithm or the knowledge of the key used. That is why image sterilization may play an important role in the field of secret communication to remove any steganographic information embedded in an image. In this paper, we propose a novel technique of image sterilization based on histogram normalization. The technique is general in the sense that it works for any LSB-based steganography algorithm and it does not need to know how the algorithm embeds information inside the image. We ran simulations over stego images created by different state-of-the-art algorithms and on average the technique succeeded in sterilizing around 77% to 91% of stego pixels depending on the algorithm targeted.*

**Keywords:** Information Hiding, Security, Histogram, Steganalysis, Steganography, Sterilization.

## 1. Introduction and Motivation

The word *Steganography* [11] originated from the Greek word “Steganos”, meaning “covered information”. It is the technique of hiding messages inside inoccuous media so that the hidden message cannot be detected by an adversary having access to the media. In [7], the history of steganography is traced from ancient Greece up to the modern times. It is reported that the Germans have successfully utilized this technology during the Second World War. Academic research in steganography has grown tremendously in last few decades and currently many steganographic algorithms exist in the literature.

*Steganalysis* [15], [2] is the art and science to defeat steganography. Steganalysis is performed without the prior knowledge of the steganographic algorithm or the secret key used for embedding the information into the cover media. This is why determining whether the secret message exists in the media is a difficult and challengable task.

Steganography can be applied to a variety of multimedia contents like images, audio, video, text etc. The media before embedding any secret information is called *cover* and after inserting the information is called *stego*. In this paper, we focus on steganography in digital image which is a very popular cover media for steganography.

Here we are proposing a new approach to sterilize the hidden information from the stego media based on image histogram. Our objective in this work is to develop an algorithm to revert as many stego pixels of an image as possible to their original cover form, which we call *image sterilization*. Image sterilization may have an important application in defense and security domain. For example, suppose that a spy wants to inform his team about the venue of performing a bomb blast in some target place using some image based steganographic technique. During the time of transmission, if sterilization of the stego information is performed by the security personals, then the attackers would be completely unaware about the venue and their plan may be jeopardized. One obvious method of performing sterilization would be to replace the least aignificant bits (LSBs) of all the pixel intensities by zero (or one). But this immediately gives a clue to the recipient that the image might has been modified by an adversary. Our algorithm does not leave any such signature and preserves the pseudo-randomness of the sequence of the LSBs in an image.

## 2. Steganographic Techniques in Spatial Domain

Before going into the details of our sterilization technique, we briefly describe here the major categories of steganographic algorithms in the spatial domain. Different algorithms operating in the spatial domain can be considered as different methods for selecting the pixel positions. These are classified into three categories: non-filtering algorithms, randomized algorithms and filtering algorithms [8].

### 2.1 Non-filtering Algorithm

The non-filtering steganographic algorithm [8] is the most popular and the most vulnerable steganographic technique based on LSB. The embedding process is done as a sequential substitution of each LSB of the pixel for each bit of the message. Hence a large amount of information can be stored into the cover media.

The only requirement in this method is sequential LSB reading, starting from the first pixel in order to extract the secret message from the cover media (viz. image). As the message is embedded in the initial pixels of the image, leaving the remaining pixels unchanged, this technique gives an unbalanced distribution of the changed pixels.

### 2.2 Randomized Algorithm

This technique solves the limitation of the previous technique. Each of the sender and the receiver has a password denominated stego key which is generated through a pseudo-random number generator [8]. This creates an index sequence to denote the pixel positions. The message bit is embedded in the pixel of the cover media following the index sequence produced by the pseudo-random number generator.

The two main features of this technique are: **(i)** use of password to have access to the message and **(ii)** well-spread message bits over the image, that are difficult to detect compared to the previous case.

### 2.3 Filtering Algorithm

The filtering algorithm [8] filters the cover image by using a default filter and hides information in those areas that get a better rate. The filter is used to the most significant bits of every pixel, leaving the less significant bits to hide information. The filter gives the guarantee of a greater difficulty of detecting the presence of hidden messages. The retrieval of information is ensured because the bits used for filtering are not changed.

Each of the aforesaid three categories of steganographic techniques in spatial domain is susceptible to image sterilization described in subsequent sections.

## 3. Image Sterilization

A 24-bit color image [3] consists of a number of pixels and each pixel contains three intensity values (of 8 bits each), one for each of red, green and blue color components.

One of the most popular steganography techniques is the Least Significant Bit (LSB) insertion [13]. Since changing the LSB changes the pixel intensity value by at most 1 (see Table 1), if one changes the LSBs of some pixels, the resulting picture would be visually indistinguishable from the original image.

| 128 | 64 | 32 | 16 | 8 | 4 | 2 | 1 |
|---|---|---|---|---|---|---|---|
| ⇓ | | | | | | | ⇓ |
| MSB | | | | | | | LSB |

Table 1: Weights of different bit positions of the pixel intensity value.

Histogram of a digital image in spatial domain can be defined as:

$$H_{r_k} = n_k \tag{1}$$

where, $r_k$ is the $k^{th}$ intensity value and, $n_k$ is the number of pixels with intensity $r_k$ $[0 \leq r_k \leq 255]$. Now, consider Fig. 1 that shows the sample histogram of a portion of a stego image.

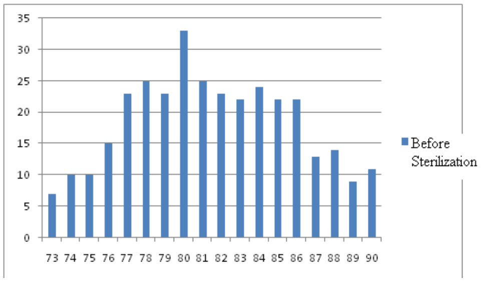


Fig. 1: A sample histogram of a portion of a stego image

We can normalize the histogram in order to remove the hidden information from stego-images based on following concept.

Table 2: LSB embedding in a pixel.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 48 | 48 | 49 | 48 | 48 | 77 | 76 | 76 | 77 |
| 48 | 49 | 49 | 49 | 48 | 48 | 48 | 77 | 77 |
| 48 | 48 | 77 | 77 | 76 | 77 | 76 | 76 | 77 |
| 48 | 76 | 76 | 76 | 77 | 77 | 77 | 48 | 49 |
| 48 | 77 | 77 | 76 | 77 | 77 | 76 | 48 | 49 |
| 48 | 48 | 49 | 77 | 77 | 77 | 77 | 49 | 49 |
| 49 | 49 | 48 | 77 | 77 | 77 | 77 | 49 | 48 |
| 49 | 49 | 49 | 49 | 77 | 77 | 77 | 77 | 49 |

Consider Table 2 that shows some sample intensity values of one component (either blue, red or green) of an arbitrary image. There are two groups - one shown in red , another in blue. Suppose that the intensity of a pixel in the original image is 48. After stego insertion, the value may either remain as 48 (if 0 is inserted) or be changed to 49 (if 1 is inserted), as shown below in Table 3.

Table 3: Embedding message bit 1 into the LSB of a pixel with intensity value 48.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 48 | = | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | |
| | | | | | | | | | ⇓ | |
| | | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | = 49 |

The LSB flipping function [4] for a stego image is defined by $F_1 = 0 \leftrightarrow 1$, 2↔3, 4↔5, . . . , etc. We form groups of $r_k$ values based on this flipping function, where the intensity values $2j$ and $2j + 1$, for $0 \leq j \leq 127$, belong to the same group. So the maximum possible number of groups for each component of an image is 128. Suppose that an image contains $N$ pixels with $c$ groups. Let $n_i$ be the number of pixels in the $i^{th}$ group, $1 \leq i \leq c$. Thus, $N = \sum n_i$. The set of pixels (based on their intensity values) for the $i^{th}$ group is represented by

$$G_i = \{x_{i,k} : 1 \leq k \leq n_i\},$$

such that

$$x_{i,j} - x_{i,m} \epsilon \{-1, 0, +1\}, 1 \leq j \neq m \leq n_i. \qquad (2)$$

Fig. 2 shows the normalized version of histogram shown in Fig. 1 using Algorithm 1.

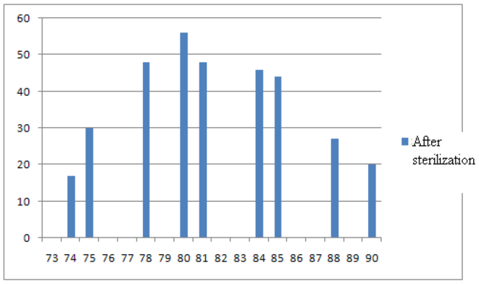


Fig. 2: Normalized version of the histogram shown in Fig. 1

In Algorithm 1, we present our procedure for image sterilization, called *StegoSterilize*. The basic idea behind our image sterilization technique is to replace an intensity value X with some other value Y such that one cannot extract the hidden message from the cover media. This technique can be applied to both 24-bit color images as well as 8-bit gray-scale images. Each group in the histogram contains at most two intensity values, of the form $2j$ (we call them *even* pixels) and $2j + 1$ (we call them *odd* pixels). Let $n_o$ and $n_e$ be the number of even and odd pixels in a group. If $n_e \geq n_o$, we replace all $2j + 1$ intensity values by $2j$, otherwise we do the opposite replacement. In other words, we force all pixels in a group to be either odd or even depending on the majority of the pixels being odd or even.

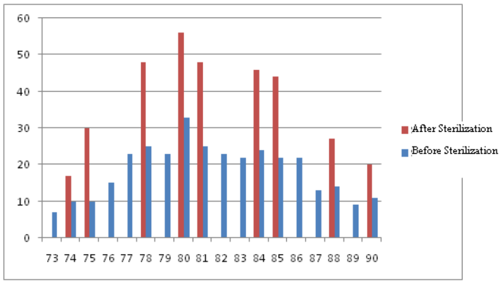


Fig. 3: Comparative representation of Histogram (before and after sterilization of the same sample)

```
Input: A stego image.
Output: The sterilized version of the input
        stego image.

Read the intensity values from the the stego
image;
Draw the histogram of the stego image;
for each color component do
    Form the groups based on Equation (2);
    for each group do
        Count the odd and even pixels with
        intensity values of the form $2j + 1$
        and $2j$ respectively from the
        histogram of the stego image; Let $n_o$
        and $n_e$ be the respective counts;
        if $n_e \geq n_o$ then
            Replace all $2j + 1$ intensity
            values by $2j$;
        end
        else
            Replace all $2j$ intensity values by
            $2j + 1$;
        end
    end
end
Output the transformed image;
```

**Algorithm 1:** StegoSterilize

Fig. 3 shows the changes in the histogram (blue color for before sterilization and red color for after sterilization)

## 4. Accuracy Measurement

To estimate the accuracy of our technique, we need to take as inputs some sample stego images for which we know which pixel values are actually changed due to the LSB embedding. For any stego image $I$, let

$S$ := the number of stego pixels,
$S'$ := the number of stego pixels (out of $S$) having different intensity value from their cover counterpart,
$S''$ := the number of recovered stego pixels (out of the $S'$) due to the sterilization process.

The **accuracy of sterilization** for this image is defined as :

$$Acc_{steri}(I) = \frac{S''}{S'} \qquad (3)$$

We have used a database of 200 (24-bit) color images in BMP format and 100 gray-scale images (freely available from [14] and many other internet sources). We have also prepared 50 different text files containing the story of

Sherlock home's (downloaded from[1]). Each pixel of a 24-bit color image contains three components, viz. red, green and blue. So using LSB embedding, at the most three bits of data can be embedded in a pixel. If the dimension of an image is $m \times n$, then maximum number of data bits possible to be inserted in the 24 bit color image can be $m \times n \times 3$. Since a character is of eight bits, the maximum number of characters (including white spaces) of the text would be $\lfloor \frac{m \times n \times 3}{8} \rfloor$. Thus, each of the 50 text files consists of at most $\lfloor \frac{m \times n \times 3}{8} \rfloor$ characters depending upon the values of the dimensions $m$ and $n$. Similarly for gray-scale image, each text file should have $\lfloor \frac{m \times n}{8} \rfloor$ characters. We have used MATLAB 7.7.0 as a software tool for implementation.

Fig. 4 shows a stego (on the left) and the corresponding sterilized (on the right) versions of the famous image of Cameraman in gray-scale. Similarly, Fig. 5 shows the stego and sterilized versions of a 24-bit color image on the left and right respectively.

In Table 4, we give an example of how the text extracted from a sterilized image may differ from the original text that was embedded and is expected to be extracted from the unsterilized stego counter part. We observe that after sterilization, the actual message is scrambled enough so that it cannot be reliably recovered.

Table 4: Sample text embedded in an image before and after sterilization

I was the means of introducing to his notice
that of Mr Hatherleys thumb and that of
Colonel Warburtons madness.

Embeded Messege (taken from [12]) before Sterilization

⇓

J!ybtsgf!ofbmtÃşnmssnevcjnhp!gjtÃşmnshb
dsgbtÃşnhÃşNs!GbsgfqmfytÃşsgvnb!bme!sh?
uÃşnfÃşBnmpmfm!X?savqsnmtÃşnbdmftr

Embeded Messege (using algorithm 1) after Sterilization

Table 5 shows the ability to sterilize the stego image using the algorithm described in the previous section for three different embedding techniques. The first algorithm is the naive LSB based sequential embedding technique; we call it algorithm A. Another technique is taken from [10], which we refer as algorithm B. The third method, denoted by algorithm C, uses random pixel selection and segmentation mechanisms [12].

# 5. Other Performance Parameters

In addition to the accuracy, we compute some other performance measures as explained below.

## 5.1 Mean Squared Error (MSE) and Peak Signal to Noise Ratio (PSNR)

The imperceptibility of hidden information in a stego image is measured by the image quality in terms of Mean Squared Error (MSE) and Peak-Signal-to-Noise Ratio (PSNR) in dB [9], [16]. Consider a discrete image $A(m, n)$, for $m = 1, 2, 3, \ldots, M$ and $n = 1, 2, 3, \ldots, N$, which is treated as a reference image. Consider a second image $B(m, n)$, of the same spatial dimension as $A(m, n)$, that is to be compared to the reference image.

Table 5: Accuracy(minimum, maximum, average and standard deviation) of sterilization over hundred gray-scale and two hundred 24-bit color images for three different algorithms A,B,C.

| | | Grey scale | 24 bit color image | | |
|---|---|---|---|---|---|
| | | | R | G | B |
| Minimum % | A | 72.50 | 68.01 | 68.9 | 69.75 |
| | B | 79.27 | 74.41 | 76.2 | 75.57 |
| | C | N.A | 84.29 | 84.36 | 85.89 |
| Average% | **A** | **78.09** | **77.16** | **76.64** | **78.34** |
| | **B** | **79.31** | **81.20** | **80.35** | **81.68** |
| | **C** | **N.A** | **91.15** | **89.28** | **91.43** |
| Maximum % | A | 87.74 | 90.85 | 91.01 | 90.02 |
| | B | 83.60 | 88.6 | 82.72 | 91.44 |
| | C | N.A | 96.12 | 95.35 | 96.07 |
| Standard Deviation | A | 0.0351 | 0.0622 | 0.0769 | 0.0729 |
| | B | 0.0222 | 0.0483 | 0.0226 | 0.0562 |
| | C | N.A | 0.0392 | 0.0336 | 0.0431 |

The MSE is given by

$$MSE = \frac{1}{MN} \sum_{M,N} \left( (A(m,n) - B(m,n) \right)^2 ,$$

where $M$ and $N$ are the number of rows and columns in the input images and PSNR is given by

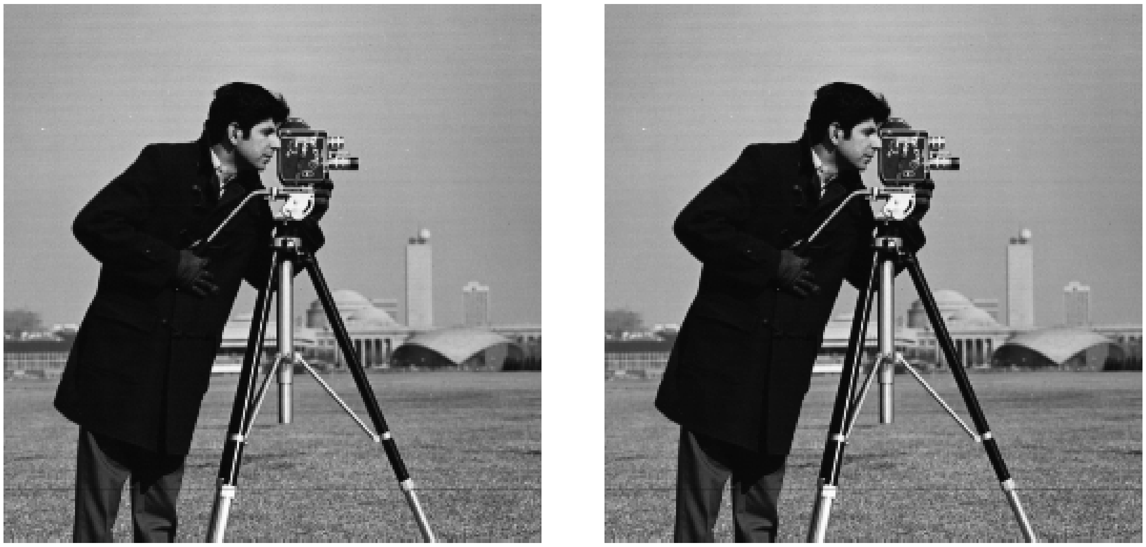
$$PSNR = 10 \log_{10} \left( \frac{T^2}{MSE} \right),$$

where $T$ is the maximum intensity value of all pixels. The MSE represents the cumulative squared error between the two images. The mean square error measure is very popular because it can correlate reasonably with subjective visual tests and it is mathematically tractable.

Lower MSE and higher PSNR imply that the difference between the original image and the test image is small, i.e., it is usually not possible to distinguish whether the image is a stego one or a sterilized one. In our experiments, we have obtained quite low MSE (0.4179 for gray-scale images and 0.1807 for color images) between cover and sterilized images. Similarly, the PSNR of cover and sterilized images are high (27.82 dB for gray-scale images and 35.12 dB for color images). Fig. 6 shows the MSE and PSNR of some selected imges. These results indicate that our technique is successful in hiding the fact that the image has been sterilized.

## 5.2 Histogram Analysis

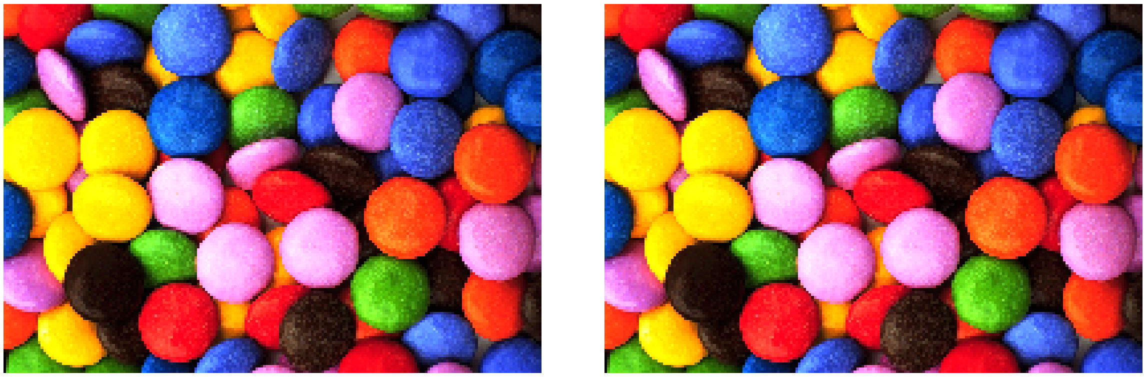
The main purpose of histogram analysis [5] in our context is to detect significant changes in frequency of appearance



Stego version

Sterilized version

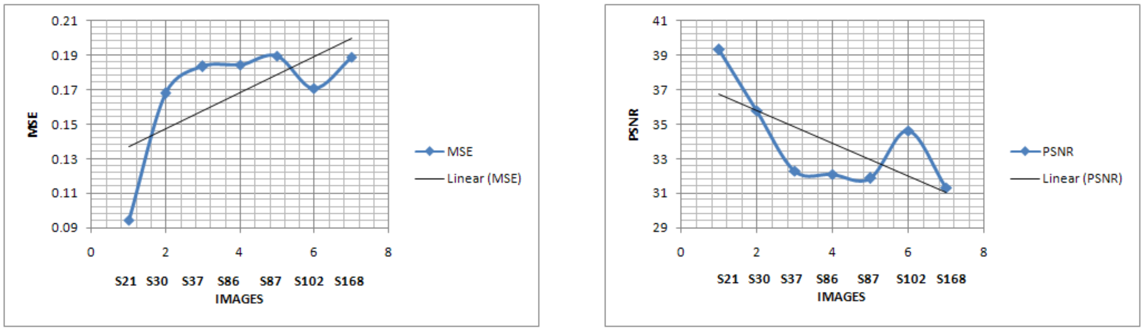
Fig. 4: Stego and sterilized version of the gray-scale image (Cameraman.bmp)



Stego version

Sterilized version

Fig. 5: Stego and sterilized version of a 24-bit color image



Mean Squared Error

Peak Signal to Noise Ratio

Fig. 6: MSE and PSNR of some selected images

of each color component in an image by comparing the cover version of the image with its stego and sterilized counterparts.

Fig. 7 shows the histograms of the Cameraman image in three stages: before stego insertion (on the left), after stego insertion (in the middle) and after sterilization (on the right). We see that our sterilization algorithm does not detectably distort the histogram of the input image.

## 6. Conclusion

In this paper, we have provided a novel concept of image sterilization. We have achieved on an average 77% to 91% success rate to sterilize the stego information of an image (the average rate varies with the steganography algorithm used to create the stego images). We would like to emphasize that the goal of our technique is not hidden message recovery, rather we aim at annihilating stego information transmission without distorting the image visibly. Our approach is generic and applied to any LSB based steganography algorithm. There has been some work [6] on double bit sterilization from the uncompressed image. Future directions may include multi-bit sterilization with the goal of sterilizing three or more bits in a pixel.

## References

[1] http://221bakerstreet.org

[2] R. Chandramouli and K.P. Subbalakshmi, "Current Trends in Steganalysis: A Critical Survey", Control, Automation, Robotics and Vision Conference, 2004, pages 964-967.

[3] http://www.digicamsoft.com/bmp/bmp.html

[4] J. Fridrich, M. Goljan and R. Du, "Reliable Detection of LSB Steganography in Color and Grayscale Images", Proceedings of the 2001 workshop on Multimedia and security: new challenges, pages 27 -30

[5] R. C. Gonzalez and R. E. Woods, "Digital Image Processing", Pearson Education, 2008.

[6] I. Mukherjee and G. Paul. "Double Bit Sterilization of Stego-images", Accepted in Track: Security and Management, WORLDCOMP, July 18-21, 2011, vol.-1, pages 743-746, Las Vegas, U.S.A.

[7] N. F. Johnson, "Steganography", Technical Report, November 1995. Available at http://www.jjtc.com/stegdoc

[8] S. Katzenbeisser and F. Petitcolas, "Information hiding techniques for steganography and digital watermarking", Artech House Books, 1999.

[9] http://www.mathworks.com/access/helpdesk/help/toolbox/vipblks/ref /psnr.html

[10] J. Mielikainen, "LSB Matching Revisited", IEEE Signal Processing Letters, vol. 13, no. 5, May 2006, pages 285-287.

[11] K. Nozaki, M. Niimi and Eiji Kawaguchi, "A large capacity steganography using color BMP images", Third Asian Conference on Computer Vision Hong Kong, China, January 8âĂŞ10, 1998, Proceedings, Volume I, pages 112-119, Lecture Notes In Computer Science, Vol. 1351/1997, Springer.

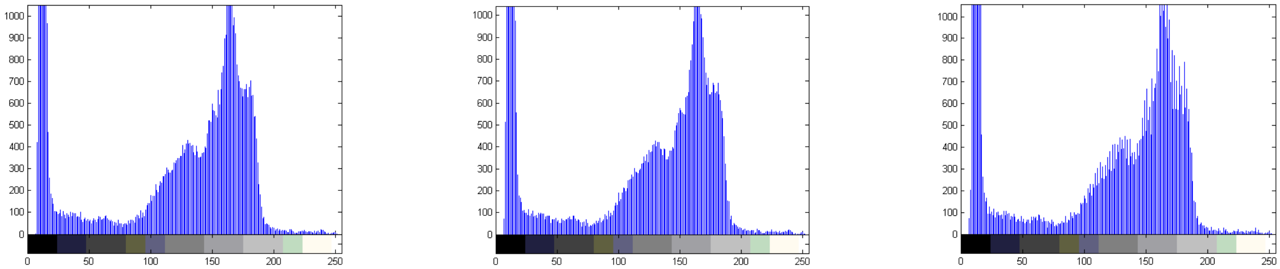
[12] R. Rana and Er. D. Singh, "Steganography-Concealing Messages in Images Using LSB Replacement Technique with Pre-Determined Random Pixel and Segmentation of Image", International Journal of Computer Science & Communication, vol. 1, no. 2, July-December 2010, pages 113-116.

[13] J. J. Roque and J. M. Minguet, "SLSB: Improving the Steganographic Algorithm LSB", Universidad Nacional de EducaciÃşn a Distancia (Spain) , Available at http://www.fing.edu.uy/inco/eventos/cibsi09/docs/Papers/CIBSI-Dia3-Sesion9(1).pdf

[14] www.webshots.com

[15] A. Westfeld and A. Pfitzmann,"Attacks on Steganographic Systems", Proceedings of the Third International Workshop on Information Hiding 1999, pages 61-76, Lecture Notes In Computer Science, vol. 1768, Springer.

[16] http://en.wikipedia.org/wiki/Peak_signalto-noise_ratio



Cover version

Stego version

Sterilized version

Fig. 7: Histograms of Cameraman.bmp at different stages

# Infrared Technology Pinpoints Energy Loss in Buildings

Timmy Nguyen
Computer Science
Wentworth Institute of Technology
Boston, MA, USA
nguyent15@wit.edu

Leonidas Deligiannidis
Computer Science
Wentworth Institute of Technology
Boston, MA, USA
deligiannidisl@wit.edu

## ABSTRACT

In this paper we present an undergraduate project where we use an infrared camera to survey our campus buildings in order to find areas of energy loss. We surveyed the buildings of our campus using infrared (IR) technology then assessed for faults and cavities. We extend our study and claim that by sealing these faults and cavities, buildings with highly sensitive data and/or government personal will have an additional layer of protection around them against terrorist attacks. These energy leaks can be used as prime points in which terrorists inject lethal chemicals into the buildings. Also if a terrorist is biologically raiding an area and using IR technology, they will have the knowledge of which buildings are vulnerable to the attacks. Pictures taken by an IR camera were textured mapped onto 3D buildings within Google Earth. Using Google Earth as our main visualization platform allowed us to easily analyze in 3D the entire campus giving us a better perspective of the buildings and their faults compared to looking at flat two-dimensional images.

**Keywords**- Infrared technology; Google Earth

## 1. INTRODUCTION

Infrared Technology is used in many areas by the government as well as by professionals. It is used to find the temperatures of surfaces. It is an easy, non-invasive, technique to detect energy waste, moisture, construction, electrical, and many other problems related to heat energy [13]. It does not require a lighting source. It works in the dark and from a distance, which makes it an attractive technology to inspect hazardous locations that are difficult to reach [7]. It measures the amount of radiation emitted by a surface then it creates a 2-dimensional thermo-graphic image related to the temperature distribution [3]. It's instantaneous when reading temperatures. In the context for speed in diagnostics “it took the user 30 seconds to find the source of moisture, while it would have taken an engineer 45 minutes without a camera” [8]. For our study, we were able to expose exterior heat leaks.

Its applications range from heat loss [13], finding mold [5], water damage [5], plumbing, mechanical and electrical problems [5][9], air conditioning [4] and moisture condensation problems, as well as missing or non-performing insulation in buildings[10]. IR technology can also help us improve concrete structures [15] by pinpointing areas of high stress [12], and even design better wall material for buildings and houses [1].

IR technology could also be used to prevent terrorist attacks [16]. Biological threat agents such as anthrax and smallpox and chemical threat agents such as tabun and arsine could all potentially be used in attacks [14]. These could all seep through crevices and faults of a building not properly insulated. By surveying the exterior infrastructure with an IR camera, points of interests in heat leaks can be properly addressed and would prevent against these chemical threats. For example, a terrorist with access to the outside of buildings could place chemical agents, airborne or in liquid form that can evaporate, to enter the buildings via pathways that air enters or escapes from the buildings.

## 2. INFRASTRUCTURE AND STANDARDS TO PREVENT TERRORIST ATTACKS

As a direct reaction from 9/11, federal agencies and standards increased their concerns for building protection. American Society of Heating, Refrigerating and Air-conditioning Engineers (ASHRAE) addressed that ventilation systems could be used to contaminate buildings [11]. A key suggestion by the ASHRAE Presidential Ad Hoc Committee on Homeland Security on how to protect buildings is that “air intakes should be located as high as possible” [2]. This would prevent ground level people from tampering with the airflow of the buildings. Using IR technology on the inside would allow fixing any insulation problems. So in the event that chemical or biological lethal agents enter a building it could all be isolated. As for the outside, finding construction faults could prevent further propagation, prolonging the life of the building as well as sealing entrances to lethal chemicals [6].

The ways that terrorists could strike could be through poorly insulated windows or construction faults. These openings could allow anthrax, tabun and other threats to enter a building [14]. Protection of US Embassies on foreign soil and even buildings with sensitive data should be constantly checked on the exterior to keep the buildings' safety up to date. Ideally the buildings' exterior would be monitored 24/7, but that would become too costly in resources. It is more realistic to have a yearly check-up. If a region was bombed by chemicals, buildings with poor infrastructure would be at risk at being exposed inside. Terrorists on land using an IR camera could find vulnerable places on buildings such as windows that don't seal well, holes, etc. where the heat escapes. They would then use that point of interest to contaminate the building's air. Depending on the wind direction, chemical and biological threats can enter inside a building from the same pathways that air enters and escapes from a building.

## 3. IMPLEMENTATION

Wentworth Institute of Technology buildings were surveyed for heat leaks. Lack of insulation around windows was a common theme when surveying the buildings as shown in figures 1 and 2.

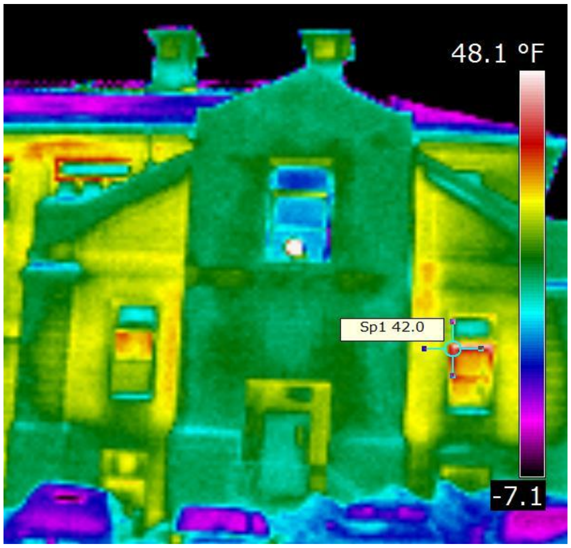


**Figure 1.** Heat escaping around the window on the right due to broken seal

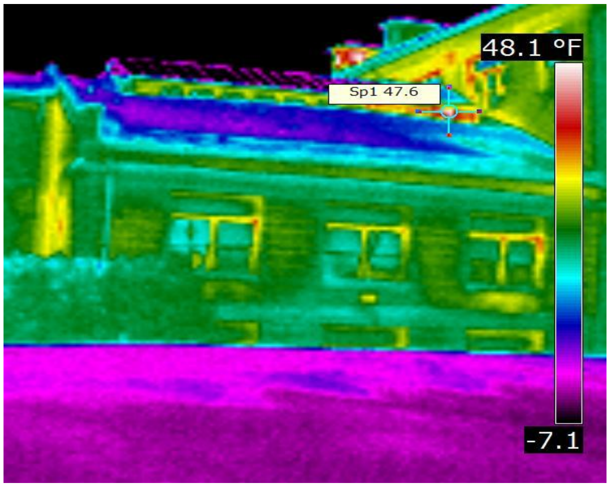


**Figure 2.** Heat escaping from the top window on the right due to lack of proper insulation

We utilized Google Earth as our visualization platform. We loaded the default 3D models of the campus that were part of the Google Earth's repository, which we downloaded and used as a template. We then went around the campus with our FLIR IR camera and took pictures of every side of all the buildings of our campus. The default temperature range was from -7 to 48 degrees Fahrenheit; the IR pictures were taken in the month of January. Creating the 3D models with the infrared faces was implemented using Google's Sketch-

Up tool. Every picture was resized and adjusted geometrically to line up with the corresponding building face.

After all the models with the IR imagery mapped onto the buildings were created, they were exported as a single KMZ file which could be loaded by Google Earth, as shown in figure 3.

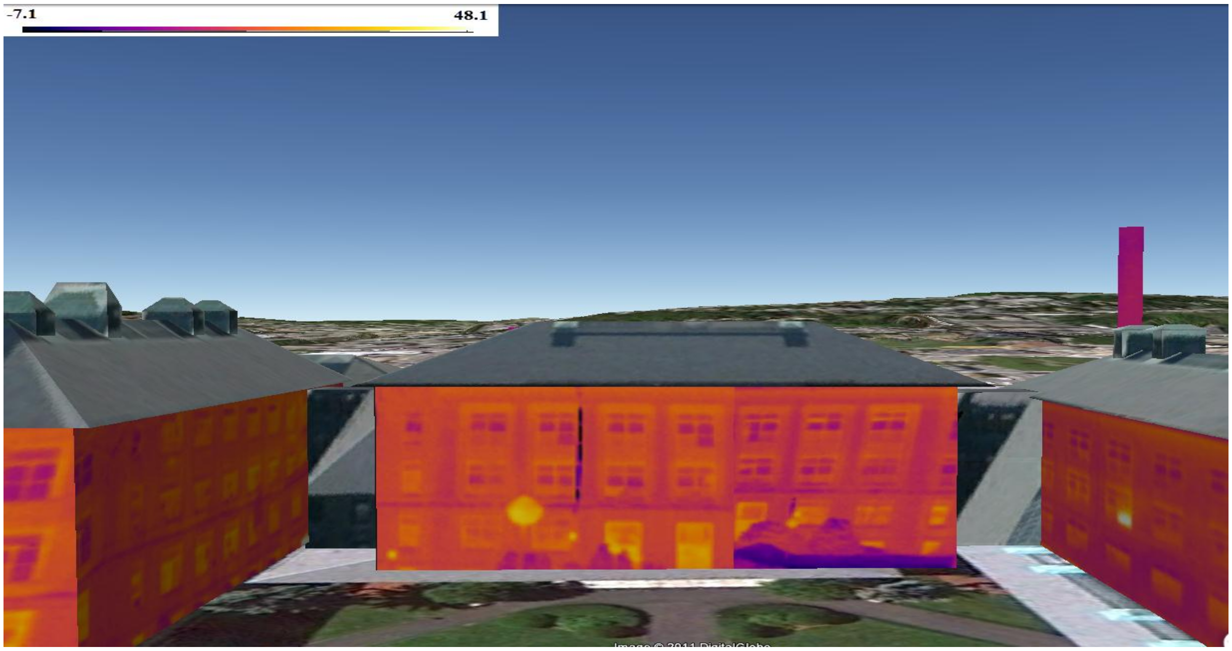


**Figure 3**. Main campus view within Google Earth. The user can switch between real and IR imagery by clicking a single button within Google Earth's navigation pane.

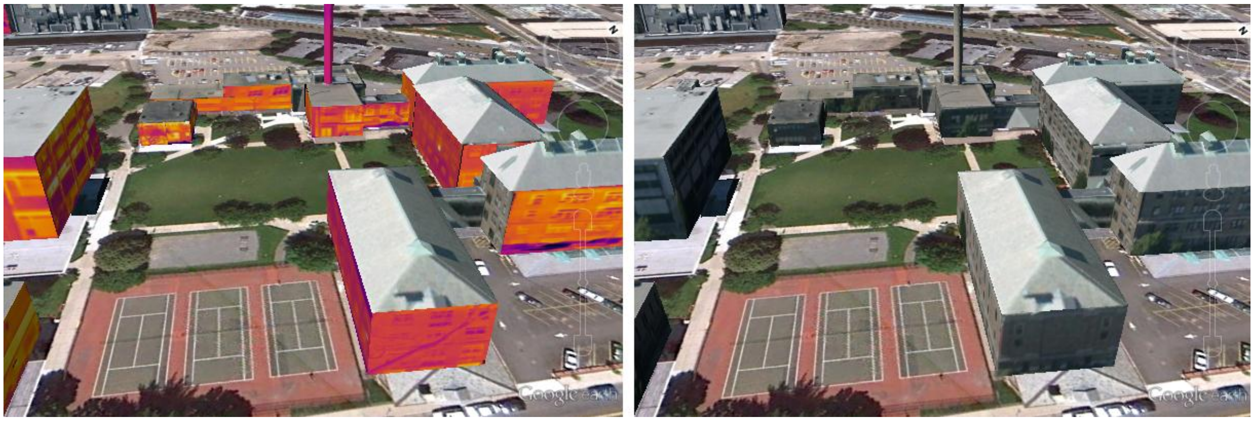


**Figure 4.** Wentworth’s main campus with IR imagery (left) and without IR imagery (right)

Figure 4 shows a snapshot of the campus with IR imagery (left) and without IR imagery (right).

Every model contains the longitude and latitude of their current real-world position and a user can switch between the real and the IR imagery by clicking a single button on Google Earth's navigation pane.

To add the screen overlay (the scale - shown in figure 3 at the top left corner), the KMZ file was unzipped, creating a KML file and an image directory. The desired image logo was placed inside the “files” directory and the KML was modified through a text-editor. This code was added inside the 'Document' and 'kml' attributes:

```
<ScreenOverlay>
        <name>Infrared Wentworth</name>
        <Icon>
           <href>files/scale.png</href> //scale.png overlay image
        </Icon>
        <overlayXY x="0" y="1" xunits="fraction"
yunits="fraction"/>
        <screenXY x="0" y="1" xunits="fraction"
yunits="fraction"/>
</ScreenOverlay>
```

A virtual tour was set up using Google Earth's built-in tour functionality. A complete campus tour was created, switching from regular models to IR models autonomously. Free 3D virtual roaming is also possible through the use of the mouse and scroll-wheel.

Both Google Earth and Sketch-Up are freely available from google.com. Images were converted to the rain palette using the FLIR QuickReport 1.2 SP1 application that came with the FLIR Camera. When choosing a color scheme for analyzing pictures, the rain palette was used over the iron palette because the rain palette has more color gradients.

## 4. DISCUSSION AND CONCLUSION

The practical uses of using Google Earth alongside IR imagery mapped onto 3D building models, is that it gives users a good space visualization of where they need to go to address heat escaping as well as potential vulnerable places for terrorist attacks. If there were only 2D IR images to view, it would take a considerable amount of time to index through the images of a certain building. Then once you found the images regarding a building you would have to guess what face of the building it was, if it was on the first floor or second, and if it was accessible from the outside. These problems are solved using 3D models.

Without the use of IR cameras, poor insulation and heat leaks would not be found as easily. IR technology has made it possible to pin-point the location where the heat is escaping. And in turn, these locations could be used to prevent biological and chemical attacks from terrorists. Google Earth has made it possible to effectively analyze buildings in 3D space. The benefits of this are indexing through buildings is easy and perspective is enhanced. ASHRAE made appropriate changes in standards in response to 9/11. These standards made safer and more protected buildings, and with the integration of IR technology we could potentially make buildings even safer.

## REFERENCES

[1] Darius Arasteh, Howdy Goudey, Christina Kohler, "Highly Insulating Glazing Systems Using Non-Structural Center Glazing Layers, ASHRAE Transactions pp. 188-198 2008.

[2] F.E. Yeboah, F. Chowdhury, S. Ilias, H. Singh, L. Sparks, "Protecting Buildings Against Bioterrorism--Review of Guidance and Tools." ASHRAE Transactions 113.1 (2007): 263-272. Academic Search Premier. EBSCO. Web. 1 Mar. 2011.

[3] Hahn, J. B. "Seeing-Eye Technology." Canadian Underwriter 76.12 (2009): 58-59. Business Source Complete. EBSCO. Web. 1 Mar. 2011.

[4] Hydronics by Mark Eatherton, "The world of thermal imaging - Part 2" Contractor pp47 April 2010.

[5] J.B. Hahn, "Seeing-Eye Technology", Canadian Underwriter, pp.58-59 Dec. 2009.

[6] M. Kogut, et al. "Thermographic non-destructive testing damage detection for metals and cementitious materials." Imaging Science Journal 48.1 (2000): 33. Academic Search Premier. EBSCO. Web. 1 Mar. 2011.

[7] Maintenance, "Taking IR predictive maintenance to new heights", Engineering & Manufacturing magazine, pp34 Oct. 2010.

[8] Manickam, Raj. "Thermography ACTS AS WATER DETECTIVE." Claims 54.6 (2006): 27-28. Business Source Complete. EBSCO. Web. 1 Mar. 2011.

[9] Mark B. Goff, Tennesse Valley Authority, "Infrared Thermography Monitors Bushing Health", Preventive Maintenance, pp32-38 , July 2009.

[10] Martins Vilnitis, Juris Noviks, "Research of heat transfer processes in walls made from new generation autoclaved aerated concrete", Constraction Science pp.96-104 2007-08.

[11] Miró, Chuck. "ASHRAE and Homeland Security." ASHRAE Journal 46.10 (2004): 70. Academic Search Premier. EBSCO. Web. 1 Mar. 2011.

[12] Monica A. Starnes, Nicholas J. Carino, Eduardo A. Kausel, "Preliminary Thermography Studies for Quality Control of Concrete Structures Strengthened with Fiber-Reinforced Polymer Composites", Journal of Materials in Civil Engineering,pp.266-273 May-June 2003.

[13] Monitoring Performance "Thermal Cameras for Building Diagnostics", The Building Economist, pp27-29, June 2010.

[14] Nakano, Victor M., William J. Croisant, and Dulcy M. Abraham. "Methodology to Assess Building Designs for Protection against Internal Chemical and Biological Threats." Journal of Computing in Civil Engineering 23.1 (2009): 14-21. Academic Search Premier. EBSCO. Web. 1 Mar. 2011.

[15] Taylor, Ian. "Feel the heat." Engineer (00137758) 283.7756 (2008): 56. Business Source Complete. EBSCO. Web. 1 Mar. 2011.

[16] Timmy Nguyen, Leonidas Deligiannidis, "Utilizing Infrared Technology to Prevent Terrorist Attacks". In Proc. of IEEE Intelligence and Security Informatics - Cyber-Physical-Social System Security (ISI'11), pp.229, July 10-12 2011, Beijing, China.

# Methods for Restricting Access in a Web Application

**N. Nassar**[1], **G. Miller**[2]
[1]IBM, Somers, NY, US
[2]IBM, Arvara, CO, US

**Abstract** - e-Commerce Applications, such as those used for the check-out process, could be in a position of not providing a fair chance to all consumers. This is especially true when a commerce site offers hot inventory items where many traders are competing to get a limited supply item. What happens is the e-Commerce sites security is compromised when some of the traders utilize pre-formatted scripts/ spiders to place orders, thus giving them an unfair advantage. The problem is: how to eliminate scripts/spiders in e-Commerce applications by using cognitive pattern recognition for security access prior to check out. In this paper, we introduced three methods using CAPTCHA. We described the architecture of each technique, and the security advantage of using it.

**Keywords:** CAPTCHA e-commerce bot cognitive

## 1 Introduction

CAPTCHA[1], Completely Automated Public Turing test to tell Computers and Humans Apart, are automated tests that are designed so that humans can pass, but current computer programs cannot pass [1]. The goal is to create a test that is presented on a web page that prevents anything besides a human to correctly pass the test. There are common CAPTCHA tests that are used. There are many examples of CAPTCHAs have been put into practice. Examples of these are shown and critiqued in Figures 1-3. [3]

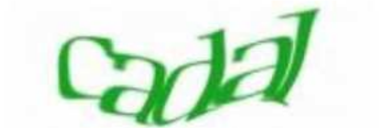


Figure 1

Figure 1 shows a Meshed characters CAPTCHA, where characters are meshed and requires human eye to identify the boundary of each character. This methods is a very common method used for security access.

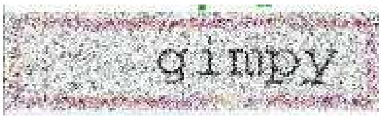


Figure 2

Figure2, Background effect CAPTCHA, backgrounds plays part of the distortion, this is a typical CAPTCHA technique where

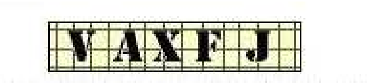


Figure 3

Figure 3 shows an example of a PayPal challenge: alphabetic characters and numerals are chosen at random and then typeset spaced widely apart, and finally a grid of dashed lines is overprinted. The wide character spacing invites segmentation attacks.

While the term "CAPTCHA" was coined in 2000 at Carnegie Mellon University, the idea originated in 1997 when Andrei Broder and his colleagues at the DEC Systems Research Center developed a scheme to block the abusive automatic submission of URLs to the AltaVista web-site[2]. Their approach was to challenge a potential user to read an image of printed text formed specially so that machine vision (OCR) systems could not read it, but humans still could. Since that time, inspired also by Alan Turing's 1950 proposal of methods for validating claims of artificial intelligence [7], many such CAPTCHAs |are used to tell Computers and Humans Apart | have been developed, including CMU's EZ-Gimpy [BAL00, HB01], PARC's PessimalPrint [5] and BaeText [8], Paypal's CAPTCHA[3] and Microsoft's CAPTCHA [9].

This paper describes a new set of tests that can be used as a CAPTCHA that employ a cognitive test for querying the user as opposed to simply reading a blurred text message that has been presented.

## 2 Background

This is section provides background information on the proposed methods proposed in this paper.

### 2.1 Preliminaries

The goal of a CAPTCHA system, in part, is to prevent bots and other automated systems from navigating through a website and negotiating a transaction system in a time that gives it an unfair advantage over slower human counterparts[6] .

[1] http://en.wikipedia.org/wiki/CAPTCHA
[2] AltaVista's “Add-URL" site: altavista.com/sites/addurl/newurl, protected by the earliest known CAPTCHA
[2] www.paypal.com

The common solution to this issue is to provide a distorted image of letters and numbers used to prevent automated use of websites, a.k.a CAPTCHA. These solutions require a person to read the distorted letters and type them into a field, something a bot cannot do. This proves that the page is being accessed by a person. The existing solution counts on generating a combination of letters and numbers that are distorted and displayed as image that a person must interpret and re-type into a field. The approach can be problematic for users as they cannot always read the letters or numbers because of too much distortion. The main drawback is that hackers developed smart spiders where it builds a library of images that would allow figuring out the content of the CAPTCHA based on its size [4].

New methods for testing who, or what, is interacting with a system allows for the interaction with images as a possible alternative instead of text typing that is commonly used today. Computer-based recognition algorithms require the extraction of color, texture, shape, or special point features, which can be complicated by manipulating the image. Humans are capable of understanding the original concept depicted in the images even manipulated [10].

Adding cognitive pattern recognition for security access prior to check out eliminates the automation the process so that the spider will not be able to bypass the cognitive security check. Based on the implementation of our proposed solution, spiders building a library of images will not help bypassing this security feature as identifying the image alone is not the goal, rather recognizing the cognitive pattern which must be done by a human [2].

## 3. Proposed Solutions

While there are many ideas in the CAPTCHA space, we are introducing three methods that are novel approaches presented by IBM. Two of the methods are cognitive based, and the third is a coordinate based method. In this paper, we will introduce each method, Implementation of the security feature, and discuss the advantage of each.

### 3.1.1 Cognitive Based Security

The first method is to provide a unique way of discriminating a human involvement in a transaction that would occur on a computerized experience that utilizes internet or network connected interactions that require an endurance of actual human intervention and not a "pseudo intervention" that could be performed by an automated system or application in the realm of computer technology. This methodology requires human interaction that can not be performed by an application program. Cognitive perception and reaction are performed by the human end-user to verify that a human is actually interacting with the information technology system and verifying the intentions of the user initiating the transaction.

The intent of this methodology is to challenge the end-user with a query that only a human could cognitively understand, and with which artificial intelligence is not capable of determining. It is proposed to do this challenge through a series of images where a user must select the correct image. The first image gives an instruction on what the task is, e.g. pick the picture of a car. The next set of images, say 3 images, would be of 3 different objects, one being a car. To further challenge a bot from merely storing images in a library, the images could be rotated some angle, or flipped along an axis. The first image with the instruction and the test images would all be randomly selected from a pool of available images. The order of the images, and location of the correct image, would also be randomly assigned.

Figure 4 shows a possible layout of images selected for presentation to a user.

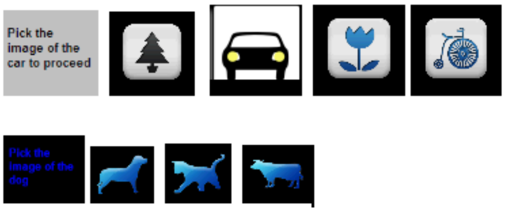

Pick the image of the car to proceed
Pick the image of the dog

Figure 4

The purpose is to create an extremely high number of possible combinations of test images and correct images that would make it extremely difficult for a bot to query and store correct answers. In addition, if the system noticed that a "user" was methodically querying a page for access, then that instance could be denied, blocked, etc. There is different business existing rules (e.g. incoming IP address, username, etc.) that could be applied on how to block a suspected bot from continually accessing a test [8].

### 3.1.2 Flow of Tasks Based Security

The second method challenges the end-user with a query that only a human could cognitively understand, and with which artificial intelligence is not capable of determining is the intend of using cognitive patterns recognition. It is proposed to do this challenge through a series of symbols with which the majority all have something in common, but of the series, only one is not in common or associated with the rest of the symbols. An example would be having four graphical symbols that represent three numbers and one letter of the alphabet as shown below in figure 5:

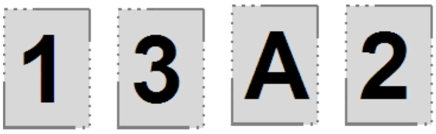


Figure 5

The correct challenge response would be to select the letter "A" graphic.

The concept of cognitive association or dissociation does not have to be constrained to numbers and letters, but could include globally accepted symbols such as material objects and organic objects such as shown in figure 6.

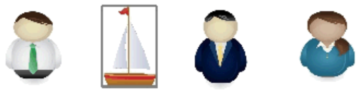


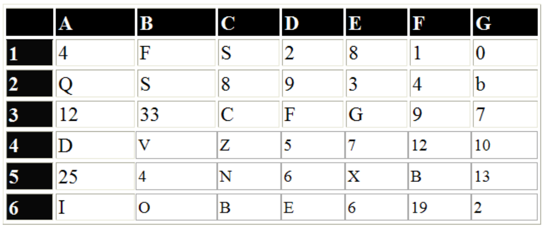
Figure 6

The intent is for the end-user to be queried for the symbol or graphic that does not fit within the total set of symbols or graphics and those selections becomes the successful challenge answer for any transaction that would take place. That transaction could include, but is not exclusive to, banking, commerce, download activity, etc.

### 3.1.3 Coordinates Based Security

The final method challenges the end-user with a query that only a human could understand, identify and with which artificial intelligence is not capable of determining. It is proposed to do this challenge through a series of coordinates (vertical and horizontal) loaded as an image where a user must select the correct value represented by certain coordinates. The question to the user would be what are the values of A1, C4, B3, D5? There will be an input field where the user can enter the values corresponding to the coordinates in the challenging question.

The coordinates map image, as depicted in figure 7, could be generated using server logic in a way that each new request generates a new, random, mapping represented to the end user as an image. What is A1, C4, B3, D5?



|  | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | 4 | F | S | 2 | 8 | 1 | 0 |
| 2 | Q | S | 8 | 9 | 3 | 4 | b |
| 3 | 12 | 33 | C | F | G | 9 | 7 |
| 4 | D | V | Z | 5 | 7 | 12 | 10 |
| 5 | 25 | 4 | N | 6 | X | B | 13 |
| 6 | I | O | B | E | 6 | 19 | 2 |

Figure 7

The purpose is to create an extremely high number of possible combinations of test images and correct images that would make it extremely difficult for a bot to query and store correct answers. In addition, if the system noticed that a "user" was methodically querying a page for access, then that instance could be denied, blocked, etc. There is different business existing rules (e.g. incoming IP address, username, etc.) that could be applied on how to block a suspected bot from continually accessing a test.

## 3.2 Functionality

In this section we describe the architecture details of each solution with various implementations. The ultimate goal is to cater the challenges in sliced images as well as the challenging question so that any page reader or machine driven intelligence would not be able to neither comprehend the question asked, nor build any intelligence around the answer.

### 3.2.1 Cognitive Based Security

Fundamentally, the approach works by storing images in a database, then randomly selecting instruction images, a related image, and a non-related images from that database to present to a user.

A user is presented a web page that requires to them to select the image from instructions that are presented in an image. Selecting the correct image is required to move forward to the next page, e.g. a shopping cart or checking out. To present the images, the logic would pick an instruction image and a related image to that instruction (correct answer). Then 2 or more (randomly selected at run time) or more non-related images are retrieved from a database. Any of the images selected can be rotated some amount in 10 degree increments, giving 36 possible combinations for an image. Whether or not they are rotated and the amount of rotation is randomly selected. The image could also be flipped, again by a random selection, to be flipped along the x-axis or y-axis. How the images are identified are defined by an administrator who sets up the database. The images can be any item identifiable by a person. The related images (all images of a tree) can be stored in one table, or by using another database construct. Ideally, a large number, 1000+, related images are stored, and there is a large number of sets of related images.  The following high-level flow, shown in figure 8, demonstrates the process.

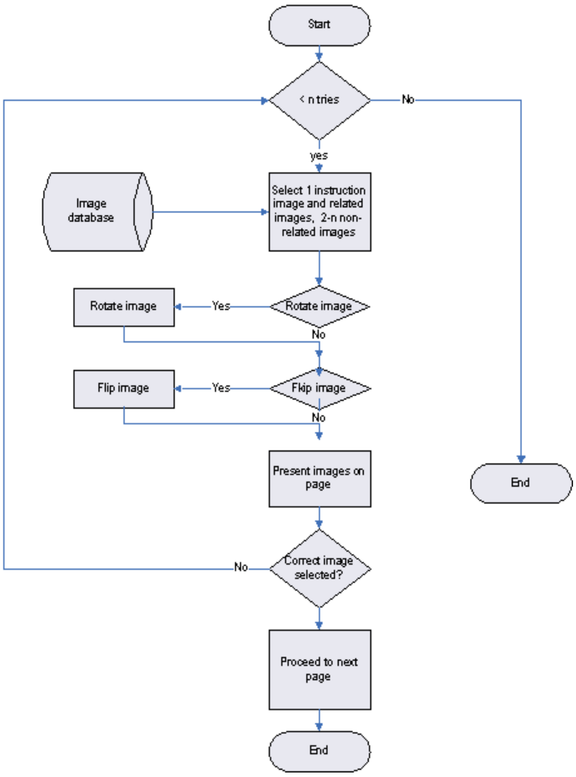


Figure 8

The approach starts by randomly selecting a set of related instruction and response images, then randomly selecting 2 to n images from different sets. These images are presented on the screen along for the user to select the to select the image that matches the instruction. If the user selects the correct image, then the process flow continues to the next page. If the user does not select the correct image, then they are required to try again. In this case, a new set of images is selected and presented as described. The administrator can define how many times the user is allowed to attempt to select the correct image.

Accessibility is also something that is taken into consideration. This method would utilize a common enterprise accessibility solution. Users with visual accessibility and would be provided with an audio sounding the question and name of the each selection mixed in noise – just like regular CAPTCHA solution. The audio would still require a cognitive approach to solving the choice of image. The audio samples would be created by tags stored along with the images in the database. This way they could easily translated. The tag would not be part of the page so that a bot could not detect it. The accessibility solution should not present a soft spot for spiders or malicious scripts to compromise the site security.

### 3.2.2 Task Flow Based Security

By the same token, this approach works by storing images in a database, then randomly selecting related images, and a non-related image from that database to present to a user. The following high-level flow demonstrates the process.

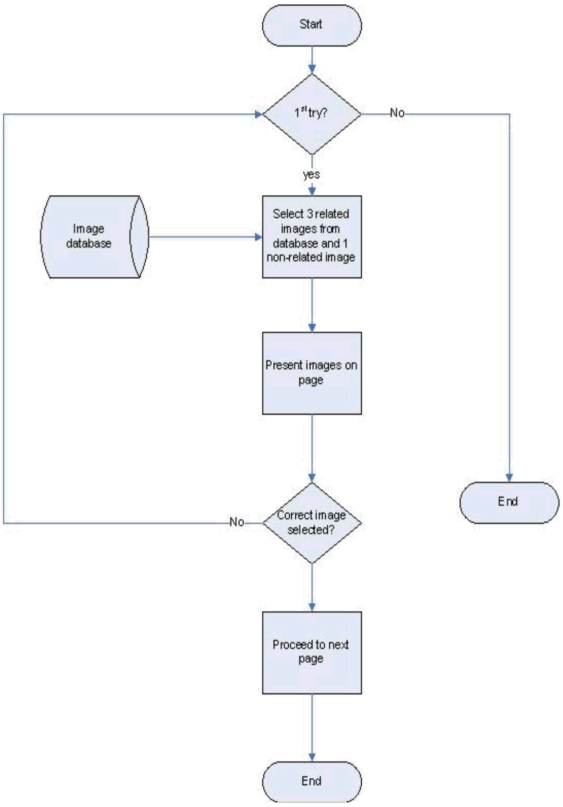


Figure 9

To start, a user is presented a web page that requires to them to select the image from what is shown that does not belong with the others. Selecting the correct image is required to move forward to the next page, e.g. a shopping cart or checking out. To present the images, the engine picks 3 related images from a database. How the images are related is defined by an administrator who sets up the database. The images can be related because they are plants, birds, transportation items, etc. Tags, latent semantic groups, or taxonomy can be used to relate objects to one another. The related images can also be stored in one table, or another database construct. Ideally, a large number, 100+, related images are stored, and there is a large number of sets of related images.

The approach starts by randomly selecting a set of related images, then randomly selecting 3 images from that set. Next, the logic randomly selects a different set and then randomly selects one image from that set (this is the non-related image). These images are presented on the screen along with the instruction for the user to select the non-related image. If the user selects the correct image, then the process flow continues to the next page. If the user does not select the correct image, then they are required to try again. In this case, a new set of images is selected and presented as described. The administrator can define how many times the user is allowed to attempt to select the correct image.

The flow pictured in figure 10 shows a little more detail for the selection of an image.

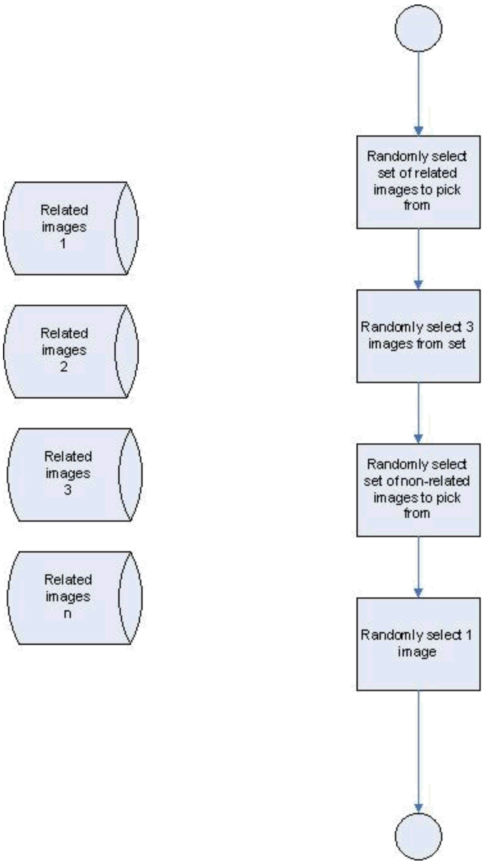
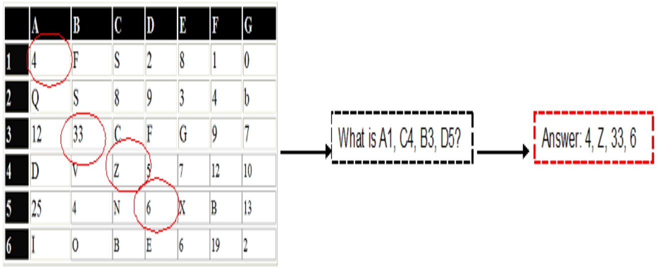

Related images 1
Related images 2
Related images 3
Related images n
Randomly select set of related images to pick from
Randomly select 3 images from set
Randomly select set of non-related images to pick from
Randomly select 1 image

Figure 10

### 3.2.3 Coordinates Based CAPTCHA

Fundamentally, the approach works by creating NxM array where each element in the array has an alphanumerical value. The array is converted to a visual image with header rows and columns and stored into the database. This image is mapped in the database to a question about particular set of coordinates the array. By the same token, the value of this particular set of coordinates is stored in the database mapped to both the question and to the image for validation. The order of the answer should match the order of the requested coordinates in the question.

In the example shown in figure 11, the answer is stored as 4.Z.33.6. which is mapped to the coordinates of A1, C4, B3, and D5 accordingly.



|  | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | 4 | F | S | 2 | 8 | 1 | 0 |
| 2 | Q | S | 8 | 9 | 3 | 4 | b |
| 3 | 12 | 33 | C | F | G | 9 | 7 |
| 4 | D | V | Z | 5 | 7 | 12 | 10 |
| 5 | 25 | 4 | N | 6 | X | B | 13 |
| 6 | I | O | B | E | 6 | 19 | 2 |

What is A1, C4, B3, D5?

Answer: 4, Z, 33, 6

Figure 11

**Implementation**

There are two ways to implement this method.

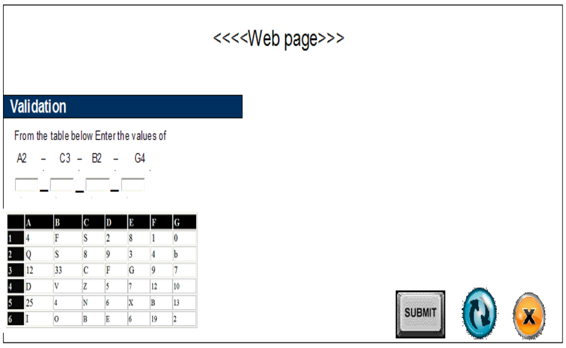
A. Real time approach- In this case, when the browser sends a request to render a page, the back-end logic would generate a random NxM array and assign its value, then a governance module would

- Generate the image, and store it in the db as CLOB
- Identify a random set of coordinates in the array and store them in the question column in the table
- Fetch corresponding values of these coordinates and store them into answer column in the table
- Serve the image and its mapped question to the user's browser page

B. Library approach- In this implementation, an off line process would run to generate thousands of Arrays

- Identify a random set of coordinates for each array
- Fetch corresponding values of the coordinates for each array
- Generate the images and store both question and mapped answer in the database
- When a browser sends a request, a server logic manages which image to display so that the images would never repeat.
- In both implementations would make it almost impossible for any pot to detect the correct value of the coordinates because both coordinate and corresponding answers changes with every request.

To start, a user is presented a web page that requires to them to select the image from instructions that are presented in an image. Selecting the correct value of the listed coordinate is required to move forward to the next page, e.g. a shopping cart or checking out. To present the images, the invention picks the question/ instruction and a related image to that instruction. The image selected can be rotated some amount in 10 degree increments, giving 36 possible combinations for an image. Whether or not they are rotated and the amount of rotation is randomly selected. The image could also be flipped, again by a random selection, to be flipped along the x-axis or y-axis. How the images are identified is defined by an administrator who sets up the database. The images can be any item identifiable by a person.



<<<<Web page>>>

Validation

From the table below Enter the values of

A2 – C3 – B2 – G4

|   | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | 4 | F | S | 2 | 8 | 1 | 0 |
| 2 | Q | S | 8 | 9 | 3 | 4 | b |
| 3 | 12 | 33 | C | F | G | 9 | 7 |
| 4 | D | V | Z | 5 | 7 | 12 | 10 |
| 5 | 25 | 4 | N | 6 | X | B | 13 |
| 6 | I | O | B | E | 6 | 19 | 2 |

Figure 12

If the user clicks the refresh button, a new image is display using either of the approaches.

Accessibility is also something that is taken into consideration. This method would utilize a common enterprise accessibility solution. Users with visual accessibility and would be provided with an audio sounding the question and name of the each selection mixed in noise – just like regular CAPTCHA solution. The audio would still require a coordinate mapping recognition approach to identify the value of the requested coordinates. The audio samples would be created by tags stored along with the images in the database. This way they could easily translated. The tag would not be part of the page so that a bot could not detect it. The accessibility solution should not present a soft spot for spiders or malicious scripts to compromise the site security.

## 4. Conclusions

In conclusion, applying security to web applications can take various flavors. CAPTCHA is an essential way to enforce human interaction with the web site. We demonstrated three unique methods to apply such a technology. What stand out in our methodology is that each method provides technically a substantial number of permutations that makes it extremely difficult for any spider to be able to bypass the security imposed. By the same token, we introduced a new method of displaying the actual challenging question in an undetectable format by converting it into a set of sliced images and serve it to the browser. As a result, spider will not be able to neither identify the question nor recognize the challenges presented to the end user. In return, compromising this security level is very highly unlikely

## 5. References

[1] Von Ahn, et. al. Proceeding UROCRYPT'03 Proceedings of the 22nd international conference on Theory and applications of cryptographic techniques, 2003

[2] Chew, M. et al. "Image recognition CAPTCHAs ". The Information Security, Conference, LNCS 3225, 268-279, 2004

[3] Ahn, L. et al. "Telling humans and computers apart automatically". Communications of The ACM, 47(2), pp. 56–60, February, 2004

[4] IBM Corp. " Method and system to generate human knowledge based CAPTCHA". IP.com number: IP-COM000l84977D, July, 2009

[5] A. L. Coates, H. S. Baird, and R. Fateman, "Pessimal Print: a Reverse Turing Test," Proc., IAPR 6th Intl. Conf. on Document Analysis and Recognition, Seattle, WA, pp. 1154-1158, September 2001

[6] Elson, J. et al. Asirra: "A CAPTCHA that exploits interest-aligned manual image categorization", ACM, CCS'07, October-November, 2007

[7] A. Turing, "Computing Machinery and Intelligence," Mind, Vol. 59, pp. 433-460, 1950

[8] M. Chew and H.S. Baird, "BaffleText, a Human Interaction Proof", 10th SPIE/IS&T Document Recognition and Retrieval Conference (DRR'03), Santa Clara, CA, USA, pp. 305-316, January, 2003

[9] P. Y. Simard, R. Szeliski, J. Benaloh, J. Couvreur, I. Calinov, "Using Character Recognition and Segmentation to Tell Computer from Humans" Proc., IAPR Int'l Conf.on Document Analysis and Recognition, Edinburgh, Scotland, August 2003

[10] IBM Corp, "A new method for telling humans and computers apart automaticall". IP.com number: IP-COM000188716D, October, 2009

# A Bayesian Network Analysis of System Failure in the Presence of Low-Probability External Correlating Events

Jack K. Horner
P.O. Box 266
Los Alamos NM 87544 USA
email: jhorner@cybermesa.com

**Abstract**

*Using a simple Bayesian network model of the electrical power backup for the Fukushima Daiichi reactor control systems as an example, I show that systems with multiple independent backup modes (MIBMs) can be disastrously sensitive to seemingly low-probability external events, even when the intrinsic joint failure rate of the backup subsystems is practically zero. This counterintuitive behavior frames a design rubric which I call the "External Correlator Test" (ECT): a determination of the acceptability of the cumulative probability of system failure in the presence of an external correlating distribution.*

**Keywords**: autonomous systems, Fukushima, power-law distribution

## 1.0 Introduction

On 11 March 2011, a magnitude 9 earthquake generated tsunami waves that struck the Fukushima-Daiichi nuclear facility on the east coast of Japan. At least one of these waves was estimated to be 14 meters high and overwhelmed the Fukushima defenses, which were only designed to withstand waves of a maximum 5.7 meters high. Electrical power to the reactor controls, including electricity from all emergency electrical backup systems at the site, ceased. The resulting facility blackout caused the loss of all instrumentation and control systems in Reactors 1-4. Loss of coolant was followed the release of radiation into the surrounding area. A region with a radius of approximately 30 miles, centered on the complex, is now uninhabitable ([5]).

Fukushima reactor control was largely autonomous. By convention, autonomous systems are expected to respond in desirable ways to all external stimuli; they often manage pathological events through multiple independent backup modes (MIBMs). Catastrophic failures in systems that contain MIBMs would seem to have very low probabilities, yet occur with unexpectedly high frequency. How can we understand this sensitivity of such system failures to low-probability disaster scenarios, and what can be done during system development to help mitigate their occurrence?

Significant aspects of risk management can be modeled as a "betting" regime ([7]). Any rational betting regime must at least be consistent with probability theory ([7]). Any probabilistic system can be modeled as a Bayesian network (BN; [2],[8]). A BN is a system of conditional probabilities ([3], p. 23) mapped onto a directed graph ([4]) of system entities. BNs are widely used in decision-assistance systems, including

- health-allied diagnosis
- automotive Built-In-Test
- spam filters
- intelligence analysis

## 2.0 Method

I first model the Fukushima-Daiichi electrical power backup systems with BNs, then abstract a criterion of design adequacy from that example.

## 3.0 Results

### 3.1 A closer look at Fukushima Daiichi

The Fukushima Daiichi nuclear power complex contained six General Electric (GE) Mark I boiling water reactors (BWRs). The Mark I has been extensively tested and ~30 are in use around the world ([16]).

Power to the reactor controls is normally generated within the site. The site has three independent electrical power backup sources: an external commercial electrical supply, diesel-powered generators, and batteries. The reactor controls fail if all three electrical sources fail.

None of these three backup systems depends on any other. Each of the backup systems has a nominal intrinsic probability of failure of ~$5 \cdot 10^{-4}$ per year, given preventive maintenance. Thus, by the law of independent events, the probability of all three failing from intrinsic events is $(5 \cdot 10^{-4})^3$ = ~$10^{-10}$ per year -- for all practical purposes, "zero".

Figures 1 and 2 depict BNs showing the essentials of the backup electrical power system for reactor controls in the Fukushima Daiichi complex. The models include representations of

- some failure modes of the system
- sources which could supply electrical power to the reactor controls
- reactor control
- probabilistic (“quasi-causal”, [6]) relations among the above

Figure 1 is a “naïve”, and Figure 2, a “tsunami-augmented”, model of that system. The models are implemented in the *Netica* ([1]) BN development and runtime framework.

In Figures 1 and 2

- boxes represent system entities of interest
- upper portion of a box indicates the name of the entity
- lower portion of a box identifies the probability, expressed as a percentage, that the entity is in the named state
- an arrow from Box A to Box B means the probability of the states of entity B depends on the states of entity A
- prior probabilities ([9], Section 1.3) of the system entities are defined in tables (not shown)

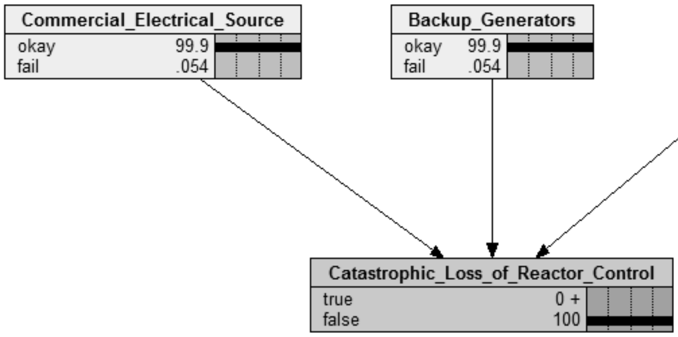


**Figure 1. "Naïve" model. In this model, there is no correlation of the failures of the electrical sources; each source has an intrinsic failure probability of 0.054%. The probability of catastrophic loss of control is therefore $(5.4 \times 10^{-4})^3$ = ~$10^{-10}$ per year, which may be acceptable.**

Figure 2, the "tsunami-elaborated" model, is Figure 1, plus a tsunami probability explicitly modeled as a Pareto (power-law; [3], p. 193) probability density function (pdf)

$$P(h) = (a/b)(b/h)^{a+1} \qquad \text{Eq. 1}$$

where

P(h) is the probability of a tsunami of height h (in meters), per year

a is a distribution "shape" parameter, here set to 1.0

b is a distribution "location" parameter, here set to 0.01

Eq. 1, with parameters set as noted, is a reasonable fit to tsunami occurrences on the Japanese east coast ([10]).

The probability of reactor control failure, given a tsunami with height > 6 meters, can be determined by evaluating the joint *cumulative* distribution function (cdf) corresponding to Eq. 1 and the MIBMs, for h > 6 meters.

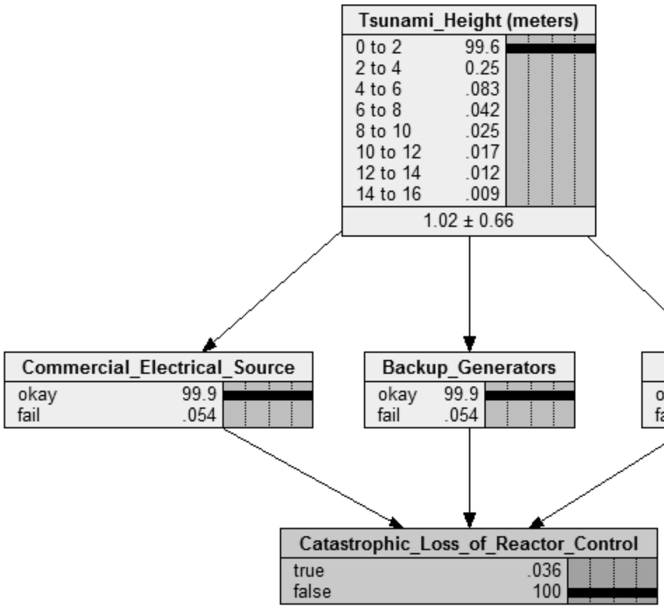


**Figure 2. "Tsunami-augmented" model. The"naïve" model, extended with a (Pareto) tsunami-height distribution, forces correlation of the failure of all three electrical sources. The prior probabilities of the "okay/fail" distribution of the electrical source nodes are defined to be the *same* in the naïve and tsunami-elaborated models. The resulting system probability of catastrophic failure is ~10-$^{4}$ per year, a 6-order-of-magnitude increase over that probability in the naive model, which may be unacceptable. The bottom of the "tsunami box" shows the mean ± one standard deviation of the tsunami distribution.**

Once the seawall has been topped, the probability that the electrical backup power sources will fail rises sharply. In the model shown in Figure 2, the backup systems were assumed to have a relatively high probability of surviving a 6-meter tsunami, but a very low probability of surviving a 15-meter tsunami (see Figure 3).

| Prob(individual backup system failure) | Tsunami_Height |
|---|---|
| 1.1e-5 | 0 to 2 |
| 1e-5 | 2 to 4 |
| 1e-5 | 4 to 6 |
| 0.1 | 6 to 8 |
| 0.5 | 8 to 10 |

| | |
|---|---|
| 0.9 | 10 to 12 |
| 0.99 | 12 to 14 |
| 0.99999 | 14 to 16 |

**Figure 3. Probability of individual backup system failure as a function of tsunami height assumed in the model shown in Figure 2.**

The probability of system failure of the system shown is ~0.0001 per year, sometimes interpreted as "a 1000-year event " (i.e., a 10% chance of failing in 1000 years). A backup system that would fail only once per 1000 years would seem robust enough. The reactor control failure probability in the presence of this tsunami distribution is *$10^6$ times greater,* however, than the probability of reactor control failure in the absence of the tsunami distribution. The cdf of what seem to be extremely rare events, therefore, can hugely amplify the probability of system failure, even though the probability of individual scenarios in the associated pdf is acceptably small.

Worse is true. Given the model described above, in 50 years of reactor operation -- the nominal design lifetime -- the probability of system failure at Fukushima Daiichi due to a 6+ meter tsunami is ~0.05. It is likely that most people would regard that probability as unacceptably high.

For comparison, as of August 2011, there had been ~$10^{-3}$ radiation-releasing accidents per reactor-year in the nuclear power industry worldwide ([19]). The observed incidence of radiation-releasing accidents at sites with GE Mark I BWRs, excluding Fukushima Daiichi, is ~$10^{-5}$ per reactor-year ([19]). The probability of system failure at Fukushima Daiichi due to tsunamis is therefore ~100 times the empirically expected system failure rate for sites with GE Mark I BWRs, excluding Fukushima Daiichi,.

### 3.2 A criterion of design adequacy

The models illustrated in Figures 1 and 2 are easily adapted to other systems that have MIBMs subject to external correlating distributions. Bayesian analyses similar to those of Section 3.1 have been applied to other systems with similar backup systems (e.g., the Space Shuttle ([12]), unmanned aerial vehicles (UAVs) operated in the absence of routine preventive maintenance ([13]), the Three Mile Island Accident ([18]), global threats to amphibians ([20]), and the proposed Keystone Pipeline ([11],[17])), with similar results.

What lessons can be learned from such examples?

First, an event, E, external to a system, S that has MIBMs can force a correlation of the failures of S's MIBMs. Second, the probability of system failure is determined by the joint *cdf* for E and the MIBMs. This cdf can induce a system failure probability that is several orders of magnitude larger than the probability of individual events in E's pdf. These considerations frame a criterion of design adequacy, which I call the External Correlator Test (ECT):

(ECT) Let S be a system with MIBMs and E be an external correlating distribution for S. The design of S is robust only if the system failure probability for S is acceptable, given the joint cdf of E and the MIBMs.

## 4.0 Discussion and conclusions

The considerations of the preceding sections motivate several observations:

1. In general, enumerating all possible, or even the most likely, candidate external correlating distributions for a given system S is not a mechanical task, and which candidates should be considered will depend on the nature of S and on cost/benefit trades. However, there are several commonly occurring categories of external correlating events worth considering by default, including

a. natural disasters (e.g., hurricanes, tornados, floods, earthquakes, fire, solar flares, and tsunamis)
b. vibration
c. strong electromagnetic fields
d. temperature and humidity extremes
e. dust
f. accidental vehicle crashes
g. sabotage
h. control (e.g., of utilities, especially of electrical power) delivered through the Internet
i. whether robust systems engineering processes ([15]), including human factors considerations, were used during development and operation

2. The effects on system failure probability of *power-law* external correlating distributions are particularly susceptible to underestimation because the values of the *pdf* for *individual* backup-failure events often seem too small to matter -- the probability of a 15-meter tsunami, for example, is miniscule. But value of the cdf -- the integral (or in the case of a discrete distribution, the sum) of the pdf -- over the entire range of backup-failure scenarios can amplify system failure probability by several orders of magnitude, compared to the effect on system probability failure of individual events in the pdf of the external correlating distribution. Many natural disasters are power-law distributed ([14]), and thus their effects on system failure are easily underestimated.

3. As part of system design review (SDR, [15]), system safety design should be analyzed for system-failure-amplifying correlations of low-probability states/scenarios, in accordance with the ECT. BNs can cost-effectively support this kind of analysis.

## 5.0 Acknowledgements

This work benefited from discussions with Tony Pawlicki. For any problems that remain, I am solely responsible.

## 6.0 References.

[1] Norsys Software Corporation. *Netica* v4.16. http://www.norsys.com. 2011.
[2] Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Revised Second Printing. Morgan Kaufmann. 1991.
[3] Hogg RV, McKean JW, and Craig AT. *Introduction to Mathematical Statistics.* Sixth Edition. Prentice Hall. 2005.
[4] Diestel R. *Graph Theory*. Springer. 1997.
[5] International Atomic Energy Agency. *IAEA International Fact Finding Expert Mission of the Nuclear Accident Following the Great East Japan Earthquake and Tsunami. Preliminary Summary*. http://www.iaea.org/newscenter/focus/fukushima/missionsummary010611.pdf. 1 June 2011.
[6] Pearl J. *Causality: Models, Reasoning, and Inference.* Second Edition. Cambridge. 2009.
[7] Kemeny JG. Fair bets and degree of confirmation. *The Journal of Symbolic Logic* XX (1955), 263-273.
[8] Jensen FV. *Bayesian Networks and Decision Graphs*. Springer. 2001.
[9] Sivia DS. *Data Analysis: A Bayesian Tutorial*. Oxford. 1996.
[10] University of Southern California. The Tsunami Research Center. http://www.tsunamiresearchcenter.com/.
[11] Horner JK. A Bayesian network assessment of earthquake risk to the Keystone Pipeline. Unpublished manuscript. 2011.

[12] Columbia Accident Investigation Board. Final Report. http://caib.nasa.gov/. 2003.

[13] Science Applications International Corporation. Global Hawk/Reaper Advanced Diagnostic Expert System Demo. Power Point briefing. 16 August 2007.

[14] Bak P. *How Nature Works*. Springer. 1996.

[15] International Standards Organization. *Reference Standard ISO/IEC 15288. Systems and software engineering —System life cycle processes.* Second edition. 1 February 2008.

[16] General Electric. *The Mark I Containment System in BWR Reactors.* http://www.gereports.com/the-mark-i-containment-system-in-bwr-reactors/. 2011.

[17] US Department of State. Keystone Pipeline Project. Final Environmental Impact Statement. Project Description. Section 2.2. http://www.keystonepipeline-xl.state.gov/clientsite/keystonexl.nsf/05_KXL_FEIS_Sec_2.0_Project_Description.pdf?OpenFileResource.

[18] The President's Commission on the The Accident at Three Mile Island. *Final Report of The President's Commission on The Accident at Three Mile Island.* 30 October 1979. http://www.pddoc.com/tmi2/kemeny/.

[19] European Nuclear Society. Nuclear power plants, worldwide. http://www.euronuclear.org/info/encyclopedia/n/nuclear-power-plant-world-wide.htm.

[20] Horner JK. A Bayesian network model of global threats to amphibians. Submitted to the *2012 International Conference on Bioinformatics and Computational Biology*.

# Tackling Internet abuse in Great Britain: Towards a framework for classifying severities of 'flame trolling'

**J. Bishop**

Electronic Law for Internet Empowerment Group
Centre for Research into Online Communities and E-Learning Systems
The European Parliament. Square de Meeus 37, 4th Floor
Brussels B-1000, Belgium

**Abstract -** *While trolling has existed as a term since the 1990s and as a reality even earlier there has been an exponential increase in the prevalence of the abusive kind - 'flame trolling'. Mistakenly the media calls these flame trollers, 'trolls', when in fact there are more often than not 'Snerts' and 'E-Vengers'. The justice system in Great Britain has taken a sporadic approach to dealing with flame trolling, and the wide range of legislation that has existed since the 1980s has no strategic method to assign its usage on the basis of the nature of the flame trolling as its use often depends on the whim of different police forces. This paper hopes to change this. After a brief presentation of the background of Internet trolling in Great Britain and in general a new framework is presented. This allows prosecutors to easily classify flame trolling based on the facts of the case and pick the appropriate level based on the severity.*

**Keywords:** A Maximum of 6 Keywords

## 1 Introduction

Despite being one of the greatest innovations of all time, the Internet is less safe than one might have expected it to be considering one is often simply sitting in front of a computer screen at one's home. Internet abuse (IA) is rife and despite differences in how it is defined and the specific criteria used, there is general agreement that IA can be explained in terms of the negative effect of Internet use, that is, Internet use that causes disturbances in an individual's life [1]. It has been argued that Internet abuse is a mild form of Internet addiction [2], but there is a lack of evidence to support this claim. While technological approaches have been taken to combating IA [3], it was clear for many governments at the start on the 21s century that legislation was also needed to reinforce these techniques. Recently, the term 'trolling' has become a catch-all term for this sort of Internet abuse [4], but this is a simple view of the situation. Technically the abusive element of trolling could be called 'flame trolling' and the more positive aspects could be called 'kudos trolling'. This paper will deal mainly with the former.

It was once mistakenly forecast that the rise of Internet use would be impaired through over-regulation of cyberspace [5]. Whilst there has never been a shortage of parliamentarians wanting to legislate in any number of areas of people's lives, this paper shows how in all but name there is enough legislation on the statute books in Great Britain to deal with the problem of flame trolling,' if it was used. Such overlegislating is a price to pay for MPs failure to pay attention to the generalisability of law and its flexibility in not requiring the enacting a new law for every change and or convergence in technology [6] . If one were to look at Figure 1, it can be seen that the marginal increase in interest in trolling and cyberbullying is proportional to the increase in interest in social networking and social media. Correspondence in the Liverpool Daily Echo showed that one of the MPs calling for additional laws on flame trolling assumed that trolling only started when the term, 'social media' became popular [7], which was when the political classes started taking interest in non-text based social networking to "connect with voters" and 'get their message out', most notably by US president Barrack Obama.

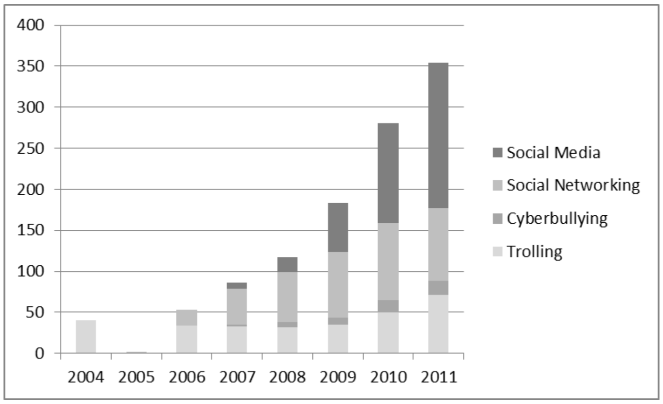


Figure 1 Trends of search terms on Google and Google News

### 1.1 Trolling in Great Britain

In fact, the term 'trolling' was probably first published in the Internet dictionary, Netlingo [8], which it was defined as an 'act of posting a message in a newsgroup that is obviously exaggerating something on a particular topic'. The pressure in 2011 to take extra measures to tackle 'trolling' was as a result of some high profile episodes of Internet abuse and cyberbullying. A few high profile cases brought attention to trolling as a term to reflect all cyberbullying. Celebrity Jade Goody's vandalised memorial page was the first case to be prosecuted under the MCA, with He was also banned from using online public communications services for five years. The second high profile case was that of Georgia Varley, a Liverpool teenager whose family was subject to RIP Trolling after a memorial page was set up to remember her. The local MP representing Gerogia, Steve Rotheram, vowed to introduce legislation to "see a greater conviction rate for those guilty of this vile practice" leading to her mother, Paula Varley saying in support, "We are hopeful that Steve Rotherham can push the legislation through. If something comes out of this then it could save another family the heartache we have had to go through". This shows the problem of lack pf awareness of the available legislation in relation to trolling, as even though the term has doubled in use since 2006, the connection between the 'new term' and 'old crime' has not been easily made. Something that will be addressed by this paper.

For instance, the Telecommunications Act 1984 can be widely applied beyond its original scope and in fact the Malicious Communications Act 1988 has seen a surge in use since the introduction of Facebook, being used to penalise many people for 18 weeks in prison for defacing memorial pages to people who had died, known as 'RIP Trolling'. It may be the case that the Internet was not high on the radar of British central government for decades after its potential was envisaged around this time. Indeed, even as the Worldwide Web was being launched on 6 May 1994 Hansard reports Welsh Secretary John Redwood apparently downplaying the role of the Internet, as follows: 'My Department is considering the use of the Internet and other electronic services initially for the transmission to non-media outlets of press notices. The possibility of including other departmental information would be considered after this first stage.' Even though it was apparent there was a lack of understanding of the Internet by the British Government in the 1990s, that didn't stop pressure being put on them by MPs. For instance, Hansard on 15 October shows future Home Secretary David Blunkett asking the then Home Office Minister what action the British Government was taking to 'outlaw the downloading and publication of obscene material from the Internet which would be illegal if produced in printed form.' This pressure was eventually realised with the passing of the Criminal Justice and Police Act 2001 and Communications Act 2003 (CA2003) by Mr Blunkett's New Labour Government.

## 2 Towards a framework for classifying flame trolling

Please Great Britain has an extensive array of laws that can be used to tackle flame trolling. At the moment it is quite sporadic. In March 2012 the CA2003 had only been used twice to successfully prosecute flame trolling, whereas the older Acts such as the Telecommunications Act 1984 (TA) and Malicious Communications Act 1988 (MCA) had been used more frequently.

Table 1. A Framework for identifying severities of trolling

| Trolling type | Severity | Flow / Involvement | Appropriate legal provision |
|---|---|---|---|
| Playtime | Minor | High Flow / Low Involvement | Fixed penalty notice £75 |
| | Major | Med Flow / Low Involvement | Fixed Penalty Notice £150 |
| Tactical | Minor | Med Flow / High Involvement | Common law detention for breach of the peace |
| | Major | High Flow / High Involvement | ASBO under s.1 the Crime and Disorder Act 1998. |
| Strategic | Minor | High Flow, Med involvement | Fine under s.43 of the Telecommunications Act 1984 |
| | Major | High Flow, Med Involvement | 18 weeks for each guilty act under s.1 of the Malicious Communications Act 1988 |
| Domination | Minor | Low Flow, Med Involvement | Restraining Order under s.5 of the Protection from Harassment Act 1997 |
| | Major | Low Flow, High Involvement | 6 months for each guilty act under s.127 of the Communications Act 2003 |

There is a strong case therefore for using these and other Acts more strategically so that those with less severe sentences are used for Minor flame trolling and then Major flame trolling is subject to harsher sentencing on an equivalent basis.

In this context Minor refers to a guilty act where there is a less serious guilty injury (malum reus), such as a standard flame.

Major on the other hand refers to a guilty act where there is a more serious guilty act, such as inciting racial hatred as in the case of R v Stacey (2012) who was a student at Swansea University, or otherwise where the actions are persistent (i.e. meet pertinax reus), or both. Table 1 above provides a possible strategic framework for determining when to use particular statutes for trolling.

### 2.1 Applying the framework to British cases

The Applying Table 1 to actual cases of flame trolling, one can see how better judgments could have been made had this framework been in place and accepted by the public prosecutor (CPS) in Great Britain.

In the case of minor Playtime Trolling, a clear case would be that of Jamie Counsel who posted and then removed a page on Facebook calling for riots. He was jailed for his Twitter post, yet would only be fined £75 if Table 1 was implemented, as he took it down straight after realising his in-the-moment mistake for which no one was harmed.

Tactical Trolling is where someone may be caught up in the situation, but they could chose to avoid posting if they want to, as it requires more effort than in Playtime Trolling. An example of Major Tactical trolling would be that of Liam Stacey, who in 2012 was sentenced to jail for racist abuse. He deleted the posts after realising the consequences of his actions, which he didn't mean, making the abusive comments only following reprisals to an earlier post from other users.

Strategic Trolling is where someone goes out of their way to troll others, yet without acquiring any new equipment or substantially creating any new content in order to carry out the act. One of the earliest common forms of Strategic Trolling was the concept of 'happy slapping'. This was where people would video themselves attacking others on public transport, or pulling people off their bicycles for instance, and then posting this video to the Internet. Strategic trolling can be as mild as someone intentionally going on to a social networking service to cause mischief, to organising formally or informally with others to abuse people. This can include setting up 'flame-wars' where they subject people to a tirade of abuse.

Domination Trolling on the other hand is where someone goes out of their way and puts in a huge amount of effort to troll another person. A recent UK example of this was that of British broadcaster Richard Bacon. Mr Bacon was tormented by a flame troller who was obsessed with doing all they could to attack him. They sought out many avenues for using the Internet to abuse him, including creating a dedicated website called, '"Richard Bacon is a [expletive]". Another example, that of Sean Duffy was not confined to one person, but going to massive efforts to seek out and abuse people who had died for their families to see. Duffy, from Reading, trolled the 'R.I.P. pages' of Natasha MacBryde and Sophie Taylor, both teenagers. In both cases he was charged under the Malicious Communications Act 1988, which is best used for Strategic trolling, which is often a one off yet severe incident. As he was harassing many individuals the Protection from harassment Act 1997 may not be appropriate meaning the Communications Act 2003 would be most effective with its ability to impose long jail terms to keep the public safe. It was used for this in the case of R v Counsel (2011) where the defendant was sentenced to 4 years for trying to incite riots on Facebook.

## 3 Conclusions

While trolling has existed as a term since the 1990s and as a reality even earlier there has been an exponential increase in the prevalence of the abusive kind - 'flame trolling'. Mistakenly the media calls these flame trollers, 'trolls', when in fact there are more often than not 'Snerts' and 'E-Vengers'. Despite being one of the greatest innovations of all time, the Internet is less safe than one might have expected it to be considering one is often simply sitting in front of a computer screen at one's home. Internet abuse is rife and despite differences in how it is defined and the specific criteria used, there is general agreement that IA can be explained in terms of the negative effect of Internet use, that is, Internet use that causes disturbances in an individual's life. The justice system in Great Britain has taken a sporadic approach to dealing with flame trolling, and the wide range of legislation that has existed since the 1980s has no strategic method to assign its usage on the basis of the nature of the flame trolling as its use often depends on the whim of different police forces.

. The pressure in 2011 to take extra measures to tackle 'trolling' was as a result of some high profile episodes of Internet abuse and cyberbullying. A few high profile cases brought attention to trolling as a term to reflect all cyberbullying. Celebrity Jade Goody's vandalised memorial page was the first case to be prosecuted under the MCA, with He was also banned from using online public communications services for five years. The second high profile case was that of Georgia Varley, a Liverpool teenager whose family was subject to RIP Trolling after a memorial page was set up to remember her. The local MP representing Gerogia, Steve Rotheram, vowed to introduce legislation to "see a greater conviction rate for those guilty of this vile practice" leading to her mother, Paula Varley saying in support, "We are hopeful that Steve Rotherham can push the legislation through. If something comes out of this then it could save another family the heartache we have had to go through". This shows the problem of lack of awareness of the available legislation in relation to trolling, as even though the term has doubled in use since 2006, the connection between the 'new term' and 'old crimes' has not been easily made.

From this it is clear to see that there is not enough known about the laws that exist to deal with trolling, what to use them for and when to use them. By providing a framework to assess the severities of flame trolling this paper has gone some way to providing solid guidance to law enforcement authorities so that the appropriate action is take, and so that the Great British justice system is not used to 'set examples', but to deal with the problem fairly as one would expect in a true democracy.

# 4 References

[1] J. A. Cannataci & J. P. M. Bonnici. "The UK 2007–2008 data protection fiasco: Moving on from bad policy and bad law?"; International Review of Law, Computers & Technology, 23., 1, 47-76, 2009.

[2] A. Calcutt. "White noise: an AZ of the contradictions in cyberculture". Palgrave MacMillan, 1999.

[3] N. Lacey, C. Wells & O. Quick. "Reconstructing criminal law: text and materials". Cambridge Univ Pr, 2003.

[4] T. Fitzpatrick. "Critical theory, information society and surveillance technologies"; Information, Communication & Society, 5., 3, 357-378, 2002.

[5] L. Lessig. "The law of the horse: What cyberlaw migsht teach"; Harvard law review, 113., 2, 501-549, 1999.

[6] J. Ubena. "Legislative Techniques and ICT: In the Wake of Law Keeping Pace with Technology". National Research Institute of Legal Policy Research Communications 105 2008. p.171. .

[7] S. Rotheram. "We need to stop this vile trolling". Liverpool Daily Post 2012. p.18. .

[8] E. Jansen & V. James. "NetLingo: the Internet dictionary". Netlingo Inc., 1995.

# SESSION

# COMPUTER AND INFORMATION SECURITY

## Chair(s)

### Dr. Kathy Liszka

# A Theoretical Model for Probabilistically Based Detection and Mitigation of Malware Using Self Organizing Taxonomies

Gregory Vert
Department of Computer Information Systems
Texas A&M Central Texas
(206) 409-1434
greg.vert@ct.tamus.edu

Anitha Chennamaneni
Texas A&M University Central Texas
(254)519-5463
anitha.chennamaneni@ct.tamus.edu

S.S. Iyengar
FIU School of Computing and Information Sciences
Florida International University
Miami, Florida
Iyengar@cis.fiu.edu

**Abstract** – Initial identification of attacks on computer systems is crucial to defending against them. The wide range of malware attacks available makes detection and defense a difficult prospect when looking at software vulnerabilities only. However, if the data space is reduced to only look at response from hardware based system state variables, it is possible to detect a wide range of unknown malware, simply by looking at how malware affects state variables describing hardware operation. In this paper, we present a developing theoretical model for the detection of unknown malware by probabilistic time series based modeling. Our model lends itself to a new method for responding to an attack based on fuzzy similarity matrices that lead to dynamic classification of unknown malware using self-organizing taxonomies. Such methods can be made adaptive to reflect general trends in system state variable data over time.

*Key-Words:* - Taxonomies, Probabilistic, Self Organizing, Computer Security, Complex Systems

## I. INTRODUCTION

With the proliferation of malware, increasingly more sophisticated approaches have been needed for its mitigation. It has been estimated that about 10-20 new viruses appear daily [1].

Many information resources are available that will notify users of new security holes on a subscription basis [2,3,4,5]. There are also several security databases that will let users browse the vulnerabilities for various software packages [6,7]. Companies such as Symantec, Security Focus and CERT keep large databases of known attacks [6,7,11,12]. Symantec has over 50,000 entries for known internet security related threats [13]. With the proliferation of new viruses daily, these databases will soon become unwieldy. Yet, with all of the above, the malware problem only seems to be growing in size instead of going away. Part of the reason for this is that security, even with all the information about threat and vulnerability is still a very manual operation in response and mitigation. Few systems currently have the ability to mount an automated mitigation to malware activity partly because of the wide range of software vulnerabilities malware attempts to exploit.

## II. PROBLEM FORMULATION

There has been research attempting to classify different types of attacks, from Unix specific vulnerabilities [8,9] to network attack assessment [10]. This research has been important and useful, but their classification has focused on a specific class of attacks. Even more key classification is that classification often is focused on software vulnerabilities. This has lead to a high dimension data space that describes the key attributes of an attack. This is often referred to as the needle in the haystack problem. What has been needed is a method to reduce the high dimension data space of software vulnerabilities to a lower dimension equivalency. Such a reduction can be found by looking at state variables describing the operation of software and malware on the underlying hardware. This is a new approach developed in the model theory presented in subsequent sections.

## III. PROBLEM SOLUTION

Attack databases, such as Security Focus' Vulnerability Database [6] and NIST's ICAT [7], list information about reported attacks. Once an attack is known, it often becomes clear about how to mitigate its effects the next the malware operates. The problem is trying to figure out and anticipate the unknown malware that next arrives. With the exploding proliferation of malware anticipation of the unknown is a steep challenge. The key is to find a method to for dynamic classification of unknown malware. Dynamic classification implies self organization in classification schemes. Very little work has been done in this area because of the question of how to create classification instances and sub instances in an automated fashion. One approach that is part of our model, describes a methodology that can potentially be used to classify unknown attacks and subsequently respond and mitigate their threat with the use of self organizing taxonomies.

Hierarchies are a good paradigm for organizing data. Taxonomy is an important tool in organizing data. Taxonomy makes it easier to navigate, access and maintain data. However, the construction of taxonomies manually is time consuming, cumbersome, expensive and inefficient. Besides, in a dynamic setting, where change is a constant, taxonomy needs to adapt constantly. The first part of our model proposed a good taxonomy for characterizing security hardware vulnerabilities that is unique, distinctly different from other taxonomies found in the extant literature which focuses on software vulnerabilities. As stated previously focus on hardware and its state variables is believed to reduce the dimension of the data space describing malware operation against software. Given the expensive and unmountable nature of manual intervention, later work presented in this paper may have to found a way to construct the taxonomy in a completely automated fashion.

*3.1 Attack Attributes:* The first step in taxonomy development was to come up with a naming notation for hardware based state variables. We have developed a list of attributes in previous work to describe hardware state variables and relate them in a fashion that would support taxonomic trees development. An example of the notation is:

*Network.Protocol.TCP.InPorts*

For attributes that have multiple values, we separate the values by commas, and define ranges using the "En dash" character (–). For example, a TCP port scan that targets ports 25 (SMTP), 80 (HTTP), and 1024 through 6000 would be defined as:

*Network.Protocol.TCP.InPorts = 25, 80, 1024–6000*

In previous work our list of state variables included 22 separate hardware state variable that are thought to change upon the execution of malware on the hardware. . For the sake of brevity this list is not included in this paper.

*3.2 Node Actionable Response Rules (NARRs):*Complex systems theory makes a general supposition that complex behaviors can be created through the composition and action of a series of smaller, simpler rules. In this model actionable rules are associated with state variables mentioned in section 3.1, and are simple in form. Our taxonomy is built by clustering state variables into taxonomic nodes with simple actionable rules to mitigate the abnormal operation of a variable. As a taxonomy is processed with current system state information to classify malware operation, these simple rules at each node in the tree are applied to system operation if abnormal system operation has been detected. These are referred to as Node Actionable Response Rules (NARR) and will be integrated into our taxonomic model later in this paper. For example a rule in the Network taxonomy (Figure 1.) might be:

*Network.TCP.InPort n= High: NARR = block n*

or for the CPU

*CPU Usage = High: NARR = Kill process with high use*

These action rules are progressively applied to the system as processing drops through layers of the taxonomies. After each application system operation is evaluated to see if it has returned to normal. If not, NARR's are applied as processing and classification continues deeper into the taxonomy. Effectively, this model starts trying to mitigate or counter attack malware operation. More will be presented about this in section 4.

*3.3 Attack Taxonomies:*As stated in section 3.1, the first type of taxonomy developed is based on common attributes of attacks that originated through a remote connection across a network as shown in figure 1.

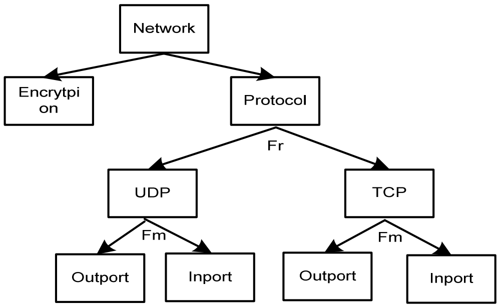


Figure 1: Network Attribute Taxonomy

Figure 1 presents the *network attribute taxonomy*. In our hierarchy, child nodes inherit all the attributes and descriptive properties of their parent nodes, as well as having node specific attributes. The network attributes specified in the tree help to define attacks based on the protocol, bandwidth, and action characteristics of the attack.

In our research we also found an entire category of attacks on files and file systems as mentioned above. The *file system taxonomy* (Figure 2) was developed to structure and organize this data into a taxonomic model. The attributes in this tree define what files on the victim's machine are created, changed, and deleted.

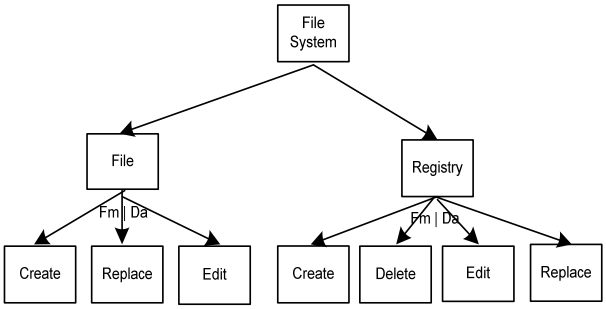


Figure 2: File System Attribute Taxonomy

Another category of attacks are based on system exploits. Figure 3 presents the *exploit attribute taxonomy* which was referenced in section 3.1. The exploit tree defines the vulnerability that an attacker may use on a victim's machine. This taxonomy models common programming errors, improper configurations, and user errors.

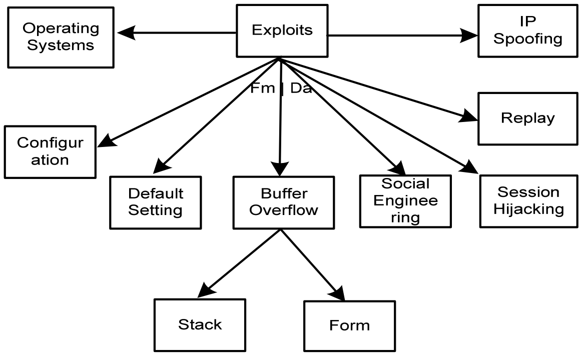


Figure 3: Exploit Attribute Taxonomy

Finally, attacks exist that use services and drivers to gain elevated system privileges [14]. The *kernel taxonomy*, mentioned above, is presented in Figure 4 and shows the types of attacks that are possible using kernel privileges.

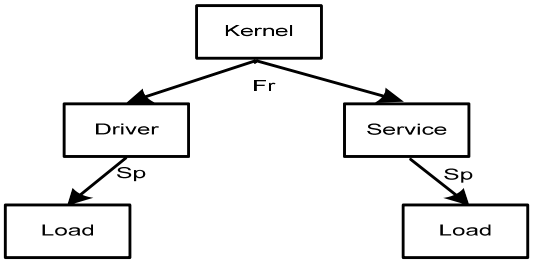
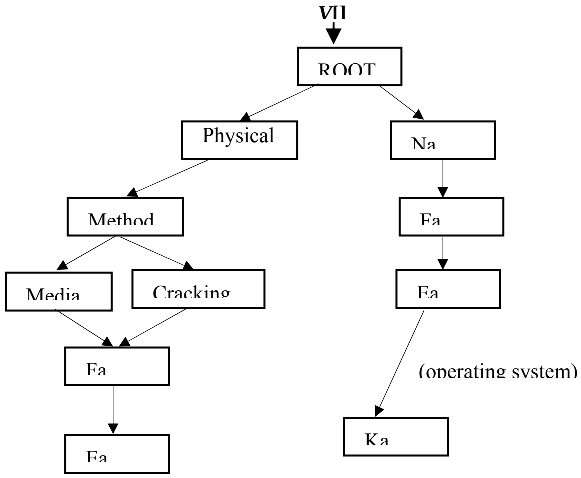


Figure 4: Kernel Attribute Taxonomy

Our initial effort was to consolidate all of the above hard state variable based taxonomies into a single taxonomy produces what might be referred to as a taxonomic graph. This taxonomy is shown in Figure 5, where each box represents the sub taxonomies presented in Figure 1 through Figure 4.



Ka – Kernel attribute tree
Ea – Exploit attribute tree
Na – Network attribute tree
Fa – File attribute tree

Figure 5: Consolidated Taxonomic Graph

In Figure 5 all leaf nodes are connected to the next subtaxonomic tree's root except for the right subtree connection from Ea to Ka. In this case, only the "operating system" node of the Exploit subtree is connected to the root of the Kernel subtree.

Input vector ***V*** is an *n*-dimensional feature vector whose attributes describe the current system data variable data used to populated the sub taxonomies. It might reflect normal operation or it might reflect malware operation. This vector contains the same attributes as those used in the subtrees when selecting and moving to the next child node. At this point in the development of our research several thing were realized:

*i) an attack can actually branch to two or more child nodes in a subtree or two or more subtrees in the consolidated taxonomic graph. The reason is that attacks are typically multi-pronged in their approaches.*

*ii) the structure of the taxonomic graph could be different between attacks and needs to be able somehow to dynamically self organize.*

## IV APPROACH

As mentioned when examining the above taxonomy in figure 5, it was realized that it has limitations, among them are *that the static nature of the taxonomy may limit its ability to detect unknown types of malware operation on the system*. What was needed was a taxonomic model that had the following properties:

*i) the taxonomy algorithm can self organize*

*ii) the method of malware operation detection*
*could be probabilistic and quantitative*

*iii) the method should be adaptive*

Borrowing on the work developed, it was realized that the approach of looking at hardware and state variables describing hardware operation was a good approach because it limited the data space that needed to be evaluated for malware operation. In other words there can be numerous types of vulnerabilities in software but the fundamental operation of hardware was probably going to have a typical type of signature for *classes* of vulnerabilities thus making detection simpler if it was focused on hardware response to malware. This relationship can be mapped as the following:

$$|h_r| < |s_v|$$

where:
$|h_r|$ is the cardinality of hardware malware state variable response for a given attack
$|s_v|$ is the cardinality of software vulnerabilities response for a given attack

In other words, we felt that it is simpler to monitor for malware activity by watching for state changes in hardware state variable rather than anticipate every possible way software can operate when malware is affecting it. What has been developed based on this concept is based on time series analysis, standard deviation over time for hardware state variables and affinity analysis in the creation of self organizing taxonomies.

*4.1 Probabilistic Malware Detection:*The first component of the theoretical model argues that standard deviations of a state variable ($s_v$) (same as a hardware state variable) over time is an indication of malware in operation, ergo, malware starting to operate. An example of this would be cpu usage might typically have an average (μ) utilization of 75% and 38.2% of the time falls within .05 standard deviation of μ = .75.

$$38.2\% = \mu \pm .05\sigma$$

For a given hardware state variable ($h_s$) a time series of standard deviations can be calculated with a time window such that the series can produce σ as it changes over time. Note the times series is expected to produce a normal distribution unless there is a malware attack in progress. This could look like the following

| $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ …….. $t_n$ |
|---|---|---|---|---|
| σ | | | | |
| 1.2 | 1.4 | 1.1 | 1.2 | 1.25…… |

*where:*
$t_n$ - is σ at time n

*and:*

the calculation of σ is done for $h_s$ (e.g. cpu usage) using a system specified time slice that ranges in the form of:

$$t_n \pm t_s$$

where:
$t_s$ - is a specified time slice e.g. 60 seconds

In the above, the if it is 1:10pm, the sampling time slice would run for example from 1:09 – 1:10. During which perhaps every six seconds the cpu usage would be collected and a standard deviation calculated over this population of 10 samples.

Using the above concept of sampling to calculate standard deviation of given state variable over a time window, for all $s_v$'s defined in a system to following would be an examples of many state variables being sampled and the standard deviation for a time slice then being stored::

| $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ …….. $t_n$ |
|---|---|---|---|---|
| $s_{vi}$ σ | | | | |
| 1.2 | 1.4 | 1.1 | 1.2 | 1.25…… |
| $s_{vj}$ σ | | | | |
| 6 | 4.5 | 5 | 5 | 5.7…… |
| $s_{vz}$ σ | | | | |
| $z_1$ | $z_2$ | $z_3$ | $z_4$ | $z_5$…… |

In the above example the calculated σ stay relatively constant around a similar values for a given $s_{vi}$ with minor predictable jitter.

Using the above time series during malware initiation and activity e.g. cpu usage, the time series model would look very different. If cpu usage had a μ of 50% ± .05σ and malware operation drove the usage to 97% the calculated value over the time window hypothetically will change. As an example:

| $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ …….. $t_n$ |
|---|---|---|---|---|
| σ | | | | |
| 1.2 | 1.4 | 1.1 | 5 | 1.25…… |

could be what is observed indicating initialization of malware operation at $t_4$. From the previous table where multiple hardware state variables were being watched we can imagine the following pattern where the malwares original signature initiates in the cpu and them at the next time moment moves into driving disk usage up. Such an example might look like the following:

| $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ …….. $t_n$ |
|---|---|---|---|---|
| $s_{v\ spu\ utilization}$ σ | | | | |
| 1.2 | 1.4 | 1.1 | 5 | 5…… |
| $s_{v\ disk\ usage}$ σ | | | | |
| 6 | 4.5 | 5 | 5 | 9…… |

What the above demonstrates is the following:

*i) $t_4$ shows a change in its σ calculation for cpu usage*

*ii) $t_5$ shows a continued elevated σ for the $s_{v\ cpu\ usage}$ and there is now elevation in the σ for $s_{v\ disk\ usage}$.*

The above is the expected time based probabilistic quantitative signature of malware initiation and operation for only two hardware state variable. This method could be scaled to many state variables creating a sophisticated detection mechanism in a relatively lower dimension data space.

*4.2 Taxonomic Self Organization based on σ and Fuzzy Similarity Matrices:* In the previous section, the model demonstrates the potential ability to detect malware operation in a probabilistic and time based dynamic fashion. Thus, when the time series σ state variable analysis detects suspicious changes in system system, our model suggests a way malware can be mitigated using the response rules mentioned previously for the state variables and self organizing taxonomies.

The second component of the model argues for self organization of the state variables found in the previously presented taxonomies so that the response rules (NARR) can be applied to counter the attack. This can be done also quantitatively and adaptively by the creation of a similarity matrix. In this model the state variables from the previous static taxonomies, no longer are owned by a sub tree. Instead they are clustered by association into taxonomic nodes forming the various levels of the taxonomic graph. This is self organization and can be accomplished via the use of similarity analysis of state variables and their σ changes flagging malware operation.

A similarity matrix is used often in fuzzy set theory to indicate that some variable has a degree of relation to another variable in a set [16]. This theory organizes data into sets by the degree of relation. Set membership traditionally is denoted by a characteristic function of the form:

$$u(n) = \left\{ \begin{matrix} 0| & \text{if n} \neg \text{memberof set} \\ 1| & \text{if n is fullmember,crisp}| \\ [0..1]\| & \text{membershipfunction} \end{matrix} \right\}$$

In section 4.1, the detection of malware operation was done with the use of state variable changes in the above time series detection of malware operation using state variable σ. This can then be utilized in conjunction with fuzzy similarity based matrices to create self organizing taxonomies.

The first step in this process is to dynamically build (train) a similarity matrix based on the degree that a change in one state variables σ correlates with a change in another state variables σ. Using the fuzzy function above, this results in the following logic:

*i)* $\Delta s_i \rightarrow \Delta s_j \quad | \; u(s_i,s_j) = 1$

*ii)* $\Delta s_i \neg \rightarrow \Delta s_j \quad | \; u(s_i,s_j) = 0$

*iii) ii)* $\Delta s_i \approx\!\rightarrow \Delta s_j \quad | \; u(s_i,s_j) = [0,,1]$

In more basic terms the above means that if a change in $s_{vi}$ *always* results in a change in $s_{vj}$ then the similarity matrix value is 1, if is *sometimes* results in a change then it is results in a similarity matrix value of:

$$0 < u(s_i,s_j) < 1$$

*where:*

$u(s_i,s_j)$ = can be calculated based on the
percentage of times when $s_{vi}$ changes that
there is a corresponding change in $s_{vj}$

Building on the above foundation, the similarity matrix for $s_{\text{cpu usage}}$ and $s_{\text{disk usage}}$ might look like the following:

Table I
A sample similarity matrix for state variables where similarity is calculated base on the percent of times an change in one state variable results in a change in another variable

| | $s_{v1}$ | $s_{\text{disk usage 2}}$ | $s_{\text{cpu usage 3}}$ | $s_{v4}$ | $s_{v5}$ | $s_{v6}$ | $s_{vj}$ |
|---|---|---|---|---|---|---|---|
| $s_{v1}$ | 1 | 0 | 0 | .1 | 0 | 0 | 0 |
| $s_{\text{cpu}}$ usage 2 | .86 | .9 | 1 | .9 | .82 | .6 | .5 |
| $s_{\text{disk}}$ usage 3 | .81 | 1 | .2 | .7 | .3 | 0 | 0 |
| $s_{v4}$ | .7 | 0 | .8 | 1 | 0 | 0 | 0 |
| $s_{v5}$ | … | | | | | | |
| $s_{v6}$ | … | | | | | | |
| $s_{vi}$ | … | | | | | | |

Table I is a partially populated table of state variable similarity values. In the above table, state variables always have a value of 1 as membership with them selves. The above does illustrate a few curious findings:

i) $\Delta s_{\text{cpu usage}} \approx\!\rightarrow \Delta s_{\text{disk usage}} = .9$ (%) $\therefore$ $s_{\text{disk usage}}$ R $s_{\text{cpu usage = strong}}$

ii) i) $\Delta s_{\text{disk usage}} \approx\!\rightarrow \Delta s_{\text{cpu usage}} = .2 \therefore s_{\text{disk usage}}$ R $s_{\text{cpu usage = weak}}$

The above illustrates the fact that relationships (R) in similarity are *not symmetric*. From the above table, an increase in cpu usage usually leads to an increase in disk usage (.9), however, the opposite is not true (.2) that an increase in disk usage would result in an increase in cpu usage.

The values in the table are calculated based on *u(n)* and reflect the % of times a change in one state variable results in a change of another state variable.

*4.3 Taxonomic Creation and Application to Offensive Mitigation of Malware Threat:* The values in table 1 meet the adaptivity criteria of our model in that it can be calculated and recalculated dynamically as the system operates. The above values then become the basis for the creation of the self organizing taxonomy once the time series σ analysis of state variables suggests the need to create a taxonomy with NARR rules in it.

Reiterating, from the previous discussion about the Node Actionable Response Rules (NARR), each particular state variable can have a very simple rule associated with it, and the state variables around found in the static taxonomic nodes. For instance:

*NARR 1: if cpu usage > σ r1 = kill largest process*

*NARR 2: if disk usage > .5 σ r2 = don't allow deletion of files (e.g. malware could be doing a recursive delete of the file system)*

*where:*
*NARRn - node actional rule*
*$r_n$ – the specific action to apply for the rule*

The similarity matrix shows how to construct dynamically a self organizing taxonomy with the NARR rules associated with each state variable. In this method, state variables are not statically assigned to nodes as they were in the previous section. Instead state variables are clustered into nodes at various levels in the tree based on their similarity relations. NARR's are also clustered with their state variables at various nodes in the self organizing taxonomy. The following is the clustering algorithm that builds the taxonomy:

```
//create the root node of the taxonomy
if $s_{vi}$ change > system set σ for $s_{vi}$
    gather all $s_v$'s from the similarity matrix > .9
      from the row $s_{vi}$
    place them in the taxonomic root node
      with their NARR's and response rules
```

```
//create the children nodes, stepping for similarity
//values classes  n < level < m,  e.g. .9top - .8bottom
// level 1,...
for 0 < similarity values < 1
     calculate level boundaries for top and bottom
        values of the class level

     create new child node$_i$
     gather all $sv_j$ values in the row $sv_i$ that fall within
       current level top and bottom ranges

     assign all collected NARR for $sv_{j's}$ to node level$_i$
```

The above algorithm, given the similarity matrix in table 1, might have the following structure upon completion. The initiation of the trees construction starts when the time series probabilistic analysis of a given $S_{cpu\ usage2}$ σ > system threshold σ (triggering tree construction):

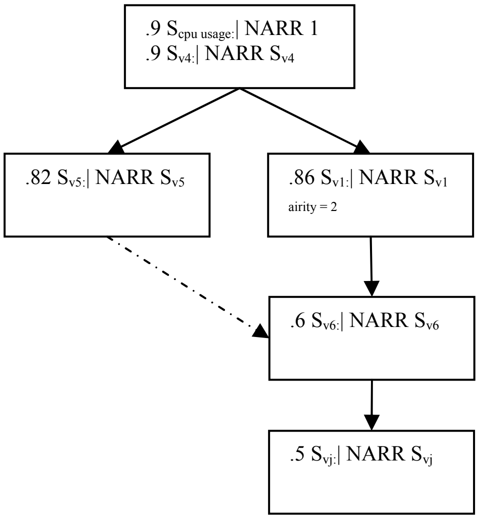

.9 $S_{cpu\ usage:}$| NARR 1
.9 $S_{v4:}$| NARR $S_{v4}$
.82 $S_{v5:}$| NARR $S_{v5}$
.86 $S_{v1:}$| NARR $S_{v1}$
airity = 2
.6 $S_{v6:}$| NARR $S_{v6}$
.5 $S_{vj:}$| NARR $S_{vj}$

Figure 7. Sample taxonomy constructed from similarity matrix (airity = 2)

The above tree may be augmented with what informally we are calling airity. For the sake of this research, airity is defined to be:

*Definition: Airity is the number of child nodes at the same level in a taxonomy where the level is a similarity range and that range is sub divided mathematically into sub ranges and NARR response layers. All child nodes at a given level have a sub class NARR set applied, system measured and then the next airity of a level in the tree is applied.*

Of note in figure 7, is the use of the airity concept at level 2 in the tree to create two nodes. Airity divides a levels range and used to partition the response rules such that the rules first applied have the highest similarity value. In this case the level 2 node with the values ranges .85 < similarity < .9, the right most child node would be applied first in the follow sections algorithm .

A matter of further research is how does a node at one level link to the child node of a sibling's at the save level in the tree. This is shown with a dashed arrow in Figure 7.

Additionally multiple state variables may have abnormal σ's from the time series probabilistic analysis method developed earlier. It is a matter of further research about if new stand alone taxonomies should be created for these case or if they are hooked somehow into an existing taxonomy using the magnitude σ's to determine where and how trees are joined together.

*4.4 Application of the taxonomy to Threat Mitigation:* Once the similarity based, self organized taxonomy has been created, it has in theory has the ability to start mitigation efforts against the malware through progressive application of state variable NARR rules as deeper and deeper nodes in the tree are procesed. The idea is to *only apply enough NARR rules to return the systems operation to normal as measured by σ*. Thus the general method for this is to:

i*) apply a nodes, NARR's*

*ii) measure system response*

*iii) if normal, abort further NARR application,*
*otherwise proceed to next child node and iterate*

The pseudo code algorithm for this might be of the form:

```
// general algorithm for application of the nodes
//NARR rules
for root node to child node$_j$
     execute NARR rules for taxonomic node$_j$
      measure system response
     if system response (σ) normal
               exit
     else
          proceed to next child node on same level or
            next child node
        iterate
```

Interestingly the above model should create various organizations of state variables in different nodes over time as they respond to different types of unknown malware attacks. This is where it is believed that the model can detect and respond to previously unknown malware. The summing multiple taxonomies for unknown and mitigated software over time has the potential to create elaborate, self adaptive, system specific powerful taxonomies..

## V CONCLUSION

The wide range of malware attacks available makes detection and defense a difficult prospect when looking at software vulnerabilities. However, if the data space is reduced to only look at response from hardware based system state variables, the data space can be reduced and offers a possibility to detect a wide range of unknown malware, simply by looking at how it affects state variables describing hardware operation.

Identification of state variable behavior and mal-behavior can be done probabilistically, with time series based monitoring of standard deviations for a given state variable. In conjunction, this information, when questionable state operation data is found σ, coupled with a similarity matrix, the model offers a method for dynamic creation of malware classification taxonomies. Borrowing from complex systems behavior, nodes in the taxonomy can have very simple rules associated with them that can create complex responses to threat and offer the possibility of a system that has the capacity to mitigate malware operation in an automated and adaptive fashion.

This initial work is being further refined and developed. It is at a theoretical point currently with many further research questions to be investigated. The future includes building a small prototype to determine how well the model actually works an to refine its theory of operation.

*References:*

[1] Ducklin, Paul. The ABC of Computer Security. Retrieved April 12, 2003, from http://www. sophos.com/virusinfo/whitepapers/abc.html

[2] Symantec Corporation. Security Response. Retrieved March 15, 2003, from http:// securityresponse.symantec.com/

[3] SecurityFocus. What is BugTraq? Retrieved March 15, 2003, from http://www. securityfocus.com/popups/forums/bugtraq/intro.shtml

[4] NTBugTraq. NTBugTrack Home. Retrieved March 16, 2003, from http://ntbugtraq.ntadvice. com/

[5] SANS Institute. Computer Security Education and Information Security Training. Retrieved March 20, 2003, from http://www.sans.org/

[6] SecurityFocus. Vulns Archive. Retrieved March 12, 2003, from http://www.securityfocus.com/ bid

[7] National Institute of Standards and Technology. ICAT Metabase. Retrieved March 13, 2003, from http://icat.nist.gov/icat.cfm

[8] Taimur Aslam. A Taxonomy of Security Faults in the Unix Operating System. Master's Thesis, Purdue University, Department of Computer Sciences, August 1995

[9] M. Bishop. A taxonomy of unix system and network vulnerabilities. Technical Report CSE-9510, Department of Computer Science, University of California at Davis, May 1995.

[10] Shostack, Adam and Scott Blake. Towards a Taxonomy of Network Security Assessment Techniques, July 1999. Retrieved March 29, 2003, from htttp://razor.bindview.com/publish/ papers/taxonomy.html

[11] CERT. CERT® Advisory CA-2003-07 Remote Buffer Overflow in Sendmail. Retrieved April 2, 2003, from http://www.cert.org/advisories/ CA-2003-07.html

[12] Symantec Corporation. Backdoor.FTP_Ana.D. Retrieved April 13, 2003, from http:// securityresponse.symantec.com/avcenter/venc/ data/backdoor.ftp_ana.d.html

[13] Symantec Corporation. Security Response. Retrieved April 21, 2003, from http:// securityresponse.symantec.com/avcenter/search.html

[14] SANS Institute, Knark: Linux Kernel Sub-version. Retrieved April 24, 2003, from http://www.sans.org/resources/idfaq/knark.php

[16] Yen, John, Langari, Reza. *Fuzzy Logic, Intelligence, Control and Information*, Prentice Hall, 1999.

# Identifying Rootkit Infections Using a New Windows Hidden-driver-based Rootkit

**Woei-Jiunn Tsaur**[1] **and Lo-Yao Yeh**[2]

[1]Department of Information Management, Da-Yeh University, Changhua, Taiwan

[2]Network and Information Security Division, National Center for High-Performance Computing, Tainan, Taiwan

**Abstract** - *It can be observed that most sophisticated kernel mode rootkits implement hiding tasks via loading drivers in Windows. Also, more and more malware writers are taking advantage of rootkits to shield their illegal activities. Therefore, the role of a detector for effectively detecting Windows driver-hidden rootkits is becoming extremely important. In our previous work, we focused on developing a new Windows driver-hidden rootkit with five tricks based on the technique of DKOM (Direct Kernel Object Manipulation), which has verified that it can successfully avoid a variety of well-known rootkit detectors. In this paper, we extend our previous work by employing what we learn from the proposed new driver-hidden rootkit to explore remedies and solution for identifying not only the new threat but also other existing rootkits. It is expected that this research will contribute to the development of rootkit detection methods for unknown Windows hidden-driver-based rootkits.*

***Keywords:*** System security, Rootkit, Malware, MS Windows, Kernel mode

## 1 Introduction

The term "rootkit" represents stealth techniques which can hide information about computer resources furtively, and prevent itself from being discovered by system administrators. The evolution of rootkits can be roughly divided into four stages. At the first phase, a rootkit can alter Unix's log files to hide the presence of certain users. At the second phase, it not only alters log files but also replaces some important programs such as "*ls*", "*ps*" or "*netstat*" to hide users' activities. The main drawback of substituting programs is that it can be easily found by integrity checking programs such as Tripwire [1]. To overcome this drawback, the third phase is that rootkits aim to intercept user mode function requests and hide desired information in memory. It is not necessary to modify or replace files at this stage. In order to pursue higher stealth level, rootkits and detectors have transferred to the kernel mode at the fourth phase. The targeted operating system is also changed from the UNIX system to Microsoft Windows system. It can be observed that most of the sophisticated kernel mode rootkits are implemented as driver-style programs to execute hiding tasks in Windows.

More and more malware writers are integrating rootkits to shield their illegal activities [2]-[5]. Any computer security products that are not equipped with the anti-rootkit functionality may not identify this kind of threat. Therefore, the role of a detector for effectively detecting Windows driver-hidden rootkits is becoming extremely important. Though much research [6]-[17] has been focused on kernel data to develop schemes for finding malicious behaviors and undoubtedly they can effectively detect hooking based or virtual machine based rootkits in the Linux or Windows system, they cannot foresee what the result is when meeting unknown Windows DKOM (Direct Kernel Object Manipulation) based rootkits that are more difficult to be detected than hooking based ones [7], [18]-[20]. As King et al.'s research process [9], in our previous work [21] we first assume the perspective of the attacker, who is trying to run malware and avoid detection. Our previous work was focused on developing a new Windows driver-hidden rootkit with five tricks based on DKOM, which has verified that it can successfully avoid a variety of well-known rootkit detectors. By assuming the above perspective, in this paper we then hope to help defenders understand and defend against the threat posed by not only the proposed new type of rootkit [21] but also other existing ones. That is, this paper aims to present detective measures to guard the avenues of attack by developing a new detection technique for identifying driver-hidden rootkits using DKOM in the Windows system.

In summary, the contributions of our related works are threefold. First, we demonstrate the vulnerability of well-known Windows rootkit detectors by exploring the design and implementation of the proposed driver-hidden rootkit in our previous work [21]. Second, we extend our previous work by employing what we learn from the proposed new driver-hidden rootkit to explore remedies and solution for identifying not only the new threat but also other existing rootkits. Lastly, our efforts for identifying rootkit infections will be useful for stimulating detector developers to upgrade their software to enhance their detection techniques against such a wide variety of new threats in the Windows system.

The remainder of this paper is organized as follows. Section 2 surveys current rootkit creation and detection techniques. Section 3 presents the detection method for identifying a variety of Windows driver-hidden rootkits. Section 4 depicts the experimental results of testing the proposed rootkit's detection ability, and further discusses the advantages existing in the proposed detection approach. Finally, some concluding remarks are presented in Section 5.

# 2 Related work

In this section, we first present several existing rootkit stealth techniques, and then introduce the current useful rootkit detection techniques.

## 2.1 Rootkit stealth techniques

A kernel-mode rootkit can hide computer resources essentially by two different techniques. One is hooking that intercepts the requests of accessing resources. The other is Direct Kernel Object Manipulation (DKOM) that manipulates the data used by operating systems to keep track of resources. When using DKOM, rootkit writers need to clearly understand the data structures in kernel, but it can be more furtive than using hooking [7], [18]-[20]. The DKOM technique was first used in the FU rootkit and then used in the FUTo rootkit to hide drivers [18], [20]. In 2007, the DKOM-based Unreal rootkit was created and shown off that all of the famous detectors cannot detect it. At present, several well-known detectors like GMER and Rootkit Unhooker are capable of effectively detecting the above-mentioned three driver-hidden rootkits.

In 2010, Tsaur and Chen [21] proposed a new Windows hidden driver based rootkit to explore the weaknesses of several well-known rootkit detectors. Though much research has been focused on kernel data to develop schemes for finding malicious behaviors and undoubtedly they can effectively detect hooking based or virtual machine based rootkits in Linux or Windows, they cannot foresee what the result is when meeting unknown Windows DKOM (Direct Kernel Object Manipulation) based rootkits. Therefore, Tsaur and Chen [21] developed a new Windows driver-hidden rootkit with five tricks based on DKOM, and have verified that it can successfully avoid a variety of well-known rootkit detectors. Although Tsaur and Chen [21] indicated the weaknesses of current detectors and briefly discussed possible remedies and solution for detecting the proposed subtle rootkit, they do not implement effectively detective measures to defend against the threat posed by not only their proposed new type of rootkit but also other existing ones.

## 2.2 Rootkit detection techniques

As for identifying rootkits, there are two main approaches to develop rootkit detectors. The first detection approach targets hiding mechanisms by detecting the presence of API hooking [6], [7]. It is similar to the signature-based detection [22], and thus it cannot catch unknown rootkits whose signatures of hiding mechanisms do not exist in its signatures repository in advance. The second approach targets hiding behaviors by detecting any discrepancies between the original and the fake. BlackLight [23], invented by F-Secure, is designed to be a driver to detect hidden processes and files. It collects resources information from two different storage places of processes and files, and then compares each other to find rootkits. This approach is to belong to the cross-view rootkit detection [11]. It is noticed that in this approach both targeted information cannot be modified simultaneously by rootkits, otherwise a detector using this approach cannot distinguish the differences between the two places storing targets and then it must be useless. Although this method has the drawback, it does not need to maintain a signature database as used in the signature-based detection method, and can find unknown rootkits. Our proposed effective driver-hidden rooktit detector belongs to this category.

# 3 Proposed detection method for identifying Windows driver-hidden rootkits

In this section, we will introduce unknown DKOM-based rootkit identification techniques which may be taken into consideration by rootkit detection researchers and be studied by the security community, in order to help defenders strengthen their detection techniques for identifying a wide variety of unknown Windows driver-hidden rootkits.

Current rootkits and detectors are almost always through drivers in the kernel mode to execute their tasks. Thus, such a way is also adopted in the proposed solutions for detecting the sophisticated driver-hidden rootkit. In our previous work [21], the first four tricks included in the proposed driver-hidden rootkit are to independently hide drivers by respectively modifying the List_Entry structures of Object Directory, Object Driver, Object Device and PsLoadedModuleList. If a detector counts on these List_Entry structures, it is consequently unable to identify our proposed Windows driver-hidden rootkit, because the hidden driver has already disappeared from the modified doubly linked lists of the List_Entry structures and therefore it is impossible for a detector to easily find the hidden driver from these List_Entry structures. As Schuster's research process [24], we can search for drivers from Microsoft Windows memory dumps. This approach does not depend on the List_Entry structures, and is consequently able to resist the attack of DKOM modification to provide trusted information. Therefore, we can be based on such an idea to get reliable information about drivers. In order to detect a variety of new rootkits, we can employ the cross view based detection method that can find unknown rootkits and does not need to maintain a signature database as used in the signature based detection. It is noted that this method is to compare two information gathered from different sources to discriminate the discrepancies, where the two information represent the original one which is untouched or unknown by rootkits, and the fake one which may be altered or known by rootkits, respectively. The thoughts on proposing solutions for effectively detecting a variety of subtle driver-hidden rootkits are to collect drivers information from two different sources, and then to compare them to find the driver-hidden rootkit. One source is by searching kernel memory, and the other is through the PsLoadedModuleList structures.

Many utilities rely on the kernel function ZwQuerySystemInformation to get information about drivers. The function ZwQuerySystemInformation returns a list of loaded drivers in the kernel, and can also get the list of loaded drivers from a set of PsLoadedModuleList Objects which are necessary targets for rootkits to manipulate for hiding desired drivers. We traverse the set of PsLoadedModuleList Objects to get the list of loaded drivers as the fake information. After gathering the original information and the fake information, the cross view based rootkit detection can be employed to detect a variety of sophisticated driver-hidden rootkit. The details about these four detection steps are described as follows:

Step 1: As the proposed detector is loaded into memory, we can get its Object Driver that will be utilized in Step 2 for finding all possible drivers in the memory.

Step 2: The proposed detector finds all possible Object Drivers, and records to a list their driver addresses stored in the DriverStart member of each Object Driver. In the following, we will demonstrate how to find a possible Object Driver. Since the detector itself is also a driver, we can verify whether a candidate is an original Object Driver by employing the following rules:

Rule 1: Verifying whether the value of the Object_Header Type member at a candidate is equal to that at the detector, as shown in Fig. 1.

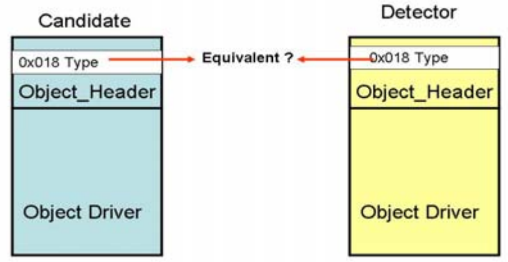


Figure 1. Rule 1 for finding Object Drivers

Rule 2: Verifying whether the value of the Object Driver Type member at a candidate is equal to 0x004 and whether the value of the Object Driver Size member at a candidate is equal to 0x0a8, as shown in Fig. 2.

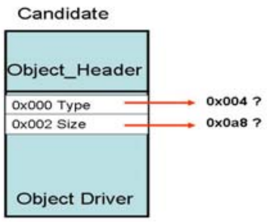


Figure 2. Rule 2 for finding Object Drivers

Rule 3: Verifying whether the value of the Object Driver HardwareDataBase member at a candidate is equal to that at the detector, as shown in Fig. 3.

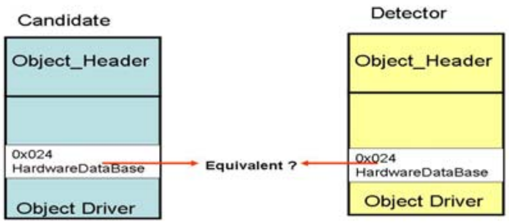


Figure 3. Rule 3 for finding Object Drivers

Rule 4: Verifying whether the value of the Object Driver DriverExtention member at a candidate minus the address of the Object Driver is equal to 0xa8.

Rule 5: Verifying whether the value of the Object Driver DriverSection member at a candidate is between 0x80000000 and the end point of the defined searching memory range, as shown in Fig. 4.

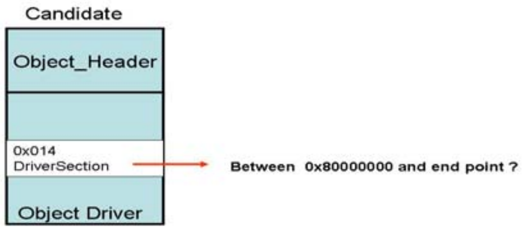


Figure 4. Rule 5 for finding Object Drivers

Rule 6: Verifying whether the value of the Object Driver DeviceObject member at a candidate is between 0x80000000 and the end point of the defined searching memory range, as shown in Fig. 5.

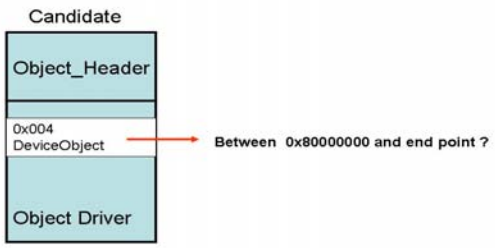


Figure 5. Rule 6 for finding Object Drivers

Rule 7: Verifying whether the value of the Object Device Type member at a candidate is equal to 0x003, as shown in Fig. 6.

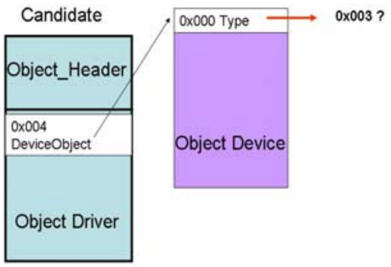


Figure 6. Rule 7 for finding Object Drivers

Rule 8: Verifying whether the value of the Object Device DriverObject member at a candidate is equal to the address of the Object Driver, as shown in Fig. 7.

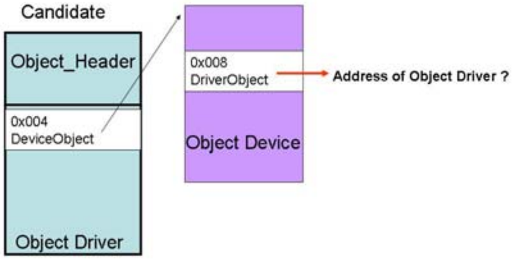


Figure 7. Rule 8 for finding Object Drivers

Rule 9: Verifying whether the value of the PsLoadedModuleList DriverStart member is equal to the value of the Object Driver DriverStart member at a candidate, as shown in Fig. 8.

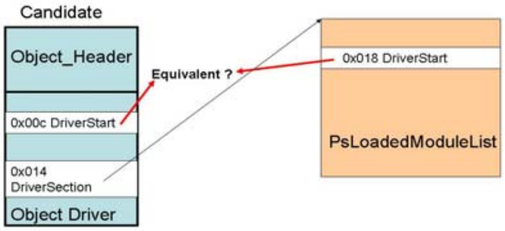


Figure 8. Rule 9 for finding Object Drivers

Rule 10: Verifying whether the value of the PsLoadedModuleList DriverInit member is equal to the value of the Object Driver DriverInit member at a candidate, as shown in Fig. 9.

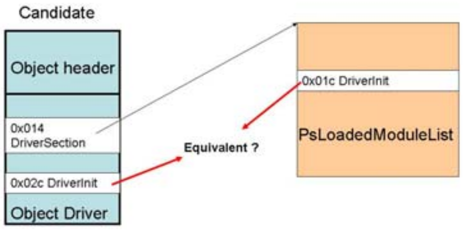


Figure 9. Rule 10 for finding Object Drivers

Rule 11: If PsLoadedModuleLists 1 and 2 are the values of PsLoadedModuleList 0 TypeList member's Flink and Blink at a candidate, respectively, then we can verify whether the value of PsLoadedModuleList 1 TypeList member's Blink and the value of PsLoadedModuleList 2 TypeList member's Flink are equal to the address of PsLoadedModuleList 0, as shown in Fig. 10.

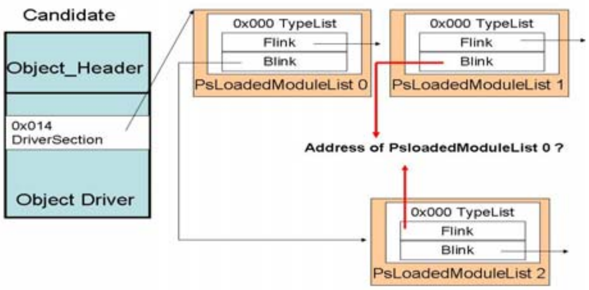


Figure 10. Rule 11 for finding Object Drivers

Step 3: Since the DriverSection member of an Object Driver is a pointer to its PsLoadedModuleList structure whose first 8-bit is a List_Entry structure which the detector can traverse for getting the PsLoadedModuleList of each Object Driver, the value of the DriverStart member in the PsLoadedModuleList structure is the driver address that will be recorded to a list for later use.

Step 4: We contrast the first list created by Step 2 with the second list produced by Step 3. If a driver address is listed in the first list but not displayed in the second list, then it will be tagged to be a hidden driver.

# 4 Experimental results and discussion

We have illustrated the proposed detection method for identifying Windows driver-hidden rootkits in Section 3. In the following, we will demonstrate the experimental results of evaluating the detection ability. Within the Windows family of operating systems, Windows XP is by far the most popular [25], and thus we execute the test of rootkits detection on Windows XP. In addition to our previously proposed rootkit [21], we select a variety of well-known hooking-based and DKOM-based rootkits, as shown in Table 1, to verify whether the proposed detector can identify them or not. Table 2 is the comparison of detection ability among a variety of detectors. In Table 2, the reason for choosing detectors DarkSpy, GMER, IceSword, Rootkit Unhooker, RootkitBuster and Tucan to compare with the proposed one is depicted as follows. According to the recommendations of the Antirootkit website [26], we find kernel-level detectors DarkSpy, GMER, IceSword and Rootkit Unhooker all have a rating with five stars. Detectors RootkitBuster [27] and Tucan [28], developed by famous Trend Micro and Panda respectively, are not rated, but they also possess a driver-hidden detection capability. Moreover, as stated in the literature [13], [15], [16], [19], [20], [29], we can find DarkSpy, GMER, IceSword and Rootkit Unhooker are highly effective detectors for identifying DKOM-based systemic threats to kernel data. Therefore, in order to effectively verify the detection ability of the proposed detector, we select the above-mentioned six rootkit detectors. In Table 2, we can find only the proposed detector can identify all of the listed DKOM-based and hooking-based rootkits. In particular, although detectors RootkitBuster, Tucan, GMER and Rootkit Unhooker can effectively detect the known sophisticated DKOM-based Unreal rootkit, they are still unable to find Tsaur and Chen's DKOM-based rootkit [21]. Accordingly, our proposed detection method is superior to the other six detectors on the whole.

The proposed detection approach, including eleven valuable rules, is to detect driver-hidden rootkits. Although we employ signatures and searching memory to find all possible loaded drivers, the searching procedure to find targets does not need to meet all of the eleven rules. That is, we adopt clustering concepts and set a threshold value to efficiently find all Object Drivers. When a candidate meets a rule, one will be added to the original value of a counter; if the current value of the counter is equal or bigger than the threshold value, the candidate will be marked as a true Object Driver. It can identify not only normal drivers but also variable drivers whose appearances are altered by rootkit writers to let them look different as compared with normal ones so that they are no longer driver formats and thus escape driver signature based detectors to achieve the purpose of hiding. Moreover, we can also find that much research [7], [11], [30] have been focused on kernel data to provide schemes to monitor them for finding malicious behaviors, but if their schemes are executed after a rootkit is installed in a compromised computer, they may not be capable of identifying the rootkit. By contrast, our proposed approach has not this kind of constraint.

Table 1. The rootkits for testing the detection ability

| Rootkit name | Stealth technique used |
|---|---|
| AFXRootkit2005 | Hooking-based |
| Migbot | Hooking-based |
| Klog | DKOM-based |
| Unreal | DKOM-based |
| Rootkit.Win32.Agent.ff | DKOM-based |
| Rootkit.Win32.Agent.eii | DKOM-based |
| Hacktool.Rootkit | DKOM-based |
| RootKit.Win32.Mnless.akx | DKOM-based |
| Tsaur and Chen's rootkit [21] | DKOM-based |

# 5 Conclusion

In this paper, we develop a new detection technique for identifying driver-hidden rootkits using DKOM in the Windows system to defend against the threat posed by not only the proposed new type of rootkit [21] but also other existing ones. We affirm our research is valuable for rootkit detector researchers to discover the weaknesses of their detectors, and can be a great inspiration to defenders to effectively improve the current techniques of detecting unknown Windows driver-hidden rootkits.

Based on the experience in developing a driver-hidden rootkit detector, we are continuing our research on developing technologies for detecting other kinds of Windows kernel-mode rootkits. It is expected that in the future our proposed detector will give developers a different aspect of detection so that they can combine their detection method with ours to form better solutions.

# Acknowledgment

This work was supported by the National Science Council of Republic of China under contract numbers NSC 100-2219-E-212-001, NSC 101-2219-E-212-001 and NSC 100-2218-E-492-026.

Table 2. The comparison of detection ability among a variety of detectors

| Detectors<br>Rootkits | Proposed detector | RootkitBuster | Tucan | DarkSpy | GMER | IceSword | Rootkit Unhooker |
|---|---|---|---|---|---|---|---|
| AFXRootkit2005 | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Migbot | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Klog | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Unreal | ○ | ○ | ○ | × | ○ | × | ○ |
| Rootkit.Win32.Agent.ff | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Rootkit.Win32.Agent.eii | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Hacktool.Rootkit | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| RootKit.Win32.Mnless.akx | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Tsaur and Chen's rootkit [21] | ○ | × | × | × | × | × | × |
| Note: 「○/×」 represents the detector can/cannot identify rootkits, respectively. | | | | | | | |

# References

[1] G. H. Kim and E. H. Spafford, "The design and implementation of Tripwire: a file system integrity checker," in *Proc. the 2nd ACM Conference on Computer and Communications Security*, pp. 18-29, 1994.

[2] K. Chian and L. Lloyd, "A case study of the Rustock rootkit and spam bot," in *Proc. USENIX First Workshop on Hot Topics in Understanding Bonets*, pp. 10-18, 2007.

[3] E. Florio, "When malware meets rootkits," *White Paper*, Symantec, 2005.

[4] J. G. Levine, J. B. Grizzard and H. L. Owen, "Detecting and categorizing kernel-level rootkits to aid future detection," *IEEE Security & Privacy*, vol. 4, no. 1, pp. 24-32, 2006.

[5] McAfee, "Rootkits, Part 1 of 3: the growing threat," *White Paper*, McAfee, 2006.

[6] A. Baliga, P. Kamat and L. Iftode, "Lurking in the shadows: identifying systemic threats to kernel data," in *Proc. the 2007 IEEE Symposium on Security and Privacy*, pp. 246-251, 2007.

[7] A. Baliga, L. Iftode and X. Chen, "Automated containment of rootkits attacks," *Computers & Security*, vol. 27, Issues 7-8, pp. 323-334, 2008.

[8] M. Christodorescu, S. Jha, S. Seshia, D. Song and R. Bryant, "Semantics-aware malware detection," in *Proc. the 2005 IEEE Symposium on Security and Privacy*, pp. 32-46, 2005.

[9] S. T. King et al., "SubVirt: implementing malware with virtual machines," in *Proc. the 2006 IEEE Symposium on Security and Privacy*, pp. 314-327, 2006.

[10] C. Kruegel, W. Robertson and G. Vigna, "Detecting kernel-level rootkits through binary analysis," in *Proc. the 20th Annual Computer Security Applications Conference (ACSAC'04)*, pp. 91- 100, 2004.

[11] N. L. Petroni Jr., T. Fraser, A. Walters and W. Arbaugh, "An architecture for specification-based detection of semantic integrity violations in kernel dynamic data," in *Proc. the 15th USENIX Security Symposium*, pp. 289-304, 2006.

[12] N. L. Petroni Jr. and M. Hicks, "Automated detection of persistent kernel control-flow attacks," in *Proc. the ACM Conference on Computer and Communications Security (CCS)*, pp. 103-115, 2007.

[13] J. Rhee, R. Riley, D. Xu and X. Jiang, "Defeating dynamic data kernel rootkit attacks via VMM-based guest-transparent monitoring," in *Proc. the 4th International Conference on Availability, Reliability and Security*, pp. 74-81, 2009.

[14] Y. Wang, D. Beck, B. Vo, R. Roussev and C. Verbowski, "Detecting stealth software with Strider GhostBuster," in *Proc. the 2005 International Conference on Dependable Systems and Networks (DSN'05)*, pp. 368-377, 2005.

[15] Y. Wen, J. Zhao, H. Wang and J. Cao, "Implicit detection of hidden processes with a feather-weight hardware-assisted virtual machine monitor," in *Proc. the 13th Australasian Conference on Information Security and Privacy*, LNCS 5107, pp. 361-375, 2008.

[16] Y. Wen, J. Zhao and H. Wang, "Implicit detection of hidden processes with a local-booted virtual machine," *International Journal of Security and Its Applications*, vol. 2, no. 4, pp. 39-48, 2008.

[17] C. Xuan, J. Copeland and R. Beyah, "Shepherding loadable kernel modules through on-demand emulation," in *Proc. the 6th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, pp. 48–67, 2009.

[18] G. Hoglund and J. Butler, Rootkits: Subverting the Windows Kernel, Addison-Wesley, 2005.

[19] C. Ries, Inside Windows Rootkits, VigilantMinds Inc, 2006.

[20] L. Stevenson and N. Altholz, Rootkits for Dummies, Wiley Publishing, 2007.

[21] W.-J. Tsaur and Y.-C Chen, "Exploring rootkit detectors' vulnerabilities using a new Windows hidden driver based rootkit," in *Proc. the 2010 IEEE International Conference on Information Privacy, Security, Risk and Trust*, pp. 842-848, 2010.

[22] C. Kreibich and J. Crowcroft, "Honeycomb: creating intrusion detection signatures using honypots," *ACM SIGCOMM Computer Communication Review*, pp. 51-56, 2004.

[23] BlackLight, http://www.f-secure.com/blacklight/, Retrieved June 10, 2011.

[24] A. Schuster, "Searching for processes and threads in Microsoft Windows memory dumps," *Digital Investigation: The International Journal of Digital Forensics & Incident Response*, vol. 3, no.1, pp. 10-16, 2006.

[25] W3Schools, "OS Platform Statistics," Retrieved April 20, 2011 from http://www.w3schools.com.

[26] Antirootkit, http://www.antirootkit.com, 2012.

[27] RootkitBuster, http://www.trendmicro.com/, 2012.

[28] Tucan, http://www.pandasecurity.com/, 2012.

[29] E. U. Kumar, "Battle with the unseen — understanding rootkits on Windows," in *Proc. the 9th AVAR International Conference*, pp. 82-97, 2006.

[30] N. L. Petroni, T. Fraser, J. Molina and W. A. Arbaugh, "Copilot - a Coprocessor-based Kernel Runtime Integrity Monitor," in *Proc. the 13th USENIX Security Symposium*, pp. 179-194, 2004.

# BLAST Your Way through Malware
# Malware Analysis Assisted by Bioinformatics Tools

**Jay Pedersen, Dhundy Bastola, Ken Dick, Robin Gandhi, William Mahoney**
School of Interdisciplinary Informatics
College of Information Science and Technology
University of Nebraska at Omaha
Omaha, Nebraska
{jaypedersen, dkbastola, kdick, rgandhi, wmahoney} @unomaha.edu

**Abstract**—*As a new strain of computer malware is discovered, it triggers a meticulous process of analyzing its behavior and developing appropriate defenses. A systematic process which identifies regions of commonality and variability with known samples can ease the burden of malware analysis. We address this challenge using an interdisciplinary approach which applies biological sequence analysis methods to computer malware. Specifically, we have developed a method which has the goal of classifying a digital artifact (possibly malware) based on its similarity to known digital artifacts (or known malware samples) using methods and tools of bioinformatics. Our approach is analogous to classifications of biological sequences, which are routinely performed using online databases of known biological sequences.*

***Keywords:*** clustering, classification, malware, plagiarism

## 1. Introduction

Consider the evolution of a biological pathogen, which can be tracked using DNA markers. This operation often involves examining nucleotide sequences in DNA using powerful bioinformatics tools to identify regions of local or global similarity, or interactions with specific enzymes, which may be a consequence of functional, structural, or evolutionary characteristics of the pathogen's genetic makeup. This is biological stylometry at work!

In field of computer security, there is significant interest in understanding malware behavior to develop effective detection, prevention and recovery mechanisms [1] [2]. Unfortunately, malware analysis is still much of an acquired tradecraft, and the results depend heavily on the quality of personnel involved. Malware analysis typically involves reverse engineering compiled digital artifacts, configuration files, metadata or foraging through other information. The results from such analysis provide clues for malware origin, behavior, locating other variants, signature patterns, and proper malware classification (e.g. Trojan, worm, virus, zombie, fork bomb, bot, etc.).

In the field of bioinformatics, the Basic Local Alignment Search Tool (BLAST) tool discovers areas of local similarity between DNA or protein sequences [5] [8] [10] [14] [15] [16]. Local similarity comparisons have advantages over global similarity in studying the functional and evolutionary relationships among specimens. For example, ape and human DNA shares several areas of local similarity. When compared globally the DNA similarity is harder to discover due to years of evolutionary changes. In particular, BLAST compares a given DNA nucleotide sequences to a database of known sequences which can be from a wide variety of organisms. Based on the results of discovered local similarities, a given specimen can be inferred to have functional and evolutionary relationships with a known sample. Such local alignments help identify related members of gene families.

Computer malware analysis has interesting parallels with the study of biological pathogens. The similarity does not just stop with bio inspired names of computer malware and their high-level behaviors, but also extends to how we analyze and study them. The objective of our work is thus to utilize an interdisciplinary approach to determine the pedigree of a digital artifact of unknown origin. In this paper, we implement bioinformatics inspired methods in the study of three application areas: (a) document clustering using similarity detection (b) rapid malware classification, (c) plagiarism detection.

In the present research we apply BLAST to study synthetic DNA sequences that represent digital artifacts. The ability to use bioinformatics tools to study digital artifacts opens up several avenues of interesting studies ranging from literary stylometry; digital forensics; sources code clone detection; malware functional characteristics and evolutionary relationships; and most importantly attribution of digital artifacts to compilers, platforms, chipsets, versions, and possibly the author!

We assume that the reader is not familiar with bioinformatics tools, and in section two include an overview of BLAST and the technology involved. Our methods for the analysis of digital artifacts are in section three and include the description of how the data is manipulated to appear as DNA to the bioinformatics tools. The results of several experiments are contained in section four. These results include the three areas described above: malware, file type, and plagiarism. The final section includes our ideas for additional research as well as our conclusions for our work thus far.

## 2. Bioinformatics Background

DNA is the biological blueprint used for building proteins and other cellular components of living organisms. It is comprised of a long stretch of adenine (A), guanine (G), cytosine (C), and thymine (T) molecules, commonly referred to as "bases" due to their chemical nature. They are also referred to as nucleotides. DNA is represented computationally by character strings containing only the characters A, G, C and T. The seminal paper of Watson and Crick in April 25 of 1953 [3] described the molecular structure of DNA as a double helix. This discovery revolutionized the study of Biology. In double stranded DNA each strand runs anti-parallel to the other and each strand can be used as a template to construct the other strand using Chargaff's base pairing rule [3] [4] which states that Adenine (A) will only pair with Thymine (T), and Guanine (G) will only pair with Cytosine (C).

Each strand has an associated direction, which is indicated by its 5' (5 prime) and 3' positions; the direction is from 5' to 3'. The positions of the 5' and 3' ends of the strands are opposite, and thus they are anti-parallel. The following shows a representation of a small piece of double-stranded DNA. It shows the complementary base pairing and the anti-parallel nature of the strands.

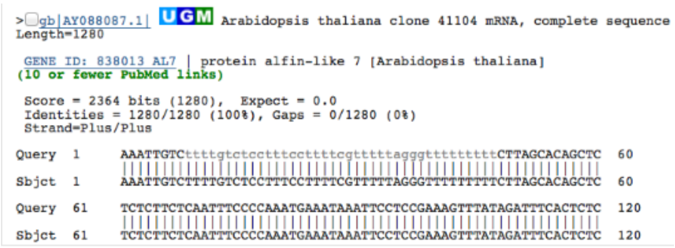
```
5′ GAATTCGGCC 3′
   ||||||||||
3′ CTTAAGCCGG 5′
```

The computational representation of DNA only includes one of the strands; and the other strand is implied and can be computed as necessary using the base-pairing rules. The strand direction is also implied -- the 5' position is at the beginning of the string and the 3' position is at the end of the string.

DNA is the key information source in many bioinformatics research projects, including those trying to determine the relatedness of two organisms by a comparative study. GenBank is an international nucleotide sequence database and currently holds sequences from about 407,000 organisms.

BLAST is a widely used bioinformatics tool [5] [8] [10] [14] [15] [16] that compares a given DNA sequence with other known DNA sequences (e.g. GenBank sequences) that reside in a BLAST database and determines similarities between them. A BLAST database is collection of known biological sequences, optimized for similarity querying by the BLAST tool. The result set returned in response to a given query sequence includes local alignments between the query sequence and "subject" sequences in a BLAST database; an example alignment is shown in Figure 1. A local alignment indicates a region of similarity between two sequences. The regions involved can be in any part of either sequence. Within these regions, every base is aligned to exactly one base in the other sequence or to a gap position inserted between bases in the other sequence. Gaps are introduced to represent deletions or insertions of bases, which may have occurred over time. A local alignment is distinguished from a global alignment, which is an alignment of two entire sequences (rather than alignments of arbitrary regions within two sequences).

For each determined alignment, BLAST returns the name of the query and subject sequence and the positions within the sequences that were aligned. BLAST also returns a statistical measure of the likelihood that the identified alignment is a randomly expected occurrence; this is called the expect value (E-value). An E-value near zero indicates a nearly zero probability that the alignment represents a random occurrence [6]. A BLAST parameter allows you to specify a threshold E-value. Specifying an E-value near zero asks BLAST to return only highly similar alignments.



```
>□gb|AY088087.1| UGM Arabidopsis thaliana clone 41104 mRNA, complete sequence
Length=1280

 GENE ID: 838013 AL7 | protein alfin-like 7 [Arabidopsis thaliana]
(10 or fewer PubMed links)

 Score = 2364 bits (1280),  Expect = 0.0
 Identities = 1280/1280 (100%), Gaps = 0/1280 (0%)
 Strand=Plus/Plus

Query  1      AAATTGTCttttgtctcctttccttttcgttttagggttttttttttCTTAGCACAGCTC  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1      AAATTGTCTTTTGTCTCCTTTCCTTTTCGTTTTAGGGTTTTTTTTTTCTTAGCACAGCTC  60

Query  61     TCTCTTCTCAATTTCCCCAAATGAAATAAATTCCTCCGAAAGTTTATAGATTTCACTCTC  120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61     TCTCTTCTCAATTTCCCCAAATGAAATAAATTCCTCCGAAAGTTTATAGATTTCACTCTC  120
```

Fig. 1. A BLAST alignment between two highly similar DNA sequences, which has an expect value (E-value) of zero.

## 3. METHODS

The premise of this project is that a digital artifact may be represented by a "synthetic" DNA sequence and that BLAST should be able to find similarities between that sequence and a set of sequences representing other digital artifacts, which are stored in a BLAST database. (Note that BLAST has previously been used to examine sequences, which do not represent actual biological sequences. A previously reported use was examining journal papers [7])

### *A. DNA representation of an arbitrary digital artifact*

BLAST supports both nucleotide (DNA) and protein sequences. However, BLAST attaches biological significance to the amino acids in a protein sequence (BLOSUM and PAM scoring matrices [12] have this logic encoded in them). On the other hand, when BLAST is analyzing DNA sequences there is minimal scoring logic related to chemical properties of the nucleotides. For this study the DNA format was chosen, so that chemical properties would not be a significant factor in the BLAST analysis. As a result, all digital artifacts are transformed into a corresponding DNA representation for processing by BLAST. This transformation is obtained by following a mapping between digital bits and characters representing nucleotides.

A digital artifact is considered to be a sequence of byte values. The initial step is thus to convert the sequence of bytes from an arbitrary digital artifact into a DNA representation. The conversion is completely reversible. Four DNA characters are created for each byte in the digital artifact. Each DNA character represents two bits of the byte. The four characters represent bits six and seven, four and five, two and three and zero and one, respectively. The following mapping is used:

```
00 ←→ T
01 ←→ G
10 ←→ C
11 ←→ A
```

There are twenty-four possible ways to perform such a mapping. Any of those mappings could be used to provide a consistent and comparable DNA representation of a digital artifact. This mapping has the property that the values for G and C; and T and A are complementary, considering bit values of zero and one to be complements.

A method was developed which uses this mapping to allow for the comparison and clustering of arbitrary digital artifacts. Steps in the method are guided by the alignments discovered by BLAST among the DNA representations of those artifacts. The method includes three steps: 1) preprocessing, 2) sequence analysis, and 3) visualization. The steps can be applied for several use cases.

Digital artifact clustering: A set of digital artifacts to be clustered is converted into DNA representations. A pairwise comparison by BLAST produces alignments between artifacts with similar structures. The alignments are used to build a clustered graph representation of the similarities between the artifacts. The visualization is performed using Cytoscape (a popular bioinformatics graph visualization tool) [9] [13].

Digital artifact identification: Consider an artifact of unknown origin and a BLAST database of sequences of known digital pedigree. The unknown artifact is converted to a DNA format, and compared to the sequences in the database using BLAST. The resulting alignments are used to identify the most likely type of that artifact. Again, the results are visualized for foraging through the various reported alignments.

### A. *Preprocessing*

The input to this step is an arbitrary set of digital artifacts to be examined. The artifacts are converted to a DNA format and a BLAST database is created using the steps shown in Figure 2.

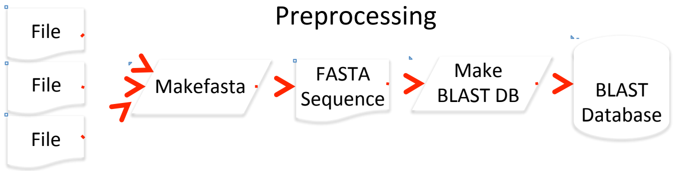


Fig. 2. Preprocessing Step for Digital Artifact Analysis. The various digital files are converted to DNA sequences and are merged into a FASTA sequence file. This is converted into the database used by BLAST.

The specific type of file that is created by this step is a FASTA format file [11]. This is a flat text file, which can contain multiple DNA sequences. Each sequence is introduced by an identification line which has the ">" character as the first character on the line and then has information which identifies the sequence. Each identification line is followed by one or more lines containing the DNA characters which define the sequence. The following shows the beginning of a FASTA file representing the digital artifact "zeus_005_f04.exe":

```
>lcl|/home/jayp/bigtest/zeus_005_f04.exe
GTAGGGCCCGTTTTTTTTTATTTTTTTTTTTTTTGTTTTTTT
TTTTTTAAA
```

Given a FASTA file containing one or more DNA sequences -- the "makeblastdb" tool from NCBI's BLAST tools [15] [16] can be used to create a BLAST database containing those sequences.

### B. *Sequence Analysis*

Once the database has been created, the artifacts can be analyzed for pairwise similarity using BLAST. We have used NCBI's BLAST version 2.25 [16] for our analyses. Bioinformatics practitioners call this type of search an "all versus all" BLAST comparison, used in Biology to look for orthologs (similar genes) across multiple species [14]. In our case, we are looking for similarities among a set of digital artifacts. The result of this step is a BLAST report of the determined alignments between the artifacts.

The same FASTA sequence file, which was used to create the BLAST database, is now used to query against the database. Thus, BLAST will determine the similarities between each sequence and every sequence in the database; which is the same set of sequences. Thus, the "all versus all" comparison. The overall flow is depicted in Figure 3.

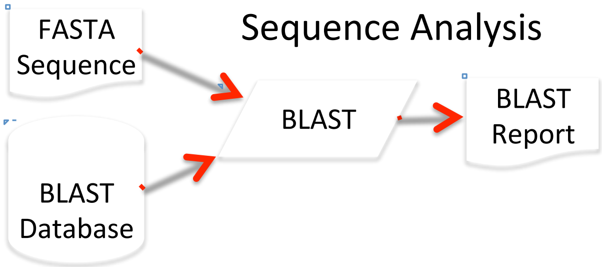


Fig. 3. Sequence Analysis Step. In this middle step the FASTA file and the BLAST database are examined and the sequence alignments between the original artifacts are reported.

As described in Section 2, BLAST has an "E-value" parameter. Specifying an E-value close to zero asks BLAST to return only highly similar alignments. We examined E-values ranging from $10^{-6}$ to $10^{-300}$ with "all versus all" BLAST comparisons. As the negative exponent decreased towards -300, the number of small alignments reduced dramatically, while the larger alignments remained stable . This indicates that BLAST considers longer alignments to be less likely to be random occurrences.

A BLAST report includes information concerning each alignment. This includes:

- Names of the sequences involved in the alignment
- Starting and ending positions of the alignment.
- Measures of the statistical significance including E-value

### C. *Visualization*

The final step is the visualization of the results of the BLAST alignment. The visualization step consists of examining the BLAST report and creating a graph that represents those files where the sequences aligned with each other. The BLAST report is parsed and the alignment information is saved in a Simple Interaction Format (SIF) file. This graph format is used as input by the Cytoscape visualization tool [9] [13]. The overall flow of the visualization step is depicted in Figure 4.

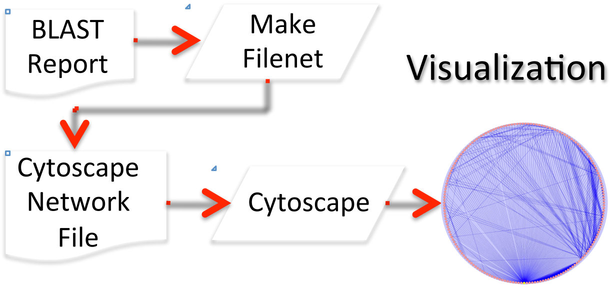


Fig. 4. Visualization of Similar Files. The output from the BLAST step is used to create input for Cytoscape. This provides various visualizations for the alignments of the DNA sequence

The visualization graph is constructed by considering every digital artifact to be a node in the graph, and edges representing the BLAST alignments between the artifacts. Such a graph will contain components that can be considered as clusters. The size of the clusters and the density of relationships among nodes in a cluster will vary depending on the E-value, which was used when performing sequence analysis.

This step optionally creates files, which show the alignments in "original format" (instead of DNA characters). This can be useful to examine what BLAST is determining aligns in its original form (rather than as DNA characters). Such inspection is particularly useful for text-based digital artifacts such as documents and source code.

## 4. Results

Several experiments were designed to evaluate the usefulness of this approach including document clustering, malware classification and plagiarism identification

### A. *Document Clustering*

A set of 1,202 digital artifacts of fifteen different types were collected to test the document clustering process. The artifacts included text and binary files and both benign and malicious executables and benign and malicious JavaScript files. The types and counts of artifacts were as follows:

| | | |
|---|---|---|
| 14 executable files | 33 JavaScript | 319 Java |
| 229 Java ".class" | 203 natural language | 192 C |
| 45 Scala | 46 Perl | 9 CGI (Perl) |
| 15 Python | 31 C# | 24 HTML |
| 18 PNG (image) | 17 MP3 (audio) | 7 ZIP |

Among the 14 executable files, were 7 benign files and 7 malware files (4 Zeus Trojans and 3 Zeus Version Two Trojans as identified by MalwareDomainList.com). Also included were 33 JavaScript source files, of which 25 were malicious including 13 obfuscated and 12 de-obfuscated files. The malware JavaScript examples were obtained from http://redleg-redleg.blogspot.com/p/examples-of-malicious-javascript.html

The preprocessing, sequence analysis and visualization steps defined in the method section were followed. The E-value parameter of BLAST was set to $10^{-300}$. This resulted in the creation of a graph, which contained 9,932 edges between the artifacts. The visualization of much of the graph, including its largest clusters is in Figure 5.

The graph contained clusters with very similar files, some of the clustering highlights included:

- A cluster with 23 HTML files (of the 24 in the data set)
- A cluster with 18 C# files (of the 31 in the data set)
- A cluster with 16 MP3 files (of the 17 in the data set)
- A cluster with 127 Java class files from a single project; another cluster with 16 Java class files from a different project; another cluster with 22 Java class files from three highly related projects (all implementing the same class assignment)
- A cluster with 10 Windows executable files including 4 Zeus Trojan executable files (but none of the Zeus version two executable files).
- A cluster with all 3 Zeus version two executable files was generated
- A cluster with 45 Java files from the same project, another with 19 Java files from a different project, one cluster with 14 Java files from one project.
- One cluster with a mixture of 50 Java and Scala source files (29 Java, 21 Scala); the Scala files had been converted directly from the project the Java files belonged to and were thus highly similar.
- One cluster contained 61 C files and header files from the same assignment

One cluster with 3 deobfuscated JavaScript files and another with 2 deobfuscated JavaScript files. (of the 12 deobfuscated JavaScript files)

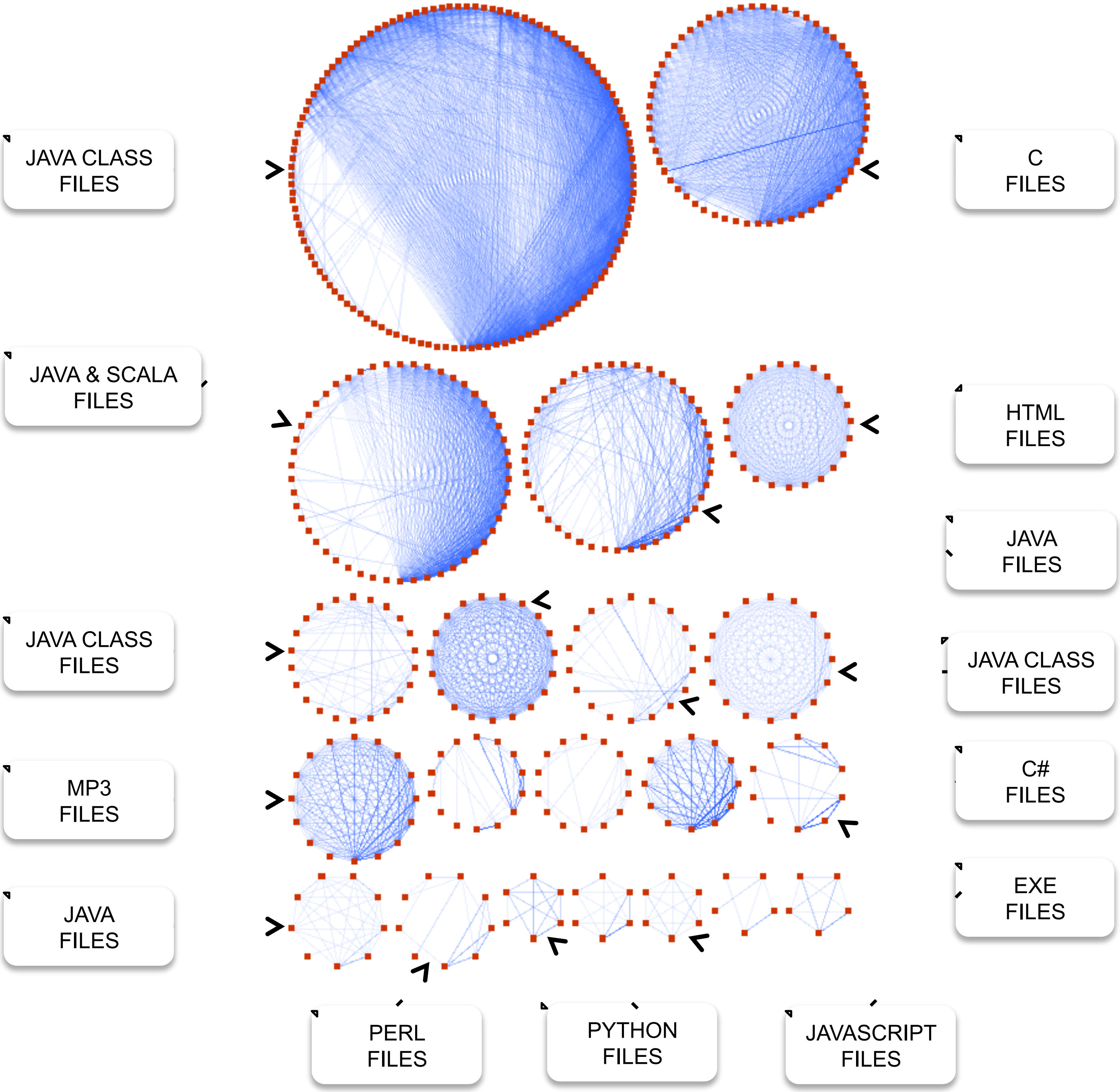


Fig. 5. Document Clustering at E-value $10^{-300}$. Each almost exclusively consisted of files of the same type. For example, Java source files from different projects cluster together, while MP3 files and HTML cluster into separate groups.

The following was also observed:

- Of the 1202 files, there were 453 which had no alignments with any other files (approximately 38%). This is not unexpected due to the very high local similarity requirement enforced by using E-value $10^{-300}$. Of the non-clustered files, there were 182 natural language files. If natural language files are excluded from consideration, there were 27% of the remaining files, which remained un-clustered.
- Obfuscated JavaScript had few alignments to other files.
- The following file types had almost no alignments to any other files: ZIP files, PNG files and natural language files.

In general, the clusters that formed had strong similarity of file type and frequently consisted of files from the same programming project. There were cases where a single Java programming project resulted in multiple clusters, but the clusters always exclusively consisted of Java source files. There were a significant number of files, which did not align and join with other files in a cluster. This appeared to be especially prevalent with natural language files.

The four Zeus Trojan executable files clustered with benign executable files, but Zeus Version Two Trojans executable files did not.

Additional testing was performed to see how the results would differ when examining the same 1202 files but using "looser" E-values of $10^{-250,}$ $10^{-200,}$ and $10^{-150}$, which reduced the amount of local similarity required.

The following were some of the differences observed when clustering at E-value $10^{-200}$:

- There were 30,680 edges in the network (compared to 9,932 previously).
- A cluster of all 14 executable files was created; including the Zeus and Zeus version two Trojan executable files. The Zeus version two executable files were no longer distinguished from other executable files).
- A large cluster of 259 files was generated consisting of 189 Java class files but also contains 35 related Java files and 35 Scala files. All files were related to implementing the same programming project.
- A cluster of 143 Java files from the same project was created.
- A cluster containing 91 C files from multiple programming assignments was created (of the 192 total C files).
- A cluster of 6 of the 8 benign JavaScript files was created (compared to two separate clusters which contained 5 benign JavaScript files previously).
- There were several clusters that were nearly identical to those of the $10^{-300}$ E-value case:
  - A cluster containing 23 C# source files (of the 31 total C# files)
  - A cluster containing 23 HTML files (of the 24 total HTML files)
  - A cluster with 19 Java files from the same project.

When further increasing the E-value to $10^{-150}$, the following was observed:

- There were 177,011 edges in the network
- There were 235 files, which had no alignments to any other files; of these, 169 were natural language files. If we exclude natural language files, then 6.6% of the remaining files were un-clustered.
- One cluster contained 453 files which included Java class files but also some C source code, Java source code and Scala source code and C# source code.

The observations indicate that the specificity of the clustering based on file type starts to break down at a higher E-value setting. In summary, as the E-value increased from $10^{-300}$ to $10^{-150}$, the number of alignments returned by BLAST increased substantially, and the level of similarity between files in clusters appeared to be reduced. For example, Java files from different projects, which were separated into different clusters at E-Value $10^{-300}$ were being clustered together at E-Value $10^{-150}$. Similar results were seen for C source files and Java class files (byte code files).

### *B. Rapid Malware Classification*

This experiment repurposed the BLAST database created by the document clustering test. It relies on the fact that there are Zeus and Zeus version two malware executable files within the database. The premise is that Zeus and Zeus version two malware executable files found "in the wild" should align closely with their counterparts in the database.

Recall that the 1,202 digital artifacts in the BLAST database were of the following types:

| | | |
|---|---|---|
| 14 executable files | 33 JavaScript | 319 Java |
| 229 Java ".class" | 203 natural language | 192 C |
| 45 Scala | 46 Perl | 9 CGI (Perl) |
| 15 Python | 31 C# | 24 HTML |
| 18 PNG (image) | 17 MP3 (audio) | 7 ZIP |

The experiment was to find another Trojan executable and see if it could be identified as such by examining its BLAST alignments with that BLAST database.

A malware executable was obtained on March 7, 2012, from a reference at MalwareDomain.com, which identified it as a Zeus Trojan. Its size and content differed from the four Zeus Trojan executable files in the BLAST database.

A biological representation of this executable was generated and BLAST was used to determine its alignments with the 1,202 files in the BLAST database.

At E-value level $10^{-300}$, BLAST generated 31 alignments which were all to the four Zeus Trojan executable files in the database. This was viewed as a positive result.

At E-value level $10^{-200}$, BLAST generated 2,364 alignments, of which the highest scoring 2355 (99.6%) were to the 4 Zeus executable files in the database. Of the remaining 9 lowest scoring alignments -- 7 were to benign executable files and 2 were to Zeus version two executable files. This was also viewed as a positive result.

### *C. Plagiarism Detection*

A separate investigation was undertaken to investigate possible plagiarism in student program submissions in a C programming class taught by one of the authors. The examination was done using E-value $10^{-300}$. Several sets of programs files were examined. The topic of the assignment, and whether or not any "boilerplate" code (i.e. assignment bootstrapping code) was given to the class influenced the results significantly.

In cases where there was boilerplate code given to the class as part of the programming assignment, the student submitted programs all tended to cluster together. The alignments were observed primarily due to the boilerplate code that was common to all of the student files. Even in this circumstance, there were seen clusters with more inter-file alignments. There was not a clean separation of into components, but the number of alignments between files was a strong indicator of those files, which were suspiciously similar in regions other than the boilerplate.

When no boilerplate code was given, there tended to be fewer alignments. In this case, the alignments that were found showed suspiciously similar code between different student submissions.

## 5. Further Research and Conclusions

The specific uses of bioinformatics tools in this project gives only a taste of what the tools could be used for in the future. There are tools for classification of biologic objects, such as Restriction Enzymes, which may be of use in classifying computer artifacts – given a biological representation of those artifacts. One caveat is that there may be strong Biological assumptions made by those tools that would not be satisfied by biological representations of arbitrary digital artifacts

An intriguing possibility is to create a BLAST database containing sequences representing known malware including strains of malware executable files and JavaScript malware. This could be used as a possible rapid malware identification mechanism. A digital artifact whose biological representation aligns closely with any sequences in that database could be considered to be a likely malware executable. The "Rapid Malware Detection" test that was done indicates that this could be an effective identification mechanism.

It may be found that it is beneficial to have different BLAST databases for different types of artifacts. Just as there are different biologic databases for nucleotides and proteins, perhaps it may be useful to have databases that are specific to viruses as opposed to malware JavaScript source files.

The specific outputs produced by this project can be improved in various ways:

- The Cytoscape visualization can be improved to change the size of a node or the width of an edge based on the size of the alignment. Multiple edges between two artifacts might be able to be condensed into a single edge using a color scheme to indicate the number of alignments
- A visual map of the alignments of two digital artifacts could be produced. This would be analogous to a "homolog map" in Biological domain which shows the positions of related genes in the genomes of two species.
- A Cytoscape plugin could be created to allow an viewing the original format of any alignment and remove those not of concern or highlight those of concern.

This paper presented a novel method for clustering digital artifacts, identifying a digital artifact as similar to known malware, and detecting plagiarism by using a synthetic DNA representation of the digital artifacts and using the bioinformatics BLAST tool. It demonstrates that the classification power of bioinformatics tools that are used in biology problems can also be used in other domains.

The success of using BLAST to examine nucleotide representations of computer artifacts can be partly attributed to the fact that BLAST does not make strong assumptions about the chemical differences between nucleotides when processing a nucleotide database. This is not true when using BLAST with protein databases. It may be possible to use BLAST with synthetic protein definitions by providing specialized scoring matrices to BLAST which do not make biologic assumptions that would not hold.

## 6. References

[1] Distler, Dennis, and Charles Hornat. "Malware Analysis: An Introduction." *Sans Reading Room*. Sans, 14 Dec. 2007.

[2] Kendall, Kris. "Practical Malware Analysis." Blackhat, 2007. Web. 11 Jan. 2012. <http://www.blackhat.com/presentations/bh-dc-07/Kendall_McMillan/Paper/bh-dc-07-Kendall_McMillan-WP.pdf>.

[3] Watson JD, Crick FH, "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid", *Nature, 1953, Vol* **171**

[4] Elson D, Chargaff E, "On the deoxyribonucleic acid content of sea urchin gametes".Experienti, 1952, Vol 8

[5] Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. Basic local alignment search tool, 1990, J. Mol. Biol, Vol 215

[6] Pagni M, Jongeneel CV, Making sense of score statistics for sequence alignments, Briefings in Bioinformatics, 2001, Vol 2

[7] Krauthammer M, Rzhetsky A, Morozov P, Friedman C, Using BLAST for identifying gene and protein names in journal articles,

[8] Altschul, S. et al, Gapped BLAST and PSI-BLAST; Nucleic Acids Research, 1997, Vol. 25, No.17,

[9] Shannon, P. et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks; Genome Research, 2003, Vol. 13, Pgs. 2498-2504

[10] McGinnis S and Madden T, BLAST: at the core of a powerful and diverse set of sequence analysis tools, Nucleic Acids Research, 2004,

[11] Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison, 1988, Proc. Natl Acad. Sci. USA,

[12] Mount, D., Comparison of the PAM and BLOSUM Amino Acid Substitution Matrices, 2008, Cold Spring Harbor Protocols, doi:10.1101/pdb.ip59

[13] Cline, M. et al, Integration of Biological Networks and Gene Expression Data using Cytoscape, Nature Protocols, 2007, Vol 2, Pgs 2366-2382

[14] Moreno-Hagelsieb G and Latimer K, Choosing BLAST options for better detection of orthologs as reciprocal best hits, Bioinformatics, 2008, Vol 24

[15] HTTP://BLAST.NCBI.NLM.NIH.GOV/

[16] ftp://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST

# Evaluating Intrusion Detection and Prevention Systems Using Tomahawk and Wireshark

**David Mudzingwa and Rajeev Agrawal**
Department of Electronics, Computer and Information Technology
North Carolina A&T State University, Greensboro, NC, USA

**Abstract** - *The increase in the security breach of computer systems and computer networks has led to the increase in the number of security tools that seek to protect these asserts. Among these tools are intrusion detection and prevention systems (IDPS). An IDPS is a security system that is used to detect and prevent security violations. Evaluating the effectiveness of IDPS is complicated and there has not been much work done on ways IDPS users can follow to evaluate the IDPS. Most of the work on evaluating IDPS is focused on developing new testing methodologies. This paper seeks to offer a practical approach to evaluate both hardware and software based IDPS using publicly available open source tools Tomahawk and Wireshark.*

**Keywords:** Intrusion detection and prevention system (IDPS), Tomahawk, Wireshark, PCAP file

## 1 Introduction

Intrusion detection and prevention systems are security tools that are used to detect and prevent security threats to computer systems and computer network systems. Although these systems are widely adapted and continue to grow, they continue to be very difficult to evaluate. This is due to the lack of publicly available research and test data sets. The available test data sets are dated [1]. Test data sets have to be current and public available so that interested parties can evaluate the quality of the data sets and effectiveness of IDPS. The available work in this area mainly focus on improving the methodologies used in evaluating IDPS instead of offering a simple way to evaluate the IDPS. There are non-commercial tools that can be used to test and validate the effectiveness of an IDPS, but there are still challenges in acquiring test data that contains live exploits needed to perform credible tests. There are also other challenges in correctly evaluating IDPS such as setting up a test environment and properly conducting the evaluation. This paper presents a simple but effective way to evaluate both the appliance/hardware and software based IDPS. This paper focuses on setting up an evaluation lab using publicly available tool Tomahawk for replaying network traffic that is used in the evaluation.

## 2 Related Work

The first research that looked the claims of the intrusion detection systems was done in 1998. This work put forth a framework for thoroughly testing an intrusion detection system and also offered data sets for use in evaluations. This early work is evaluated and checked for accuracy in [2]. This work challenged the continual use of the data sets produced in IDPS tests. This work puts on the argument that these data sets are out dated and that the procedures used to generate the data sets were not representative of a production network [3]. These limitations in the evaluation of IDPSs were addressed by the Lincoln Adaptable Real time Information Assurance Test bed (LARIAT) [8]. This was a better testing application that used a graphical user interface instead of the command line and was easier to use but was only available to the United States government [8]. Trident evaluation was another work that tried to improve on the early works by introducing ways to add new background and attack traffic to the test data sets [9]. This work offered a way to account for evasion techniques during an evaluation of an IDPS. A comparison of two IDPS methodologies to evaluate them is given in [5]. This work describes how to set up a test bed that can be used to evaluate an IDPS using a Snort and Spade. Before an IDPS is evaluated for effectiveness, its underlying detection methodologies need to be understood. An IDPS can be based on any of the four main detection methodologies. The signature based, anomaly based, stateful protocol analysis based, and the hybrid based detection methodology [16].

## 3 Background

### 3.1 IDPS Methodologies

There are many different methodologies used by IDPS to detect changes on the systems they monitor. These changes can be external attacks or misuse by internal personnel. Among the many methodologies, four stand out and are widely used. These are the signature based, anomaly based, Stateful protocol analysis based, and hybrid based. Most current IDPS use the hybrid methodology which is a combination of other methodologies to offer better detection and prevention capabilities. All detection methodologies use the same general model and the differences among them is mainly on how they process information they gather from the monitored environment to determine if a violation of the set policy has occurred. Figure 1 shows a broad architecture of which these systems are based on. This architecture was developed by the Intrusion Detection Working Group and has four functional blocks, the Event blocks which are the event boxes that gathers event from the monitored system and will be analyzed by other blocks, then the Database blocks which are the database boxes which stores the events from the Event blocks, then the Analysis blocks that processes the events and sends an alert, and final the Response blocks whose purpose is to respond to an intrusion and stop it [15].

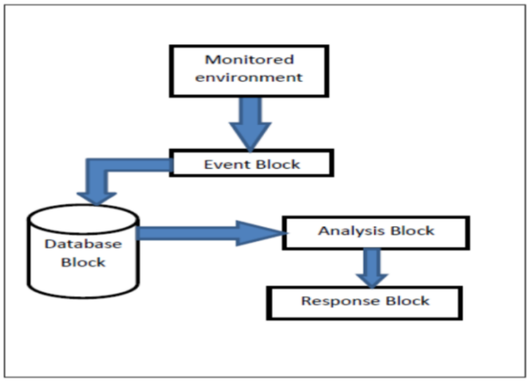


Figure 1. General architecture for IDPS systems

### 3.1.1 Anomaly Based Methodology

Anomaly based methodology works by comparing observed activity against a baseline profile. The baseline profile is the learned normal behavior of the monitored system and is developed during the learning period were the IDPS learns the environment and develops a normal profile of the monitored system. This environment can be networks, users, systems and so on.

### 3.1.2 Signature Based Methodology

Signature based methodology works by comparing observed signatures to the signatures on file. This file can be database or a list of known attack signatures. Any signature observed on the monitored environment that matches the signatures on file is flagged as a violation of the security policy or as an attack. The signature based IDPS has little overhead since it does not inspect every activity or network traffic on the monitored environment. Instead it only searches for known signatures in the database or file.

### 3.1.3 Stateful Protocol Analysis Based Methodology

The Stateful protocol analysis methodology works by comparing established profiles of how protocols should behave against the observed behavior. The established protocol profiles are designed and established by vendors. Unlike the signature based methodology which only compares observed behavior against a list, Stateful protocol analysis has a deep understanding of how the protocols and applications should interact/work.

### 3.1.4 Hybrid Based Methodology

The hybrid based methodology works by combining two or more of the other methodologies. The result is a better methodology that takes advantage of the strengths of the combined methodologies.

## 4 Setting Up IDPS Testing Environment Using Tomahawk

Tomahawk is an open source network tool that can be used to generate background network traffic, replay network traffic, and manipulate network traffic using captured network traffic files. The captured traffic can then be replayed during the evaluation of an IDPS. For the evaluation to be effective the traffic capture should come from the environment where the IDPS will reside. This produces a more accurate test. Once the traffic is captured, it can be replayed in a controlled environment where more experimentation can be done without negatively affecting the production environment. Within the controlled environment the captured traffic can be used as background traffic while known exploits are introduced to the monitored workstation/server.

The following are the minimum hardware and software requirements for using Tomahawk to test an IDPS:

Hardware and Software Requirements

- Workstation or server with a minimum of two network cards (running Linux/Unix flavor operating system)
- Second workstation or server with a minimum of two network cards
- Captured network traffic file in the libpcap format (cleaned)
- Network traffic capture tool (Wireshark)
- A minimum of a 2.0 GHz Pentium equivalent processor
- A minimum of 1GB of RAM (2 or more is recommended)
- Network switch (optional)
- Management pc (optional)
- Third network card on the other workstation/server (optional)

- Network cables (crossover optional)

## 4.1 Hardware Setup

There are three ways of configuring the hardware for testing an IPDS using Tomahawk. A basic way with just two computers with a minimum of two network cards each, a medium setup which is a basic setup with a hardware IDPS, and an advanced way which is a medium setup with an addition of a network switch and a management computer. These setups do not include an internet connection as a way to control the network traffic during the evaluation of an IDPS using tomahawk. If desired an internet connection can be added to either the attacker or the attacked computer.

### 4.1.1 Basic Set Up

A basic setup is the simple way to test a software based IDPS. As shown in figure 2 only two machines with two network cards and two crossed-over network cables are required. One of the machines is configured as the attack machine and Tomahawk is installed on it. The other machine is configured as the attacked machine and the software based IDPS is installed on it. The two machines are connected to one another with the crossover cables. Tomahawk only runs on Unix or Linux based operating systems and as a result the attack machine will require a Linux or Unix based operation system. The attacked machine can run any current operating system.

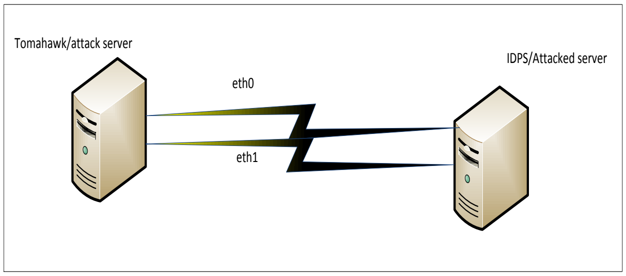


Figure 2. Basic Setup

### 4.1.2 Medium Setup

The medium set up is also simple but adds a hardware/appliance based IDPS. It requires two machines with two network cards each and three network cables. The machines are connected through the IDPS as shown in figure 3. In this setup the IPDS saves as a network switch connecting the two computers. The computer that tomahawk will be installed on has to have a Linux or Unix based operating systems. Also the computer that has tomahawk running on it must have two network cards that will used by tomahawk. The other computer that saves as the attacked one can have any of the current operating systems.

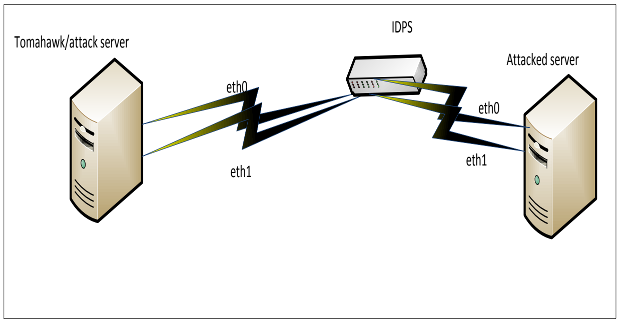


Figure 3. Medium Setup

### 4.1.3 Advanced Setup

The advanced setup is more involved than the other two as it adds a third network card, a switch, five network cables, a management machine, and the internet. The machines are connected through the switch and the IDPS as shown in figure 4. One of the machines with three network cards is configured as the attack machine and Tomahawk is installed on it. The other machine is configured as the attacked machine and sets on behind the IDPS and a third management machine is connected to the switch. This the ideal setup for testing IDPSs as it allows for different configuration changes to be made. For example more computers can be added to the test by adding another switch between the IDPS and the attacked computers or by adding more computers on both sides. This would allow for evaluating the IDPS behavior under high network traffic. .

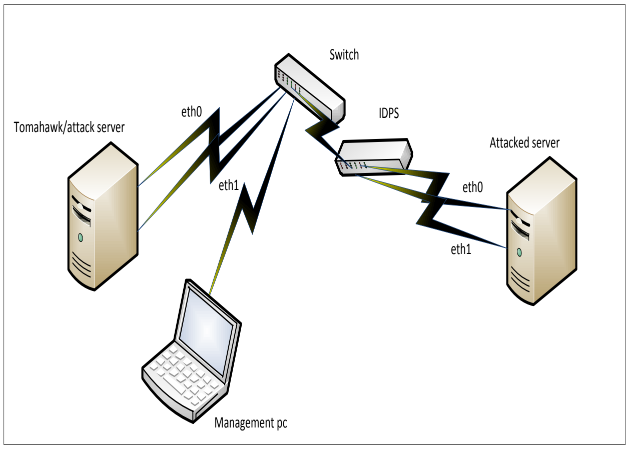


Figure 4. Advanced Setup

# 5 Advantages of using Tomahawk

Tomahawk was chosen for this setup due to the advantages it offers over other tools that replays captured network packets. Tomahawk uses simple commands and flags that can be teamed together to easily manipulate the traffic going to the attacked computer. It can take a small packet and manipulate it to produce the desired traffic flow.

Some advantages of using tomahawk include:

- Tomahawk is free and publicly available
- It is simple to use
- It is very stable and mature
- Does not require a lot of resources to run it
- Can evaluate both software and hardware based IDPS

## 5.1 PCAP File

A PCAP file is a file that contains captured network activity and saved in the libpcap format with a PCAP extension. This format and extension allows the file to be used by multiple network related tools on most current operating systems. Tomahawk works by replaying and manipulating PCAP files. Tomahawk does not create its own PCAP files but they can be created by other network monitoring tools such as Wireshark. Wireshark is an open source network monitoring tool that has multiple functions and it runs on most current operating systems. PCAP files can also be downloaded from the Internet from trusted sources. There are advantages to creating own PCAP file for use with Tomahawk. Using own PCAP files allows to use traffic that is representative of the environment where the IDPS will reside and protect. It also allows the capture of traffic at different times with diverse load situations. This facilitates different mixes of traffic volumes and applications on the network. Using a downloaded pcap may not present a true picture of the environment being tested which can led to a wrong IDPS been chosen.

# 6 Testing Methodology

Tomahawk can be configured and used in variety ways to support different test configurations. The three setups described above are examples of setting up different test environments for evaluating an IDPS using Tomahawk. Tomahawk works the same way regardless of the configuration of the setup and it supports both software and hardware based IDPS. Tomahawk works by replaying captured network packets that are saved as a pcap file in a bi-direction fashion and breaking the PCAP file into two pieces and then assigns these pieces to the client and the other to the server [13]. Using this system allows Tomahawk to keep track of the PCAP file as it is replayed. By breaking the PCAP file into packets allows Tomahawk to assign the first IP address it encounters in the PCAP file to the client and the second IP address to the server. This process is repeated until the whole pcap file is replayed entirely. Once the PCAP file is broken down into packets and the client and server IP addresses are assigned, tomahawk starts replaying the packets. The client packets are sent out on eth0 and the server packets are send out on eth1. Tomahawk has a default of 0.2 seconds for re-transmissions of lost packets and it also auto manages other network related tasks such as MAC addresses, client and server IP address. If the IDPS detects and blocks the pcap file that contains an attack, tomahawk will report a time out. If tomahawk reports that the pcap containing an attack completed without any errors, then the IDPS will have missed the attack [13].

## 6.1 Capturing the PCAP files

We use Wireshark to capture and create four PCAP files that we use in our test environment. Creating PCAP files with Wireshark is documented in [12]. The PCAP files we use are created using default settings in Whireshark. We started Wireshark and started recording the network traffic and then initiated the attacks/exploits that way we capture all the packets related to the attack. The following PCAP files were created:

### 6.1.1 PCAP1

This is a simple file that contained normal network traffic and no attack traffic. This file is a capture of traffic browsing a server. This capture will be used to test how the IDPS handles normal traffic and establish some baselines.

### 6.1.2 PCAP2

This file is a capture of a known OS exploit and will be used to test if the IPDS will detect and respond to the attack.

### 6.1.3 PCAP3

This file contains a DOS attack on the server and will be used to test how the IDPS detects and responds to the attack.

### 6.1.4 PCAP4

This file is a capture of an exploit and a DOS attack while there is high volume of traffic on the network. This file will be used to verify how the IDPS reacts under different situations.

## 6.2 Using Tomahawk

Tomahawk is a command line utility that runs on Linux based operating systems. Tomahawk commands can be used to run a basic evaluation on an IDPS. To use Tomahawk just type *tomahawk* on the command prompt followed by any of the flags. A detailed explanation of Tomahawk's every command and flag is detailed in [13]. In our test setup we used the Medium setup described above with the following hardware and software:

- The attack workstation- IBM Workstation running SUSE Linux
- A switch that has DHCP and IDPS capabilities
- The attacked server- An IBM Workstation running SUSE Linux
- Four network cables

The attack workstation and the attacked server were connected through the switch/IDPS. Care was taken to make sure that all the traffic from the attack workstation to the attacked server passed through the IDPS.

The first test involved the PCAP1 been replayed against the attacked server and these are the Tomahawk commands we used:

```
tomahawk -l 2 -f pcap1.pcap
```

This command replayed the pcap1 file twice and produced the following output:

```
Beginning test
Completed 1 loop of trace pcap1.pcap
Completed 1 loop of trace pcap1.pcap
Finished 2 loops of trace pcap1.pcap Completed: 2, Timed out: 0 Retrans: 0 Sent: 1686 Recv: 1686
```

This output shows that the both replays finished without being blocked. If the pcap1 file was blocked/dropped by the IDPS then the loop will have not completed.

```
tomahawk –l 2 –f pcap2.pcap
```

The above command replayed PCAP2 file which contained a known exploit against the server. The IDPS blocked this attack and the packets were dropped. As a result the loops did not complete.

```
tomahawk –l 2 –f pcap3.pcap
```

The above command replayed PCAP3 file which contained a DOS attack against the server. The IDPS blocked this attack and the packets were dropped. As a result the loops did not complete.

```
tomahawk –l 2 –f pcap4.pcap
```

The above command replayed PCAP4 file which contained an exploit and a DOS attack while there is high volume of traffic on the network. These replay packets were dropped by the IDPS and as result the loops did not complete.

## 7 Conclusion and future work

There are two main problems in evaluating the effectiveness of an IDPS. One is the lack of previous work that lays down a simple and straight forward way for setting an evaluation lab that can evaluate both a software and hardware based IDPS. The other problem is the lack of publicly available datasets that can be used in the evaluation of an IDPS. This work targeted both problems by presenting a basic setup for evaluating an IDPS using Tomahawk and Wireshark. The evaluation setup presented three ways to setup the hardware and software involved in testing an IDPS and how it can be used to evaluate both the software and hardware based IDPS.

Future work entails building out a test environment based on the setups presented here and carrying out some experiments with different IDPS products currently on the market. The work will also involve documenting how to create PCAP files for use with Tomahawk using Wireshark.

## 8 References

[1] J. W. Haines, R.P. Lippmann, D.J. Fried, M.A. Zissman, E. Tran, and S.B. Boswell. 1999 DARPA Intrusion Detection Evaluation: Design and Procedures. MIT Lincoln Laboratory: Lexington, MA, 2001.

[2] J. McHugh. "Testing Intrusion Detection Systems: A Critique of the 1998 and 1999 DARPA Intrusion Detection System Evaluations as Performed by Lincoln Laboratory", Proc. ACM TISSEC 3(4) 262-294, 2000.

[3] Mahoney, M. V. & Chan, P. K. (2003), An Analysis of the 1999 DARPA/Lincoln Laboratory Evaluation Data for Network Anomaly Detection, In Proceedings of the Sixth International Symposium on Recent Advances in Intrusion Detection, Springer-Verlag, 220-237, 2003.

[4] Brugger, S. & Chow, J. An assessment of the DARPA IDS Evaluation Data set using Snort, Technical report, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, 2007.

[5] Cardenas, A.A., Baras, J.S., Seamon, K.: A framework for the evaluation of intrusion detection systems. In: Proc. of the IEEE Symposium on Security and Privacy (IEEE security'06), Washington, DC, USA,IEEE Computer Society, 63–77, 2006.

[6] Sommers, J., Yegneswaran, V., & Barford, P. Toward comprehensive traffic generation for online ids evaluation. University of Wisconsin: Tech Rep, 2005.

[7] Corsini, J. Analysis and Evaluation of Network Intrusion Detection Methods to Uncover Data Theft, Master's thesis, Edinburgh Napier University, Edinburgh, UK, 2009.

[8] Athanasiades, N., Abler, R., Levine, J., Owen, H. & Riley, G. Intrusion detection testing and benchmarking methodologies', Proceedings of First IEEE International Workshop on Information Assurance, IWIAS2003, 63-72, 2003.

[9] Sommers, J., Yegneswaran, V. & Barford, P. Toward Comprehensive Traffic Generation for Online IDS Evaluation, Technical report, Department of Computer Science, University of Wisconsin, Madison, 2005.

[10] Walsh, J & Koconis, D. Cleaning Packet Captures for Network IPS cleaning, 2006. http://www.icsalabs.com

[11] Sannella, M. J. Constraint Satisfaction and Debugging for Interactive User Interfaces. Doctoral Thesis. UMI Order Number: UMI Order No. GAX95-09398, University of Washington, 1994.

[12] Wireshark 2012. http://www.wireshark.org/

[13] Tomahawk 2012. http://tomahawk.sourceforge.net/

[14] T. Brugger. KDD cup'99 dataset (network intrusion) considered harmful,15Sept, 2007. http://www.kdnuggets.com/news/2007/n18/4i.html.

[15] Mudzingwa, D & Agrawal. R. A Study of Methodologies used in Intrusion Detection and Prevention Systems, Proceedings of the IEEE SoutheastCon2012, Mar 2012.

# Towards Effective Integrated Access Control Lists in Internet Networks

*Kamal A. Ahmat*
*Department of Information Technology*
*City University of New York*
*New York, NY 11101*
kamal.ahmat@live.lagcc.cuny.edu

*Ammar Elnour*
*School of Information Technology*
*The University of Sydney*
*NSW 2006, Australia*
ammar@it.sydney.edu.au

## Abstract

Access Control Lists (*ACL*s) represent a traditional way in filtering packets in routers. In modern complex enterprise networks that provide a vast array of services, there is an ever increasing need for verifying the integrity of *ACL*s to detect any potential security holes and improve the network performance. This paper concerns the integrity of routers' *ACL*s in large enterprise networks. We first investigate the integrity of the *ACL*s of two routers by describing a bottom-up approach for detecting redundancies in *ACL*s of two routers. We then extend our study to multiple routers and provide a heuristic algorithm for detecting redundant *ACL*s in multiple routers. We validate the practicality of our algorithm through real-life and synthetic router *ACL* groups of large networks. Performance results show that our heuristic algorithm don't only improve the performance by reducing the number of comparisons overhead, but also helps in discovering potential security holes that cannot be discovered by considering the *ACL*s of each router individually.

## I. Introduction

In today's complex high performance internetworks, there is increasing need to implement packet forwarding and routing efficiently and effectively by organizations according to their own defined policies. Such implementations do not only need to be in a way that goes beyond traditional issues related to routing protocols, but should also, most importantly, address the security concerns of these organizations. Even though firewalls serve as the first line of defense against cyber attacks, relying solely on the protection provided by firewalls is not sufficient to secure networks' infrastructures. Consequently, many router products allow a firewall capability, such as Access Control Lists (*ACL*s) to filter packets and provide additional protection line within the network infrastructure.

The *ACL*s are mainly created for each router on an individual basis and stored on a file. Then, network administrators usually use TFTP servers to grab the preconfigured *ACL* file and perhaps modify it to configure each router's security policy. The *ACL*s on these policy files are often have different moment, some grant the access for certain applications, others for IPs and protocols, third ones make accessible the certain interfaces, etc. Furthermore, for large-scale complex networks, these files can be modified by multiple administrators over an extended period of time, which makes the *ACL*s prone to conflicts, redundancy and security holes. The situation even becomes more complex and critical if the integrity of *ACL*s of all routers in the network infrastructure, which is often ignored, is considered.

Once written, the *ACL* is experiencing a constant modification, which becomes painstaking with the course of time. The number of *ACL*s in a group increases constantly which doesn't only raise serious scalability and performance issues, but also highly increases the probability of anomalies among these *ACL*s, which in turn raises serious security concerns.

The anomalies have double effect: on the one hand, they can compromise organization for letting confidential information out, on the other hand, they expose the network to external attacks, the consequences of which are hard to predict [1]. Despite the importance of automating the process of discovering *ACL* integrity at the network level, we are not aware of any commercial platform available on the market today that offer tools for analyzing the integrity or discovering and eliminating anomalies in routers' *ACL*s at the network level. The situation becomes even more complex with the presence of extensions to the standard *ACL*s such as Dynamic, Reflexive, and Time-based *ACL*s.

Even though firewall policy rules and routing *ACL* groups may look similar, there are several fundamental differences between them:

1) Firewall policies filter the network inbound/oubound traffic, while router's *ACL*s control all types of traffic including the traffic that is originated from within the network. Thus, routers are used to secure the network infrastructure against insider threats. While firewalls can be also installed within the network, their effect remains limited comparing to the routers.
2) A typical firewall has one incoming and one outgoing interfaces, and the firewall decides to either accept or deny the traffic. However, routers have multiple interfaces and the traffic control can vary based on the source and destination interfaces. Thus, routers are generalization of firewalls.
3) Firewalls don't deploy any other strategies besides policies to filter the traffic, while routers take the decision based on routing strategies (i.e. protocols) which coordinate the communication between routers.
4) The vast proliferation of virtualization, and consequently virtual networks, resulted in unique set of open issues that are distinctly related to routing and associated policies. For instance, the impact of features such as Router Grafting [8] on the security of network infrastructures remains vague.

This paper represents an initial exploration step in studying this unknown field, posing more questions than it answers. We hope presenting this paper will draw the attention from the research community on this challenging and important problem. In this paper, we address the problems related to the integrity of routers' *ACL*s in the network's infrastructure. First, we describe a bottom-top method for verifying the integrity of two router's *ACL* lists and find any potential redundant *ACL*s. Then, we generalize our approach to multiple routers. After that, we validate the practicality of our algorithms by conducting experiments on multiple large network infrastructures as well as on synthetically generated data. Our results demonstrate that our approach does not only find and eliminate redundant *ACL*s, but also improves the security level of the network by discovering potential security holes that cannot be discovered by considering the *ACL*s of each router individually.

### A. Organization of the Paper

The rest of the paper is organized as follows. The next section describes the system model used in this paper and problem formulation. We subdivide Section III into two subsections. First, in Subsection III-A we present an algorithm to find the set of redundant *ACL*s in a pair of *ACL* groups. Second, in Subsection III-B we describe an algorithm for discovering redundant *ACL*s in multiple groups. Our environment setup and experimental results are described in Section IV. Finally, Section V concludes the paper.

## II. System Model and Problem Formulation

In this section, we describe the model we adopted and present formal description of the problem addressed in this paper.

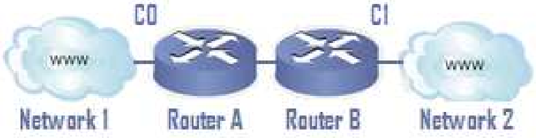


Fig. 1. A typical network with two routers. For the traffic originating at Network 1 and destinated at Network 2, Router A is upstream while Router B is downstream.

### A. Routers and Access Control Lists

A router is a network device that is designed to control data traffic between different computer networks based on what is known as a routing table. It reads every data packet that it receives from a network and directs it through the right connection, so that it reaches the right destination based on what is known as *Routing Table*[15]. The routing table stored on a router has information about the IP addresses of all computers in a network. It also has a set of priority rules for data transport, which the router follows. Furthermore, routers can filter network traffic by deciding whether packets will be *accepted* or *denied* based on what is known as the Access Control List (*ACL*). An *ACL* is typically assigned to a specific router interface and it defines the parameters against which each connection is compared, resulting in a decision on what action to take. As soon as a network packet matches an *ACL*, the corresponding action is applied, and processing stops. If the action is acceptance, then the packet is forwarded to the appropriate destination based on the routing table, otherwise, it is dropped. If no matches are found when the router reaches the end of the list, the traffic is denied. For this reason, each *ACL* group should have the frequently hit entries at the top of the list. Also, there is an implied deny for traffic that is not permitted. A single-entry *ACL* with only one deny entry has the effect of denying all traffic. Consequently, there must be at least one permit statement in an *ACL* group or all traffic is blocked.

The standard *ACL* is typically a five-tuple filter that has the following format:

(order, **src_ip**, **src_mask**, **dst_ip**, **protocol**, **action**)

The first attribute is the order assigned to the *ACL* and represents the relative order of the *ACL* within the given *ACL* group. To improve the performance and scalability of routers, the *ACL*s that have higher hit probability are typically listed on the top of *ACL* group so that the number of comparisons is kept minimum.

The remaining attributes represent the source IP address, source IP mask, destination IP address, protocol, and action, respectively. Additional parameters such as the port number, timeout, traffic type (e.g. in or out) are also often used. However, in this paper, we assume the most basic model consisting of the six attributes described above.

To be able to build a useful model for filtering *ACL*s, we need to determine all the relations that may relate two or more packet filters. In this section we state all the possible relations that may exist between filtering rules. Since router *ACL*s can be looked at as a superset of firewall filtering policies, all types of relationships between firewall rules are also inherited in a router *ACL*s. Any two firewall rules can be either (1) disjoint; (2) exactly matched; (3) inclusively matched; (4) partially disjoint; or (5) correlated [3]. Besides the relationships listed above, we also define the sibling relationship for two *ACL*s on the same router as follows.

*Definition 2.1:* [SIBLING-*ACL*s]
Two *ACL*s $a_l$ and $a_2$ are siblings if and only if the following conditions are satisfied:

1) $a_1$ and $a_2$ are completely disjoint; and
2) If $a_1$ succeeds $a_3$ in the *ACL* group, denoted $a_1 \succ a_3$, then $a_2 \succ a_3$.

□

We use the notion $\bowtie_i$ as a function that lists the relationships between two *ACL*s in a given group $i$. For example, in the *ACL* group described in Table 1, $a_5 \bowtie_1 a_6$ means $a_5$ and $a_6$ are siblings since they both are disjoint and that $a_5 \succ a_3$ and $a_6 \succ a_3$, while $a_6 \bowtie_1 a_3$ means $a_6 \succ a_3$.

Since router *ACL* groups inherit the relationship types from the firewall rules, the anomalies are also implicitly inherited. These anomalies include shadowing, correlation and redundancy. Throughout this paper, we assume that the firewall policy anomalies are not present. That is, each *ACL* group is optimally configured, and we focus on the optimization of *ACL* groups integrity. This assumption is practical as the research community has addressed practical methods to resolve them [3], [4], [6], [7], [10]. However, the model for router groups can be much more complex. For example, both *ACL* groups for Router A and Router B contain the *ACL* $a_4$ which may seem as a redundancy anomaly. But by looking carefully at both routers' *ACL* groups it can be observed that a data packet can be treated differently by each router's *ACL* group. For example, an *IP* packet that is originated at *IP* address 10.1.1.1 will be denied by the group 1 while the same packet will be accepted by the group 2. Thus, the comparison at the *ACL* level may not give an accurate overview about the anomalies in *ACL* groups at the network level.

| **Id** | **Source IP** | **Mask** | **Dest. IP** | **Protocol** | **Action** |
|---|---|---|---|---|---|
| 1 | 10.1.1.1 | 0.0.0.0 | 10.0.0.1 | IP | deny |
| 2 | 192.168.0.0 | 0.0.255.255 | 10.0.0.1 | IP | allow |
| 3 | 10.1.1.0 | 0.0.255.255 | 10.0.0.1 | TCP | allow |
| 4 | 10.1.2.0 | 0.0.255.255 | 10.0.0.1 | UDP | allow |
| 5 | 10.1.1.0 | 0.0.255.255 | 10.0.0.* | SMTP | allow |
| 6 | 10.1.2.0 | 0.0.255.255 | 10.0.0.* | FTP | allow |
| 7 | 192.168.0.0 | 0.0.255.255 | 10.0.0.1 | TCP | allow |
| 8 | 192.168.0.0 | 0.0.255.255 | 10.0.0.1 | UDP | deny |

TABLE I
THE ACL GROUP FOR ROUTER A IN FIG. 1.

| **Id** | **Source IP** | **Mask** | **Dest. IP** | **Protocol** | **Action** |
|---|---|---|---|---|---|
| 2 | 192.168.0.0 | 0.0.255.255 | 10.0.0.1 | IP | allow |
| 3 | 10.1.1.0 | 0.0.255.255 | 10.0.0.1 | TCP | allow |
| 4 | 10.1.2.0 | 0.0.255.255 | 10.0.0.1 | UDP | allow |
| 5 | 10.1.1.0 | 0.0.255.255 | 10.0.0.* | SMTP | allow |
| 6 | 10.1.2.0 | 0.0.255.255 | 10.0.0.* | FTP | allow |
| 7 | 192.168.0.0 | 0.0.255.255 | 10.0.0.1 | TCP | allow |
| 8 | 192.168.0.0 | 0.0.255.255 | 10.0.0.1 | UDP | deny |
| 9 | 192.0.0.0 | 255.255.192.0 | 10.0.0.1 | UDP | deny |

TABLE II
THE ACL GROUP FOR ROUTER B IN FIG. 1.

In our model, we state that an *ACL* is redundant among multiple routers' *ACL* groups if and only if all packets are treated similarly by a specific *ACL* that belongs to all these groups. For instance, *ACL*s $a_7$ and $a_8$ are redundant in both group l and group 2, while, as stated earlier, *ACL* $a_4$ is not redundant.

In today's large enterprise networks, the network may comprise tens or even hundreds of routers. In such networks, it is important to look at the global view rather than considering the *ACL* of each router individually. Furthermore, the situation becomes much more complex when taking into account practical considerations such as the impact resulted in the interaction of the *ACL*s groups with Border Gateway Protocol (BGP) and other protocols. One major anomaly in distributed *ACL*

| Id | Source IP | Mask | Dest. IP | Protocol | Action |
|---|---|---|---|---|---|
| 2 | 192.168.0.0 | 0.0.255.255 | 10.0.0.1 | IP | allow |
| 3 | 10.1.1.0 | 0.0.255.255 | 10.0.0.1 | TCP | allow |
| 5 | 10.1.1.0 | 0.0.255.255 | 10.0.0.* | SMTP | allow |
| 6 | 10.1.2.0 | 0.0.255.255 | 10.0.0.* | FTP | allow |
| 7 | 192.168.0.0 | 0.0.255.255 | 10.0.0.1 | TCP | allow |
| 8 | 192.168.0.0 | 0.0.255.255 | 10.0.0.1 | UDP | deny |

TABLE III
THE REDUNDANT *ACL*S FOR THE *ACL* GROUPS IN TABLE 1 AND TABLE 2.

groups is *ACL* redundancy. Since redundant *ACL*s may affect the performance of the network infrastructure security management, no redundant rules should be present in routers' *ACL* groups in an optimal router group configuration. Consequently, we define a *simple optimal configuration* as configurations of routers *ACL*s such that no redundant *ACL* is present in the given configuration. We use the notion *simple optimal configuration* to distinguish the problem addressed in this paper from more optimally sophisticated configurations that take other factors into account.

To this end we define a notion of the *deletion-safe ACL* as the *ACL* that its deletion from a group doesn't result in accepting traffic that would have been denied otherwise; or denying traffic that would have been accepted otherwise. For example, the *ACL* $a_8$ is deletion-safe since removing it will not affect the router's actions while the *ACL* $a_1$ in Table 1 is not deletion-safe since all packets originated at IP address 10.1.1.1 will be accepted otherwise. It is a best practice to deny the packets at the upstream router (see Fig 1) rather than allowing the packet to enter the network and then deleting it at the downstream router. Consequently, the priority is given to eliminating deletion-safe *ACL*s from upstream routers.

### B. Problem Formulation

Given a collection of *ACL* group configurations each associated to a router in the network, our goal is to discover and remove all deletion-safe redundant *ACL*s from the given set of configurations. Formally, the *simple optimal configuration* problem is defined as follows.

*Definition 2.2:* [SIMPLE-OPTIMAL-CONFIGURATION]
Given a collection $\mathcal{G} = \{G_1, G_2, \ldots, G_n\}$ of *ACL* group configurations, construct another collection $\grave{\mathcal{G}} \subseteq \mathcal{G}$ such that $\grave{\mathcal{G}}$ doesn't contain any redundant *ACL*s and the both collections have identical output on any given input set. □

```
Algorithm Discover-Redundant-ACLs($G_1$, $G_2$)
Input: Two sets $G_1$ and $G_2$ of ACL groups
Output: Set $R$ of redundant ACLs
(* Finds redundant ACLs in a pair of ACLs groups *)
set $R \leftarrow \phi$;
(* Find all ACLs that belong to both groups *)
for every $a_i \in G_1$
    if $a_i \notin G_2$
        set $G_1 \leftarrow G_1 \setminus a_i$;
        set $G_2 \leftarrow G_2 \setminus a_i$
(* Compare relationship type in both groups *);
for every $a_i$ and $a_j \in G_1$
    if $a_i \bowtie_1 a_j = a_i \bowtie_2 a_j$
        if $a_i$ is deletion safe
            set $R \leftarrow R \bigcup a_i$;
        if $a_j$ is deletion safe
            set $R \leftarrow R \bigcup a_j$;
return $R$;
```

Fig. 2. A formal description of the Discover-Redundant-ACLs procedure.

## III. Redundancy Detection Algorithm

In this section we first describe an algorithm for discovering and eliminating *ACL* redundancies in two router configurations. Then, we extend our algorithm to multiple routers. After that, we analyze the complexity of our algorithms and provide an example.

### A. Discovering Redundant ACLs in Two ACL Groups

The Discover-Redundant-ACLs procedure receives as an input two groups of router *ACL*s. It then eliminates from each group the *ACL*s that are not present in the other group. At this stage, both *ACL* groups contain the same set of *ACL*s but these ACLs could have different relations. The algorithm then finds the relation between each pair of *ACL*s from the first group and checks whether the same relationship is maintained between this pair in the second group eliminating all *ACL*s that have different relationship from groups. The algorithm then returns the set $R$ of redundant *ACL*s. The formal description of the Discover-Redundant-ACLs procedure is given in Fig. 1.

To derive the time complexity of the algorithm we observe that the first stage of the algorithm, which finds similar *ACL*s in both configuration, compares each *ACL* from the first group to each *ACL* from the second group, resulting of an $O(n^2)$ time complexity, where $n$ is the number of *ACL*s in the smaller group. The second phase finds the type of relationships between each pair *ACL*s in the same group, then compares the results with the ones obtained from the other group. Consequently, we derive that the time complexity of the algorithm is $O(n^2)$.

```
Algorithm Network-Redundant-ACLs($\mathcal{G}$)
Input: A collection $\mathcal{G}$ of ACL group sets
Output: A collection $\mathcal{G}$ of redundant-free ACL group
        sets
(* Eliminates redundant ACLs from ACL groups *)
find $G_i$, $G_j$ $s.t.$ $G_i \bigcap G_j$ is maximum;
set $R$(i,j) ← Discover-Redundant-ACLs($G_i$, $G_j$);
if $R_i$ is $upstream\ router$
        set $G_j \leftarrow G_j \setminus R(i, j)$;
    else set $G_i \leftarrow G_i \setminus R(i, j)$;
return $\mathcal{G}$;
```

Fig. 3. A formal description of the Network-Redundant-ACLs algorithm.

### *B. Discovering Redundant* ACL*s for Multiple* ACL *Groups*

Unfortunately, generalizing the Discover-Redundant-ACLs algorithm to multiple groups will result in an algorithm that has the complexity of O($n^k$), which is exponential. We suspect that finding redundant *ACL*s in multiple groups is an NP-hard problem. Consequently, we develop a heuristic for eliminating redundancies from *ACL* groups even though it doesn't guarantee finding the optimal collection that doesn't contain any *ACL* redundancy.

Our algorithm, which is termed Network-Redundant-ACLs, takes as an input a collection of groups such that each group consists of a list of *ACL*s. The algorithm follows the greedy approach and starts by finding two groups such that these groups share the maximum number of exact matching *ACL*s. Then, it applies the Discover-Redundant-ACLs procedure and finds and eliminates the redundant *ACL*s. These steps are repeated until all redundancies are eliminated. Observe that the result obtained from the Network-Redundant-ACLs algorithm may depend on the order in which the group pairs are processed. For instance, in a collection that contains three groups, processing $G_l$ and $G_2$ first may result in a different result than processing $G_l$ and $G_3$ first, which supports our hypothesis that the problem is NP-hard.

To derive the time complexity of the Network-Redundant-ACLs algorithm we observe that the Discover-Redundant-ACLs procedure is used to find redundant *ACL*s between each pair of configurations. The step is repeated $k$ times where $k$ is the number of groups (routers) in the network. Consequently, the total cost of removing all redundant *ACL*s from the network is $O(kn^2)$, where $k$ is the number of routers in the network and $n$ is the number of *ACL*s in the smallest group. We strongly believe that more efficient methods for redundancy discovery and elimination can be developed.

*Example 3.1:* Consider the *ACL* groups depicted in Table 1 and Table 2. The algorithm will first eliminate from both $G_1$ and $G_2$ all *ACL*s that are not present in both groups which will result in $\grave{G}_1 = \{a_2, a_3, a_5, a_5, a_6, a_7, a_8\}$ and $\grave{G}_1 = \{a_2, a_3, a_5, a_5, a_6, a_7, a_8\}$. We observe that $a_5^1 \bowtie a_6^1 = a_5^2 \bowtie a_6^2$ ($a_5^1$ and $a_6^1$ are siblings, and $a_5^2$ and $a_6^2$ are siblings). Furthermore, $a_3^1 \prec a_5^1$, $a_3^1 \prec a_6^1$, $a_3^2 \prec a_5^2$, and $a_3^2 \prec a_6^2$. Consequently, all $a_3$, $a_5$, and $a_6$ will be added to $R$. After that, since $a_4^1 \bowtie a_3^1 \neq a_4^2 \bowtie a_3^2$, $a_4$ is not redundant and consequently, it will not be added to $R$. Also, $a_2$, $a_7$, and $a_8$ will be added to $R$. Thus, the set $R(G_1, G_2)$ is the group of $ACL$s depicted in Table III. □

## IV. Implementation and Experimental Results

We implemented the algorithms described in this paper in the context of Verizon Internet Security Suite's (Beta) system using platform-independent C++. In this section, we present our evaluation study of the scalability and performance of our method. To evaluate the performance of our algorithms, we conducted two sets of experiential tests. In the first set, In the first set, we obtained measurements of original network routers' *ACL* configurations for thirty-day period from six distinct networks and then ran our algorithms on the same networks and obtained measurements for thirty days and calculated average performance results for both sets. In the second set, we conducted further stress tests on synthetic routers' configurations that have very large number of *ACL*s. Since the goal of this paper is to discover network-level *ACL* anomalies rather than router-level anomalies, the *ACL*s of each router were first ordered and conflicts were removed using the methods described in [3], [7].

To test the performance of our methods on real-life networks, we deployed our implementation component in six distinct networks that consisted of 11, 24, 52, 107, 131, and 158 routers, respectively. For the stress test, we ran six test sets on very large *ACL* groups. For each test set, we generated random network topologies using *BRITE* brite and then generated the *ACL*s of each router using a modified version of the method described in [13]. Then, we submitted the *ACL* group files to our component and calculated the average for each test set. The performance results are shown in Tables IV and V, respectively. One of our main observation is that network administrators have the tendency to often add, remove, and edit *ACL*s that control active traffic in the network, which implies that the majority of redundant *ACL*s among multiple routers are mostly related to the active traffic in the network. Thus, de-

| **Network** | **No. of Routers** | **Avg. No. of ACLs** | **Avg. No. of Redundant ACLs** | **Imp. Ratio** |
|---|---|---|---|---|
| **1** | 11 | 35 | 3.7 | 11.3% |
| **2** | 24 | 34 | 2.6 | 8.2% |
| **3** | 52 | 61 | 7.1 | 6.7% |
| **4** | 107 | 42 | 9.8 | 9.2% |
| **5** | 131 | 66 | 17.8 | 21.7% |
| **6** | 158 | 71 | 12.4 | 17.8% |

TABLE IV
PERFORMANCE RESULTS FOR REAL-LIFE *ACL* GROUPS.

| **Network** | **No. of Routers** | **Avg. No. of ACLs** | **Avg. No. of Redundant ACLs** | **Imp. Ratio** |
|---|---|---|---|---|
| **1** | 200 | 500 | 85.4 | 43.1% |
| **2** | 300 | 700 | 121.7 | 24.6% |
| **3** | 400 | 1000 | 64.2 | 35.1% |
| **4** | 500 | 1500 | 213.2 | 21.9% |
| **5** | 1000 | 2000 | 262.1 | 36.8% |
| **6** | 2000 | 2500 | 394.5 | 64.1% |

TABLE V
PERFORMANCE RESULTS FOR SYNTHETIC *ACL* GROUPS.

tecting and removing these redundant *ACL*s result in considerable performance improvement. Obviously, the average comparison numbers depend on the traffic type and the complexity of *ACL*s which we didn't consider at this stage. We observed that the performance ratio of our method is not affected by the increasing number of routers and the numbers of *ACL*s, which validates the high scalability and efficiency of our approaches.

In general, we have discovered that our algorithm is sufficiently fast and highly scalable for all practical purposes. As can be seen in our results, the performance improvement ratio was quite high and our method helped in discovering many redundancies as well as potential security holes that were undiscovered by network administrators in real-life configurations.

## V. Conclusion

Can more sophisticated network-level techniques provide significant improvement on security and performance gains over the conventional device-level methods used currently in large networks? Even though the initial results we obtained are very encouraging, we believe more research needs to be conducted to find the precise answer to this question. This paper addresses a new unexplored direction in networking security by considering integrating security configurations for all routers in the network. We provide initial exploration of the problem and the motivations behind it. Then, we study the problem of discovering and eliminating redundant *ACL*s from multiple routers' configurations and describe efficient methods for removing such redundancies. We have implemented our methods and validated its practicality on both real-life as well as synthetically generated access control lists on several large-size networks. The experimental results support our theoretical analysis and show that our proposed method improves the performance of firewall performance significantly and can also discover potential security holes in network infrastructures.

## References

[1] C. Benecke, *A Parallel Packet Screen for High Speed Networks*, In Proc. Proceedings of the 15th Annual Computer Security Applications Conference (ACSAC '99), 1999.

[2] BRITE, *http://www.cs.bu.edu/brite*. Boston University, 2002.

[3] E. Al-Shaer and H. Hamed, *Firewall Policy Advisor for Anomaly Discovery and Rule Editing*, IFIP/IEEE Eighth International Symposium on Integrated Network Management, 2003, pp. 17-30.

[4] E. Al-Shaer, H. Hamed, R. Boutaba, M. Hasan, *Conflict Classification and Analysis of Distributed Firewall Policies*, IEEE Journal on Selected Areas in Communications 23(10): 2069-2084, 2005.

[5] D. Decasper, Z. Dittia, G. Parulkar, B. Plattner, *Router Plugins A Software Architecture for Next Generation Routers*, In Proc. of SIGCOMM, 1998.

[6] E. W. Fulp, *Optimization of Network Firewall Policies using Directed Acyclical Graphs*, In Proc. IEEE Internet Management Conference (IM '05), 2005.

[7] H. Gobjuka and K. Ahmat, *Fast and Scalable Method for Resolving Anomalies in Firewall Policies*, In Proc. of Global Internet Symposium (In conjunction with IEEE INFOCOM), 2011.

[8] E. Keller, J. Rexford, J. van der Merwe, *Seamless BGP Migration With Router Grafting*, In Proc. of USENIX NSDI, 2010.

[9] T. V. Lakshman and D. Stidialis, *High Speed Policy-based Packet Forwarding Using Efficient Multi-dimensional Range Matching*, In Proc. of SIGCOMM, 1998.

[10] A. Liu, C. R. Meiners, Y. Zhou, *All-Match Based Complete Redundancy Removal for Packet Classifiers in TCAMs*, In Proc. IEEE INFOCOM, 2008, pp. 574-582.

[11] S. McCanne, and V. Jacobson, *The BSD Packet Filter: A New Architecture for User-level Packet Capture*, In Proc. of the 1993 Winter USENIX Technical Conference, 1993.

[12] L. Qiu, G. Varghese, S. Suri, *Fast firewall implementations for software-based and hardware-based routers*, In Proc. of SIGMETRICS, 2001.

[13] D. Rovniagin, A. Wool, *The geometric efficient matching algorithm for firewalls*, In Proc. of IEEE Convention of Electrical and Electronics Engineers in Israel (IEEEI), 2004, pp. 153156.

[14] V. Srinivasan, G. Varghese, S. Suri, and M. Waldvogel, *Fast Scalable Level Four Switching*, In Proc. of SIGCOMM, Sep. 1998.

[15] A. Tanenbaum, *Computer Networks, Fourth Edition*. Prentice Hall Ptr., 2002.

# Polymorphic Worms Detection Using A Supervised Machine Learning Technique

Mohssen M. Z. E. Mohammed, H. Anthony Chan,
Neco Ventura
Dept. of Electrical Engineering, University of Cape Town
Rondebosch, South Africa
m_zin44@hotmail.com; h.a.chan@ieee.org;
neco@crg.ee.uct.ac.za

Mohsin Hashim, Eihab Bashier
Faculty of Mathematical Science, University of Khartoum
Khartoum, Sudan
mohsinhashim@yahoo.com; eihabbashier@gmail.com

*Abstract— Polymorphic worms are considered as the most dangerous threats to the Internet security, and the danger lies in changing their payloads in every infection attempt to avoid the security systems. We have designed a novel double-honeynet system, which is able to detect new worms that have not been seen before. To generate signatures for polymorphic worms we have two steps. The first step is the polymorphic worms sample collection which is done by a Double-honeynet system. The second step is the signature generation for the collected samples which is done by using a Support Vector Machines (SVMs) technique. The system is able to generate accurate signatures for single and multiple worms.*

*Keywords-honeynet; worms; machine learning algorithm.*

## 1. Introduction

An Internet worm is a self-propagated program that automatically replicates itself to vulnerable systems and spreads across the Internet. Worms take the attack process one step further by self-replicating. Once a worm has compromised and taken over a system, it begins scanning again, looking for new victims. Therefore a single infected system can compromise one hundred systems, each of which can compromise another one hundred more systems, and so on. The worm continues to attack systems this way and grows exponentially. This propagation method can spread extremely fast, giving administrators little time to react and ravaging entire organizations. Although only a small percentage of individuals can identify and develop code for worms, but once the code of a worm is accessible on the Internet, anyone can apply it. The very randomness of these tools is what makes them so dangerous. A polymorphic worm is a worm that changes its appearance with every instance [1].

It has been shown that multiple invariant substrings must often be present in all variants of worm payload. These substrings typically correspond to protocol framing, return addresses, and in some cases, poorly obfuscated code [8].

Intrusion detection systems serve three essential security functions: they monitor, detect, and respond to unauthorized activities. There are two basic types of intrusion detection: host-based and network-based. Host-based IDSs examine data held on individual computers that serve as hosts, while network-based IDSs examine data exchanged between computers [15, 16].

Our research is based on Honeypot technique. Developed in recent years, honeypot is a monitored system on the Internet serving the purpose of attracting and trapping attackers who attempt to penetrate the protected servers on a network. Honeypots fall into two categories. A high-interaction honeypot such as (Honeynet) operates a real operating system and one or multiple applications. A low-interaction honeypot such as (Honyed) simulates one or multiple real systems. In general, any network activities observed at honeypots are considered suspicious [1, 9].

Supervised machine learning is the search for algorithms that reason from externally supplied instances to produce general hypotheses, which then make predictions about future instances. In other words, the goal of supervised learning is to build a concise model of the distribution of class labels in terms of predictor features. There are several applications for Machine Learning (ML), the most significant of which is data mining. People are often prone to making mistakes during analyses or, possibly, when trying to establish relationships between multiple features. This makes it difficult for them to find solutions to certain problems. Machine learning can often be successfully applied to these problems, improving the efficiency of systems and the designs of machines. Every instance in any dataset used by machine learning algorithms is represented using the same set of features. The features may be continuous, categorical or binary. If instances are given with known labels (the corresponding correct outputs) then the learning is called supervised, in contrast to unsupervised learning, where instances are unlabeled. By applying these unsupervised (clustering) algorithms, researchers hope to discover unknown, but useful, classes of items. Another kind of machine learning is reinforcement learning. The training information provided to the learning system by the environment (external trainer) is in the form of a scalar reinforcement signal that constitutes a measure of how well the system operates. The learner is not told which actions to take, but rather must discover which actions yield the best reward, by trying each action in turn [17].

This paper is organized as follows: Section 2 reviews General issues of supervised learning algorithms. Section 3 discusses the related work regarding automated signature generation systems. Section 4 introduces the proposed system architecture to address the problems faced by current automated signature systems. Signature generation algorithm for Polymorphic Worm will be discussed in section 5. Section 6 concludes the paper.

## 2. General issues of supervised learning algorithms

Inductive machine learning is the process of learning a set of rules from instances (examples in a training set), or more generally speaking, creating a classifier that can be used to generalize from new instances. The process of applying supervised ML to a real-world problem is described in Figure 1.

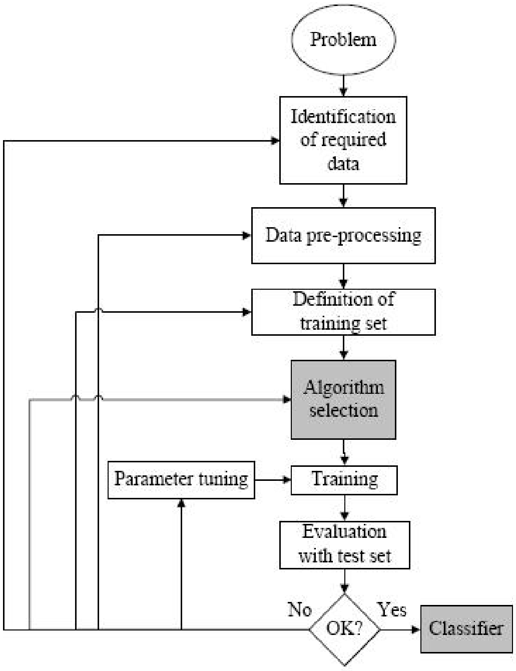


Figure 1. The process of supervised ML

The first step is collecting the dataset. If a requisite expert is available, then s/he could suggest which fields (attributes, features) are the most informative. If not, then the simplest method is that of "brute-force," which means measuring everything available in the hope that the right (informative, relevant) features can be isolated. However, a dataset collected by the "brute-force" method is not directly suitable for induction. It contains in most cases noise and missing feature values, and therefore requires significant pre-processing .

The second step is the data preparation and data preprocessiong. Depending on the circumstances, researchers have a number of methods to choose from to handle missing data. Hodge & Austin [22] have recently introduced a survey of contemporary techniques for outlier (noise) detection. These researchers have identified the techniques' advantages and disadvantages. Instance selection is not only used to handle noise but to cope with the infeasibility of learning from very large datasets. Instance selection in these datasets is an optimization problem that attempts to maintain the mining quality while minimizing the sample size. It reduces data and enables a data mining algorithm to function and work effectively with very large datasets. There is a variety of procedures for sampling instances from a large dataset.

Feature subset selection is the process of identifying and removing as many irrelevant and redundant features as possible. This reduces the dimensionality of the data and enables data mining algorithms to operate faster and more effectively. The fact that many features depend on one another often unduly influences the accuracy of supervised ML classification models. This problem can be addressed by constructing new features from the basic feature set. This technique is called feature construction/transformation. These newly generated features may lead to the creation of more concise and accurate classifiers. In addition, the discovery of meaningful features contributes to better comprehensibility of the produced classifier, and a better understanding of the learned concept [17].

## 3. Related Work

Honeypots are an excellent source of data for intrusion and attack analysis. Levin et al. described how honeypot extracts details of worm exploits that can be analyzed to generate detection signatures [4]. The signatures are generated manually.

One of the first systems proposed was Honeycomb developed by Kreibich and Crowcroft. Honeycomb generates signatures from traffic observed at a honeypot via its implementation as a Honeyd [5] plugin. The longest common substring (LCS) algorithm, which looks for the longest shared byte sequences across pairs of connections, is at the heart of Honeycomb. Honeycomb generates signatures consisting of a single, contiguous substring of a worm's payload to match all worm instances. These signatures, however, fail to match all polymorphic worm instances with low false positives and low false negatives.

Kim and Karp [6] described the Autograph system for automated generation of signatures to detect worms. Unlike Honeycomb, Autograph's inputs are packet traces from a DMZ that includes benign traffic. Content blocks that match "enough" suspicious flows are used as input to COPP, an algorithm based on Rabin fingerprints that searches for repeated byte sequences by partitioning the payload into content blocks. Similar to Honeycomb, Auto-graph generates signatures consisting of a single, contiguous substring of a worm's payload to match all worm instances. These signatures, unfortunately, fail to match all polymorphic worm instances with low false positives and low false negatives.

S. Singh, C. Estan, G. Varghese, and S. Savage [7] described the Earlybird system for generating signatures to detect worms. This system measures packet-content prevalence at a single monitoring point such as a network DMZ. By counting the number of distinct sources and destinations associated with strings that repeat often in the payload, Earlybird distinguishes benign repetitions from epidemic content. Earlybird, also like Honeycomb and Autograph, generates signatures consisting of a single, contiguous substring of a worm's payload to match all worm instances. These signatures, however, fail to match all polymorphic worm instances with low false positives and low false negatives.

New content-based systems like Polygraph, Hamsa and LISABETH [8, 10 and 11] have been deployed. All these systems, similar to our system, generate automated signatures for polymorphic worms based on the following fact: there are multiple invariant substrings that must often be present in all variants of polymorphic worm payloads even if the payload changes in every infection. All these systems capture the packet payloads from a router, so in the worst case, these systems may find multiple polymorphic worms but each of them exploits a different vulnerability from each other. So, in this case, it may be difficult for the above systems to find invariant contents shared between these polymorphic worms because they exploit different vulnerabilities. The attacker sends one instance of a polymorphic worm to a network, and this worm in every infection automatically attempts to change its payload to generate other instances. So, if we need to capture all polymorphic worm instances, we need to give a polymorphic worm chance to interact with hosts without affecting their performance. So, we propose new detection method "Double-honeynet" to interact with polymorphic worms and collect all their instances. The proposed method makes it possible to capture all worm instances and then forward these instances to the Signature Generator which generates signatures, using a particular algorithm.

An Automated Signature-Based Approach against Polymorphic Internet Worms by Yong Tang and Shigang Chen[9] described a system to detect new worms and generate signatures automatically. This system implemented a double-honeypots (inbound honeypot and outbound honeypot) to capture worms payloads. The inbound honeypot is implemented as a high-interaction honeypot, whereas the outbound honeypot is implemented as a low-interaction honeypot. This system has limitation. The outbound honeypot is not able to make outbound connections because it is implemented as low-interaction honeypot which is not able to capture all polymorphic worm instances. Our system overcomes this disadvantage by using double-honeynet (high-interaction honeypot), which enables us to make unlimited outbound connections between them, so we can capture all polymorphic worm instances.

## 4. Double- Honeynet System

We propose a double-honeynet system to detect new worms automatically. A key contribution of this system is the ability to distinguish worm activities from normal activities without the involvement of experts.

Figure 2 shows the main components of the double-honeybet system. Firstly, the incoming traffic goes through the Gate Translator which samples the unwanted inbound connections and redirects the samples connections to Honeynet 1.

The gate translator is configured with publicly-accessible addresses, which represent wanted services. Connections made to other addresses are considered unwanted and redirected to Honeynet 1 by the Gate Translator.

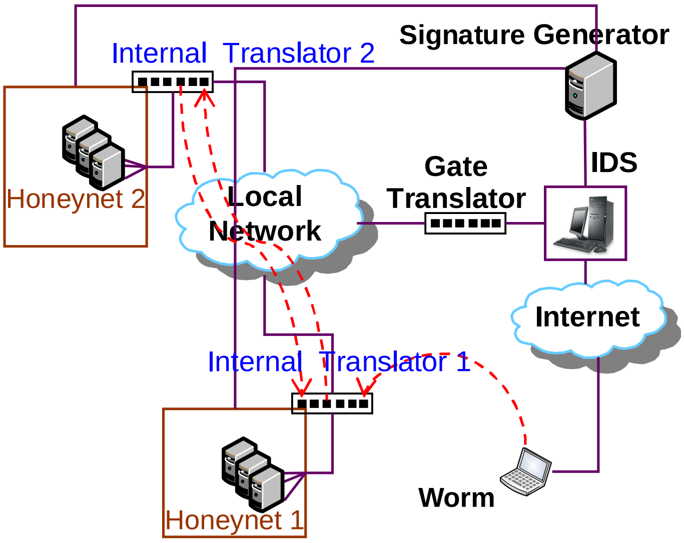


Figure 2. System architecture.

Secondly, Once Honeynet 1 is compromised, the worm will attempt to make outbound connections. Each honeynet is associated with an Internal Translator implemented in router that separates the honeynet from the rest of the network. The Internal Translator 1 intercepts all outbound connections from honeynet 1 and redirects them to honeynet 2 which does the same forming a loop.

Only packets that make outbound connections are considered malicious, and hence the Double-honeynet forwards only packets that make outbound connections. This policy is due to the fact that benign users do not try to make outbound connections if they are faced with non-existing addresses.

Lastly, when enough instances of worm payloads are collected by Honeynet 1 and Honeynet 2, they are forwarded to the Siganture Generator component which generates signatures automatically using specific algorithms that will be discussed in the next section. Afterwards, the Siganture Generator component updates the IDS database automatically by using a module that converts the signatures into Bro or pseudo-Snort format. The above proposed system implemented by using

Vmware Server 2. The implementation results are out of the scope of this paper.

For further details on the double-honeynet architecture the reader is advised to refer to our published works [13].

## 5. Signature Generation Algorithms

In this section, we describe the Support Vector Machines technique which we use it to generates signatures for polymorphic worms.

Support Vector Machines (SVMs) are the newest supervised machine learning technique [17]. SVMs revolve around the notion of a "margin"—either side of a hyperplane that separates two data classes. Maximizing the margin and thereby creating the largest possible distance between the separating hyperplane and the instances on either side of it has been proven to reduce an upper bound on the expected generalization error.

If the training data is linearly separable, then a pair

(**w**, b) exists such that

$\mathbf{W}^T \mathbf{X}_i + b \geq 1$, for all $\mathbf{X}_i \in P$

$\mathbf{W}^T \mathbf{X}_i + b \leq -1$, for all $\mathbf{X}_i \in N$

With the decision rule given by $f_{w,b}(\mathbf{X})\ \mathrm{sgn}(\mathbf{W}^T \mathbf{X} + b)$, where **w** is termed the weight vector and b the bias (or − b is termed the threshold).

It is easy to show that, when it is possible to linearly separate two classes, an optimum separating hyperplane can be found by minimizing the squared norm of the separating hyperplane. The minimization can be set up as a convex quadratic programming (QP) problem:

$$\mathbf{Minimize}_{w,b}\ \Phi(\mathbf{w}) = \tfrac{1}{2}\ \|\mathbf{w}\|^2 \qquad (1)$$

Subject to $y_i\ (\mathbf{W}^T \mathbf{X}_i + b) \geq 1$, i=1,…, *l*.

In the case of linearly separable data, once the optimum separating hyperplane is found, data points that lie on its margin are known as support vector points and the solution is represented as a linear combination of only these points (see Figure 3 ). Other data points are ignored.

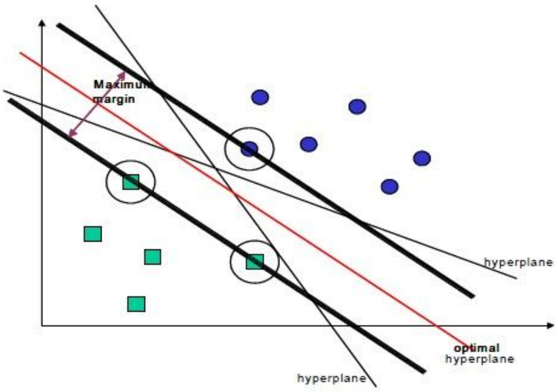
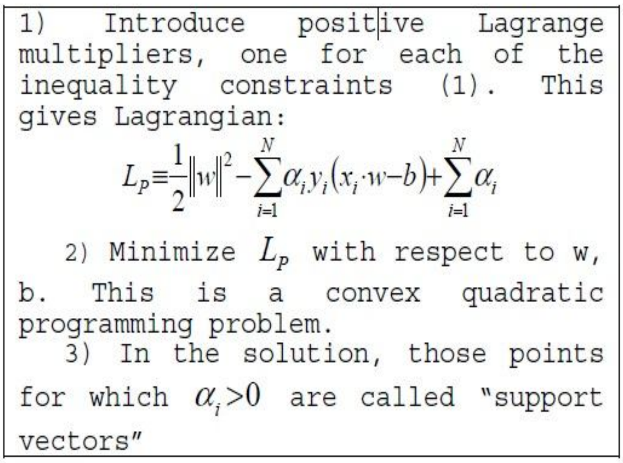


Figure 3. Maximum Margin

Therefore, the model complexity of an SVM is unaffected by the number of features encountered in the training data (the number of support vectors selected by the SVM learning algorithm is usually small). For this reason, SVMs are well suited to deal with learning tasks where the number of features is large with respect to the number of training instances.

A general pseudo-code for SVMs is illustrated in Figure 4.



```
1) Introduce positive Lagrange multipliers, one for each of the inequality constraints (1). This gives Lagrangian:
        $L_P \equiv \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N}\alpha_i y_i(x_i \cdot w - b) + \sum_{i=1}^{N}\alpha_i$
   2) Minimize $L_p$ with respect to w, b. This is a convex quadratic programming problem.
   3) In the solution, those points for which $\alpha_i > 0$ are called "support vectors"
```

Figure 4. Pseudo-code for SVMs

Even though the maximum margin allows the SVM to select among multiple candidate hyperplanes, for many datasets, the SVM may not be able to find any separating hyperplane at all because the data contains misclassified instances. The problem can be addressed by using a soft margin that accepts some misclassifications of the training instances. This can be done by introducing positive slack variables $\xi_i$ , i =1,..., N in the constraints, which then become:

$w.\, x_{i} - b \geq +1 - \xi$ for $\mathbf{y}_{i} = +1$

$w.\, x_{i} - b \leq +1 + \xi$ for $\mathbf{y}_{i} = -1$

$\xi \geq 0.$

Thus, for an error to occur the corresponding $\xi_i$ must exceed unity, so $\Sigma_i\ \xi_i$ is an upper bound on the number of training errors. In this case the Lagrangian is:

$$L_p = ½\ \|\mathbf{w}\|^2 + C\,\Sigma_i\, \xi_i - \Sigma_i\ \alpha_i\{\mathbf{y}_i\,(x_i.w - b) - 1 + \xi_i - \Sigma_i\, \mu_i\, \xi_i$$

Where the $\mu_i$ are the Lagrange multipliers introduced to enforce positivity of the $\xi_i$.

Nevertheless, most real-world problems involve non-separable data for which no hyperplane exists that successfully separates the positive from negative instances in the training set. One solution to the inseparability problem is to map the data onto a higher dimensional space and define a separating hyperplane there. This higher-dimensional space is called the transformed feature space, as opposed to the input space occupied by the training instances.

With an appropriately chosen transformed feature space of sufficient dimensionality, any consistent training set can be made separable. A linear separation in transformed feature space corresponds to a non-linear separation in the original input space. Mapping the data to some other (possibly infinite dimensional) Hilbert space H as $\Phi : R^d \rightarrow H$. Then the training algorithm would only depend on the data through dot products in H, i.e. on functions of the form $\Phi\,(x_i)\,.\Phi\,(x_j)$.

If there were a "kernel function" K such that $K\,(x_i, x_j) = \Phi\,(x_i).(x_j)$, we would only need to use K in the training algorithm, and would never need to explicitly determine Φ. Thus, kernels are a special class of function that allows inner products to be calculated directly in feature space, without performing the mapping described above. Once a hyperplane has been created, the kernel function is used to map new points into the feature space for classification.

The selection of an appropriate kernel function is important, since the kernel function defines the transformed feature space in which the training set instances will be classified. Genton [21] described several classes of kernels, however, he did not address the question of which class is best suited to a given problem. It is common practice to estimate a range of potential settings and use cross-validation over the training set to find the best one. For this reason a limitation of SVMs is the low speed of the training. Selecting kernel settings can be regarded in a similar way to choosing the number of hidden nodes in a neural network. As long as the kernel function is legitimate, a SVM will operate correctly even if the designer does not know exactly what features of the training data are being used in the kernel-induced transformed feature space.

Some popular kernels are the following:

(1) $K\,(x, y) = (x.y+1)^{p}$

(2) $K\,(x, y) = e^{-\|x-y\|2\,/\,2\,\sigma 2}$

(3) $K\,(x, y) = \tanh\,(K\,x\,.\,y - \delta)^{p}$

Training the SVM is done by solving Nth dimensional QP problem, where N is the number of samples in the training dataset. Solving this problem in standard QP methods involves large matrix operations, as well as time-consuming numerical computations, and is mostly very slow and impractical for large problems. Sequential Minimal Optimization (SMO) is a simple algorithm that can, relatively quickly, solve the SVM QP problem without any extra matrix storage and without using numerical QP optimization steps at all. SMO decomposes the overall QP problem into QP sub-problems. Keerthi and Gilbert [20] suggested two modified versions of SMO that are significantly faster than the original SMO in most situations.

Finally, the training optimization problem of the SVM necessarily reaches a global minimum, and avoids ending in a local minimum, which may happen in other search algorithms such as neural networks. However, the SVM methods are binary, thus in the case of multi-class problem one must reduce the problem to a set of multiple binary classification problems. Discrete data presents another problem, although with suitable rescaling good results can be obtained.

## 6. Conclusion

We have proposed automated detection for Zero day polymorphic worms using double-honeynet. We have proposed new detection method "Double-honeynet" to detect new worms that have not been seen before. The system is based on the Support Vector Machines technique that used to generate signatures for polymorphic worms. The main objectives of this research are to reduce false alarm rates and generate high quality signatures for polymorphic worms.

## 7. References

[1] L. Spitzner, "Honeypots: Tracking Hackers," Addison Wesley Pearson Education: Boston, 2002.

[2] Hossein Bidgoli, "Handbook of Information Security," John Wiley & Sons, Inc., Hoboken, New Jersey.

[3] D. Gusfield, "Algorithms on Strings, Trees and Sequences,", Cambridge University Press: Cambridge, 1997.

[4] J. Levine, R. La Bella, H. Owen, D. Contis ,and B. Culver, "The use of honeynets to detect exploited systems across large enterprise networks," Proc. of 2003 IEEE Workshops on Information Assurance, New York, Jun. 2003, pp. 92- 99.

[5] C. Kreibich and J. Crowcroft, "Honeycomb–creating intrusion detection signatures using honeypots," Workshop on Hot Topics in Networks (Hotnets-II), Cambridge, Massachusetts, Nov. 2003.

[6] H.-A. Kim and B. Karp, "Autograph: Toward automated, distributed worm signature detection," Proc. of 13 USENIX Security Symposium, San Diego, CA, Aug., 2004.

[7] S. Singh, C. Estan, G. Varghese, and S. Savage, "Automated worm fingerprinting," Proc. Of the 6th conference on Symposium on Operating Systems Design and Implementation (OSDI), Dec. 2004.

[8] James Newsome, Brad Karp, and Dawn Song," Polygraph: Automatically generating signatures for polymorphic worms," Proc. of the 2005 IEEE Symposium on Security and Privacy, pp. 226 – 241, May 2005.

[9] Yong Tang, Shigang Chen," An Automated Signature-Based Approach against Polymorphic Internet Worms," IEEE Transaction on Parallel and Distributed Systems, pp. 879-892 July 2007.

[10] Zhichun Li, Manan Sanghi, Yan Chen, Ming-Yang Kao and Brian Chavez. Hamsa, “Fast Signature Generation for Zero-day Polymorphic Worms with Provable Attack Resilience,” Proc. of the IEEE Symposium on Security and Privacy, Oakland, CA, May 2006.

[11] Lorenzo Cavallaro, Andrea Lanzi, Luca Mayer, and Mattia Monga, “LISABETH: Automated Content-Based Signature Generator for Zero-day Polymorphic Worms,” Proc. of the fourth international workshop on Software engineering for secure systems, Leipzig, Germany, May 2008.

[12] J. Nazario. “Defense and Detection Strategies against Internet Worms “. Artech House Publishers (October 2003).

[13] Mohssen M. Z. E. Mohammed, H. Anthony Chan, Neco Ventura. “Honeycyber: Automated signature generation for zero-day polymorphic worms“; Proc. of the IEEE Military Communications Conference, MILCOM, 2008.

[14] C. C. Aggarwal and P. S. Yu, " Outliner Detection for High Dimensional Data," Proceedings of the ACM SIGMOD Conference, Santa Barbara, CA, May 21-24, 2001.

[15] Snort – The de facto Standard for Intrusion Detection/Prevention. Available: http://www.snort.org, 1 March 2012.

[16] Bro Intrusion Detection System. Available: http://www.bro-ids.org/, 5 March 2012.

[17] S. B. Kotsiantis, " Supervised Machine Learning: A Review of Classification Techniques,", Informatica 31 (2007) 249-268.

[18] Haykin, Simon, " Neural Networks: A Comprehensive Foundation (2 ed.),". Prentice Hall. ISBN 0132733501.

[19] MacQueen, J. B. (1967), "Some Methods for classification and Analysis of Multivariate Observations," Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press. pp. 281–297.

[20] Keerthi, S. & Gilbert, E. (2002). Convergence of a Generalized SMO Algorithm for SVM Classifier Design. Machine Learning 46: 351–360.

[21] Genton, M. (2001). Classes of Kernels for Machine Learning: A Statistics Perspective. Journal of Machine Learning Research 2: 299-312.

[22] Hodge, V., Austin, J. (2004), A Survey of Outlier Detection Methodologies, Artificial Intelligence Review, Volume 22, Issue 2, pp. 85-126.

# Firewall Access Control: Static and Dynamic Aspects

Vladimir Zaborovsky, Vladimir Muliukha
Telematics department, State Polytechnical University
Saint-Petersburg, Russia
vlad@neva.ru, vladimir@mail.neva.ru

***Abstract*** *– Modern computer networks store information in the form of distributed digital resources, which have to be available for authorized use and protected against unauthorized access. A real-time control for each of these aspects is a complex technical challenge with static and dynamic interactions. The paper proposes a solution for implementing access control by firewall using dynamic access approach based on virtual connections management and algebra of filtering rules with mechanism of traffic filtering in transparent mode. Proposed security monitor architecture allows to enforce dynamic access policy depended on static set of firewall filtering rules and current condition of virtual connections and network environment.*

*Keywords – security, dynamic access control, virtual connections, traffic management*

## I. Introduction

Information in modern computer networks is stored as distributed dynamic digital resources. The most important information property becomes security. The term “information security” means protecting information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction in order to provide integrity, confidentiality and availability [1].

Usually they allocate two main aspects of an informational security management: the first one is an access right, a confidentiality of information, the second one is an access possibility, availability that is a provision of all necessary resources to implement an access to stored information. In this paper we propose a dynamic approach to access control in modern computer networks based on real-time semantics control, throughput management and algebra of filtering firewall rules. The offered solution can be implemented by a dynamic security monitor based on a stateful inspection firewall and can be applied in the existing network environment.

We consider different dynamic interactions on an access control procedure. It is known [2] that to enforce security policy we have to use some prior information about user and resource in firewall. For example when administrator has to enforce access policy on a new packet filter he forms the filtering rules list. While creating rules, he takes into account not only the access requirements but also the information about local users and their network addresses, different services addresses, ports and so on. Moreover, for a stateful firewall rules we have to use information about protocols and packet sequences. So, the firewall always has to “know” what is the “right” action in current circumstances, to make a correct decision if the packet corresponds to the security requirements.

Nowadays the situation is complicated by the dynamic nature of many network resources. Most of information security approaches were developed long ago when the information stored static and unchanged. Today many of informational resources are formed by servers for a particular users based on their requests. So now it is not enough to check the socket information while the connection is establishing. Nearly all modern application protocols base on TCP so we can’t predict the content of establishing connection.

Is the paper we divide all network connections into three levels: packet streams, technological virtual connections (TVC) and informational virtual connections (IVC). Proposed security monitor ensures access policy constantly throughout the connection. The simplest way to realize such approach is to create firewall rules for an each pair “user-resource”, but modern technical devices can’t process such amount of data in real-time. In our work we propose not to store the list of all rules but to form them for each of users and connections. Practice shows that users working together usually request the same or close informational resources. That is why we offer to optimize the set of filtering rules using special algebra [3].

In the paper we present the architecture of security monitor that is new dynamic access control system, which forms optimized set of filtering rules based on current state of virtual connections and network environment. To ensure access policy constantly throughout the connection we proposed [4] special throughput management mechanism based on randomized preemptive queueing with finite buffer size and network environment characteristics.

The paper is organized as follows: In Section II we consider main aspects of the dynamic access control enforcement in modern computer networks. In Section III we suggest architecture of security monitor. Section IV contains the comparison of our results with existing ones. Section V concludes.

## II. Dynamic Access Policy and Firewall Filtering Rules

Internet security is a main issue of modern information infrastructure. This infrastructure stores information in the form of distributed digital resources which have to be protected against unauthorized access. However, the implementations of this statement are far from simple due to the dynamic nature of network environment and users activity [5].

In modern corporate network every second hundreds of users and telematics devices establish new virtual connections with distant resources. According to the mandatory security policy if we have “N” users, “M” resources and these numbers are big, than we have to create a huge access N*M matrix. And each element of this matrix will be a vector of parameters and attributes, describing one virtual connection. Users and resources of course can be grouped according to their rights and context, but in either case such matrix is too big to be processed efficiently in real-time.

In this paper we propose the architecture of security conveyor, which consists of several filtering stages. At the first stage when a connection is established there is a classical static packet filter, which reject prior harmful traffic. The second stage enforces more accurate dynamic aspect of the access policy. Doing it the dynamic firewall have to take into account that distant network resources are not under its administration, so they can change their context any moment without informing the security conveyor. That is why we propose to use some prior information about remote users and resources in firewall to enforce security policy. Such information should be received from databases and services outside of our security monitor.

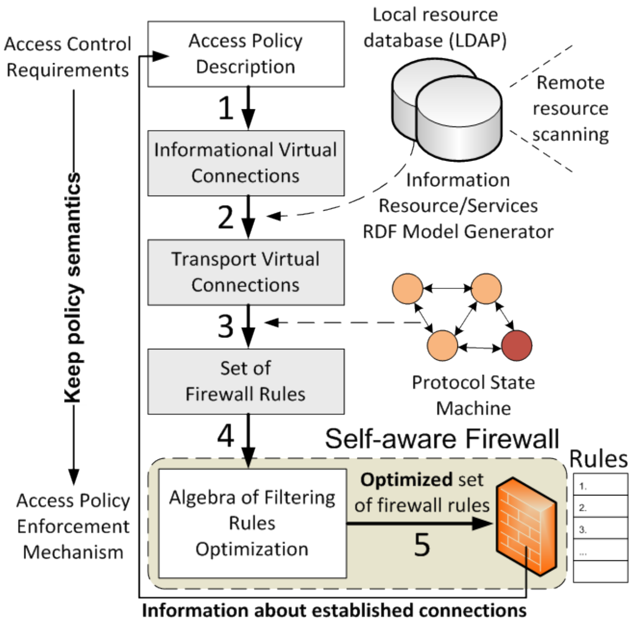

Access Control Requirements
Access Policy Description
Local resource database (LDAP)
Remote resource scanning
1
Informational Virtual Connections
2
Information Resource/Services RDF Model Generator
Transport Virtual Connections
3
Set of Firewall Rules
Protocol State Machine
Keep policy semantics
4
Self-aware Firewall
Rules
Access Policy Enforcement Mechanism
Algebra of Filtering Rules Optimization
Optimized set of firewall rules
5
1.
2.
3.
...
Information about established connections

Figure 1. Dynamic access control approach

In computer networks information is transmitted in the form of packets. Each of these packets has some part of message in it. All network devices such as routers and firewall have to operate with these packets. But people usually operate with messages and data context. That is why the information security policy is formulated towards transmitted data and application messages. Each message can be transmitted with one or several packets. According to Internet research more than 90% of data is transmitted using TCP/IP protocols. So creating dynamic security monitor we consider the specifics of existing protocols. For example, a record in security policy can we formulated like “Mr. Black shouldn’t work with Youtube”. According to Fig.1 this record can be considered as one or more informational virtual connections (1) specifying action “work” as “access”, “read”, “write” and so on. Then we have to determine what does this record mean to telematic devices, how can “Mr. Black” and “Youtube” be described. To answer these questions our security monitor has to use some prior outside information from specialized databases and services (2). Using such information we receive one or more TVCs from initial IVC. Then all these TVC’s rules should be transformed into the requirement to the packet filter. At this stage we use different transport protocols state machines to receive information about packet sequences (3). If all this procedures will be done to each established connection we’ll receive huge amount of filtering rules. That is why at the next stage we propose to optimize the set of filtering rules using specialized algebra of filtering rules (4). Only optimized set of filtering rules can be processed in real-time by firewall to enforce access policy (5). There is a detailed description of this process below.

## III. Architecture of Security Conveyor

Below we describe a new approach to configure the security network appliances that allows an administrator to overcome the semantic gap between access control requirements and the access policy enforcement mechanism of the firewall filtering rules [2]. The architecture of proposed system is presented at Fig.2.

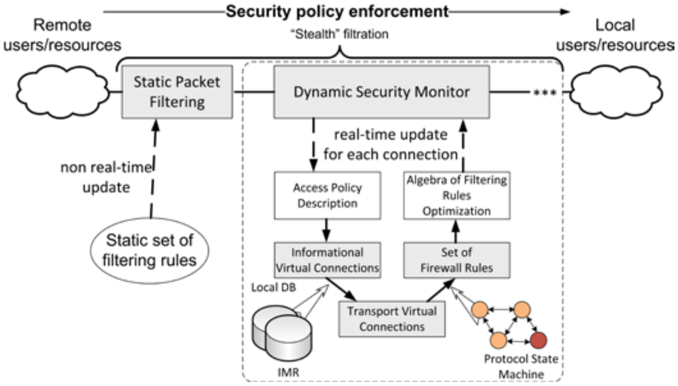

Security policy enforcement
Remote users/resources
"Stealth" filtration
Local users/resources
Static Packet Filtering
Dynamic Security Monitor
***
real-time update for each connection
non real-time update
Static set of filtering rules
Access Policy Description
Algebra of Filtering Rules Optimization
Informational Virtual Connections
Set of Firewall Rules
Local DB
Transport Virtual Connections
IMR
Protocol State Machine

Figure 2. Security conveyor architecture

Proposed security conveyor consists of two stages. The first one is a classical static packet filter, which realizes static part of security policy. For example a security administrator can set some addresses in the “black list” and all connections with such addresses have to be closed, so the static filter should drop all packets of such connections, not to spread resources of dynamic security monitor.

We proposed [6] an approach to design address invariant security appliances which basic functionality can be automatically configurable in accordance to formal specification of user access policy. Due to address invariant or "stealth" feature this security appliances have a scalable architecture and can be configured to correspond both security and performance requirements. Proposed approach is based on scalable pipeline-like architecture. Each of the filtering devices has no physical, network and other addresses so we can use two or even more of then in a row without changing network configuration. Sets of static and dynamic filtering rules could be divided into several subsets to decrease the “looking through” time. But from another point, every new filtering device will increase the delay. So creating such a multistage system we have always to solve the tradeoff between quantity of security devices and the performance of each one of them.

A dynamic security monitor forms and optimizes the set of filtering rules depending on current environment state and connection data context. The access policy enforcement by dynamic security monitor chances the current network environment state and number of connections so the traffic should be constantly watched by such security device.

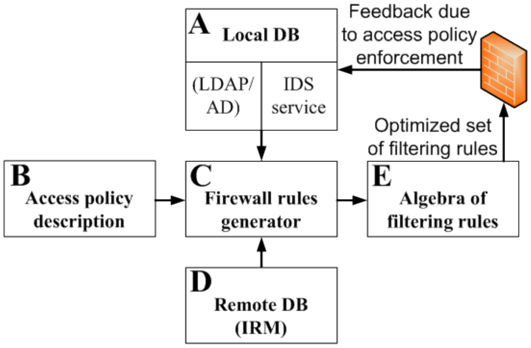

A Local DB
(LDAP/ AD)
IDS service
Feedback due to access policy enforcement
Optimized set of filtering rules
B Access policy description
C Firewall rules generator
E Algebra of filtering rules
D Remote DB (IRM)

Figure 3. Dynamic security monitor architecture

At Fig. 3 is presented the architecture of dynamic security monitor where:

## A. *Network monitor*

Network monitor controls the whole system. Network environment state consists of three main parts:

- “User activity” is the information about what computer is currently used by what user. This information can be obtained from Microsoft Active Directory (AD) by means of LDAP protocol.
- “Shared hardware resources” is the information about network infrastructure and shared internal resources that can be described by network environment state vector $X_k$
- “Network state” is the information about external network channel received from Intrusion Detection Systems (IDS).

## B. *Access policy description module*

An access policy may simply specify some restrictions, e.g., “Mr. Black shouldn’t work with Youtube” without the refinement of the nature of “Mr. Black” and “work” [7].

There is a common structure of access policy requirements, which uses the notions of subject, action and object. Thus, the informally described requirement “Mr. Black shouldn’t work with Youtube” can be formally represented as the combination of the subject “Mr. Black”, the action “read”, the object “www.youtube.com” and the decision “prohibit”. This base can also be augmented by a context, which specifies various additional requirements restricting the cases of rule’s application, e.g.: time, previous actions of the subject, attributes’ values of the subject or object, etc.

However, access rules which are based on the notions of subject, action and object are not sufficient alone to implement complex real-world policies. As a result, new approaches have been developed using the notion of a role. The role replaces a subject in access rules and it’s more invariant. Identical roles may be used in multiple information systems while subjects are specific to a particular system. Every role must be associated with some subjects as only rules with subjects can be finally enforced. During policy specification roles must be created firstly, then access rules must be specified with references to these roles, then the roles must be associated with subjects.

The OrBAC [8] model expands the traditional model of Role Based Access Control. It brings in the new notions of activity, view and abstract context. An activity is to replace an action, i.e., its meaning is analogous to the meaning of a role for a subject. A view is to replace an object. “Entertainment resources” can be an example of view, and “read” or “write” can be examples of an activity. Thus, the notions of role, activity, view and abstract context finally make up an abstract level of an access policy. OrBAC model allows to specify the access rules only on an abstract level using the abstract notions. Those are called the abstract access rules. For example, an abstract rule is “user is prohibited to read entertainment resources”, where “user” is a role, “read” is an activity, and “entertainment resources” is a view. The rules for subjects, actions and objects are called concrete access rules or informational virtual connections.

## C. *Firewall rules generator*

There is a feature common for all firewalls: they execute an access policy. In common representation the main function of access control device (ACD) is to decide whether a subject should be permitted to perform an action with an object. A common access rule “Mr. Black is prohibited to read www.youtube.com”.

As was mentioned above, “Mr. Black” is a subject, “HTTP service on www.youtube.com” is an object, and “read” is an action. So the configuration of ACD consists of common access rules that reference the subjects, actions and objects.

Although a firewall as an ACD must be configured with common access rules, each implementation uses its own specific configuration language. The language is often hardware dependent, reflecting the features of firewall’s internal architecture, and usually being represented by a set of firewall rules. Each rule has references to host addresses and other network configuration parameters. An example of the verbal description of a firewall rule may go as follows:

> Host with IP address 10.0.0.10 is prohibited to establish TCP connections on HTTP port of host with IP address 208.65.153.238.

The main complexity of this approach is to find out how such elementary firewall rules could be obtained from common access rules. Each firewall vendor reasonably aims at increasing its sales appeal while offering various tools for convenient editing of firewall rules. However, so far the problem of obtaining firewall rules from common access rules is not resolved in general.

The most obvious issue concerning this problem is that additional information beyond access rules is necessary in order to obtain the TVCs and firewall packet rules. This information concerns the expected data context, information about subject and object of the connection, configuration of network services and the parameters of network protocols that are used for data exchange. In general, such data can be received from specialized outside databases and services. For example in the simplest case we’ll have:

> Mr.Black: host with IP-address = 10.0.0.10;
>
> www.youtube.com: HTTP service (port 80) on host with IP-address = 208.65.153.238.

Thus, the TVC can be obtained by addition of the object descriptions to the IVC. It should be noted that even for small and especially for medium and large enterprises it is necessary to store and manage this prior information separately from the security policy [9].

It should also be noted that there is no need to specify any fixed rules regarding association of the network parameters with the objects. For example, HTTP port may be a parameter of an object or it may be a parameter of an action. A criterion is that the most natural representation of access policy must be achieved.

While generating the rules, the parameters of network subjects and objects can be automatically retrieved from various databases and catalogs. DNS is the best example of a world-wide catalog which stores the network addresses. Microsoft offers the network administrators the powerful means, Active Directory, to store information about users. Integration with the above mentioned technologies greatly simplifies the work of a security officer as he has only to specify the correct name of an object while forming firewall rules.

### D. *Information resource model*

Interaction between subject and object in computer network can be presented as a set of virtual connections. Virtual connections can be classified as TVC and IVC (see Fig. 4).

To implement the policy of access control, the filtering rules are decomposed in the form of TVC and IVC. These filtering rules can be configured for different levels of the data flow description based on the network packet fields at the levels of channel, transport, and application protocols.

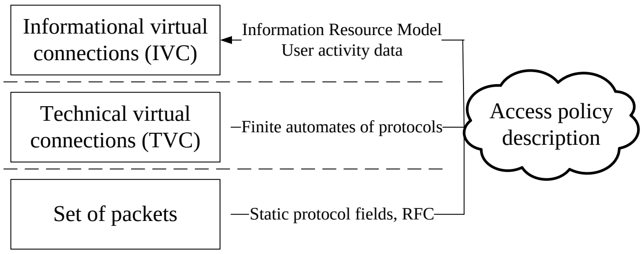

Informational virtual connections (IVC)
Information Resource Model
User activity data
Technical virtual connections (TVC)
Finite automates of protocols
Access policy description
Set of packets
Static protocol fields, RFC

Figure 4. Layers of access control policies.

At different layers of access control policy model, the filtering rules have to take into account various parameters of network environment and objects. At the packet filter layer, a firewall considers standards static protocol fields described by RFC. At the layer of TVC, firewall enforces the stateful inspection using finite machines describing states of transport layer protocols. On the upper layer of IVC firewall must consider priory information about subject and object of network interaction [2]. As was mentioned above, the information about subject can be obtained from catalog services by LDAP protocol, e.g. Microsoft Active Directory.

According to existing approach [10] a resource model can be presented in:

1) logical aspect – an N-dimensional resource space model [11];
2) representation aspect – the definition based on standard high-level description languages like XML or OWL;
3) location aspect – the physical storage model of the resource including resource address;
4) context aspect – the metadata data model based on Resource Description Framework (RDF) specifications.

All these approaches describe the network resource as a whole but don’t take into account the specific access control task. Any remote network resource can be fully classified when the connection between this resource and local user would be closed. So it is necessary to control all virtual connections in real time while monitoring traffic for security purpose.

In this paper we propose to implement a special service external to the firewall that would collect, store and renew information about remote network objects in the form of RDF context descriptions. It should automatically create information resource model, describing all informational virtual connections that have to be established to receive this resource. This service should periodically and on demand renew information about resource to keep it alive.

Firewall should cooperate with this external service to receive information resource model and enforce access policy requirements (Fig.1 (arrow 2)). Also firewall should consider TVC state models (Fig.1 (arrow 3)) that are described in different network standards and RFCs.

### *E. Algebra of filtering rules*

As was mentioned above, the information security is defined by an access policy that consists of access rules. Each of these rules has a set of attributes; the basic ones among them are identifiers of subject and object and the rights of access from one to another. In TCP/IP-based distributed systems access rules have additional attributes that help to identify flows of packets (sessions) between the client and network application server. Generally these attributes identify the network subjects and objects at different layers of TCP/IP interaction model: MAC-addresses at link layer, IP-addresses at network layer, port numbers at transport layer and some parameters of application protocols.

The access policy in large distributed informational system consists of a huge number of rules that are stored and executed in different access control appliances. The generation of the access policy for such appliances is not very difficult: information must be made available for authorized use, while sensitive data must be protected against unauthorized access. However, its implementation and correct usage is a complex process that is error-prone. Therefore the actual problem of rule generation is representation, analysis and optimization of access policy for large distributed network systems with lots of firewall filtering rules. Below we propose an approach to description, testing and verification of access policy by the means of specific algebra with carrier being the set of firewall filtering rules.

Let's define the algebra of filtering rules $\mathcal{R}$= <***R***, Σ>, where ***R*** – the set of filtering rules, Σ – the set of possible operations over the elements of ***R***. The set of filtering rules ***R*** = $\{r_j, j=\overline{1,|R|}\}$ – the carrier set of algebra $\mathcal{R}$. According to proposed approach we define a ring as algebraic structure over set of filtering rules or ***R*** [3,12]. The proposed ring consists of two operations $\Sigma=\{\varphi_1,\varphi_2\}$ over the elements of the set ***R***:

1) $\varphi_1$ – operation of addition;
2) $\varphi_2$ – operation of multiplication.

The proposed algebra is described more detailed in our paper [3, 12]. Our research shows that this algebra could significantly simplify the set of filtering rules and allow processing them in real time.

## IV. DYNAMIC ACCESS CONTROL FOR CLOUD SERVICES AND TRAFFIC MANAGEMENT

For more flexible access control and traffic management we propose to divide all TVCs into three groups:

1) permitted important connections without additional control;
2) prohibited connections;
3) other connections that are not prohibited yet but need constantly additional control throughout transmission.

The first group is the priority connections and the third one is background. All packets of virtual connections in second group are dropped by firewall and are not taken into account.

So we considered the preemptive priority queueing system with two types of packets [4]. First type has priority over the second one. The packets arrive into the buffer according to the Poisson process. The service time has the exponential distribution. The buffer has a finite size $k$ ($1 < k < \infty$) and it is shared by both types of packets. The preemptive priority in service is given to the packets of the first type. Considered system is supplied by the randomized push-out mechanism that helps precisely and accurate to manage packets of both types. If the buffer is full, a new coming packet of the first type can push out of the buffer a packet of type 2 with the probability $\alpha$ .

As it was showed in [4] by choosing $\alpha$ parameter we can change the time that packets spend in the firewall buffer, which allows to limit access possibilities of background traffic and even to block a connection if it become prohibited during the data transmission. Proposed mechanism also allows us to control TVC throughput to increase the time for the access decision without interrupting established connection.

The prototype of secure cloud environment based on proposed dynamic architecture and adopted for CAD/CAE computation tasks, is currently created at the Telematics department of the Saint-Petersburg Polytechnical University.

A distributed computing environment (cloud system) consists of the following software and hardware components:

1) virtualization nodes;
2) storage of virtual machines and user data;
3) cluster controller;
4) cloud controller.

Virtualization node is the hypervisor software which running on powerful multicore computing node. In virtualization the domain level 0 (dom0 in terms of hypervisor XEN or service console in terms of other hypervisors) and virtual computing machines (domain level U, domU) operate.

For information security and access control between virtual machines that operate under a single hypervisor, the internal ("virtual") traffic and the external traffic (incoming from other hypervisors and from public networks) have to be controlled. The solution of the access control problem could be achieved through the integration of a virtual security conveyor into the hypervisor; this dynamic system would functions under the hypervisor, but separately from the users' virtual machines. The virtual security conveyor domain can be defined as "security domain" (domS). In this task an invisible traffic filtering is an important aspect of the network monitoring because the filtering device must not change the topology of the hypervisor network subsystem. This can be achieved by using "stealth" [6] technology – a packet traffic control invisible to other network components.

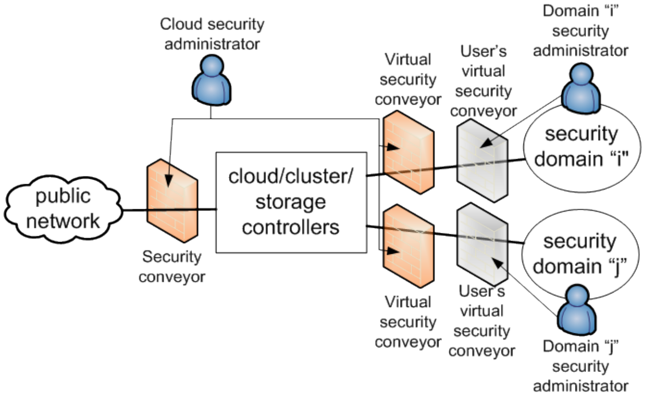

Cloud security administrator
Domain "i" security administrator
Virtual security conveyor
User's virtual security conveyor
security domain "i"
cloud/cluster/ storage controllers
public network
Security conveyor
security domain "j"
User's virtual security conveyor
Virtual security conveyor
Domain "j" security administrator

Figure 5. Secure cloud architectrure.

Fig. 5 shows the proposed architecture of a distributed cloud system with integrated security conveyor components. Some security monitors are controlled by cloud security administrator, but for each of security domains there is separate virtual security conveyor that could be controlled by domain administrator. The security domain isolates virtual machines from the hypervisor, which prevents the possibility of attack against the hypervisor inside the cloud. While the hardware security monitor isolates the private cloud components from the external threats.

## V. CONCLUSION

1. Modern computer network has lots of internal and external static and dynamic processes and influences. Each of them affects the condition and characteristics of network environment, packet streams and virtual connections.

2. Information has to be available for authorized use and protected against unauthorized access. Taking into account dynamic impacts on virtual connections we propose new dynamic access control mechanism.

3. In the paper we present the architecture of dynamic security monitor, a telematics device that receives additional external prior information about a security policy, user activities, resource context and network environment state.

4. Security policy requirements cannot be considered separately from methodology of security monitor configuration and its specified security characteristics. So we propose multistage filtering rules generation based on security policy model, informational resource model, transport protocol state machine and firewall specifics, making it possible to translate high-level abstract security requirements to low-level firewall configuration.

5. Proposed algebra of filtering rules is new mathematical description of access policy and a formal tool for firewall configuration. System approach provides possibility to prove fullness and consistency of an access policy. Proposed algebra is the base for the optimization of the set of filtering rules and of the design of dynamic firewall configuration.

It is necessary to mention that proposed solution doesn't solve all security problems of dynamic resources in computer network environment. Described above model can be merged easily with low level methods of network control and some other approaches, for example with flow-based traffic measurement. The prototype of the proposed security monitor is currently developing for a space experiment at the Telematics department of the Saint-Petersburg State Polytechnical University.

## REFERENCES

[1] William C. Barker, NIST Special Publication 800-59, Guideline for Identifying an Information System as a National Security System, 21 pages, August 2003

[2] V. Muliukha. Access Control in Computer Networks Based on Classification and Priority Queuing of the Packet Traffic, PhD. Thesis 05.13.19, SPbSPU, Russia, 2010

[3] A. Silinenko. Access Control in IP Networks Based on Virtual Connection State Models: PhD. Thesis 05.13.19: / SPbSPU, Russia, 2010

[4] V. Zaborovsky, V. Mulukha. Access Control in a Form of Active Queuing Management in Congested Network Environment // Proceedings of the Tenth International Conference on Networks, ICN 2011, pp.12-17

[5] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. 2010. A view of cloud computing. Commun. ACM 53, 4 (April 2010), pp.50-58

[6] V. S. Zaborovskii, Y. A. Shemanin and A. Rudskoy. Architecture of Distributed Network Processors: Specifics of Application in Information Security Systems // Networking - ICN 2005. Lecture Notes in Computer Science, 2005, Volume 3421/2005, pp. 681-688

[7] A. Titov, V. Zaborovsky. Firewall Configuration Based on Specifications of Access Policy and Network Environment // Proceedings of the 2010 International Conference on Security & Management. July 12-15, 2010

[8] http://orbac.org/index.php?page=orbac&lang=en

[9] V. Zaborovsky, A. Titov. Specialized Solutions for Improvement of Firewall Performance and Conformity to Security Policy // Proceedings of the 2009 International Conference on Security & Management. July 13-16, 2009. v. 2. pp. 603-608

[10] H. Zhuge, The Web Resource Space Model, Berlin, Germany: Springer-Verlag, 2007

[11] H. Zhuge, "Resource Space Grid: Model, Method and Platform," Concurrency and Computation: Practice and Experience, vol. 16, no. 14, pp. 1385-1413, 2004

[12] V. Zaborovsky, V. Mulukha, A. Silinenko, S. Kupreenko. Dynamic Firewall Configuration: Security System Architecture and Algebra of the Filtering Rules // Proceedings of The Third International Conference on Evolving Internet – INTERNET 2011, June 19-24, 2011, Luxembourg City, Luxembourg, pp. 40-45

# A statistical approach to reduce malware inside an Information System in Banking Sector

Gustavo A Valencia-Zapata, Juan C Salazar-Uribe, Ph.D.
Escuela de Estadística, Universidad Nacional de Colombia-Sede Medellín
gavalenciaz@unal.edu.co, jcsalaza@unal.edu.co

***Abstract*— The aim of this article is to illustrate the first stages of the implementation of a statistical model to dose antivirus scans in an information system in banking sector (IS). As a result, IS is strengthened by increasing malware detection and by decreasing malware attacks inside the bank. In an IS there are many components which help to build a dosage model. The IS´s components are applications such as antivirus, web filtering, Human Capital/Resource Management (HCM), and Configuration Management Database (CMDB). A CMDB provides technical information about the computer population (i.e., hard disk, operating system and so on).We can establish an analogy of some of these components with some other components from an epidemiological system which allows building the statistical model. For instance: the patients can be seen as computers into the bank network, and the malware can be seen as diseases in a population. We use this analogy to build a statistical model based on both survival analysis and data mining methods. With this modeling strategy we identify a risk profile in the IS which allows to dose the antivirus scan in a more effective way.**

**Keywords:** Malware, Survival Analysis, Missing Values, Data Mining, CART.

## I. Introduction

THIS research is intended to be a resource to improve the information security levels in banking sector (IS). The research question is: How malware incidence can be decreased in an IS? As in human epidemiologic context is necessary to apply treatments (medicine, vaccines, therapies, etc.), on a computer environment would be the application of scanning. Based on that, we can reformulate the above question as: How antivirus scans (medical tests) can be dosed, in our population (computer network), for the reduction of malware (disease) incidence in banking IS? Different approaches have been made in modeling 'disease' for Information Technology (IT) environment, using some analogies with the epidemiologic context [1]. In this sense, an exponential growth of malware has been observed in the last decades, as well as, the outlook and limitations of epidemiological concepts for malware prevention [1]. Also, malware spreading and measurements models had been elaborated [2], [3], [4]. Similarly, simulation of the networking topology influence in malware problems had been discussed by [5], [6]. Finally, epidemiological methodologies are used to estimate growth and propagation of worms in a network [7].

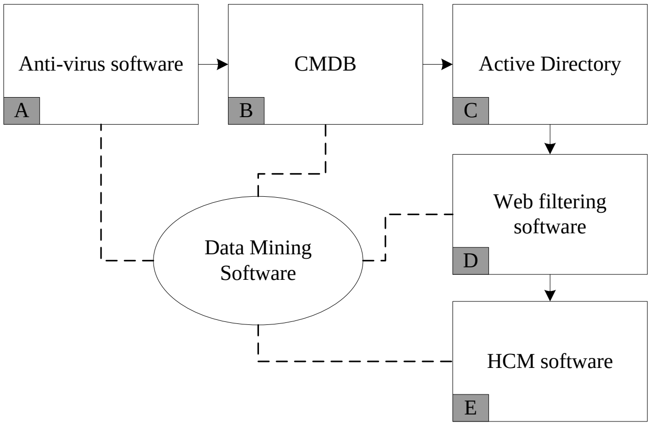


Fig. 1. Information collection stages.

TABLE I
CMDB PARAMETERS

| Variable | Meaning/value | Type | Unit |
|---|---|---|---|
| Class | Laptop, CPU or server | Nominal | NA |
| Brand | Computer brand | Nominal | NA |
| Computer_Age | Operating time | Scale | Week |
| Processor_Type | Type of computer processor | Nominal | NA |
| Processor_Clock | The speed of a computer processor | Scale | GHz |
| Processors | Number of processors | Integer | Count |
| Memory (RAM) | Memory size | Scale | GB |
| Operation_System | Operation System (OS) | Nominal | NA |
| Service_Pack | Updates to a OS | Nominal | NA |
| Hard_Disk | Hard disk size | Scale | GB |

In this paper the first stages to build the model are: information extraction (IE), handling missing values, and statistics analysis. The main information source is the bank antivirus software. Secondary information sources are: web filtering, HCM (company employees), and CMDB. Physiological computer information is provided by CMDB such as: brand, operating system, processor type, random access memory (RAM), and so on. CMDB parameters are showed in Table I. Using data mining software (through Open Database Connectivity, ODBC) we collect information about malware attacks over a period of eleven weeks. Then, secondary sources information is added as can be seen in

Figure 1. Around 18.22% of CMDB data (infected computers) are missing values. Classification and Regression Trees (CART) are used for handling missing values (imputation) to avoid losing valuable information. This research used nonparametric Statistical tests for checking the quality of the imputed data. Moreover, statistical analysis is conducted to select variables that will be included into the antivirus scanning dosage model.

## II. Information Collection Stages

### A. Antivirus Software

In this stage, amount and type of malware per computer are identified. Information over a period of eleven weeks about 8476 computers was collected. Antivirus software reports the active user account when malware was detected and some other technical information of the computer. The Table II shows the parameters reported by the Antivirus Software.

### B. CMDB

Here, technical information about computers is identified as well as its relationship with user´s account. Table III shows CMDB variables, percent completed valid and missing records. This information is used to assess the influence of these variables over detected malware levels.

TABLE II
Antivirus Software Parameters

| Variable | Meaning/value | Type | Units |
|---|---|---|---|
| Week | Time unit (1 to 11) | Integer | Count |
| Total_Malware | Number of malware per week | Integer | Count |
| Level_Malware | Level of malware per week | Nominal | NA |

TABLE III
CMDB DataQuality

| Variable | Percent Completed | Valid Records | Missing Records |
|---|---|---|---|
| Class | 83.483 | 7076 | 1400 |
| Brand | 83.483 | 7076 | 1400 |
| Computer_Age | 81.784 | 6932 | 1544 |
| Processor_Type | 99.493 | 8433 | 43 |
| Processor_Clock | 99.493 | 8433 | 43 |
| Processors | 99.493 | 8433 | 43 |
| Memory (RAM) | 99.493 | 8433 | 43 |
| Operation_System | 99.493 | 8433 | 43 |
| Service_Pack | 99.457 | 8430 | 46 |
| Hard_Disk | 99.493 | 8433 | 43 |

TABLE IV
Web Filtering Software

| Variable | Meaning/value | Type | Units |
|---|---|---|---|
| Adult | Number of adult websites | Integer | Count |
| Security | Number of security websites | Integer | Count |
| Entertainment | Number of entertainment websites | Integer | Count |
| Games | Number of game websites | Integer | Count |
| Unblocked | Number of websites unblocked | Integer | Count |
| Blocked | Number of websites blocked | Integer | Count |
| Browse_Time | Internet browse time | Continuous | Minute |

### C. Active Directory

At this point, the user's privileges are identified. For example: adding a user to the Local Administrator Group or some user accounts are allowed to use USB devices. In this way, privileges acquaintance is important to establish whether or not these variables have some influence over malware levels.

### D. Web Filtering Software

At this step, user account is related to web surfing, identifying variables such as: number and type of blocked websites, web surfing time, etc. The main purpose of studying this association is to establish if web surfing behavior has influence on malware levels. The parameters for the Web Filtering Software are shown in Table IV.

### E. HCM Software

In this stage, employee information is identified. Collected information such as position and work area is used to assess the influence of these variables over detected malware levels.

## III. CMDB Data Imputation

### A. Classification and Regression Trees, CART

CART model is explained in detail in [8]. The classification and regression trees (CART) method was suggested by Breiman et al. [8]. According to Breiman, the decision trees produced by CART are strictly binary, containing exactly two branches for each decision node. CART recursively partitions the records with similar values for the target attribute. The CART algorithm grows by conducting for each decision node, an exhaustive search of all available variables and all possible splitting values, selecting the optimal split according to the following criteria [9].

Let $\varphi(s|t)$ be a measure of the "goodness" of a candidate split $s$ at node $t$, where

$$\varphi(s|t) = 2P_L P_R \sum_{j=1}^{\#\,classes} |P(j|t_L) - P(j|t_R)| \qquad (1)$$

Split parameters are defined in Table V. One of the major contributions of CART was to include a fully automated and effective mechanism for handling missing values [10]. Decision trees require a missing value-handling mechanism at three levels: (a) during splitter evaluation, (b) when moving the training data through a node, and (c) when moving test data through a node for final class assignment [11].

TABLE V
SPLIT PARAMETERS

| Parameter | Meaning |
|---|---|
| $t_L$ | Left child node of node $t$ |
| $t_R$ | Right child node of node $t$ |
| $P_L$ | $\frac{\text{Number of records at } t_L}{\text{Number of records in training set}}$ |
| $P_R$ | $\frac{\text{Number of records at } t_R}{\text{Number of records in training set}}$ |
| $P(j\|t_L)$ | $\frac{\text{Number of class j records at } t_L}{\text{Number of records at } t}$ |
| $P(j\|t_R)$ | $\frac{\text{Nnumber of class j records at } t_R}{\text{Number of records at } t}$ |

TABLA VI
CONTINGENCY TABLE CLASS

| | | *E_Class* | | Total |
|---|---|---|---|---|
| | | CPU | Laptop | |
| *Class* | CPU | 5.013 | 20 | 5.033 |
| | Laptop | 14 | 2002 | 2.016 |
| Total | | 5027 | 2022 | 7049 |

TABLA VII
CHI-SQUARE TEST

| | Value | Exact Sig. (Two-sided) |
|---|---|---|
| McNemar Test | 1.058 | 0.392 |
| Nº Valid Cases | 7049 | |

Use binomial distribution

According to [11], regarding (a), the later versions of CART (the one we use) offers a family of penalties that reduce the improvement measure to reflect the degree of missingness. (For example, if a variable is missing in 20% of the records in a node then its improvement score for that node might be reduced by 20%, or alternatively by half of 20%, and so on.) For (b) and (c), the CART mechanism discovers "surrogate" or substitute splitters for every node of the tree, whether missing values occur in the training data or not. The surrogates are thus available, should a tree trained on complete data be applied to new data that includes missing values.

### *B. Handling Missing Values through Classification and Regression Trees*

Ten variables were imputed (Table III), that is, ten CART were used, a CART for each variable, which together make up a classification and regression forest. The imputation was made using PASW® Modeler (a data mining software). Model training was made with complete data and then, this trained model was applied to missing values. Table VI and Table VII show a nonparametric statistics test called *McNemar Test for Significance of Changes* [12], which evaluated model prediction by using complete data. In this case *E_Class* is the imputed value and *Class* is the real value. For instance, 5013 (99.6%) computers with *Class* equal to CPU were classified correctly by CART, and 2002 (99.3%) computers with *Class* equal to Laptop were classified correctly by the same CART.

The formulated hypotheses for McNemar test (2-sided) were:

$Null\ hypothesis, H_0 =$ Class is not changed after imputation.

$Alternative\ hypothesis, H_1 =$ Class was changed after imputation.

According to this analysis we cannot reject the null hypothesis, that is, CART doesn't change *Class* values after imputation (p-value=0.396). Figure 2 Shows a CART model for *Class* variable that was built using the software PASW® Modeler.

In particular, Table VIII shows the variables for the computer number 0022. We can identify three out of ten variables with missing values. Node 0 indicates that CPU category has the higher probability (0.7) to be selected if a random imputation is conducted. On the other hand, Laptop variable has a smaller probability (0.28) than the first one and the Server variable has null probability (0.0). As can be seen, we should slide through CART's branches according to the values of the variables shown in Table VIII.

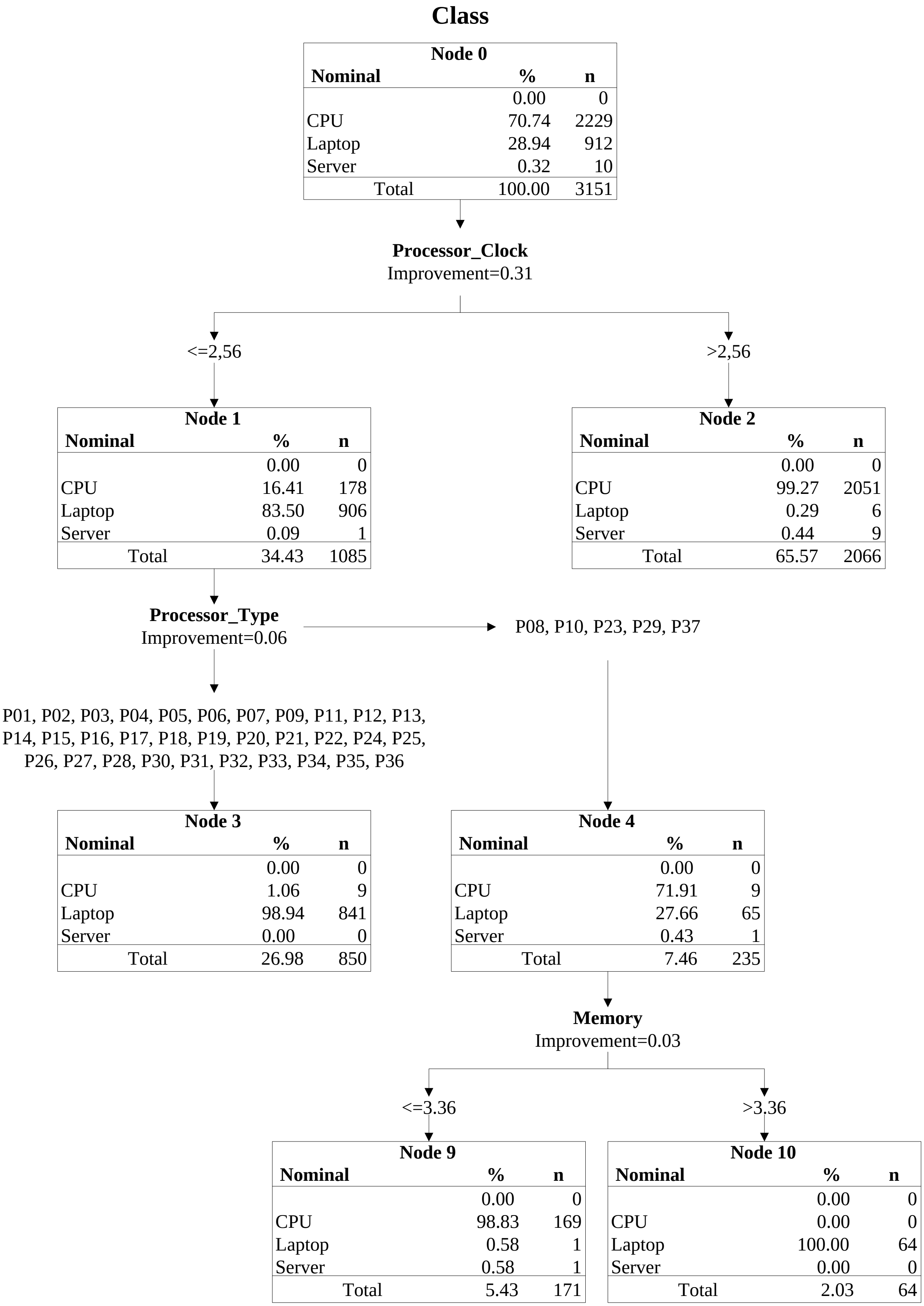


Fig. 2. CART Model for variable *Class*.

As *Processor_Clock* is less or equal than 2.56 GHz, then, Node 1 branch is selected. As *Processsor_Type* is P27, Node 3 branch is then selected. The final node (Node 3) indicates that CPU category has lower probability (0.01) and Laptop category has higher probability (0.98) than CPU. As a consequence, Class variable will be imputed as being Laptop (this imputation process is done with 99% confidence). In contrast, in random sampling imputation the result for the Class would be CPU as this category has the greater frequency.

## IV. Antivirus Scanning Dosage Statistics Model

Following the imputation process we implement statistical analysis to assess the influence of these variables over malware levels inside an IS. In this study we believe that malware level depends on variables such as: *Processors* (number of processor in the computer), *Computer_Age* and *Browse_Time*. To illustrate Table IX and Table X show the results of performing a nonparametric statistics test called Kruskal-Wallis [13], which evaluates the dependence between *Malware_Level* and *Computer_Age*. The hypotheses for this test were:

$Null\ hypothesis, H_0 =$ *Computer_Age* is not associated with *Malware_Level.*

$Alternative\ hypothesis, H_1 =$ Some levels of *Computer_Age* are associated with *Malware_Level.*

According to this test, we reject the null hypothesis (p-value=0.000), implying that *Computer_Age* has substantial influence over *Malware_Levels.*

TABLE VIII
COMPUTER 0022 – CMDB PARAMETERS

| Variable | Meaning/value | Units |
|---|---|---|
| **Class** | **Missing** | **NA** |
| Brand | Missing | NA |
| Computer_Age | Missing | Week |
| Processor_Tipe | P27 | NA |
| Processor_Clock | 2.19 | GHz |
| Processors | 2 | Count |
| Memory (RAM) | 2.14 | GB |
| Operation_System | SO_7 | NA |
| Service_Pack | SP_3 | NA |
| Hard_Disk | 80.02 | GB |

TABLE IX
KRUSKAL-WALLIS TEST

| | **Rank** | | |
|---|---|---|---|
| | Computer_Age | N | Mean Rank |
| Malware_Level | 1 | 265 | 743,36 |
| | 2 | 224 | 559.9 |
| | 3 | 134 | 576,66 |
| | 4 | 203 | 617,10 |
| | 5 | 313 | 634,68 |
| | 6 | 31 | 650,73 |
| | ≥7 | 100 | 633,06 |
| | Total | 1270 | |

TABLE X
TEST STATISTICS

| | Malware_Level |
|---|---|
| Chi-square | 37 |
| gl | 6 |
| Asymp.Sig. | 0.000 |

Grouping Variable: Computer_Age

Table XI and Table XII show a nonparametric statistic test called *Chi-Squared Test for Independence* [13], which assess the dependence between *Malware_Level* and *USB.* The formulated hypotheses for this test were:

$Null\ hypothesis, H_0 =$Variable *Malware_Level* and *USB* are independent.

$Alternative\ hypothesis, H_1 =$Variable *Malware_Level* and *USB* are not independent.

According to this test, we do not reject the null hypothesis (p-value=0.9090), implying that that *Malware_Level* and USB are independent. However, as a recommendation to improve security, disabling USB ports is an effective strategy for preventing information leakage.

TABLA XI

CONTINGENCY TABLE MALWARE_LEVEL VS USB

| | | Malware_Level | | | Total |
|---|---|---|---|---|---|
| | | High | Medium | Low | |
| USB | Yes | 170 | 834 | 210 | 1214 |
| | NO | 7 | 40 | 9 | 56 |
| Total | | 177 | 874 | 219 | 1270 |

TABLE XII

CHI-SQUARE TEST

| | Value | gl | Asymp.Sig, 2-sided |
|---|---|---|---|
| Pearson Chi-square | 0.191 | 2 | 0.909 |
| LRT | 0.195 | 2 | 0.907 |
| N | 1270 | | |

The Kaplan-Meier is used for estimating the survival function from life-time data [13]. To use this strategy we define the following outcome: Elapsed time to first malware infection in a computer. The goal is to compare the survival experiences of four processor types. Figure 3 shows the Kaplan-Meier curves associated to each the four Processor types. Survival curves show, for each time plotted on the X axis, the portion of all computers surviving at that time.

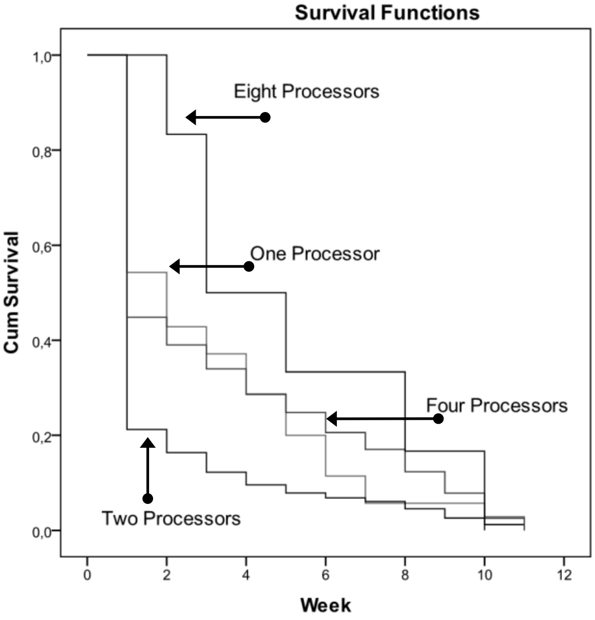


Fig. 3.Kaplan-Meier Survival curve for *Processors.*

The Table XIII shows the Log-Rank Test to compare the survival distribution of two groups (we have four groups so we have to conduct six comparisons). The null hypothesis being tested is that there is no overall difference between the two survival curves. Computers with two processors showed statistical differences when they were compared with computers that had one, four, and eight processors, respectively (p-values=0.018, 0.00, 0.05). That means the computers with two processors were infected faster than other computers with different number of processors.

TABLE XIII

PAIRWISE COMPARISONS

| | Processors | 1 Processor | | 2 Processors | | 4 Processors | | 8 Processors | |
|---|---|---|---|---|---|---|---|---|---|
| | | Chi-Square | Sig. | Chi-Square | Sig. | Chi-Square | Sig. | Chi-Square | Sig. |
| Log Rank | 1 Processor | | | 5.58 | **0.018** | 0.303 | 0.582 | 1.334 | 0.248 |
| (Mantel-Cox) | 2 Processors | 5.58 | **0.018** | | | 258.19 | **0.000** | 3.847 | **0.05** |
| | 4 Processors | 0.303 | 0.582 | 258.19 | **0.000** | | | 0.730 | 0.393 |
| | 8 Processors | 1334 | 0.248 | 3.847 | **0.05** | 0.73 | 0.393 | | |

## V. Conclusion and future work

By using both standard statistical methodology and data mining we were able to explore the association between different variables and the malware levels in an IS. It is important to highlight that we used real data to perform all the analysis. These results may affect the malware scanning policy in a bank.

On the other hand, missing values appear frequently in the real world, especially in business-related databases, such as in an IS inside a banking sector, and the need to deal with them is a vexing challenge for all statisticians and data mining modelers. One of the major contributions of CART was to include a fully automated and effective mechanism for handling missing values. CART Models, as were presented here, are more suitable for handling missing values than imputation through random sampling.

Currently, data mining efficiently integrate large amounts of data stored in repositories and allows us to discover meaningful new correlations, patterns and trends by using pattern recognition. This is a competitive research advantage for business companies. The statistical and mathematical techniques are essential to data mining for built and check models, it means, statistical tests give more confidence on the results from data mining models. Thus in our case, we evaluated the quality of CART model for handling missing values with different nonparametric statistical tests. The observed results favored this statistical strategy.

Future directions of this work include performing additional statistics analysis such as recurrence analysis and formulation of survival models through Cox-Models. This also will allow identifying significant variables to optimize the malware scanning policy in an IS as well as measure its effect size.

## Acknowledgment

The authors wish to thank Juan Carlos Correa from School of Statistics of the Universidad Nacional de Colombia at Medellín for helpful feedback that contributed to improve this manuscript. Also the authors thank the Security Team of the Bank Company for their continuous encouragement and support.

## References

[1] Weiguo J, "Applying Epidemiology in Computer Virus Prevention: Prospects and Limitations", 2010. Thesis, Computer Science, University of Auckland.

[2] Bailey, N.J.T, "The Mathematical Theory of Infectious Diseases and Its Applications" 1975, New York: Oxford University Press.

[3] Kephart J, and White S, "Directed-Graph Epidemiological Models of Computer Viruses", *IEEE Computer Symposium on Research in Security and Privacy, Proceedings*, pp. 343–359, May 1991.

[4] Kephart J, and White S, "Measuring and Modeling Computer Virus Prevalence, Research in Security and Privacy", 1993, Proceedings, 1993 *IEEE Computer Society Symposium* on, pp. 2–15, May 1993.

[5] Kephart, J, "How Topology Affects Population Dynamics" in *Langton, C.G. (ed.) Artificial Life III*. Reading, MA: Addison-Wesley, 1994.

[6] Pastor-Satorras, R. and Vespignani, A, "Epidemic Dynamics and Endemic States in Complex Networks". Barcelona, Spain: Universitat Politecnica de Catalunya, 2001.

[7] Rishikesh P, "Using Plant Epidemiological Methods To Track Computer Network Worms", 2004. Thesis, Computer Science, Virginia Polytechnic and State University.

[8] Daniel T. Larose, "Discovering Knowledge in Data. An introduction to data mining" 2005. John Wiley & Sons, Inc

[9] Leo Breiman, Jerome Friedman, Richard Olshen, and Charles Stone, "Classification and Regression Trees", 1984. Chapman & Hall/CRC Press.

[10] Vipin Kumar, "The Top Ten Algorithms in Data Mining", 2009. Chapman & Hall/Crc.

[11] Quinlan, R, "Unknown attribute values in induction". *In Proceedings of the Sixth International Workshop on Machine Learning*, 1989 pp. 164–168.

[12] Conover, "Practical Nonparametric Statistics", 1999. John Wiley & Sons, Inc

[13] Hosmer Jr, D.W. and Lemeshow. "Applied Survival Analysis: Regression Modeling of Time to Event Data", 1999. John Wiley Sons,

# SESSION

# SPECIAL TRACK ON MISSION ASSURANCE AND CRITICAL INFRASTRUCTURE PROTECTION

## Chair(s)

### Dr. Michael Grimaila

# Secure Management of Certificates for Industrial Control Systems

**Sebastian Obermeier, Ragnar Schierholz, Hadeli Hadeli,**
**Robert R. Enderlein, Ana Hristova, and Thomas Locher**
[first name].[last name]@ch.abb.com
ABB Corporate Research, Industrial Software Systems, Switzerland

**Abstract**—*In order to secure the communication of industrial automation and control systems (IACS), recent cyber security standards demand the use of certificates, but do not answer the problem of certificate management. In the IACS domain, the use cases and the corresponding implementation differ from traditional IT systems.*

*This paper points out the use cases and challenges that an efficient certificate handling technique has to address, and proposes a comprehensive certificate management approach based on existing cryptographic solutions. The paper covers all phases of an embedded device's life cycle up to the operation of the devices. Finally, the paper shows how the proposed solution is able to withstand several attack vectors.*

**Keywords:** security; industrial control systems; certificate management

## 1. Introduction

Industrial applications are often controlled and supervised by industrial automation and control systems (IACS),[1] which are highly distributed systems used to control dispersed assets, often scattered over a large geographical area. The role of an IACS is to acquire data of an industry process and allow an operator to operate and issue control commands to the associated assets if needed.

In the past, IACS were isolated systems and field devices were connected to the control system via dedicated lines. This traditional approach relied much on the special purpose components and isolation of systems to achieve security. Thus, traditional IACS technology itself had little to no protection against attacks.

Currently, a lot of open standards and technologies are in use to replace old technologies in the IACS due to maintainability issues (e.g., product discontinuations) or connectivity issues (e.g, a demand to connect Enterprise resource planning systems to the IACS).

As nowadays technologies that follow open standards are used in industrial control systems, they inherit all the vulnerabilities of those technologies as well. Consequently, this implies new challenges for the security of those systems and the associated communications. Recently, new security standards (for instance OPC UA) emphasize the need for information security, but yet they do not solve all the issues. OPC UA, and also other standards, demand X.509 certificates without specifying how the certificates are to be managed.

Securing an IACS environment using standard security protocols is challenging mainly because an industrial control system has different security requirements than enterprise information systems. In particular, in contrast to enterprise information systems, IACS follow a different prioritization regarding the relevance of security objectives, cf. [1], [2]. For instance, availability, authenticity, and integrity, are of paramount importance for IACS, while the priority of confidentiality is usually lower.

In office environments, *certificates* are used for achieving authenticity. Certificates bind a public key to identity information, and provide a convenient way to achieve mutual authentication. When certificates should be used for embedded device configuration, however, the devices' life cycle imposes several problems regarding certificate management. For example, a solution to manage the initial authentication between the device and the issuer of the certificate—which entails installing the root certificate, as well as issuing the client certificate—is required.

In addition, the requirement for constant availability, in combination with the fact that embedded devices in general have little computational power, makes complex certificate validations difficult. As in many scenarios, a breach of security can lead to a severe safety breakdown, the security properties "integrity" (making sure commands arrive as intended), "strong authentication and authorization" (making sure only those entities send commands that are entitled to do so), as well as "auditability" (being able to audit in case of an issue or to audit whether a given configuration corresponds to the policy), have the highest priority that any certificate management solution must obey.

Our **contributions** are the following. In this paper, we

- identify the life cycle of a typical embedded IACS device,
- point out use cases for embedded device certificate management including initial deployment, system integration, and certificate operation,
- specify the critical attack vectors, and
- propose and discuss solutions that are able to withstand these attacks for each use case.

[1] These systems are also known as SCADA systems.

# 2. Related Work

In a public key infrastructure (PKI) scheme, the signatures on a certificate are attestations by the certificate signer that the identity information and the public key belong together. [3] discusses a classification of various certificates.

In the embedded device scenario, however, there is no single CA involved but multiple CAs, including manufacturer and operator CAs as in the lifecycle of an industrial embedded device, different organizations provide the function of the CA. However, generally none of the organizations is available throughout the entire lifecycle. Thus, certificate management for industrial embedded devices has to consider the specific lifecycle requirements of the devices.

A restriction of the time span during which a certificate and the associated private key can be used is important for the following reasons, cf. [4]). A limited certificate lifetime

1) limits the amount of information protected by a given key that is available for cryptanalysis,
2) limits the amount of exposure if a single key is compromised,
3) limits the use of a particular algorithm to its estimated effective lifetime,
4) limits the time available for attempts to penetrate physical, procedural, and logical access mechanisms that protect a key from unauthorized disclosure.
5) limits the period within which information may be compromised by inadvertent disclosure of keying material to unauthorized entities, and
6) limits the time available for computationally intensive cryptanalytic attacks (in applications where long-term key protection is not required).

# 3. Model

## 3.1 Roles

Since the various roles in an IACS context are different from the roles in an office IT environment, we proceed by quickly describing all roles that are relevant for the management of certificates.

**Manufacturers** of the devices physically assemble the devices and install firmware and/or software on them.

**System Integrators** customize devices, integrate them into an entire system (potentially devices from multiple manufacturers), and perform commissioning. This may be the manufacturer, the asset owner, or an external company.

**Operators** monitor the system during their normal operation and respond to alarms. Typically done by the asset owner or an external company.

**Certificate Authorities** are chosen by the system integrator potentially based on operator's requirements. A certificate authority manages certificates and handles revocation during the lifetime of the plant.

**Service Units** are responsible for maintaining and repairing devices. The role can be performed by the manufacturer, the asset owner, or an external company.

## 3.2 Life Cycle of an Embedded Device

The life cycle with respect to certificate management of embedded industrial devices, e.g., industrial embedded controllers, is illustrated in Figure 1. After the device has been manufactured, it is integrated into the system within the commissioning phase. Within this phase, all devices forming the system are engineered, customized, and tested according to the operational requirements of the asset owner. Usually two tests are conducted: the "Factory Acceptance Test" (FAT) and the "Site Acceptance Test" (SAT). The goal of the FAT is to ensure that the system engineered for the customer itself works, while the SAT actually ensures that this system works in its intended environment. The asset owner usually chooses the manufacturer, a third party, or even himself to be responsible for this phase. After this phase, the operational phase begins. Again, the asset owner chooses the responsible party for the operation and maintenance of the plant. Service units repair or replace devices if the necessity occurs during operation. Finally, the devices are decommissioned at the end of their lifetime.

During the device's life cycle indicated in Figure 1, five requirements for certificate management can be identified:

**Installation of the manufacturer's default root certificate:** The manufacturer installs an initial certificate onto the device. The goal is to establish a trust relationship between the devices and the manufacturer's root Certificate Authority (CA) by installing the CA's root certificate on the device inside of a trusted device production environment. Since the ultimate destination of the devices is not known at this point, the device certificates are temporary and will have to be replaced before operation.

**Installation of system integrator root certificate:** As devices will probably operate in a multi-vendor environment, the operator will likely wish to set up (or ask a third party, e.g., the system integrator, to set up) a Certificate Authority distinct from the manufacturer CA probably even specific to a plant or at least the operator's organization. The system integrator's new root certificate will have to be installed on the devices in a secure manner in order to establish a trust relationship between the system integrator and the devices. The crux of this phase is actually to establish a trust relationship between the manufacturer's and the system integrator's respective CAs. This step can be performed any time before the Factory Acceptance Test (FAT).

**Operation-time certificate installation:** As the default certificate that the device has received at production time was not customized for operation-time and was issued by the manufacturer's CA, the new CA will issue a new "operation-time" certificate to each device. This step is functionally very

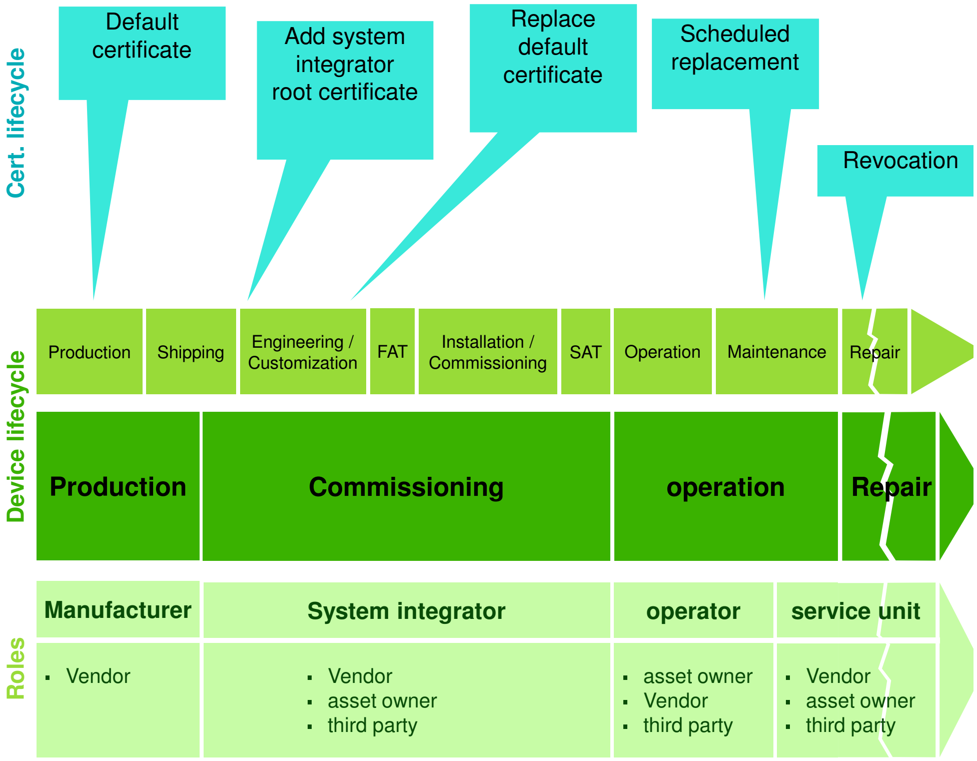


Fig. 1: Embedded Device Lifecycle

similar to the next phase, but is performed after the root certificate has been installed and before the SAT.

**Renewal of operation-time certificate:** The operation-time certificate of each device will have to be replaced periodically (because all certificates eventually expire). The operator's CA will have to keep track of which certificates need replacement and then issue the new certificates in a secure manner.

**Revocation:** A properly set-up revocation framework allows the CA to revoke a certificate that needs to be prematurely invalidated.

Within this article, necessary requirements up to a plant's operation are discussed. Certificate renewal and certificate revocation, however, are omitted due to size limitations.

Table 1 shows the mapping between a device's lifecycle phases and the phases defined in RFC4210 [5]. While RFC4210 mandates out-of-band authentication for some of these phase, the exact nature of these out-of-band channels is considered out of scope and not discussed in this paper.

Figure 2 illustrates an overview of the certificate management roles and responsibilities. In total, four CAs are assumed: the manufacturer root CA, the (device) factory CA, the system integrator root CA, and the plant CA.

## 3.3 Attack Vectors

As mentioned before, certificates protect the integrity of commands and the authenticity of all devices and involved parties. In this paper, we make the common assumption that it is not feasible to break the authentication scheme itself, i.e., we assume that the standard cryptographic algorithms are secure. Therefore, it is only possible for an attacker to launch a successful attack by somehow getting its own key onto the device. During the life cycle of a typical device in an IACS, there are several stages where an attacker may try to tamper with the device. The various possibilities for an attacker to corrupt the device are summarized in the following attack vectors:

Table 1: Mapping between RFC 4210 and Embedded Device Life Cycle

| RFC 4210 | Mapping to Device Life Cycle |
|---|---|
| CA establishment | Set up of manufacturer CA |
| End-entity initialization | Installation of manufacturer root certificate |
| Initial registration/certification | Installation of device default certificate |
| Cross-certificate request | Installation of system integrator root certificate (loosely) |
| Certificate update (+ key pair update) | Replacement of the default / operation-time certificate, certificate publication |
| Revocation request, CRL publication | Revocation |

1) During the production process, an attacker installs its own root key on the device. The goal is to later send commands signed by the attacker to the device in order

to instruct the device to perform malicious actions.

2) During the device's shipment, an attacker installs its own root key.
3) During operation, an attacker manages to install its own root key.
4) During operation, an attacker resets the device, installs its own root key, and restores the initial configuration including the plant's root key.

Thus, the goal is to ensure that the certificate management scheme mitigates exactly these security risks, assuming that the manufacturer and system integrator are not acting maliciously. Our approach is discussed in the subsequent section.

# 4. Default Certificate Installation

## 4.1 Problem Description

At this point in time, the final IP address, or other identifiers (DNS name, IEC 61850 Logical Device Name, or any other application specific naming scheme), of a device is not known. The solution must ensure that a pre-installed certificate can be securely replaced later on.

An intuitive approach would be to ship a device without any certificate. This, however, leads to the obvious problem that an attacker could secretly install a root certificate onto a device and ship it to the system integrator. Assuming the system integrator does not notice the certificate, the attacker can always access the device later on. Therefore, the chain of trust must be established right after the production process of each device.

## 4.2 Solution

The proposed solution requires the manufacturer to generate a private key for each device and create a matching default certificate (using the manufacturer CA to sign it) during device production. The private key, default device certificate, and root certificate of the manufacturer are then installed on each device while in a trusted environment. Devices freshly out of the factory need this default certificate to be able to authenticate themselves. At this stage, the device can only hold one root certificate. The problem of storing the key on a tamper proof (or at least tamper-evident), secure storage is an additional problem that is not discussed in this paper.

At this point, the final device identifiers (IP address or other distinguished names) are not known, thus the default certificates cannot be used for plant operation. However, a unique hardware identifier, such as the serial number or the MAC address of the network interface, can be used as an identifier for the devices. The devices are configured in such way that all certificates issued by the manufacturer have no privileges other than validating an "add root certificate" command and authenticating a key replacement (this authorization is not part of the certificate).

Since the contents of the subject DN (distinguished name) in X.509 certificates are not well-defined in the various standards, Table 2 shows a way to fill in the various parts of the subject in the device certificate.

Table 2: Proposed contents for X.509 certificate RelativeDN record

| RelativeDN Type | RelativeDistinguishedName |
|---|---|
| C (Country) | The country the device was manufactured in. |
| O (Organization) | The name of the manufacturer. |
| OU (Organizational Unit) | The name of the factory the device was produced in. This field could also be used for the "security domain", for instance by appending a number to the factory name. |
| CN (Common Name) | MAC address or serial number of the device; must be unique for all certificates issued by that manufacturer / that CA. |

The validity period of these default certificates must be long enough to accommodate the time span between production and the first certificate replacement. While a long key cannot be used for the certificates used in time-critical operations because the devices may have limited computational capabilities, a long key (e.g., RSA key $\geq$ 2048 bits) may be used for the default certificates, which are not used for time-critical operations. Moreover, a long key ensures that the certificate will not expire until the device is deployed.

This procedure mitigates the first attack vector as described in Section 3.3: A device that leaves the manufacturing process can only contain a single certificate root. If an attacker has installed its own root key during the production process, it is not possible to install an additional root key signed by the factory CA. In such case, the attacker would have to sign the command to install additional root keys. However, the out-of-band telephone call using SAS ensures the authenticity of the manufacturer. Thus, a malicious root key can be identified during the device's commissioning.

# 5. Installation of New Root Certificates

## 5.1 Problem Description

Once the system integrator has set up his CA infrastructure[2], the plant CA's root certificate must be installed on the devices in a secure manner. Unfortunately, since at that point there is no trust relationship between the plant CA and the rest of the world, an out-of-band method is required.

[2] The same mechanisms can be used to add an additional trusted root CA without the cooperation of the system integrator's CA (provided the manufacturer root CA was not deleted), for instance because the CA chosen by the system integrator has to be replaced.

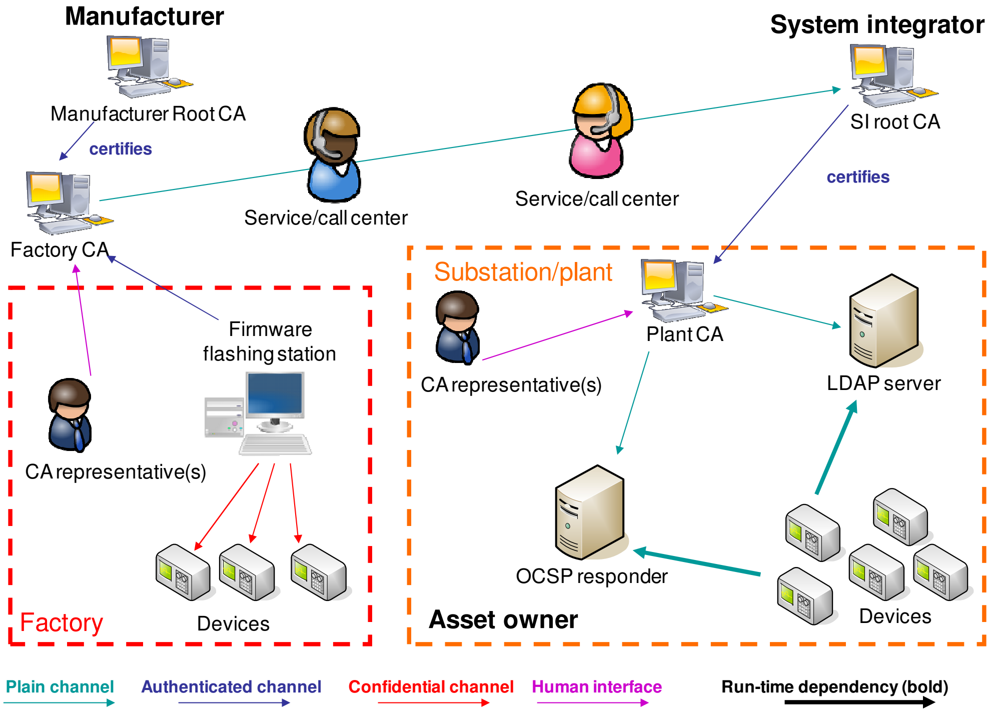


Fig. 2: Certificate Management Overview

## 5.2 Solution

A straightforward solution is to have cross-certified CAs between the manufacturer and the system integrator to extend the trust between both CAs [6]. The system integrator would be fully trusted and could immediately setup the devices. However, we cannot assume that such cross certified CAs exist for all system integrators and manufacturers. Thus, we present two options that allow the dynamic establishment of a trust relationship between manufacturer, system integrator, and plant.

The first option is for the manufacturer to "introduce" the system integrator's CA to them. For this method, an authenticated channel from the system integrator to the manufacturer is required. The second option is for the manufacturer to issue and sign a certificate for the system integrator. The latter can now install his root CA's certificate on the devices by using this certificate. This method requires a confidential channel[3] from the manufacturer to the system integrator in order to enable a secret transmission of the corresponding private key. There are several possibilities for an out-of-band authentication:

**Face-to-face meeting:** A representative of the system integrator hands over the root certificate to a representative of the manufacturer. If the CA is already up, the best possible time for this transfer would be during contract signing.

**Paper mail:** The system integrator sends printout of the cryptographic hash ($\geq$ 160 bits) of the root certificate on paper. The actual root certificate can now be sent over any channel (for instance on a USB key send with the letter, or by a separate e-mail).

**Telephone (hash):** The system integrator calls the manufacturer and transmits a cryptographic hash ($\geq$160 bits) of the root certificate over the authenticated voice channel. The actual root certificate can now be sent over any channel (for instance by a separate e-mail).

**Telephone:** The system integrator calls the manufacturer, and both parties perform a Short Authenticated Strings (SAS) protocol [7]. The SAS protocol minimizes the amount of authenticated data that has to be transmitted over the voice channel (20 bits $\approx$ 6 digits that must be transmitted). The non-authenticated part of the SAS protocol will take place on a plain data channel (e.g., TCP). To make the hash transfer user-friendly and less error prone, an agreed-upon list of words can be used instead of digits, for example the PGP word list [8], [9], which was developed for transmitting cryptographic hashes over the phone.

The manufacturer then signs an "Add root certificate" command (containing the operator's CA root certificate, authorized for the relevant security domain/list of devices), and transmits it to the system integrator, which can then relay this message to the devices. The devices then consider the operator's CA root certificate to be a trusted root. Finally, the system integrator can issue new certificates to the devices,

[3] If the system integrator generates the private key for this kind of certificate, the same situation applies as discussed the preceding paragraph, just with one level of indirection more (and more points of vulnerability).

containing all the required parameters and permissions. The system integrator may also delete the manufacturer's root certificate from the trusted store.

### 5.2.1 System Integrator Based Root Certificate Installation

An alternative is the transmission of a user certificate that is signed by the manufacturer to reconfigure the list of trusted certificates on the devices. The private key that goes along with that certificate must be transmitted to the system integrator in a confidential manner. The key must remain confidential until the root certificate of the manufacturer has been removed from the device, otherwise an adversary who finds the key can take control over the devices, for instance, by changing the root certificate, and demanding a ransom to reveal the corresponding private key.

Possible ways to establish a confidential channel:

**Face-to-face meeting:** Direct handover of the private key corresponding to the user certificate.

**Paper mail:** Since mail can be intercepted and read by a third party, it must be ensured that someone who intercepts the letter cannot access the private key. One way to proceed would be the following: first, the manufacturer sends a passphrase (at least 128 bits of entropy is recommended) in a tamper-evident envelope (as banks use to transmit the PIN code for credit cards). If that letter is lost or has been tampered with, the manufacturer can retry by sending a new letter with a different passphrase until the system integrator received the unopened letter (to be fully secure, an authenticated return channel for the system integrator is required to transmit an acknowledgement back to the manufacturer). Otherwise, if an adversary intercepts the letter, he could impersonate the system integrator telling the manufacturer that the passphrase was received properly), the manufacturer can send him the private key encrypted with the passphrase, along with the user certificate over any channel (e.g., by e-mail). All of this can happen in parallel to the device shipment, therefore latency is not a critical issue.

Having established a confidential channel, the system integrator can now add its own root certificate to the device and issue new certificates to the devices, containing all the required parameters and permissions. The system integrator should also remove the manufacturer's root certificate from the devices as otherwise anyone who finds out the private key corresponding to the user certificate could compromise the device. The devices must not allow the removal of the last root certificate with "root certificate management" privileges.

### 5.2.2 Discussion

The proposed solutions are analyzed regarding the following criteria:

| | |
|---|---|
| *Strengths* | Attributes that enhance the security of the approach. |
| *Weaknesses* | Weaknesses that may harm the security of the approach. |
| *Threats* | External conditions that may harm the security of the approach. |

**Cross-certified CAs**

| | |
|---|---|
| *Strengths* | No further interaction between manufacturer and system integrator is required. |
| *Weaknesses* | Substantial overhead beforehand. |
| *Threats* | n/a |

**Manufacturer Based – Face-to-face Meeting**

| | |
|---|---|
| *Strengths* | Authentication and authorization trivial. |
| *Weaknesses* | Requires the CA of the system integrator to be set up before contract signing. |
| *Threats* | The contract signers might not be familiar with security issues. |

**Manufacturer Based – Paper Mail**

| | |
|---|---|
| *Strengths* | Low Cost. |
| *Weaknesses* | High latency. Difficult to authenticate sender and to determine whether the letter is delivered to the authorized recipient. |
| *Threats* | Recipient might not bother checking the cryptographic hash. Liability of manufacturer if the letter was a forgery. |

**Manufacturer Based – Telephone (Hash and SAS)**

| | |
|---|---|
| *Strengths* | Low Cost. |
| *Weaknesses* | Difficult to determine if caller is authorized by the asset owner. Inconvenience of transmitting relatively complex data over the phone. This process can be facilitated by using PGP wordlists [8], [9]. |
| *Threats* | Recipient might not bother checking the cryptographic hash. SAS is a relatively new cryptographic protocol. |

**System Integrator Based – Face-to-face**

| | |
|---|---|
| *Strengths* | Authentication/authorization is trivial. |
| *Weaknesses* | If the private key of the user certificate is compromised, devices can be rendered unusable. |
| *Threats* | The contract signers might not be familiar with security issues. |

**System Integrator Based – Paper Mail**

| | |
|---|---|
| *Strengths* | Low Cost. |
| *Weaknesses* | If the private key of the user certificate is compromised, devices can be rendered unusable. "Return channel" depends on mail delivery company. |
| *Threats* | Mail delivery service must be trusted for giving the letter to the right person. |

Cross-certified CAs between manufacturer and system integrator renders the problem at hand trivial, but requires additional overhead beforehand. It can be considered the ideal choice if both manufacturer and system integrator already have a CA, i.e., the effort is already spent, or both plan to interact more frequently in the future.

If this is not the case, the use of the SAS mechanism in combination with PGP wordlists can be considered the most promising approach for a practically feasible out-of-band authentication. This allows the system integrator to call the manufacturer and transmit only a few words to the operator.

Establishing trust between manufacturers and system integrators and ensuring the verification of the certificates as described above mitigates the remaining attack vectors:
**2.** An attacker cannot install an arbitrary root key after the device has left the manufacturing process as the root key can only be installed along with a "install root key" command signed by the manufacturer's factory CA. In order to install its own root key, an attacker would have to call the manufacturer's call center and pretend to be the regular system integrator or operator. This, however, is noticed by the manufacturer during the commissioning phase when the regular customer calls the manufacturer's call center to install his root key as well. The second call will raise an alarm and instruct the customer to verify installed root keys carefully.
**3.** An attacker cannot install its own root key during operation: each new root key must be signed by the plant CA.
**4.** In order to reset a device, an attacker would have to get physical access. In addition, the failure of the device would immediately raise an alarm within the system software and instruct operators to inspect the device. In order to install the malicious root key, an attacker would have to perform the telephone authentication and successfully receive a signed device command and restore the initial configuration. As logs disappear from the device due to the factory reset, the system software detects the factory reset and issues a high priority alarm when the device is re-integrated into the system. If such an attack should be mitigated, a complete removal of the manufacturer's root key from the device can prevent the manufacturer from adding new root certificates to the device.

## 6. Summary and Conclusion

Certificate management for embedded devices faces many challenges, of which, based on the device's life cycle, the use cases "default certificate installation", "system integration", and "certificate replacement" have been discussed. Different alternative approaches have been proposed that meet the requirements for each use case with respect to industrial automation control systems. A discussion of the feasibility of different solution approaches has been presented and a solution has been recommended. Finally, an evaluation of the proposed solution has shown that it withstands critical attack vectors.

In the future, the authors plan to discuss the problem of certificate replacement and revocation to support certificates throughout the whole embedded device lifecycle.

## References

[1] D. Dzung, M. Naedele, T. von Hoff, and M. Crevatin, "Security for industrial communication systems," *Proceedings of the IEEE*, vol. 93, no. 6, pp. 1152–1177, 2005.

[2] M. Naedele, "Addressing IT security for critical control systems," in *HICSS*. IEEE Computer Society, 2007, p. 115.

[3] J. Lopez, R. Oppliger, and G. Pernul, "Classifying public key certificates," in *EuroPKI*, ser. Lecture Notes in Computer Science, D. W. Chadwick and G. Zhao, Eds., vol. 3545. Springer, 2005, pp. 135–143.

[4] E. Barker, W. Barker, W. Burr, W. Polk, and M. Smid, "NIST SP800-57: Recommendation for Key Management – Part 1: General(Revised)," Tech. Rep., March 2007.

[5] C. Adams, S. Farrell, T. Kause, and T. Mononen, "Internet X.509 Public Key Infrastructure Certificate Management Protocol (CMP)," RFC 4210 (Proposed Standard), Internet Engineering Task Force, Sept. 2005. [Online]. Available: http://www.ietf.org/rfc/rfc4210.txt

[6] V. Casola, A. Mazzeo, N. Mazzocca, and M. Rak, "An innovative policy-based cross certification methodology for public key infrastructures," in *EuroPKI*, ser. Lecture Notes in Computer Science, D. W. Chadwick and G. Zhao, Eds., vol. 3545. Springer, 2005, pp. 100–117.

[7] S. Vaudenay, "Secure communications over insecure channels based on short authenticated strings," in *Advances in Cryptology - CRYPTO 2005: 25th Annual International Cryptology Conference, Santa Barbara, California, USA, August 14-18, 2005, Proceedings*, vol. 3621. Springer, November 2005, pp. 309–326.

[8] P. Juola and P. Zimmermann, "Whole-word phonetic distances and the pgpfone alphabet," in *Proceedings of the 4th International Conference on Spoken Language Processing, Philadelphia, PA, USA*, 1996, pp. 98–101.

[9] P. Juola, "Isolated word confusion metrics and the pgpfone alphabet," in *Proceedings of the Second International Conference on New Methods in Language Processing , Ankara, Turkey*, 1996.

[10] D. W. Chadwick and G. Zhao, Eds., *Public Key Infrastructure, Second European PKI Workshop: Research and Applications, EuroPKI 2005, Canterbury, UK, June 30 - July 1, 2005, Revised Selected Papers*, ser. Lecture Notes in Computer Science, vol. 3545. Springer, 2005.

# Information Security Analysis Using Game Theory and Simulation

**Bob G. Schlicher and Robert K. Abercrombie**
Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA

**Abstract** - *Information security analysis can be performed using game theory implemented in dynamic simulations of Agent-Based Models (ABMs). Such simulations can be verified with the results from game theory analysis and further used to explore larger scale, real world scenarios involving multiple attackers, defenders, and information assets. Our approach addresses imperfect information and scalability that allows us to also address previous limitations of current stochastic game models. Such models only consider perfect information assuming that the defender is always able to detect attacks; assuming that the state transition probabilities are fixed before the game assuming that the players' actions are always synchronous; and that most models are not scalable with the size and complexity of systems under consideration. Our use of ABMs yields results of selected experiments that demonstrate our proposed approach and provides a quantitative measure for realistic information systems and their related security scenarios.*

**Keywords:** Information Security Analysis, Game Theory, Simulation, Confidentiality, Integrity, and Availability

## 1 Introduction

Title 44 of the U.S. Code [1] defines Information security as a means of protecting information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction in order to provide:

- confidentiality, which means preserving authorized restrictions on access and disclosure, including means for protecting personal privacy and proprietary information;
- integrity, which means guarding against improper in-formation modification or destruction, and includes ensuring information nonrepudiation and authenticity; and
- availability, which means ensuring timely and reliable access to and use of information

Today's security, economic, and industrial systems depend irrevocably on the security of myriad devices and the networks that connect them and that operate in ever-changing threat environments. Adversaries are applying increasingly sophisticated methods to exploit flaws in software, telecommunication protocols, and operating systems; to infiltrate, exploit, and sabotage weapon systems, command, control, and communications capabilities, economic infrastructure, and vulnerable control systems; or exfiltrate sensitive data, and to obtain control of networked systems in order to prepare for and execute attacks. Information security continues to evolve in response to disruptive changes with a persistent focus on information-centric controls. A healthy debate is needed to address balancing endpoint and network protection, with a goal of improved enterprise / business risk management.

Traditional network security solutions, typically employing firewalls and intrusion detection devices do not have a quantitative decision framework [2]. To this end, a few groups of researchers have started advocating the utilization of game theoretic approaches [2]. Game Theory provides mathematical tools and models for investigating multi-player strategic decision making. Another technique that is promising is the application of simulations [3].

### 1.1 The Problem

The motivation for this work, is highlighted by existing and emerging technologies that complement The Roadmap for Cybersecurity Research in context of survivability of time-critical systems [4] and the President's Comprehensive National Cybersecurity Initiative [5] with respect to extending cyber security into critical infrastructure domains.

The research and practicing community have been paying close attention to cyber security problems for more than two decades. However, Shiva et al. [6] state and it is generally agreed that the problem is far from being solved. In fact, some would argue that it is getting worse. As our dependence on the cyber infrastructure grows more complex and more distributed, the systems that compose it become more prone to failures and exploitation [7]. Failures in complex, tightly coupled systems can only be mitigated by collective decision making and organizational learning [8]. This is one way to view this game-theoretical approach.

The submitted manuscript has been authored by a contractor of the U.S. Government under contract DE-AC05-00OR22725. Accordingly, the U.S Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

## 1.2 Paper Organization

In this paper, we first define information security and briefly review the weakness of traditional security solutions as they do not have a quantitative decision framework. We address the motivation for this work and in Section 2 introduce the game theory in the context of information security. In that section, we will also identify the distinguishing features of our approach to the subject domain. We then document four scenarios that will be the basis for the development of the Agent-Based Model (ABM) and its testing. An overview of the alignment of computational models and the challenges with comparing models is presented. This concept is important since when comparing models, one would want to know which features or capabilities are superior to other models. To address this subject, we pattern our ABM after two works which utilize a similar base, but approach game theory from different perspectives. We set our model up according to their assumptions and produce some interesting results. In the experimental section, we address the probability of successful attacks, and the tenants of information security dealing with confidentiality, integrity and availability. We conclude with a discussion on ideas for future work.

# 2 Related Work: Game Theory In Information Security

Roy et al. provide an excellent review of the taxonomy and different approaches to game theory as it can be applied to network security [2]. Recent work analyzes information security in the subject domain of e-commerce based on game theory with the introduction of the penalty parameter of the defender and the attacker [9]. This approach encourages the defender to invest in information security. Another recent paper defines and uses an analytical framework to analyze strategic choices and identify the best strategies and corresponding defenses used in virtual coordinate systems [10].

## 2.1 Limitations of Present Research

Many of the current game-theoretic security approaches are based on either static game models (e.g. Bayesian Formulation [2]) or games with perfect information or games with complete information. However, in reality a network administrator often faces a dynamic game with incomplete and imperfect information against the attacker. Some of the current models involving dynamic game with incomplete and imperfect information are specific to mobile ad hoc networks [11] while others do not consider a realistic attack scenario [2].

In particular, Roy et al. [2] point out that some of the limitations of the present research are: (a) Current stochastic game models [12] only consider perfect information and assume that the defender is always able to detect attacks; (b) Current stochastic game models [13] assume that the state transition probabilities are fixed before the game starts and these probabilities can be computed from the domain knowledge and past statistics; (c) Current game models assume that the players' actions are synchronous, which is not always realistic; (d) Most models are not scalable with the size and complexity of the system under consideration [2].

## 2.2 Distinguishing Features of Our Approach

### 2.2.1 Hypothesis and Explanation

Information security analysis can be performed using evolutionary game theory implemented in dynamic simulations of ABMs. Such simulations can be verified with game theory analysis results and further used to explore large-scale, real world scenarios involving multiple attackers, defenders, and information assets.

### 2.2.2 Simulation Approach

The simulation is based on ABM where the active components of the model, referred to as agents, engage in interactions on scenario-by-scenario basis.

### 2.2.3 Agent-Based Model (ABM) Overview

ABMs have been used to simulate evolutionary game theory involving multiple players in both cooperative and competitive or adversarial postures [14, 15].

ABMs bring significant benefits when: (1) interactions be-tween the agents are complex, nonlinear, discontinuous or discrete; (2) space is crucial and the agents' positions are not fixed; (3) the population is heterogeneous; (4) the topology of the interactions are heterogeneous and complex; or (5) the agents exhibit complex behavior, including learning and adaptation [14, 15].

The agents in the simulation include the attacker and the defender or administrator. The agents perform actions that can change the system state of the enterprise. For each state, agents are limited in the actions they can perform. Depending on the scenario, the attacker executes one of many actions with an associated probability of deciding to do the action and a probability that the action will be successful once the decision has been committed. Within each time unit, the simulator thread visits each agent giving them the opportunity to perform an action or not.

The administrator performs actions that begin with the probability of detecting something wrong with their enterprise. Since the enterprise state is known, the simulation limits the administrator's actions, which for the most part is a possible counter action to the most current action performed by the attacker. This is a reasonable assumption in that a competent administrator is assumed to be able to recognize a problem with their system. Before the administrator performs any counter action, a detection action is required to confirm the type of attack. In the simulation, our time unit represents one minute. We executed 1,000 simulations with each simulation spanning 250 simulated minutes. Experimental results were aggregated into bins and averaged to arrive at the probabilities of attack success.

Several scenarios are considered with a description of one of many sequence s that can be realized in the simulation as depicted in Table 1 for the scenario involving httpd being hacked by an attacker and recovered by the administrator. This scenario is used to gain an understanding for the agent interactions and the probabilities associated with decision points. There are many branches that the attacker can decide. Our ABM is flexible to accommodate arbitrary topologies and enterprise states. For familiarity, we have chosen the network topology for our analysis to those similar to Lye and Wing [12] and expanded upon in [13] and Wang et al. [16] as shown in Fig. 1.

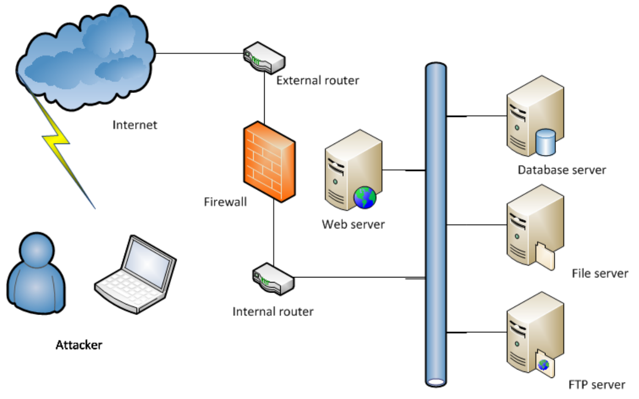


**Figure 1. Enterprise network topology showing attach paths and enterprise servers.**

In Table 1, we use P (a) to indicate the probability of taking the action and P(s) to indicate the probability of success of the action. P(s) is also the trigger for the state changes within the system. For example, at each time unit and for an attacker with the opportunity to continue_attacking, the attacker in Table 1 has a 0.5 uniform probability of deciding to perform that action and if so, has a 0.5 probability of succeeding. When the administrator performs a successful detection, in this case the detection of httpd being hacked, the payoff of -1 indicates that the recovery will not be considered until the next time unit. Hence the payoff of a negative value is interpreted as a delay in the number of time units. At that next time frame, the httpd is corrected with a 1.0 probability of action and a 1.0 probability of success. The payoff of -20 indicates a 20-minute duration to perform this action.

### 2.2.4 Overview of Challenges with Comparing Models

Computational models differ widely in their assumptions and implementations. These models must be normalized so their respective results can be analyzed to determine which model is more general and advances the field in question. This alignment is needed to determine whether two models can produce the same results, which in turn is the basis for critical experiments and for tests of whether one model can subsume another. This "alignment of computational models" has been referred to as “replicating” [17] or “docking” [18].

Table 1. Scenario 001 with Simulation Parameters

| **Scenario 001. httpd is hacked and recovered** | **Simulation parameters and notes** |
|---|---|
| 1. The attacker attacks an httpd process. | Attack_http, P(a)=0.5, P(s)=1.0 |
| 2. The attacker continues the attack to compromise the httpd. | continue_attacking, P(a)=0.5, P(s)=0.5 |
| 3. The attacker compromises the httpd system, httpd has been hacked. | State change to Httpd_hacked. |
| 4. The admin detects the hacked httpd. | detect_httpd_hacked, P(a)=0.5, P(s)=0.5, payoff = -1. |
| 5. The admin removes the compromised account and restarts httpd. | remove_compromised_account_restart_httpd, P(a)=1.0, P(s)=1.0, payoff= -20. |

Table 2. Scenario 002 Deface Website with Admin Correction

| **Scenario 002. Deface Web Site** |
|---|
| 1. The httpd is hacked, but not recovered (see Scenario 001). |
| 2. The attacker defaces the web site. |
| 3. The admin detects the defaced web site. |
| 4. The admin restores the website and removes the compromised account. |

Table 3. Scenario 003 Denial Of Service (DOS)

| **Scenario 003. Denial of Service (DOS)** |
|---|
| 1. The httpd is hacked, but not recovered (see Scenario 001). |
| 2. The attacker installs a sniffer and a backdoor program. |
| 3. The attacker runs a DOS virus on the web server. |
| 4. The network traffic load increases and degrades the system. |
| 5. The admin detects the traffic volume and identifies a DOS attack. |
| 6. The admin removes the DOS virus and the compromised account. |

Table 4. Scenario 004 File Server Data Stolen

| **Scenario 004. File Server Data Stolen** |
|---|
| 1. The httpd is hacked, but not recovered (see Scenario 001). |
| 2. The attacker installs a sniffer and a backdoor program. |
| 3. The attacker attempts to crack the file server root password. |
| 4. The attacker cracks the password; the file server is hacked. |
| 5. The attacker downloads data from file server. |
| 6. The admin detects the file server hack. |
| 7. The admin removes the file server from the network. |

Model equivalence testing is of central importance when comparing two or more computational models. The “equivalence” of models with stochastic elements must be defined in a precise statistical context. In many cases this is not trivial. There are at least two categories of equivalence beyond the initial criterion of numerical identity, (1) distributional and (2) relational equivalence. Distributional equivalence describes models that produce distributions that are statistically equivalent. Relational equivalence describes two models that produce the same internal relationship among their results [18, 19].

## 2.3 Attack Models

For certain defense techniques, some of the best attack strategies involve an inflation attack with varying percentages of malicious nodes [10].

## 2.4 Defense Models

From a defense posture, spatiotemporal and spatial outlier detections produce the best results against attacks. Temporal outlier detection in isolation is ineffective [10].

## 2.5 Experimental Plan to Test Hypothesis

We have focused our model on previous works that have documented several attack scenarios. The chosen case study was developed by interviewing network managers [12, 13]. Our enterprise network topology illustrated in Fig. 1 is quite similar to the previous papers [13, 16] and serves as our exploratory basis.

### 2.5.1 States

Our enterprise network is typical of many current configurations. Our model utilizes the following states from Lye and Wing [13] as follows:

1. normal_operation
2. httpd_attacked
3. httpd_hacked
    a. detect. hacked_detected
4. ftpd_attacked
5. ftpd_hacked
6. website_defaced
    a. detect. website_defaced_detected
7. webserver_sniffer
8. webserver_sniffer_detector
9. webserver_dos_1
    a. detect. webserver_dos_1_detected
10. webserver_dos_2
11. fileserver_hacked
    a. detect. fileserver_hacked_detected
12. fileserver_data_stolen_1
13. workstation_hacked
    a. detect. workstation_hacked_detected
14. workstation_data_stolen_1
15. network_shut_down

### 2.5.2 Actions

An action pair (one from the attacker and one from the administrator) causes the system to move from one state to another in a probabilistic manner. A single action of the attacker can be any part of his attack strategies, such as flooding a server with SYN packets or downloading a password file.

When a player takes no action, we denote the inaction as ø. Attacker consists of all the actions he can take in all the states. The actions can be described as:

- *Attack_httpd*
- *Attack_ftpd*
- *Continue_attacking*
- *Deface_website_leave*
- *Install_sniffer*
- *Run_DOS_virus*
- *Crack_file_server_root_password*
- *Crack_workstation_root_password*
- *Capture_data*
- *Shutdown_Network*

The action candidates in each state are taken from this list. For example, in the state Normal operation, the attacker has actions Attack_httpd, Attack_ftpd and ø.

In [13, 16] for similar actions taken by the administrator are mainly preventive or restorative measures. Our model uses the nomenclature provided in [13]. The actions of the administrator can be described in the following:

- *Remove_compromised_account_restart_httpd*
- *Restore_Website_remove_compromised_account*
- *Remove_virus_and_compromised_account*
- *Install_sniffer_detector*
- *Remove_sniffer_detector*
- *Remove_compromised_account_restart_ftpd*
- *Remove_compromised_account_sniffer*

The explanations of the above actions are similar to that of Lye and Wing [13], and Wang et al. [16]. Both papers assume that the administrator does not know whether there is an attacker or not, as we do. We also assume, as in [13, 16], that the attacker may have several objectives and strategies that the administrator does not know. Another realistic aspect of this model is assignment of probabilities of attack and success. Furthermore, not all of the attacker’s actions can be observed.

### 2.5.3 Parameter Modeling Set

Following the logic of our of our model of our typical enterprise network, Table 5 identifies the parameter modeling set that guided the data collection and analysis section for the attacker. Table 6 identifies the parameter modeling set that guided the data collection and analysis section for the defender administrator.

Table 5. Attacker Parameter Modeling Set

| Action Name | Prob. Action | Prob. Success | Payoff | State From | State To |
|---|---|---|---|---|---|
| Attack_httpd | 0.5 | 0.5 | 10 | 1 | 2 |
| Continue_attacking | 0.5 | 0.5 | 0 | 2 | 3 |
| Deface_website_leave | 0.5 | 0.5 | 99 | 3 | 6 |
| Install_sniffer | 0.5 | 0.5 | 10 | 3 | 7 |
| Run_dos_virus | 0.5 | 0.5 | 30 | 7 | 9 |
| Crack_file_server-root-pw | 0.5 | 0.5 | 50 | 7 | 11 |
| Capture_data_file_-server | 0.5 | 0.5 | 999 | 11 | 12 |
| Shutdown_network | 0.5 | 0.5 | 999 | 9 | 15 |

Table 6. Defender Administrator Parameter Set

| Action Name | Prob. Action | Prob. Success | Payoff | State From | State To |
|---|---|---|---|---|---|
| Detect_httpd_hacked | 0.5 | 0.5 | 1 | 3 | 3a |
| Detect_defaced_website | 0.5 | 0.5 | -1 | 6 | 6a |
| Detect_webserver_sniffer | 0.5 | 0.5 | -1 | 7 | 8 |
| Remove_sniffer | 1.0 | 1.0 | 0 | 8 | 1 |
| Remove_compromised_-account_restart_httpd | 1.0 | 1.0 | 10 | 3a | 1 |
| Restore_website_remove_-compromised_account | 1.0 | 1.0 | -10 | 6a | 1 |
| Detect_dos-virus | 0.5 | 0.5 | -1 | 9 | 9a |
| Remove_virus-and_-compromised_account | 1.0 | 1.0 | -3.0 | 9a | 1 |
| Detect_fileserver_hacked | 0.5 | 0.5 | -1 | 11 | 11a |
| Detect_fileserver_hacked | 0.5 | 1.0 | -1 | 11 | 11a |
| Remove_compromised_-account_restore_fileserver | 1.0 | 1.0 | -20 | 11a | 1 |

# 3 Experimental Results

In this section we simulate the security of the enterprise network via the above model. We initially address what constitutes a successful attack and then address the confidentiality, integrity and availability of the enterprise network.

## 3.1 Security Analysis – Probability of a Successful Attack

The probability of a successful attack is determined by the parameter modeling set defined in Table 6. Fig. 2 illustrates the successful attacks in the enterprise network at each time interval (minutes), which is not cumulative.

Fig. 3 shows the same data as a cumulative distribution indicating when the probability of successful attacks reaches 1 for the arrival rates (0.13, 0.37, 0.65 and 0.94) respectively. This particular model illustrates that the attacker has a distinct advantage as the arrival rates of the attack increases.

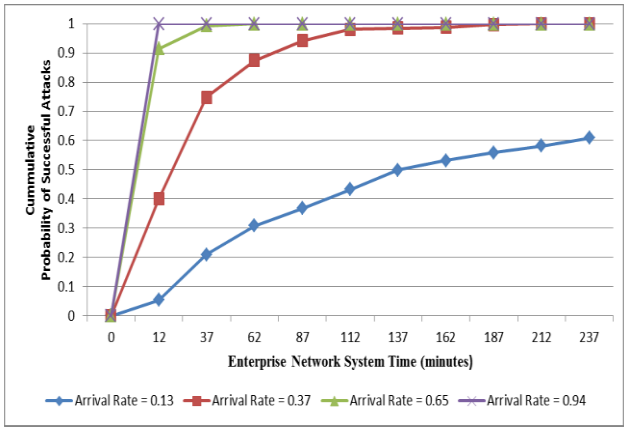


**Figure 2. The probability of successful attacks cumulatively in the enterprise network per system time**

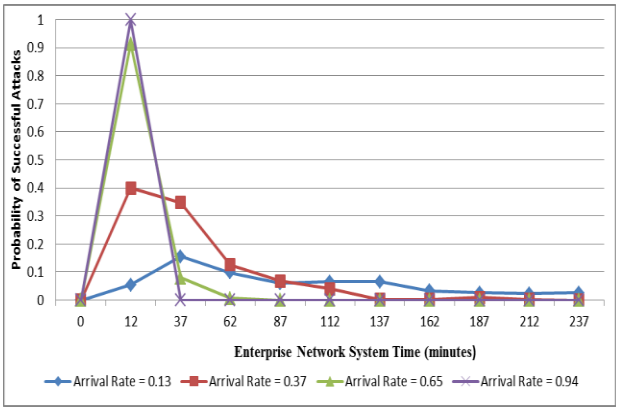


**Figure 3. The probability of successful attacks in the enterprise network per system time.**

## 3.2 Confidentiality

Wang et al. [16] define confidentiality as the absence of unauthorized disclosure of information. A measure of confidentiality is the probability that important data and information are not stolen or tampered. We adapt this logic to our model and the confidentiality can be shown as:

$$C = 1 - (P_{Fileserver\_data\_stolen} \times P_{Worstation\_data\_stolen}) \qquad (1)$$

where $P_{Fileserver\_data\_stolen}$ and $P_{Workstation\_data\_stolen}$ are the probability that the attacker succeeded in the "data stolen" category. Fig. 4 illustrates the confidentiality variation over time of the $P_{Worstation\_data\_stolen}$ to illustrate similar data trending. When compared to data in [16], it is clearly evident that our approach lends itself to the alignment of disparate models quite well.

## 3.3 Integrity

Wang et al. [16] define integrity as the absence of improper system alterations, preventing improper or unauthorized change. It is further described as the probability that the normal network services are affected or destroyed.

Our models follows Lye and Wing [13]. Integrity can be shown as:

$$I = 1\text{-}(P_{Website\_defaced} \times P_{Webserver\_DOS}) \tag{2}$$

where $P_{Website\_defaced}$ and $P_{Webserver\_DOS}$ denote the probability in our model that the attacker succeeded in defacing the website, or inserting a virus and/or shutting down the network via the actions *Website_defaced* and *Webserver_DOS*. Fig. 5 illustrates the integrity dynamics of $P_{Website\_defaced}$ over time. Again the arrival rate of attacks has a profound effect on the dynamics of the probability of the particular website being defaced.

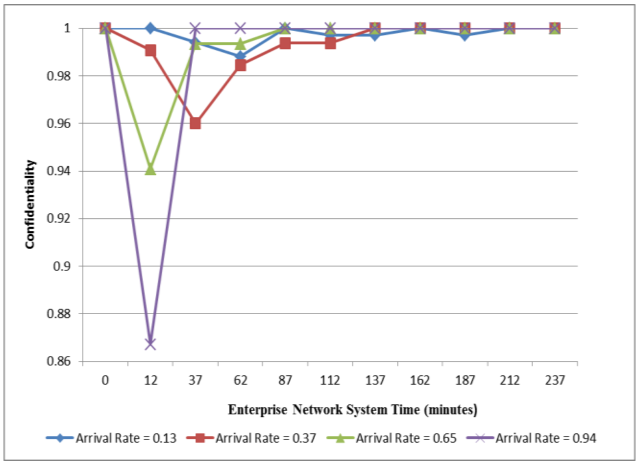


**Figure 4. Confidentiality dynamics of $P_{Worstation\_data\_stolen}$ in the enterprise network.**

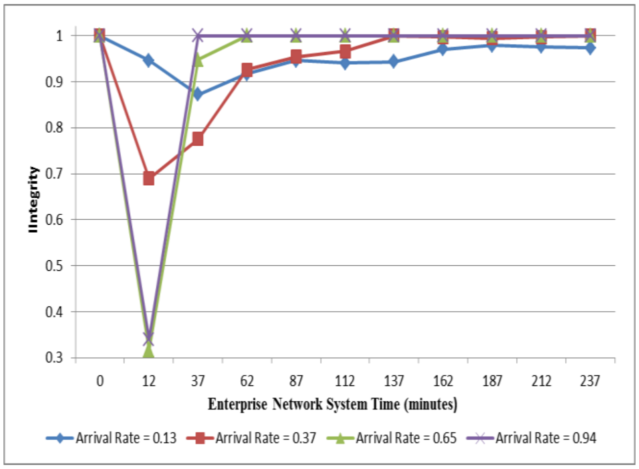


**Figure 5. Integrity dynamics of $P_{Website\_defaced}$ in the enterprise network.**

## 3.4 Availability

Wang et al. [16] define availability as systems being available when needed, and computing resources can be accessed by authorized users at any time. It is further described as whether the authorized users can access the information when necessary, when considering the probability that the normal network services are affected or destroyed.

Our model differs from Wang et al. [16] with availability expressed as:

$$A = 1\text{-}(P_{Webserver\_DOS} \times P_{Network\_shut\_down}). \tag{3}$$

Here $P_{Webserver\_DOS}$ denotes the probability the attacker succeeded in the defacing the website, or inserting a virus and/or $P_{Network\_shut\_down}$ denotes shutting down the network via the actions *Webserver_DOS* and *Network_shut_down*. Fig. 6 illustrates the availability variation.

Comparing and contrasting Figs. 4-6, we find confidentiality, integrity, and availability decrease at the beginning of the attack and then increase over time, as the administrator recovers from the attack. Therefore, it is crucial to the safety of the system that the administrator can discover the attack as early as possible.

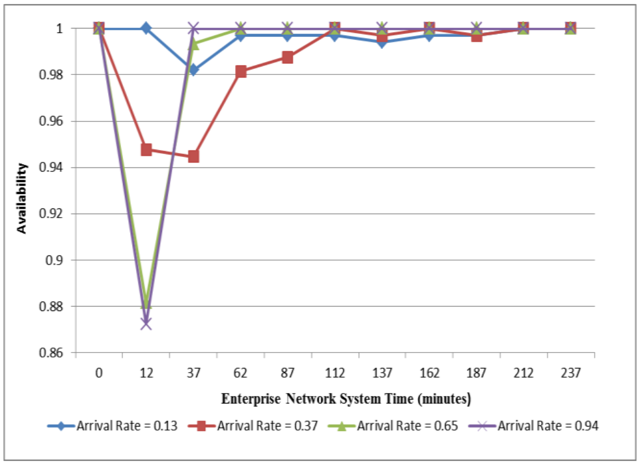


**Figure 6. Availability dynamics of $P_{Webserver\_DOS}$ in the enterprise network.**

# 4 Conclusion and Future Work

The main motivation for this work was that many of the current game-theoretic security approaches are based on either static game models (e.g. Bayesian Formulation [2]) or games with perfect information or games with complete information. In reality a network defender (the administrator) often faces a dynamic game with incomplete and imperfect information against the attacker. Some of the current models involving dynamic game with incomplete and imperfect information and others do not consider a realistic attack scenario.

In particular, Roy et al. [2] point out that some of the limitations of the present research are: (a) Current stochastic game models [12] only consider perfect information and assume that the defender is always able to detect attacks; (b) Current stochastic game models [12, 13] assume that the state transition probabilities are fixed before the game starts and these probabilities can be computed from the domain knowledge and past statistics. This second premise is the basis for our selection of the ABM approach. This allows us to expand our data collection beyond fixed probability values. Additionally, they state that current game models assume that the players' actions are synchronous, which is not always realistic and most models are not scalable with the size and complexity of the system under consideration [2]. As we embarked on the development of our ABM approach, great care was taken to duplicate exacting the underlining assumptions (transition probabilities) in the works by Lye and Wing [12, 13] and Wang et al. [16].

Our model is a simulation based on Agent Based Model (ABM) where the active components of the model, the agents, engage in interactions on scenario-by-scenario basis. The agents in the simulation include the attacker and the defender (or administrator). This is in contrast to the previous techniques that utilized nonlinear program in *MATLAB* [13] and Petri nets [16]. The agents perform actions that can change the system state of the enterprise. For each state, agents are limited in the actions they can perform. Depending on the scenario, the attacker executes one of many actions with an associated probability of deciding to do the action and a probability that the action will be successful once the decision has been committed. Within each time unit, the simulator thread visits each agent giving them the opportunity to perform an action or not. In our particular simulation, each time unit represented one minute. We executed 1,000 simulations with each simulation spanning 250 simulated minutes. Experimental results were aggregated into bins and averaged to arrive at the probabilities of attack success.

We believe that model equivalence testing (normalization) for comparing models is of central importance when comparing two or more computational models. The "equivalence" of models with stochastic elements must be defined in a precise statistical context. In some cases, this is not trivial. There are at least two categories of equivalence beyond the initial criterion of numerical identity, (1) distributional and (2) relational equivalence as described earlier. From a comparative perspective, our results matched modeling techniques that utilized nonlinear program in *MATLAB* [13] and Petri nets [16], even though our technique was quite dissimilar.

One interesting finding we discovered during the analysis of the results is in reality, damage can occur in other states, while the initial attack is being repaired. Future work will address this subject. We also plan to broaden the field of play, allowing multiple attacks to occur over the enterprise. An interesting theme will be to address unknown or zero-day attacks. The ABM approach will provide security analysts with a useful decision-making tool for information security. This tool will also provide security analysts and financial analysts a useful decision-making tool to augment analysis and investments decision making in the enterprise.

# 5 References

[1] "Public Printing and Documents," in *44 USC 3502*, ed. USA, 2009, p. 3542.

[2] S. Roy, C. Ellis, S. Shiva, D. Dasgupta, V. Shandilya, and Q. Wu, "A Survey of Game Theory as Applied to Network Security," in *43rd Hawaii International Conference on System Sciences*, Koloa, Kauai, Hawaii, 2010, pp. 1-10.

[3] H. Gintis, *Thee Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*: Princeton University Press, 2009.

[4] "A Roadmap for Cybersecurity Research," Department of Homeland Security, Washington, DC November 2009.

[5] "The Comprehensive National Cybersecurity Initiative," The White House, Washington, DC 2010.

[6] S. Shiva, S. Roy, and D. Dasgupta, "Game Theory for Cyber Security," in *Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research*, Oak Ridge, Tennessee, 2010, pp. 1-4.

[7] F. T. Sheldon, S. Prowell, R. K. Abercrombie, and A. Krings, "Cyber Security and Information Intelligence Challenges and Strategies Theme," in *Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research.*, Oak Ridge, TN, 2010, pp. 1-8.

[8] C. Perrow, *Normal Accidents: Living with High-Risk Technologies*. Princeton: Princeton University Press, 1999.

[9] W. Sun, X. Kong, D. He, and X. You, "Information Security Problem Research Based on Game Theory," in *2008 International Symposium on Electronic Commerce and Security*, Guangzhou City, 2008, pp. 554-557.

[10] S. Becker, J. Seibert, D. Zage, C. Nita-Rotaru, and R. State, "Applying Game Theory to Analyze Attacks and Defenses in Virtual Coordinate Systems," in *2011 IEEE/IFIP 41st International Conference on Dependable Systems and Networks*, Hong Kong, 2011, pp. 133-144.

[11] A. Patcha and J.-M. Park, "A Game Theoretic Formulation for Intrusion Detection in Mobile Ad Hoc Networks," *International Journal of Network Security*, vol. 2 (no. 2), pp. 131-137, 2006.

[12] K.-w. Lye and J. M. Wing, "Game Strategies in Network Security," in *Proceedings of the Workshop on Foundations of Computer Security*, 2002.

[13] K.-w. Lye and J. M. Wing, "Game Strategies in Network Security," *International Journal of Information Security*, vol. 4, pp. 71-86, 2005.

[14] E. Bonabeau, "Agent-Based Modeling: Methods and Techniques for Simulating Human Systems," *Proceedings of National Academy of Sciences*, vol. 99 Suppl 3, pp. 7280-7287, 2002.

[15] M. A. Nowak and R. M. May, "The Spatial Dilemmas of Evolution," *International Journal of Bifurcation and Chaos*, vol. 3, pp. 35-78, 1993.

[16] Y. Wang, M. Yu, J. Li, K. Meng, C. Lin, and X. Cheng, "Stochastic game net and applications in security analysis for enterprise network," *International Journal of Information Security (online First)*, pp. 1-12, October 22, 2011.

[17] U. Weldnsky and W. Rand, "Making Models Match: Replicating an Agent-Based Model," *Journal of Artifical Societies and Social Simulation*, vol. 10(4)2, 2007.

[18] R. Axtell, R. Axelrod, J. M. Epstein, and M. D. Cohen, "Aligning Simulation Models: A Case Study and Results," *Computational & Mathematical Organization Theory*, vol. 1, pp. 123-141, 1996.

[19] J. Xu, Y. Gao, and G. Madey, "A Docking Experiment: Swarm and Repast for Social Network Modeling," in *Seventh Annual Swarm Researchers Conference (Swarm2003)*, Notre Dame, 2003, pp. 1-9.

# The New Old Discipline of Cyber Security Engineering

**Thomas A. Fuhrman**
Senior Vice President, Booz Allen Hamilton, Herndon, VA, USA

**Abstract -** *Although cyber security engineering is an established and diverse engineering field, it is not widely understood, and is under-applied in practice. The large and growing need to secure IT networks has been the primary driver across society in developing the cyber security workforce from high school through college and in the continuing education programs of industry and professional societies. However, this emphasis on building the workforce skills for securing IT networks neglects the distinct technical skills needed to secure complex systems other than traditional IT systems. This paper focuses on the urgent need for the discipline of cyber security engineering and its relevance to these complex systems, using mis-use case analysis as an example of systems engineering methods that can be employed.*

**Keywords:** security engineering, systems engineering, cyber, mis-use case, tradeoff analysis

## 1 Introduction

The growing recognition of the threat that hackers pose to IT networks and the enterprise data that they hold and process has attracted a great number of professionals to the field of cyber security. This workforce is widely deployed against the difficult task of protecting IT systems and software, corporate network infrastructures, and network resources (e.g., "clouds"). Because this challenge requires a wide range of different skills, the cyber security workforce is highly diverse. Professional cyber security practitioners range from entry-level analysts to experienced System Security Engineers with multiple professional certifications. Managers often view this set of specialists as the cyber "experts" in the organization, to be brought in when problems occur on the network, sometimes without regard to their particular expertise. Assigning people with the right skill levels to the right positions is uneven in both government and industry. [1, 2, 3]

Compounding the cyber security challenge is that there are not enough cyber security professionals in the workforce. Many reports describe how the nation is critically short of people with these skills. [4, 5] Since the late-1990s, the U.S. government has made a concerted effort to increase the size and depth of this workforce by establishing numerous programs aimed at increasing the pipeline of qualified cyber security professionals. Cybersecurity scholarship programs have been set up across the civil agencies and within the Department of Defense. Additionally, the National Security Agency (NSA) and the Department of Homeland Security jointly sponsor a program to designate schools whose curriculums meet certain standards as Centers of Academic Excellence in Information Assurance Education. Yet while these and other programs are making progress in increasing the cyber workforce, the demand continues to outpace supply.

The body of knowledge for cyber security today is unquestionably centered on enterprise networks and IT systems. In fact, what is striking about the qualifications and deployment of cyber security practitioners is that only a small percentage is focused beyond IT networks. This emphasis on securing traditional IT systems is not misplaced, but it is important to realize that systems other than traditional IT also have critical and often distinct cyber security needs. Those systems are the purpose-built systems that exist to perform functions in the physical world—tasks other than pure data processing. This includes a large class of systems called by names such as closed-loop systems, embedded systems, complex systems, realtime systems, realworld systems, distributed systems, and unmanned systems. Specific examples include power grids, smart cars, aircraft, air traffic management systems, manufacturing process control systems, Supervisory Control and Data Acquisition (SCADA) systems, oil drilling platforms, nuclear power plants, autonomous underwater vehicles, Unmanned Aircraft Systems (UAS), space vehicles, healthcare tools and systems including implantable medical devices, military weaponry, and a great many others. These systems are designed to perform specific functions in the physical realm rather than in cyberspace, though certainly onboard computing and external network interfaces are almost universally critical to their functions.

In the absence of an accepted all-encompassing term, the term "mission systems" is used here in referring to this class of systems.[1]

## 2 The Cyber Challenge for Mission Systems

The cyber challenge for mission systems today has two dimensions. First, buyers and owners of mission systems often do not have sufficient appreciation of the threats facing their systems in the cyber realm and the damage they can

[1] Many such mission systems, including those that are termed "critical infrastructures," have connections to IT networks for the purpose of control and communication. In these cases, the IT network provides the automated control of the realworld system—reflecting Norbert Wiener's original usage of the term cybernetics, from which today's word cyber is derived. [6]

inflict. Second, the cyber security workforce has difficulty delivering its expertise in ways that are compatible with the main engineering effort so that the overworked adage about security being "built in, not bolted on" can be realized.

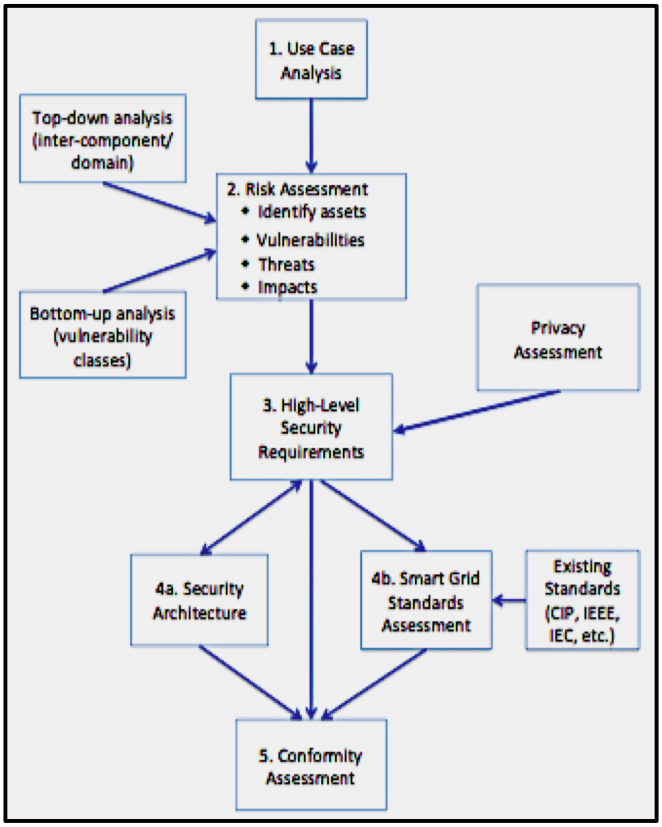
## 2.1 The Buyer/Owner Dimension

There have been many cases in recent years in which cyber vulnerabilities in mission systems were only discovered when they were exploited. Recent newsworthy examples include the 2011 case of the in-theater military UAS sensor system whose live streaming intelligence video was intercepted by the adversary using software downloaded from the Internet; the 2011 landing in Iran of a classified UAS, which at least one Iranian engineer claimed was achieved by cutting the command link and changing the vehicle's GPS position; and the widespread reporting in 2010 of a sophisticated virus that targeted computers of the Siemens product line for managing large-scale industrial control systems used by manufacturing and utility companies. Further, a 2007 test conducted by the Idaho National Laboratory proved that the so-called "Aurora Vulnerability" in a certain class of large electric generators and turbines that serve the U.S. power grid could in fact be exploited in a way that would lead to their physical self-destruction. [7, 8, 9]

These events and others like them indicate that the cyber security community often has had too small a voice in the design decisions made in the development of mission systems. But cyber security needs have not been ignored totally, and there is widespread agreement on the general concept that cyber security engineering should be part of a broader system engineering effort. In the Department of Defense, for example, cyber security for mission systems is called out in certain areas, such as in the cyber security policy for space systems, which says that Information Assurance (IA) 'shall be applied in a balanced manner by performing Information System Security Engineering (ISSE) as an integral part of the space system architecture and system engineering process to address all IA requirements in the intended operational environment.' [10]

Similarly, the National Institute of Standards and Technology (NIST) has developed draft guidelines for securing the vastly complex and emerging Smart Grid. [11] The three-volume guidelines document describes a set of tasks for assessing cyber security issues and identifying cyber security requirements. (See Figure 1.) It also contains top-level security requirements for the smart grid and defines the logical reference model for interfaces and interactions between the organizations, buildings, individuals, systems, and devices that make up the Smart Grid domains. The amount of content alone is an indication of the magnitude of the cyber challenge in this highly complex mission system.

The cyber security engineer cannot effectively work in isolation. These tasks clearly require the cyber security engineer to work side-by-side with engineers from other disciplines and domains, especially power systems specialists in this case, and to take a broad systems view of cyber risks. For mission systems, the cyber engineer needs to know the systems engineering process, the tools used, and the artifacts produced.


1. Use Case Analysis
Top-down analysis (inter-component/ domain)
2. Risk Assessment
• Identify assets
• Vulnerabilities
• Threats
• Impacts
Bottom-up analysis (vulnerability classes)
Privacy Assessment
3. High-Level Security Requirements
4a. Security Architecture
4b. Smart Grid Standards Assessment
Existing Standards (CIP, IEEE, IEC, etc.)
5. Conformity Assessment

Source: NISTIR 7628, vol 1

**Figure 1. Smart Grid Cyber Security Engineering Tasks**

One aspect of cyber security engineering that differentiates it from other engineering fields is that its focus is primarily (though not exclusively) on the potential disruption of system performance caused by the *deliberate actions of human actors intent on doing harm*. Designing for security is different in this way from designing against environmental effects, unreliable components, or external hazards. The unique value that the cyber expert can bring to an engineering effort is a technical understanding of the threat and an ability to identify potential vulnerabilities in the mission system that could be exploited by the threat, as well as the range of options for mitigating the risk posed by the threat.

Figure 2 shows some of the threat vectors that mission systems need to address. Additionally, cyber security considerations can lead to requirements for implementing special features such as a command disable function or anti-tamper technologies to guard against compromise and reverse engineering if the system is physically exploited.

- Exploitation of vulnerabilities in embedded mission platform software and firmware (e.g, Operational Flight Program) and its development and maintenance
- Exploitation of vulnerabilities of on-platform operating systems
- Exploits against the attack surface of the connected network
- Data protocol exploitation
- Insiders (both witting and unwitting)
- External interfaces/communications links
- Portable media (e.g., CDs, USB devices)
- Local "plug-in" devices (e.g., peripherals, special purpose probes, sensors, test and diagnostic tools)
- Supply chain

**Figure 2. Example Cyber Threat Vectors Affecting Mission Systems**

## 2.2 The Workforce Dimension

The workforce challenge for the cyber security of mission systems is particularly difficult. Not only are there too few cyber security professionals in total, but only a minority of those in the workforce today have the engineering training and credentials to credibly engage in the engineering process. It is still somewhat unusual to find a cyber professional with experience in mission systems engineering, and who is able to blend with an engineering team to develop meaningful requirements and operate in the trade space through which the design is evolved.

These tasks would challenge many cyber specialists today because systems engineering methods differ in important respects from the way cyber security services are typically delivered. The structure within which IT security specialists operate is the well-thought-out Risk Management Framework (RMF) described in Special Publication 800-53 of the National Institute of Standards and Technology (NIST). The framework helps the specialist define required levels of assurance, select the appropriate security controls from a comprehensive catalog, assess that the controls are implemented correctly, support a formal decision by a designated owner to authorize operation, and then continuously monitor the security of the system throughout its life cycle. [12]

While the RMF and controls catalog form an essential foundation, more is expected of the cyber security engineer. First, the systems engineering environment expects a more interdisciplinary focus and even more engineering creativity than the RMF structure fosters. For systems of any appreciable complexity, inevitably there are competing operational and technical considerations. One of the key tools of engineering for complex systems is the formal tradeoff study to examine alternatives and make design choices. Tradeoff studies, while common in the engineering of mission systems, are not normally used in the development of IT networks, and cyber security specialists are not usually expected to have this skill. Cybersecurity needs to be part of the tradespace. Advocates recognize that more formalization of the cyber security engineering career field, patterned on the features of established engineering fields, will take time.[2, 3]

Among the most mature of the efforts to advance the systems engineering approach to cyber security is the Systems Security Engineering—Capability Maturity Model (SSE-CMM) standard. Codified as an International Organization for Standardization (ISO) standard (ISO/IEC 21827:2008), SSE-CMM describes the security engineering processes that organizations need to ensure good security engineering. The standard provides a reference model for system security engineering throughout the entire system life cycle and the entire organization, including interaction with other disciplines and with other organizations. It is designed to be congruent with the Systems Engineering process. [13]

# 3 A Synthesis of Disciplines

Systems Engineering is inherently interdisciplinary. As such it provides an overarching framework in which multiple disciplines can productively operate and integrate towards a common design goal. Figure 3 summarizes some of the key features of Systems Engineering.

Systems Engineering is an interdisciplinary approach that focuses on defining customer needs and required functionality early in the development cycle, documenting requirements, then proceeding with design synthesis and system validation while considering the complete problem:

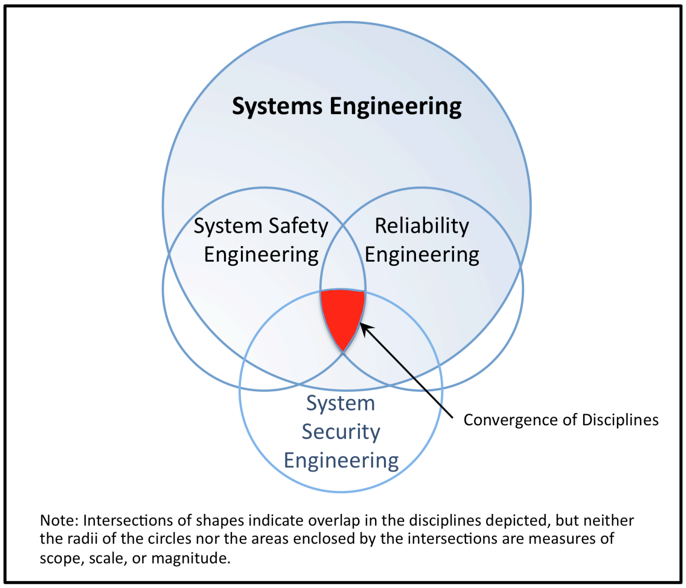
- Operations
- Performance
- Test
- Manufacturing
- Cost & Schedule
- Training & Support
- Disposal

Systems Engineering integrates all the disciplines and specialty groups into a team effort forming a structured development process that proceeds from concept to production to operation. Systems Engineering considers both the business and the technical needs of all customers with the goal of providing a quality product that meets the user needs.

*Source: International Council on Systems Engineering*

**Figure 3. What is Systems Engineering?**

Cyber Security Engineering has strong affinity with two other disciplines found in this environment—System Safety Engineering and Reliability Engineering. These disciplines have long histories and active professional communities. All three are oriented towards managing throughout the full system life cycle, and integrate very well into the overarching System Engineering framework. All three operate generally in the realm of nonfunctional requirements, with the goal of making the design inherently resistant to failures. In practice,


Systems Engineering
System Safety Engineering
Reliability Engineering
System Security Engineering
Convergence of Disciplines

Note: Intersections of shapes indicate overlap in the disciplines depicted, but neither the radii of the circles nor the areas enclosed by the intersections are measures of scope, scale, or magnitude.

*Source: Booz Allen Hamilton*

**Figure 4. Convergence of Disciplines Within the Systems Engineering Framework**

many of the tools (such as Risk Assessment) used within these disciplines are very similar to each other. They also have in common the fact that the non-functional requirements that emerge from safety, cyber security, or reliability concerns may be in tension with the performance objectives that the system is being designed to meet—and therefore may be overlooked or overcome by the pressure to deliver performance. Figure 4 shows a Venn diagram indicating the relationship among cyber security, safety, and reliability components. [14, 15, 16, 17, 18]

Good examples of the integration of these disciplines are found in two government agencies: the mission assurance program of NASA and the surety programs of the Department of Energy. Both explicitly seek to integrate safety, security, reliability, and quality across the system life cycle and have proven records of success. [19] Table 1 summarizes some of the key features and fundamental methods of Systems Engineering, Reliability Engineering, System Safety Engineering, and Cyber Security Engineering.

# 4 Cyber Security in the Tradespace: An Example

"Use case" analysis is one of the tools of Systems Engineering that has particular relevance to cyber security, used for both requirements identification and in tradeoff studies over alternative solutions. A use case is a description of the employment of the target system in an operating scenario with emphasis on its functions and interactions with the external environment including human actors. It provides

**Table 1. Summary of Four Systems Engineering Disciplines**

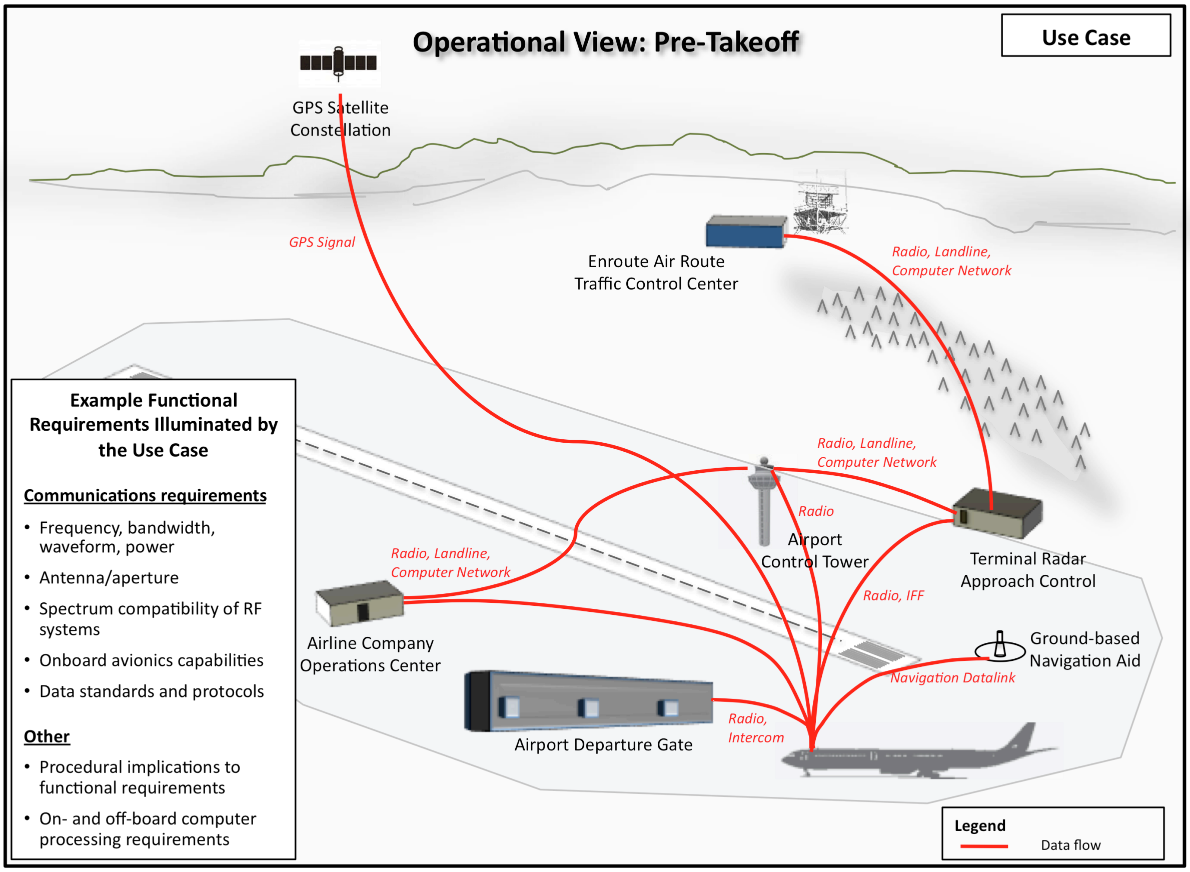
| | Background | Fundamental Methods |
|---|---|---|
| Systems Engineering | ▪ Interdisciplinary by design<br>▪ International Council on Systems Engineering (INCOSE) develops and disseminates best practices for successful systems.<br>▪ Publishes the Systems Engineering Handbook and maintains the Systems Engineering Body of Knowledge<br>▪ Certification programs [14] | ▪ Program integration and management tools<br>▪ Use Case Analysis<br>▪ Design Trade-off Analysis (Figures of Merit/Evaluation Measures)<br>▪ Life Cycle management tools |
| Reliability Engineering | ▪ Emerged in the 1950s<br>▪ Relationship to Surety Engineering and NASA Mission Assurance [15]<br>▪ Industry-recognized Certified Reliability Engineer (CRE) and Certified Reliability Professional certifications through American Society for Quality (ASQ) [16]<br>▪ IEEE Reliability Society provides numerous professional development opportunities [17] | ▪ Statistical modeling<br>▪ Reliability Physics (Physics of Failure)<br>▪ Failure Modes and Effects Analysis<br>▪ Fault Tree Analysis |
| System Safety Engineering | ▪ International System Safety Society fosters the application of systems engineering and systems management to the process of hazard, safety, and risk analysis [18]<br>▪ Certification programs | ▪ Qualitative Analysis to anticipate failure potential during the design phase<br>▪ Hazard, Safety, and Risk analyses (qualitative and quantitative)<br>▪ Designing ways to contain failures<br>▪ Safety of software as a special area of focus |
| Cyber Security Engineering | ▪ Major industry-recognized certifications through (ISC)2, SANS, ISACA, and other organizations<br>▪ System Security Engineering Capability Maturity Model (ISO/IEC 21827:2008) model for organizations [13] | ▪ Mis-Use Case Analysis<br>▪ Threat Identification and Characterization<br>▪ Risk Management Framework and controls catalog<br>▪ Continuous management of system security throughout the life cycle |

a structured way of thinking about how the system will be used in its operating environment that helps in defining the functional requirements.

In practice, use cases are usually expressed using the Unified Modeling Language (UML) that depicts both the actors and the process flow, facilitating information exchange and enabling the use of automated support tools. However, it can be helpful to begin by developing a top-level conceptual picture similar to the "operational view" of the Department of Defense Architecture Framework (DODAF). This can then provide a structured way of thinking about the problem to illuminate needs, enable creative cross-disciplinary discussion, and produce insights into the cyber security and other non-functional requirements. It can be a pre-cursor to the UML Use Case Diagrams. These insights can later be developed into UML diagrams as well, ultimately leading to additional system requirements. [20, 21]

Both analyses—use case and mis-use case—can help with the trade studies through which the design evolves in addition to their role in requirements definition.

An example of the use case and mis-use case operational views is shown in Figures 5 and 6. These figures depict a notional case in the Air Traffic Management System: the pre-takeoff preparation of the aircraft, filing of the flight plan, and the ground operations associated with starting the engines and taxiing. Coordination with the air traffic management facilities of the Federal Aviation Administration (FAA) is a necessity, as are programming the onboard navigation computer, getting authorization from the airline operations center, and obtaining taxi clearance from the control tower. These process steps are accomplished by people at a wide range of locations and facilities.[2] The operational view of the


Operational View: Pre-Takeoff
Use Case
GPS Satellite Constellation
GPS Signal
Enroute Air Route Traffic Control Center
Radio, Landline, Computer Network
Radio, Landline, Computer Network
Radio
Airport Control Tower
Terminal Radar Approach Control
Radio, Landline, Computer Network
Radio, IFF
Ground-based Navigation Aid
Navigation Datalink
Airline Company Operations Center
Airport Departure Gate
Radio, Intercom
Legend
Data flow

**Example Functional Requirements Illuminated by the Use Case**

**Communications requirements**

- Frequency, bandwidth, waveform, power
- Antenna/aperture
- Spectrum compatibility of RF systems
- Onboard avionics capabilities
- Data standards and protocols

**Other**

- Procedural implications to functional requirements
- On- and off-board computer processing requirements

*Source: Booz Allen Hamilton*

**Figure 5. Operational View of a Pre-Takeoff Use Case (Air Traffic Management)**

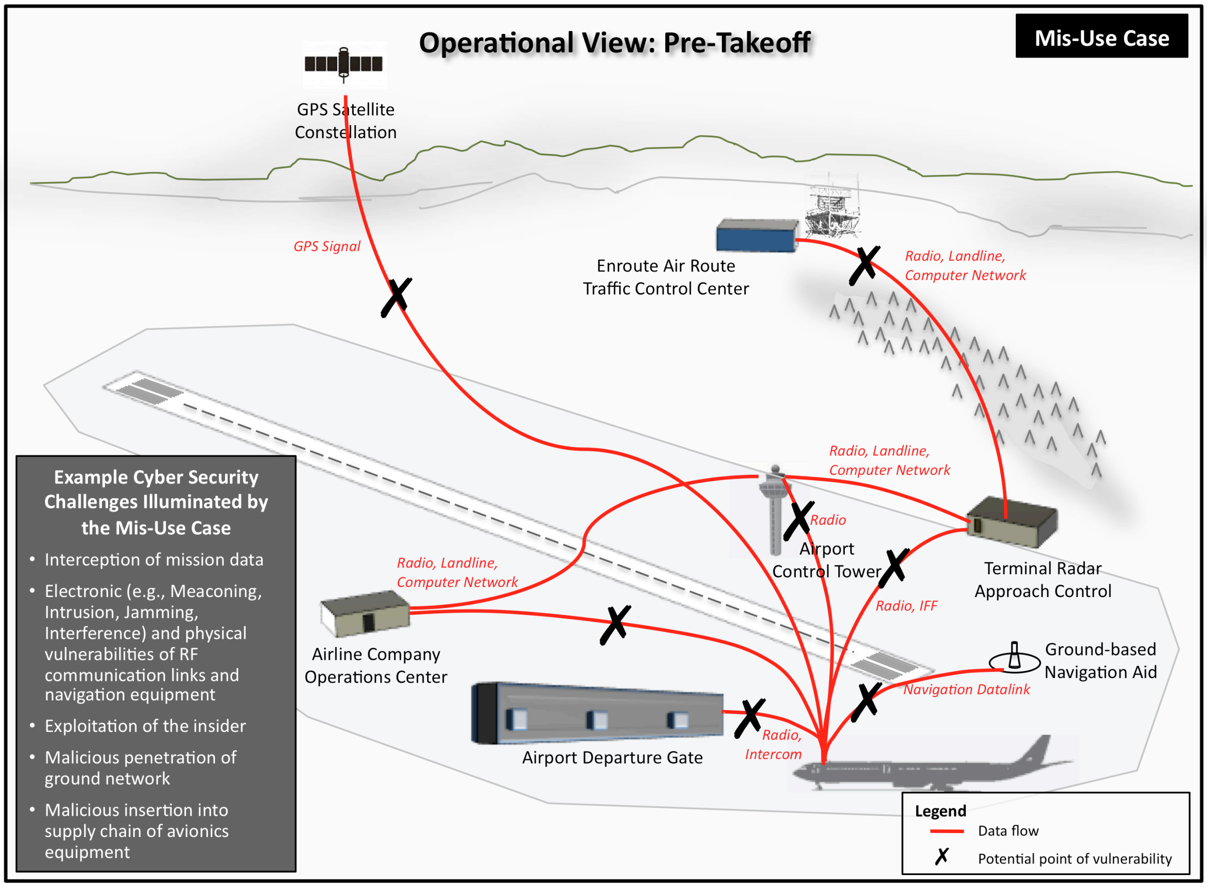
A tool that is particularly suited to the cyber security engineering challenge is "mis-use case" analysis. Initially developed in the 1990s, the mis-use case turns the use case around by focusing on what a malicious actor could do to disrupt, subvert, or negate the performance of the system. The top-level operational view can also be used for the mis-use case.

[2] This scenario is for illustration only. In reality, most of the requirements of today's Air Traffic Management System are already known and specified by standards and regulatory requirements of the FAA and other agencies. Nonetheless, specific implementation details would typically still need to be decided as part of the system engineering effort, and a regular review of mis-use cases is advisable as threats change.

use case and its associated misuse case allow all members of the systems engineering team to work together from a common starting point.

Examination of the mis-use case should involve every component and link within the system, and every relevant threat vector with the goal of illuminating the cyber security challenges. These results should be brought forward for further consideration and analysis.

In the example shown in Figure 6, possible cyber challenges suggested by the operational view include interception of mission data by intruding into the communication links in the system; exploitation of the insider leading to compromise of access controls or other critical security controls; penetration of the ground-based networks that communicate and process critical system data; and malicious exploitation of vulnerabilities in the supply chain of the avionics equipment. These insights are just the start of the process, and a full use/mis-use case analysis using accepted systems engineering tools should be the next step.


Operational View: Pre-Takeoff
Mis-Use Case
GPS Satellite Constellation
GPS Signal
Enroute Air Route Traffic Control Center
Radio, Landline, Computer Network
Radio, Landline, Computer Network
Radio
Airport Control Tower
Terminal Radar Approach Control
Radio, Landline, Computer Network
Radio, IFF
Airline Company Operations Center
Ground-based Navigation Aid
Navigation Datalink
Airport Departure Gate
Radio, Intercom
Legend
Data flow
✗ Potential point of vulnerability

**Example Cyber Security Challenges Illuminated by the Mis-Use Case**

- Interception of mission data
- Electronic (e.g., Meaconing, Intrusion, Jamming, Interference) and physical vulnerabilities of RF communication links and navigation equipment
- Exploitation of the insider
- Malicious penetration of ground network
- Malicious insertion into supply chain of avionics equipment

*Source: Booz Allen Hamilton*

**Figure 6. Operational View of a Pre-Takeoff Mis-Use Case (Air Traffic Management)**

## 5 Summary and Prescription

Although the intellectual groundwork for cyber security engineering for mission systems is solidly in place, the degree of true engagement by cyber security engineers still falls short of what it should be. Evidence indicates that acquiring organizations do not have a clear picture of the value proposition of the cyber security engineer, and, frankly, there are not enough qualified cyber security engineers to meet the needs even if the value proposition were recognized. If cyber security specialists are to have an impact on mission systems, they must have the skills to engage in the system engineering process as franchised members, not as dabblers. This will be difficult to achieve as the cyber community is already struggling to develop the workforce to address the more obvious needs of securing networks and IT systems.

More emphasis is therefore needed on the specific challenge of cyber security engineering for mission systems through existing university programs, U.S. government cyber scholarship initiatives, and professional certification programs.

Cyber security specialists themselves need to be part of the solution. They should strive to learn the practices of systems engineering, encourage their organizations to embrace SSE-CMM, and work hard at their own professional development. They should learn and internalize the unique value that the cyber security engineering community can bring to the systems engineering arena. And they should gain experience in the use of systems engineering tools.

Lastly, the similarities and strong overlaps among Cyber Security Engineering, System Safety Engineering, and Reliability Engineering should prompt those professional communities to work together in an effort to find greater synergy in the systems engineering environment. The professional societies and associations that represent these stakeholders should join together under the auspices of the International Council on Systems Engineering (INCOSE) to tackle this together to enhance the profession and produce mission systems with better performance in any environment—normal, abnormal, or hostile.

## 6 References

[1] *Cyber IN-security: Strengthening the Federal Cyber security Workforce*; Partnership for Public Service and Booz Allen Hamilton, July 2009.

[2] Brian Dutcher. "Determining the Role of the IA/Security Engineer," SANS Institute; InfoSec Reading Room. March 15, 2010, http://www.sans.org/reading_room/whitepapers/assurance/determining-role-ia-security-engineer_33508.

[3] Robert Ayoub. The 2011 (ISC)2 Global Information Security Workforce Study, Frost & Sullivan Market Survey Sponsored by (ISC)2, 2011.

[4] Karen Evans and Franklin Reeder. "Human Capital Crisis in Cybersecurity Technical Proficiency Matters," A Report of the CSIS Commission on Cybersecurity for the 44th Presidency, Center for Strategic and International Studies, November 2010.

[5] Eric Beidel and Stew Magnuson. "Government, Military Face Severe Shortage Of Cybersecurity Experts", *National Defense* (National Defense Industrial Association), August 2011, http://www.nationaldefensemagazine.org/archive/2011/August/Pages/Government,MilitaryFaceSevereShortageOfCybersecurityExperts.aspx.

[6] Norbert Wiener. *Cybernetics: or Control and Communication in the Animal and the Machine*, The Massachusetts Institute of Technology, Cambridge, MA, 1948 and 1961.

[7] Peter Neumann. Moderator, *Risks Digest*, http://catless.ncl.ac.uk/Risks.

[8] Robert McMillan. "Virus targeted at Siemens industrial control systems", IDG News Service, July 17, 2010. http://www.networkworld.com/news/2010/071710-new-virus-targets-industrial.html.

[9] Brent Kesler. "The Vulnerability of Nuclear Facilities to Cyber Attack," *Strategic Insights*, Vol. 10, Issue 1, pp. 15 – 25, Spring 2011.

[10] DoD Directive 8581.1. "Information Assurance (IA) Policy for Space Systems Used by the Department of Defense," June 21, 2005.

[11] The Smart Grid Interoperability Panel – Cyber Security Working Group. *Guidelines for Smart Grid Cyber Security*, NISTIR 7628, August 2010.

[12] NIST Special Publication 800-53 Revision 3. "Recommended Security Controls for Federal Information Systems and Organizations," National Institute of Standards and Technology, Gaithersburg, MD.

[13] ISO/IEC 21827:2008. *Systems Security Engineering—Capability Maturity Model®*.

[14] International Council on Systems Engineering (INCOSE). http://www.incose.org

[15] NASA Office of Safety and Mission Assurance. http://www.hq.nasa.gov/office/codeq/.

[16] American Society for Quality. Certified Reliability Engineer, http://prdweb.asq.org/certification/control/reliability-engineer/index.

[17] IEEE Reliability Society. http://rs.ieee.org/.

[18] The International System Safety Society. http://www.system-safety.org/.

[19] Nancy Leveson. "White Paper on Approaches to Safety Engineering." Nancy Leveson's Home Page at MIT, April 23, 2003; http://sunnyday.mit.edu/caib/concepts.pdf.

[20] Guttorm Sindre and Andreas Opdahl. *Eliciting Security Requirements by Misuse Cases*, Proceedings of TOOLS Pacific 2000, pp. 120-131, 20-23 November 2000, IEEE Computer Society Press.

[21] Ian Alexander. "Use/Misuse Case Analysis Elicits Non-Functional Requirements," *Computing & Control Engineering Journal*, Volume 14, Issue 1, pp. 40 – 45, Feb. 2003.

# A Survey of Cyberspace Mission Assurance within United States Air Force Communications Squadrons

**Mickey Evans, Michael R. Grimaila, and Robert F. Mills**
Air Force Institute of Technology, Wright-Patterson AFB, Ohio 45433, USA

**Abstract** –*The United States Air Force (USAF), like many other large enterprises, has embedded information and communication technologies into their core organizational processes as a means to increase operational efficiency, improve decision making quality, reduce delays, and/or reduce costs. Each day, USAF systems and networks are under attack from hostile actors. Recognition of these facts has led to the need to assure not only the networks and systems that make up cyberspace, but also the missions that rely so heavily on it. In this paper, we present the results of a research study focused on identifying the state of cyber mission assurance efforts within USAF communication squadrons, which operate base-level networks around the globe. A survey consisting of 41 questions was constructed to establish what extent base-level units perform cyber mission assurance activities and what factors influence their efforts. The survey was distributed to communications squadrons across all of the USAF Major Commands (MAJCOMs). Analysis of 62 respondents revealed the presence of minimal regulatory requirements; a need for consistent guidance, policy and procedures; and identified trends which complicate mission assurance at units tasked with providing communications services.*

**Keywords:** mission assurance; situational awareness; mission impact assessment

## 1 Introduction

Nearly every aspect of military operations is supported by Information and Communications Technologies (ICT) collectively called “Cyberspace.” The use of ICT in military applications has increased weapon system effectiveness and increased the margin of safety for military personnel. However, as an organization’s dependence on ICT increases so does the risk of mission failure resulting from loss or degradation of communications links and network resources [1]. The need to ensure connectivity and network services to support mission operations is clearly evident, but this is no simple task as it requires an extensive analysis of the infrastructure elements and related resources. Unfortunately, while there are strategies [2], they have not been universally adapted within the USAF [3]. Further, mission assurance is often seen as one of those tasks that can take a back seat to more pressing matters such as the daily maintenance and operations of the network. Worse, until an event arises that stresses the day-to-day operations and threatens the mission, there never seems to be enough money, time, or manpower to address mission assurance. Even when an ICT incident occurs, the post event “hot wash” tends to focus on the specific ICT event and not the broader picture of identifying what impact the event had on mission operations and capturing this information for similar future events [4]. There are several factors that make mission assurance, and in particular cyber mission assurance, planning difficult. Complexities inherent in the problem set include difficulties in dependency mapping; a lack of automated tools; and an organizational structure in constant flux with undefined lanes in the road [5-8]. In addition, the lack of standard definitions; a hodge-podge of directives and policies that give lip service to the notion of mission assurance but provide very little if any direction; a lack of accountability; and over-tasked and undermanned units that are just able to make it through each day, let alone expend energy on planning for something that may never happen [9]. Where there is a trace of mission assurance activities, it is often focused solely on the ICT systems, not on the missions they support [1,3,7]. The notion of “cyber mission assurance” is to some extent a misnomer. What good is the most robust of IT systems if, in the end, the missions they support are ineffective? Cyber mission assurance therefore, is a sub-set of an overarching mission assurance effort. Mission assurance, in order to be effective, must address all aspects of the mission, including the role played by cyberspace and the systems that define the domain [10]. The purpose of this paper is to establish an understanding of cyber assurance as implemented in the USAF.

In this paper, we present the results of a survey of USAF base-level communications squadrons conducted in 2010 in an effort to provide a clear picture of the current state of mission assurance efforts and pertinent directives and guidance. An analysis of the results provides a snapshot of the realities faced by these units involving with implementing and maintaining cyber mission assurance. The remainder of this paper is structured as follows: In section 2, we discuss the development of the survey. In section 3, we present the survey responses and provide a brief analysis of results. Finally, in section 4 we present our conclusions.

## 2 Mission Assurance Survey

As with most military endeavors, there exists strategy, doctrine, directives, and standard operating procedures that provide the sources from which tactical guidance is derived. There are also realities surrounding the tactical level implementation and the culture and atmosphere under which these practices take place. Existing mission assurance research provides an academic, and somewhat clinical, view of the current landscape. To truly appreciate the reality of cyber mission assurance at the tactical level requires a different approach. To that end, an on-line survey was produced, the goal of which, was to provide a snapshot of the current realities and atmosphere for cyber mission assurance efforts at the unit level within the USAF. The survey contained 41 questions which asked respondents to provide data involving the current state of mission assurance planning from a base, unit, and I-NOSC perspective, sources of funding for mission assurance planning, and historical incidents which have resulted in disruption of day-to-day operations. It also asked respondents to self-assess their mission assurance activities. Further details about the survey, including a list of the questions and available response options, are available in [10]. Note: In the Air Force's network management construct, network operations are performed in a distributed manner. Some base-level network management functions are performed by the Network Control Center (NCC), which falls under the local communication squadron. To improve standardization and reduce costs, Integrated Network Operations and Security Center (I-NOSC) remotely manages the networks for multiple bases [11].

### 2.1 Survey Focus

The survey only considered that part of a communications unit that dealt with core ICT services. Core services as defined by AFI 33-115V1 are:

> *"those services defined by the Air Force IT community as central components of the AF-GIG. They embody the seamless, secure, and reliable transport of timely and trusted information across the AF-GIG."* [11]

The specific core services as defined by the AFI are:

- Electronic Messaging
- Address Management
- Directory Services
- Information Assurance and Security Hardware
- Domain Name Servers (DNS)
- Exchange
- Windows Internet Naming Service (WINS)
- Domain Controllers (PDC/BDC) remote (PDC/BDC)
- Dynamic Host Control Protocol (DHCP) server
- Local Directory Service Agent (LDSA)
- Defense Message System (DMS)

For this reason, the survey focused on seven specific areas:

- The status of a base/wing mission assurance plan
- The status of a NOSC mission assurance plan
- The status of a unit mission assurance plan
- The existence of conflicts where multiple mission assurance plans exist
- Barriers and Drivers for mission assurance efforts
- Historical events that have threatened the mission
- A self-assessment of the change in mission assurance readiness

### 2.2 Survey Collection

Upon receiving approval from the Institutional Review Board, a message was sent to all MAJCOM A6s ("A6" refers to the staff entity responsible for planning and resourcing IT support functions within the MAJCOM) asking for communications squadron commanders to complete the mission assurance survey. The on-line survey opened on 18 March 2010 and closed on 31 March 2010. Of the 62 responses received, 60% (37) were from Air National Guard (ANG) and Air Force Reserve units while the remaining 40% (25) were from active duty units. The majority of the respondents, 52%, were communication squadron commanders. Another 19% were deputy commanders or directors of operations, operations flight commanders, plans flight commanders, or action officers. 29% identified as "other." The average amount of experience in the unit for respondents was 4.48 years. 82% either "agreed" or "strongly agreed" that they are "Personally Very Involved in IT Mission Assurance in their Unit."

## 3 Survey Findings

### 3.1 Base/Wing Mission Assurance

The first focus area concentrated on the base or wing's mission assurance plan. Each wing has a specific mission (e.g., airlift or air combat) and just as the communication squadron should be interested in planning for mission assurance, so should the installation on which the unit resides. 39% of all respondents said their base or wing had some sort of a plan for mission assurance in various stages of readiness, as shown in Figure 1. 31% said the plan was being developed. 30% said there was no plan currently in place.

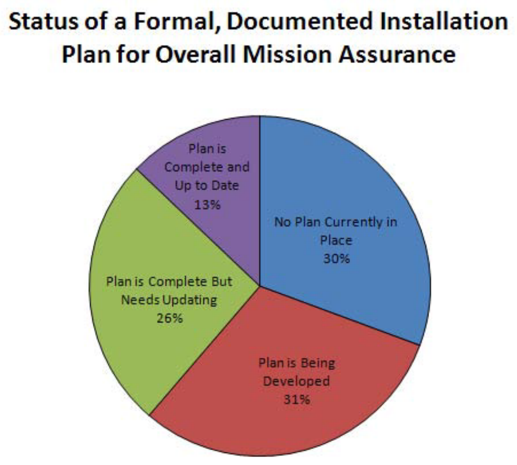


Figure 1 – Status of Installation Mission Assurance Plan

The numbers paint a worse picture for active units than for guard and reserves. 46% of active units reported no base or wing plan. On a positive note, 91% of all respondents said that where a base or wing plan existed, ICT was included.

The value of any plan is only as strong as the extent to which it was tested and verified. Beyond the basic test, to determine if it meets the desired goal, the plan should be periodically tested to see if it continues to meet the need and to train personnel on its implementation. When asked for the time frame in which the base or wing plan was last tested, 44% reported that the last test had been conducted within the last year (Figure 2). Nearly as many, 30% (7) could not identify when the last test had taken place. Of the 17 respondents that completed the question regarding the type of activity used to test the base/wing plan, 8 said the test was in the form of a table top exercise and another 8 said the test was scheduled in advance (possibly as part of an ORI). Only 1 reported that the test was completely unscheduled. Finally, as depicted in Figure 3, of the 70% (43) respondents that said there was either a plan of some sort in place or one was being developed at base/wing level, only 56% (24) said that an individual had been appointed at the installation level who was responsible for mission assurance.

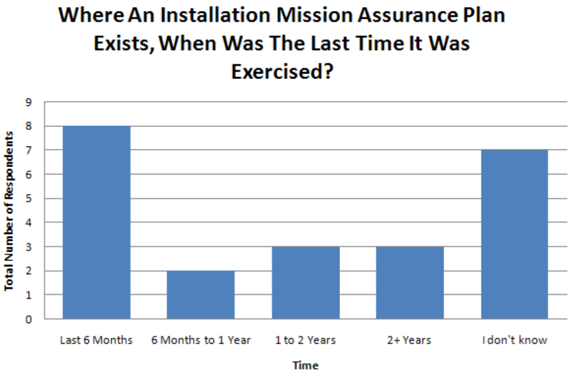


Figure 2 – Installation Plan Testing

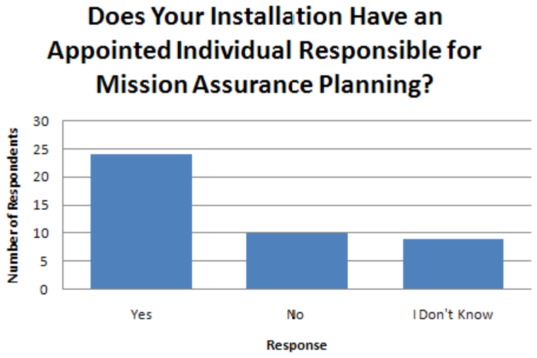


Figure 3 – Status of Installation Point of Contact

The Air Force Network Operations (AFNETOPS) construct is the latest iteration of the way in which the Air Force manages its portion of the Global Information Grid (GIG). Section 4.4.4 of AFI 33-115V1, Network Operations states that:

> *"The I-NOSC ensures Air Force networks are capable of conducting, supporting, and advancing coalition, joint, Air Force, and interagency operations. Through a common environment, the I-NOSC provides situational awareness to the AFNOSC, WFHQ, and MAJCOMs. Each I-NOSC will oversee the operation of the base level NCCs, while providing remote administration of enterprise-wide infrastructure."* [11]

It is important to note, from the definition, that part of the I-NOSC s responsibility is to provide remote administration to the enterprise-wide infrastructure. In short, the I-NOSCs provide services and support to base level communications squadrons, which is why the base level units have a vested interest in the mission assurance plans of the I-NOSCs. 10% (6 of 62) of respondents reported they do not receive services from an I-NOSC. Four of those were Guard units and one was itself an I-NOSC. The one remaining was an active duty unit. Three respondents did not provide responses on the section of the survey asking details about the I-NOSCs plan, even though they reported that they receive services from the I-NOSC.

What follows is data pertaining to the 85% (53 of 62) of respondents that reported they do receive services from an I-NOSC and provided answers questioning the status of the I-NOSC plan. Only 24% of the respondents (see Figure 4) said the I-NOSC from which they receive services had some sort of a plan for mission assurance in various stages of readiness. Further, 36% said the plan was being developed and 40% that said there was no plan currently in place; *hence, at the time of the survey, 76% of respondents reported there was no mission assurance plan in place that they could turn to should an event occur that causes a significant impact in day-to-day operations.* When asked if the services they receive from the I-NOSC are covered in the I-NOSC mission assurance plan, only 78% responded "yes."

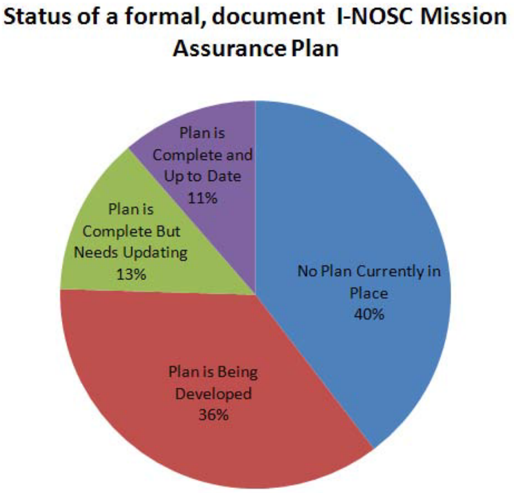


Figure 4 – Status of I-NOSC Mission Assurance Plan

As shown in Figure 5, 46% said the plan had been exercised in the last year. The remaining 54% could not identify when the plan had last been tested. Of the 6 respondents that completed the question regarding the type of exercise used to test the I-NOSC plan, 5 said the exercise was scheduled in advance. One reported that it was completely unscheduled. Only 22% responded that there was a POC responsible for mission assurance planning at the I-NOSC level. The remaining 78% said that either there was no POC or they did not know if there was one.

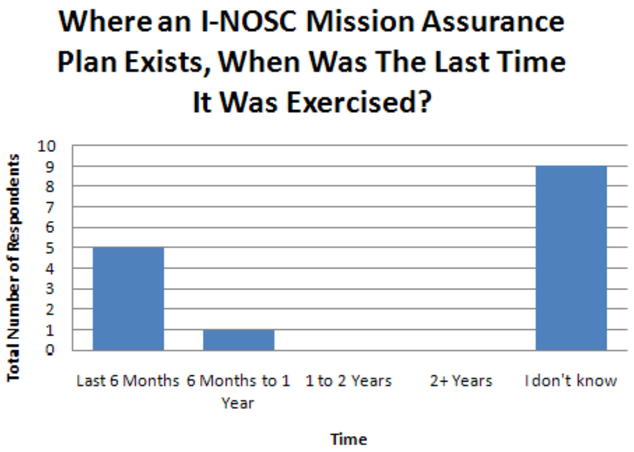


Figure 5 – I-NOSC Plan Testing

## 3.2 Unit Mission Assurance

At the unit level, 45% of all respondents said their unit had some sort of a plan for mission assurance in various stages of readiness,  27% said the plan was being developed, and the remaining 28% said there was no plan currently in place. Figure 6 shows the number of respondents, broken out but active duty and Guard/Reserve units and how they responded. The difference between active duty and Guard/Reserve units is notable. Forty percent of Guard and Reserve units reported they have a mission assurance plan that is complete and up to date, while only 24% of active duty units reported the same.

When asked for the time frame in which the unit's plan was last tested, 50% of the respondents reported the plan had been tested in the last 12 months, while 21% could not identify when the last test had taken place (see Figure 7). Of the 20 respondents that completed the question regarding the type of exercise used to test the unit plan, eight said the test was in the form of a table top exercise and nine said the test was scheduled in advance (possibly as part of an Operational Readiness Inspection). Two reported that the test was completely unscheduled and the remaining respondent reported that the type of test fell into the "other" category.

Of the 73% (45) respondents that said there was either a plan of some sort in place or one was being developed at base/wing level, only 67% (30) said that an individual had been formally appointed at the unit level as being responsible for mission assurance planning. Finally, of the 45 respondents that reported that a plan was either being developed or existed in some form, only 40% (18) reported they had been contacted to support other units' mission assurance efforts.

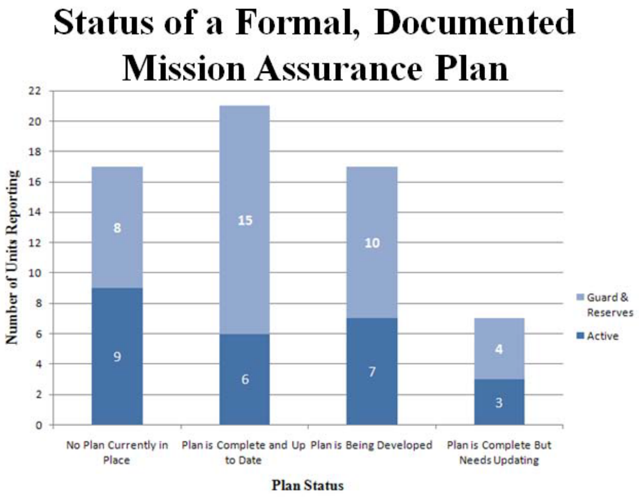


Figure 6 – Status of Unit Mission Assurance Plan

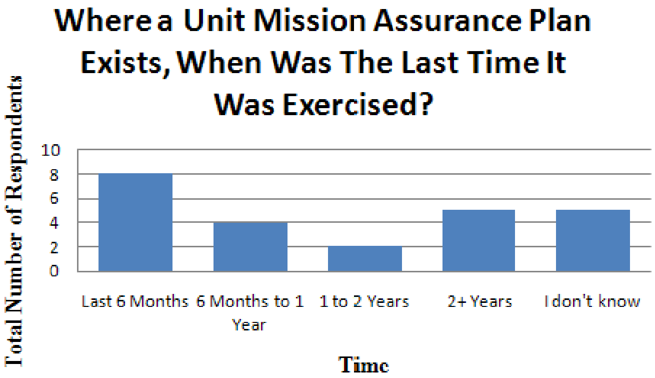


Figure 7 – Unit Plan Testing

## 3.3 Base Level Focus Areas

It is clearly evident from the data presented thus far that mission assurance planning is lacking. A series of questions was asked to try to develop an understanding of what might be causing this. The first questions asked respondents to identify the primary and secondary drivers for mission assurance planning at their base. In other words what, if anything, was motivating the base-level units to engage in mission assurance planning? As shown in Figure 8, the critical role IT resources plays in the mission of the base is by far the main reason units are involved in mission assurance planning. However, what is surprising is that the second and third major drivers are wing/MAJCOM or I-NOSC requirements.  In all cases, except one, respondents that identified the presence of a wing or I-NOSC mission assurance plan reported a corresponding requirement from that same entity for their unit. It seems to be evident that base level units are encouraged to develop mission assurance plans when it is also important to either their wing leadership or their I-NOSC. One lone respondent reported that the primary driver was the Wing/MAJCOM yet also reported that the wing had no plan of their own.

The remaining question asked units to identify the barriers to IT mission planning at their base. Respondents had the option to select up to two responses. Figure 9, provides a summary. The data shows that the primary barrier is related to technology limitations. The secondary barrier is funding. This correlates to the earlier observation that units are undermanned and that tools need to be developed to assist with mission assurance planning. When asked if their unit received funding specifically for mission assurance planning and/or implementation of mission assurance measures, 71% reported that they did not, 3% said that they did, and the remaining 26% were unsure. This data, along with the data concerning barriers to mission assurance planning suggests that funding, specifically for mission assurance planning, is needed.

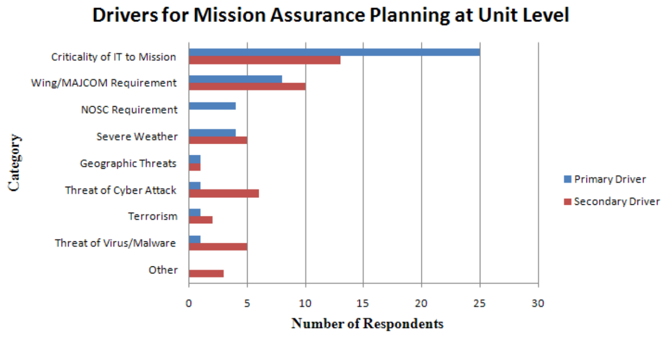


Figure 8 – Primary and Secondary Drivers of Mission Assurance Planning

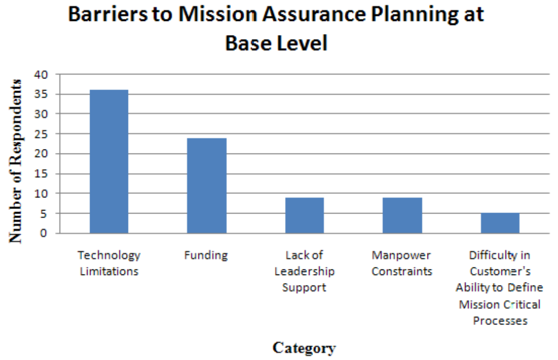


Figure 9 – Base Level Barriers to Mission Assurance Planning

## 3.4 Incident History

At its most basic level, mission assurance encompasses the component referred to as continuity of operations. At the foundation of mission assurance is the unit's ability to provide its core functions. Knowing what has happened, from an historical perspective, that caused units to deviate from normal day-to-day operations is relevant in trying to understand where to concentrate efforts that will make the most significant impact on the way ahead. What follows is data resulting from a series of questions regarding the actual incidents that have occurred in the last five years.

Figure 10 summarizes results when respondents were asked about the number of disruptions experienced in the last five years. As shown, 55% said they had experienced at least one incident in the five year time frame. In fact, 10 units reported they had experienced three disruptions, and 14 units reported they had experienced five disruptions or more.

Respondents were asked to provide additional details on the two most significant disruptions experienced in the five-year period, with results provided in Figure 11. The majority of the disruptions (75%) were caused by one of four factors: loss of electricity, hardware problems, HVAC (heating, ventilation and air conditioning) issues, and cable cut. Another 9% was caused by severe weather, and 3% (2 responses) were attributed to "cyber attack". This label is broad in scope, and there was no clarifying data to elaborate on the specific type of cyber attack.

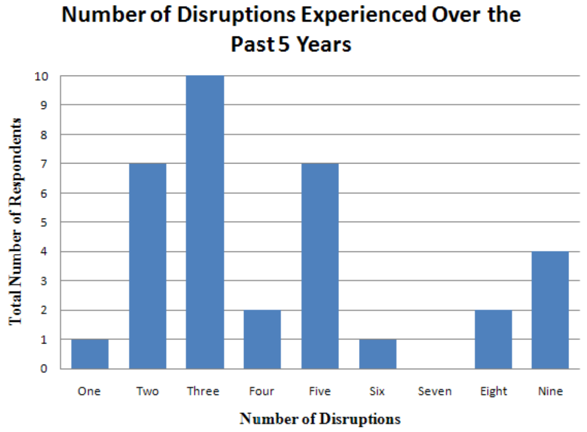


Figure 10 – Number of Disruption in Past Five Years

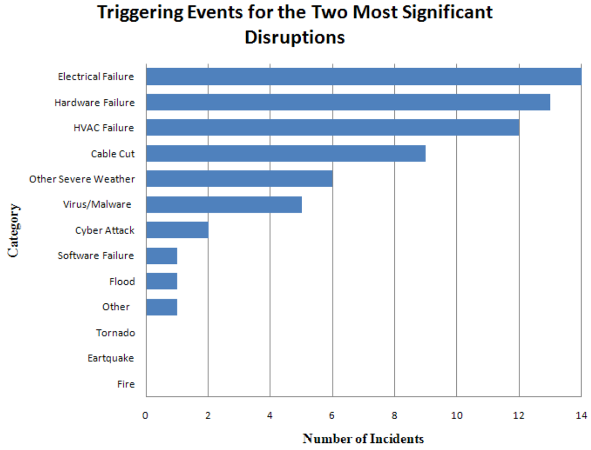


Figure 11 – Triggering Event for Most Significant Disruptions

In addition to identifying the triggering events for the two most significant disruptions, respondents were asked to identify the consequence or impact of the disruptions; Figure 12 provides a summary of the data. Respondents were not limited in the number of options they could choose; they could select all that apply. 75% of respondents said the impacts manifested themselves in the form of services being unavailable including the network itself, applications, E-mail or web sites. Damage to hardware, software, and data was only reported by 11% of the respondents.

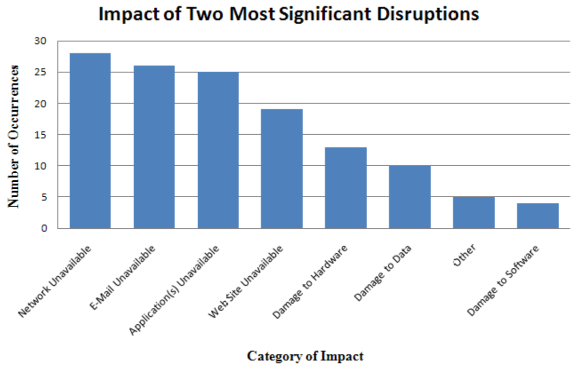


Figure 12 – Impact of the Two Most Significant Disruptions

## 3.5 Self Assessment

The final series of survey questions asked respondents to identify the importance of IT to the missions of other units on base, self-assess their unit's current ability to restore mission critical systems, compare that current ability with two years ago, and to rate their mission assurance readiness. Figure 13 summarizes the data respondents provided when asked to assess the importance of IT to the mission of other units on base. While 68% of units reported they either "disagreed" or "strongly disagreed" with the statement, "If IT services & electronic data were unavailable, other units on base would be able to carry out their mission critical tasks, 16% reported they either "agreed" or "strongly agreed." These units believe IT is not an essential part of other unit's mission capabilities. This is very hard to comprehend, given the proliferation of IT services today. What is perhaps more troubling is that 16% were neutral; one could draw the conclusion that these respondents have no concept of what, if any, services they provide are mission critical to other units.

The next questions asked respondents to self assess their unit by addressing the following question, "My unit is prepared to restore mission critical systems in the event of a disruption." Figure 14 shows that 61% of units "agreed" or "strongly agreed" that they could restore mission critical systems. That leaves 39% of respondents that were either neutral or believe they could not restore mission critical systems in the event of a disruption. Respondents were also asked to rate the level to which they agree with the following statement, "My unit is better prepared to restore mission critical systems today than it was 2 years ago." As shown in Figure 15, 69% agreed, to some level, with the statement. Another 25% disagreed, to some level; meaning they assess their ability to deal with adversity has gotten worse over a two-year period. Obviously more analysis than provided by this survey would be needed to get to the core of the responses, but they are somewhat shocking nonetheless. The remaining 16% of respondents reported their ability to restore services had neither gone up or down.

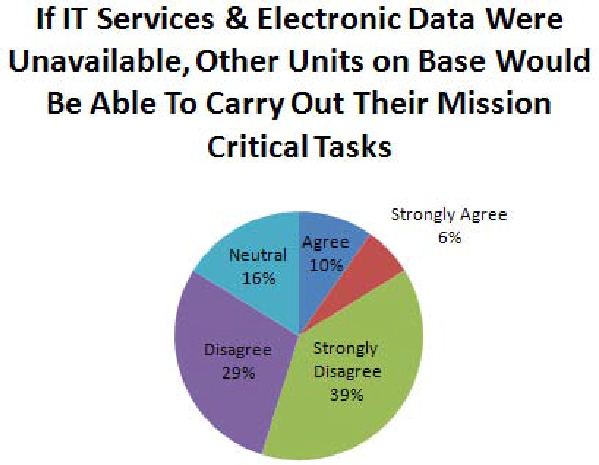


Figure 13 – Assessment of Mission Impact of Availability of IT Resources

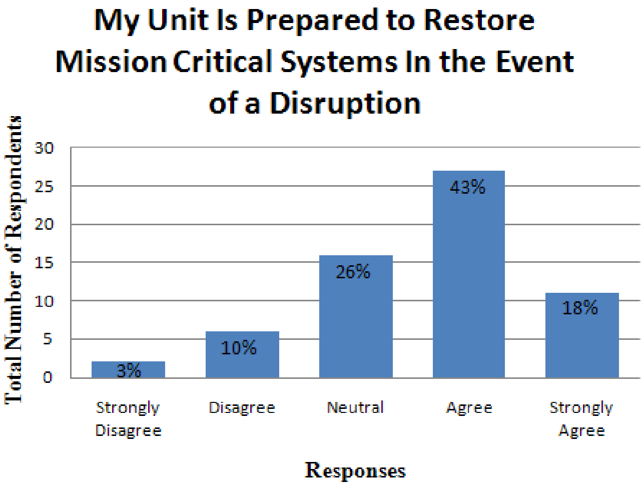


Figure 14 – Self-Assessment of Units Ability to Restore Mission Critical Systems

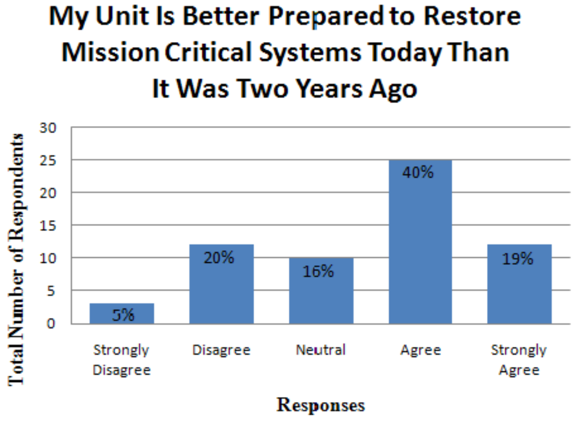


Figure 16 – Self-Assessment of Units Improvement in the Ability to Restore Mission Critical Systems

# 4 Conclusions

This research presented the findings from a survey on the cyber mission assurance culture within the US Air Force.

Analysis of the survey data indicates a desire, at base-level, for stronger leadership and better resources for conducting mission assurance efforts. It further illustrates that some very basic activities can be undertaken to increase the mission assurance capabilities at base-level, such as modernizing supporting infrastructure, documenting and testing current procedures, and developing a strong and consistent message with base leadership that illustrate the importance of mission assurance planning.

## 5 Acknowledgements

This work was supported by a research grant from the Air Force Research Laboratory (F4FBBA9067J001).

## 6 Disclaimer

The views expressed in this paper are those of the authors and do not reflect the official policy or position of the United States Air Force, the Department of Defense, or the U.S. Government.

## 7 References

[1] Grimaila, M.R. and Fortson, L.W., "Towards an Information Asset-Based Defensive Cyber Damage Assessment Process," Proc. of the 2007 IEEE Computational Intelligence for Security and Defense Applications (CISDA 2007); Honolulu, HI; April 1-5, 2007, pp. 206-212.

[2] Alberts, C.J. & Dorofee, A.J., "Mission Assurance Analysis Protocol (MAAP): Assessing Risk in Complex Environments," Carnegie Mellon University Networked Systems Survivability Program Report, 2005.

[3] Grimaila, M.R., Mills, R.F., and Fortson, L.W., "An Automated Information Asset Tracking Methodology to Enable Timely Cyber Incident Mission Impact Assessment," Proc. of the 2008 International Command and Control Research and Technology Symposium (ICCRTS 2008), Bellevue, WA, 17-19 June 2008.

[4] Grimaila, M.R., Fortson, L.W., and Sutton, J.L, "Design Considerations for a Cyber Incident Mission Impact Assessment (CIMIA) Process," Proc. of the 2009 International Conference on Security and Management (SAM09), Las Vegas, Nevada, July 13-16, 2009.

[5] Hale, B., "Mission Assurance: A Review of Continuity of Operations Guidance for Application to Cyber Incident Mission Impact Assessment (CIMIA)," Master's Thesis, Department of Systems Engineering and Management, Air Force Institute of Technology, June 2010.

[6] Anderson, E., Choobineh, J., Fazen, M., and Grimaila, M.R., "Mission Impact: Role of Protection of Information Systems," Proc. of the 2010 Intl. Conf. on Information Warfare and Security (ICIW 2010), WPAFB, OH, April 8-9, 2010.

[7] Grimaila, M.R., Fortson, L.W., Sutton, J.L, and Mills, R.F., "Developing Methods for Timely and Relevant Mission Impact Estimation," Proc. of the 2009 SPIE Defense, Security and Sensing Conference (SPIE DSS 2009), Orlando, Florida, April 13-17, 2009.

[8] Grimaila, M.R., Schechtman, G., and Mills, R.F., "Improving Cyber Incident Notification in Military Operations," Proc. of the 2009 Institute of Industrial Engineers Annual Conference, Miami, FL, May 30, 2009 - June 3, 2009.

[9] Donley, M.B., "Problems of Defense Organization and Management," Joint Forces Quarterly (JFQ), Summer 1995.

[10] Evans, Mickey, "An Informational Analysis and Communications Squadron Survey of Cyberspace Mission Assurance," Master's Thesis, Air Force Institute of Technology, AFIT/IDE/ENV/10-J01, June 2010.

[11] NETWORK OPERATIONS (NETOPS). Washington DC : U.S. Air Force, 2006. AFI 33-115V1, http://www.af.mil/shared/media/epubs/AFI33-115V1.pdf.

# SESSION

# DISCUSSION SESSION/TRACK

## Chair(s)

### TBA

# Impact of Security Measures and Algorithms on Operations at Transportation Facilities

Marguerite Zarrillo and Elia El Lazkani

*Abstract*-- Central to the successful deployment of Electronic Transportation Payment Systems, ETPS, is a sound understanding and systematic evaluation of deployment system vulnerabilities to security and privacy breaches. The expectation is that such an evaluation will lead to new fundamental knowledge and further acceptance by the public of such payment systems. Public acceptance has been determined as a major factor associated with the success of electronic payment systems. Provided that privacy and security objectives are achieved, these payment systems will in turn help generate increased revenue to finance transportation infrastructure improvements using innovative pricing schemes such as congestion pricing, carbon based fees, and vehicle mile and weight based charges. Questions of interest in this evaluation study include the extent of the breach occurrence in ETPS and the impact on traffic flow operations whenever security protective measures are adopted.

## I. Introduction

The application domain of cryptographic protocols for Electronic Transportation Payment Systems, ETPS, requires secure, private, and trusted transactions that also need to be efficient, low-cost, usable and reliable. ETPS includes payment for trains, subways, buses, and ferries as well as toll collection for roads, border crossings, bridges, tunnels and payment for parking and recharging of electric cars. Deploying electronic payment systems in transportation offers important economic benefits for revitalizing our aging transportation infrastructure and fostering new green technologies.

Next-generation payment systems for transportation should support seamless travel by integrating various forms of transportation as well as parking and small scale commerce. Moreover, information gathered from these systems can facilitate advanced traveler information, traffic management, travel time estimation, emergency management, congestion pricing, and emissions control. Mileage-based fees, recently proposed in several U.S. states, will not be possible without such systems. Hence, ETPS will become an even more important component for financing critical transportation infrastructure with security and privacy issues that must be balanced with cost, efficiency and reliability.

Historically, toll collection systems have been designed as one of the primary mechanisms of highway financing. Tolls are direct-user charges collected to fund construction and maintenance of highway roads, bridges, and tunnels. Over time, toll collection evolved from manual toll collection systems to electronic toll collection, ETC, systems. Today, the most widely used technology is the dedicated short range communication, DSRC. Most recently, new Global Positioning System, GPS, technologies are emerging as capable technologies of the future. Compared to ETC, conventional manual collection significantly lowers speeds at toll plazas, thus limiting highway capacity and creating an environmentally poor atmosphere. ETC technologies allow drivers to make electronic payments from prepaid toll accounts. As a vehicle approaches the toll plaza, an electronic transponder installed in the vehicle communicates information to the receiving antenna at the toll plaza. Although these technologies are promising, they also bring new problems connected with security and privacy.

ETPS projects such as the CharlieCard and MetroCard for transit or ETC such as E-Pass and SunPass in Florida and the E-ZPass and FastLane in the northeast lack sufficient mechanisms to protect their security and the privacy of their users. Future ETPS, aiming towards unified modes of transportation and mileage-based fees (which require a fine-grained tracking of users) further increase concerns about privacy and security. Privacy is an especially challenging problem in this context since it spans cryptographic theory, engineering, policy and sociology. Although there is a wealth of cryptographic literature proposing secure and anonymous payment schemes, the application to the transportation domain differs significantly in terms of both security requirements and engineering constraints. Furthermore, the extent of the problem worldwide has not been determined, nor has there been much investigation into the impact on traffic operations whenever security measures are adopted.

## II. Breach Event Survey

This research study had a two-fold focus. The first objective was to systematically investigate the extent of the existing breach problem by developing an interactive website that produced a survey of reported breaches across the globe. With a secure login, researches at

Marguerite Zarrillo, Corresponding Author, is a Professor of Physics within the College of Engineering at the University of Massachusetts Dartmouth, N Dartmouth, MA 02747, USA (phone: 508-999-9268; fax: 508-999-9115; email: ~~mzarrillo@umassd.edu~~).

Elia El Lazkani is a graduate student with the Department of Electrical and Computer Engineering, University of Massachusetts Dartmouth.

DOI: 9790

selected universities and selected tolling and transit agencies are encouraged to upload breach events and news articles. Users are asked to submit basic information about the breach such as location, date, severity, and whether the breach was transit or toll related.

Contributors to the survey identify the type of breach reported by selecting from five defined categories of security breaches and five categories of privacy breaches. Security is defined as actions that need to be taken to protect any system from violation. The five types of security breaches are *Confidentiality*, *Integrity*, *Availability*, *Authenticity*, and *Nonrepudiation*. [1] Privacy is defined as the right of individuals to have control over when, how, to what extent, and for what purpose their personal information is being disclosed. [2] The five key attributes of privacy that need to be protected to ensure individual's information is kept private include *Anonymity*, *Pseudonymity*, *Unlinkability*, *Identity Management*, and *Unobservability*. [3]

The results of this survey are continuously monitored and statistically analyzed. The initial survey of 180 news events indicates that the *Integrity* type is the most frequently occurring security breach, and the *Unlinkability* type is the most often occurring privacy breach. There also seems to have been an across the board surge in reported breaches in 2008.

## III. Case Study Simulation

A second goal has been to investigate the value and effectiveness of traffic engineering simulation tools for this application. Specifically the question asked whether these tools could be used to determine the impact on traffic operations when security measures are implemented.

The case study involved simulation of traffic operations within the ticketing toll plaza off the Massachusetts turnpike, I-90 exit 11A, in Westborough, using two widely adopted traffic simulation tools, S-Paramics from SIAS Limited and VISSIM from PTV America. [4,5] The plaza was simulated for existing peak hour traffic without security measures as well as various scenarios of assumed transaction delay after security cryptographic measures were deployed. In order to reflect a realistic traffic flow simulation, vehicles were divided into four categories, each with customized characteristics. These included the *Truck* category and *Passenger Car* category, each of which were subdivided into, *vehicles making cash manual payments* and *vehicles paying electronically*. The percentage of all approaching vehicles to the plaza that belonged to each category was provided by the Massachusetts Turnpike Authority. The input to VISSIM included these percents as well as typical vehicle headways or seconds between vehicles (bumper to bumper); all based on measured values at the plaza. [6] In the simulated environment, drivers are proficient; they know the road network and always drive towards their destination. Therefore the modeled effect of security protocols on traffic operations may be underestimated. However, a legitimate comparison can be made of traffic operations occurring at the plaza as a function of increasing added transaction time.

Preliminary results using VISSIM indicate that the impact on operations at the Westborough plaza due to additional transaction times for toll collection is negligible for protection measures that add milliseconds of transaction time. However, for added transaction times in a range of seconds, there is a more significant impact. In addition, a threshold value, between 5 and 10 seconds of added transaction time, results in a sudden decrease in the operational performance at the plaza. A sudden breakdown in traffic flow occurs.

Finally, simulations of two other plazas on the Turnpike indicate that plaza buffer zone geometry plays a critical role in the actual breakdown threshold value of added transaction time. In other words, the sudden decrease in operational performance may be a result of buffer zones becoming jammed which causes electronic paying vehicles to lose access to lanes dedicated to ETC toll collection. By obstructing the electronic paying vehicles from paying their toll, these dedicated lanes are underutilized and the plaza capacity is reduced. In addition, ETC paying vehicles are now processed at a reduced rate comparable to that of the slow cash paying rate. Thus, whenever security protocols result in additional payment transaction time delay beyond a few seconds, it is recommended that plazas are designed with larger buffer zones to reduce the risk of jamming.

## References

[1] W. Stallings, *Cryptography and network security,* 4th ed. Prentice Hall, 2005, pp. 12 - 21.

[2] A. Tsohou, C. Lambrinoudakis, S. Kokolakis and S. Gritzalis, "The Importance of Context-Dependent Privacy Requirements and Perceptions to the Design of Privacy-Aware Systems," *UPGRADE: The European Journal for the Informatics Professional*, vol. 11, pp. 32-37, 2010.

[3] Andreas Pfitzmann and Marit Hansen, "A terminology for talking about privacy by data minimization: Anony mity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management," TU Dresden Privacy and Data Security website, vol. 0.34, 2010. http://dud.inf.tu-dresden.de/Anon_Terminology.shtml

[4] PTV America, http://www.ptvamerica.com/

[5] SIAS Limited, http://www.sias.com/ng/home

[6] Zarrillo, Marguerite L. and A. Essam Radwan, "Methodology SHAKER and the Capacity Analysis of 5 Toll Plazas," ASCE *Journal of Transportation Engineering*, vol. 135, Issue 3, March 2009, pp 83-93. http://tris.trb.org/view.aspx? id=885585

## Acknowledgement

This work was supported in part by the National Science Foundation under grant no. 0964292.

# PortSentry for Special Occasions

**L. Markowsky**
School of Computing and Information Science, University of Maine, Orono, ME USA

**Abstract –** *Using PortSentry with an iptables/netfilter firewall is a well-known means of implementing an active firewall that can provide a robust defense against many broad-spectrum port scans. More targeted scans, however, are more difficult to prevent: in practice, certain services must be available, and although some may be restricted to authorized users, others, such as web and mail servers, are usually unrestricted. This poster describes two novel uses of PortSentry to enhance security, providing an efficient, active, enumeration-resistant layer that permits access to both restricted and unrestricted services while protecting those services from targeted enumeration.*

**Keywords:** PortSentry, active intrusion detection system, firewall, iptables, enumeration, port scanning

## 1 The role of port scanning and enumeration in a cyberattack

Attackers often scan IP address blocks and then enumerate target machines before carrying out an attack. During the port-scanning phase, the attacker might use a tool such as nmap to identify potential target hosts in a specified IP address block. Then, during the enumeration phase, the attacker might use a tool such as Nessus to collect more detailed information about a particular target system's vulnerabilities.

PortSentry (in its default configuration) paired with a firewall provides an active intrusion detection system that can block attackers who attempt to scan and enumerate specific ports not used by the system to provide services. In addition, Linux firewalls can themselves provide passive defense from many broad-spectrum TCP scans, including Null, Xmas, ACK, and FIN scans, using simple netfilter/iptables rules. These measures, while extremely useful, still leave the restricted and unrestricted services vulnerable to targeted enumeration by attackers. Methods to detect targeted enumeration and take preventative action are essential to protect hosts from would-be attackers in the earliest phases of an attack.

## 2 PortSentry and prerouting: booby-trapping restricted services

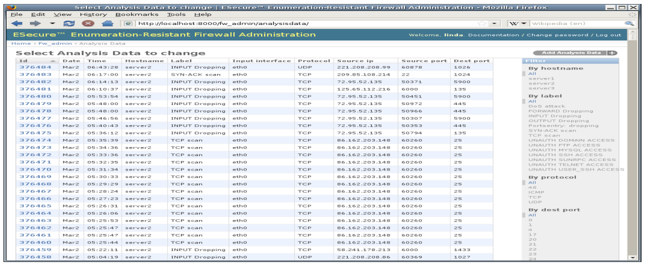
Extensive use of prerouting with PortSentry and configurable classes of users can effectively booby-trap services, such as ssh, sftp, POP mail, and SQL servers, that might be restricted to a group of authorized users. For these users, the available service will be provided, but for all others, requests to the service's port will be redirected to another port running PortSentry. In effect, the booby-trapped ports simultaneously run two services: the expected service for authorized users, and PortSentry for all others, including attackers who attempt to enumerate the service running on that port.

## 3 SmartSentry: using machine learning to detect targeted enumeration of unrestricted services

Unrestricted services , such as a web server or mail server, cannot be protected from targeted enumeration using prerouting based on source IP addresses. To protect these services, a Support Vector Machine (SVM) can learn to detect enumeration of the unrestricted services. Once the SVM is trained, the system will invoke PortSentry when it detects targeted enumeration and will add an iptables rule and/or an entry in /etc/hosts.deny to actively refuse all subsequent connection attempts from the attacker. The SVM is designed to run in real time after initial training.

## 4 Implementation and results

A prototype of the method described in Section 2 was designed, implemented, and tested against a second system specifically configured for comparison. A partial listing is shown in Figure 3. To facilitate analysis, an administrative interface was implemented using Django and graphical reports were generated using ReportLab (Figures 1 and 2).



Select Analysis Data to change

| Id | Date | Time | Hostname | Label | Input interface | Protocol | Source ip | Source port | Dest port |
|---|---|---|---|---|---|---|---|---|---|
| 376484 | Mar2 | 06:43:28 | server2 | INPUT Dropping | eth0 | UDP | 221.208.208.99 | 60878 | 1026 |
| 376483 | Mar2 | 06:17:00 | server2 | SYN-ACK scan | eth0 | TCP | 209.85.108.214 | 22 | 1024 |
| 376482 | Mar2 | 06:14:13 | server2 | INPUT Dropping | eth0 | TCP | 72.95.52.135 | 50371 | 5900 |
| 376481 | Mar2 | 06:10:37 | server2 | INPUT Dropping | eth0 | TCP | 125.65.112.216 | 6000 | 135 |
| 376480 | Mar2 | 05:53:54 | server2 | INPUT Dropping | eth0 | TCP | 72.95.52.135 | 50451 | 5900 |
| 376479 | Mar2 | 05:48:00 | server2 | INPUT Dropping | eth0 | TCP | 72.95.52.135 | 50972 | 445 |
| 376478 | Mar2 | 05:48:00 | server2 | INPUT Dropping | eth0 | TCP | 72.95.52.135 | 50966 | 445 |
| 376477 | Mar2 | 05:46:56 | server2 | INPUT Dropping | eth0 | TCP | 72.95.52.135 | 50307 | 5900 |
| 376476 | Mar2 | 05:40:43 | server2 | INPUT Dropping | eth0 | TCP | 72.95.52.135 | 50353 | 445 |
| 376475 | Mar2 | 05:36:12 | server2 | INPUT Dropping | eth0 | TCP | 72.95.52.135 | 50794 | 135 |
| 376474 | Mar2 | 05:35:39 | server2 | TCP scan | eth0 | TCP | 86.162.203.148 | 60260 | 25 |
| 376473 | Mar2 | 05:34:36 | server2 | TCP scan | eth0 | TCP | 86.162.203.148 | 60260 | 25 |
| 376472 | Mar2 | 05:33:36 | server2 | TCP scan | eth0 | TCP | 86.162.203.148 | 60260 | 25 |
| 376471 | Mar2 | 05:32:35 | server2 | TCP scan | eth0 | TCP | 86.162.203.148 | 60260 | 25 |
| 376470 | Mar2 | 05:31:34 | server2 | TCP scan | eth0 | TCP | 86.162.203.148 | 60260 | 25 |
| 376469 | Mar2 | 05:30:33 | server2 | TCP scan | eth0 | TCP | 86.162.203.148 | 60260 | 25 |
| 376468 | Mar2 | 05:29:29 | server2 | TCP scan | eth0 | TCP | 86.162.203.148 | 60260 | 25 |
| 376467 | Mar2 | 05:28:24 | server2 | TCP scan | eth0 | TCP | 86.162.203.148 | 60260 | 25 |
| 376466 | Mar2 | 05:27:23 | server2 | TCP scan | eth0 | TCP | 86.162.203.148 | 60260 | 25 |
| 376465 | Mar2 | 05:26:31 | server2 | TCP scan | eth0 | TCP | 86.162.203.148 | 60260 | 25 |
| 376464 | Mar2 | 05:26:06 | server2 | TCP scan | eth0 | TCP | 86.162.203.148 | 60260 | 25 |
| 376463 | Mar2 | 05:25:53 | server2 | TCP scan | eth0 | TCP | 86.162.203.148 | 60260 | 25 |
| 376462 | Mar2 | 05:25:47 | server2 | TCP scan | eth0 | TCP | 86.162.203.148 | 60260 | 25 |
| 376461 | Mar2 | 05:25:47 | server2 | TCP scan | eth0 | TCP | 86.162.203.148 | 60260 | 25 |
| 376460 | Mar2 | 05:25:44 | server2 | TCP scan | eth0 | TCP | 86.162.203.148 | 60260 | 25 |
| 376459 | Mar2 | 05:22:11 | server2 | INPUT Dropping | eth0 | TCP | 58.241.178.213 | 6000 | 1433 |
| 376458 | Mar2 | 05:04:19 | server2 | INPUT Dropping | eth0 | UDP | 221.208.208.86 | 60369 | 1027 |

Figure 1: Administrative Interface

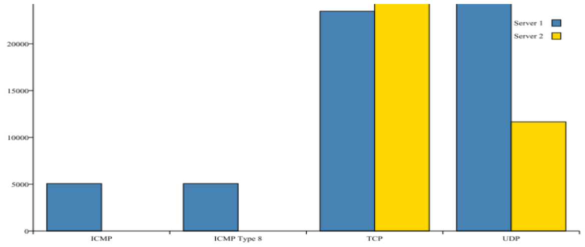


Figure 2: Graphical Reports

```
### Define chains

# Declare user-defined chains (nat table).
# Similar rules (omitted) define these chains for the filter table.
/sbin/iptables -t nat -N admin-nat
/sbin/iptables -t nat -N user-nat

# Administrator IP addresses (nat table)
/sbin/iptables -t nat -A admin-nat -s $ADMIN1 -d $SERVER -j ACCEPT
/sbin/iptables -t nat -A admin-nat -s $ADMIN2 -d $SERVER -j ACCEPT

# User IP addresses (nat table)
/sbin/iptables -t nat -A user-nat -s $USER1 -d $SERVER -j ACCEPT
/sbin/iptables -t nat -A user-nat -s $USER2 -d $SERVER -j ACCEPT

### PREROUTING chain

# Restrict the ssh daemon to administrators and authorized users.
# Attempted access by others is regarded as hostile--redirect to portsentry.
/sbin/iptables -t nat -A PREROUTING -p tcp --dport ssh -j admin-nat
/sbin/iptables -t nat -A PREROUTING -p tcp --dport ssh -j user-nat
/sbin/iptables -t nat -A PREROUTING -p tcp --dport ssh -j REDIRECT --to-ports $PORT_SEN_SSH

# Restrict the MySQL daemon (tcp) to administrators.
# Attempted access by others is regarded as hostile--redirect to portsentry.
/sbin/iptables -t nat -A PREROUTING -p tcp --dport mysql -j admin-nat
/sbin/iptables -t nat -A PREROUTING -p tcp --dport mysql -j REDIRECT –to-ports $PORT_SEN_MYSQL

### INPUT chain

# ADMINISTRATOR-IPS ACCEPT CHAIN
/sbin/iptables -t filter -A INPUT -p tcp --dport ssh -j admin-input
/sbin/iptables -t filter -A INPUT -p tcp --dport mysql -j admin-input

# USER-IPS ACCEPT CHAIN
/sbin/iptables -t filter -A INPUT -p tcp --dport ssh -j user-input
```

Figure 3: Using PortSentry and Prerouting to Simultaneously Run
Two Services on Each of Two Ports: 22 (ssh) and 3306 (mysql)

Tests of the prototype show that the method described in Section 2 provides an effective, lightweight defense against scanning and enumeration. All TCP connection attempts of the protected services activated PortSentry. The only means of defeating the method was to scan or enumerate the system from an IP address protected by the firewall or the PortSentry configuration file. In addition, a version of this method was used by the author in an in-class cyberwar at the University of Maine. The method was effective: none of the other students were able to enumerate the author's server during the cyberwar.

## 5 Future work

Additional testing and analysis using large datasets is needed to confirm the results determined for the method described in Section 2. Implementation of the SVM described in Section 3 is in progress.

## 6 Acknowledgments

The author thanks Dr. James Fastook, Dr. Phil Dickens, Dr. Bruce Segee, and Dr. George Markowsky of the University of Maine and Dr. Danny Kopec of the CUNY (City University of New York) Graduate Center and Brooklyn College for their advice and support and Trefoil Corporation of Orono, Maine, for donating and hosting two test servers to collect data for this project.

## 7 Selected references

[1] O. Andreasson, *Iptables Tutorial 1.2.2*, iptables-tutorial.frozentux.net/iptables-tutorial.html, iptables-tutorial.frozentux.net/scripts/rc.firewall.txt.

[2] D. Bandel. “Taming the Wild Netfilter,” *Linux Journal*, Sep. 2001, pp 64-72, www.linuxjournal.com/article/4815.

[3] H. Burgiss, *Security Quick-Start HOWTO, Step 3: Firewalls and Setting Access Policies*, tldp.org/HOWTO/Security-Quickstart-HOWTO/firewalls.html.

[4] I. Dubrawsky, *PortSentry for Attack Detection, Parts One and Two*, http://www.securityfocus.com/infocus/1580, http://www.securityfocus.com/infocus/1586.

# A Study on Secure User Authentication Protocols in Smartwork Systems

**Yu-jong Jang and Jin Kwak**
[1]Def, of Information Security Engineering, Soonchunhyang University, Korea

**Abstract**—*As the demand for connecting user-authenticated Smartwork systems to open networks increases, the study of Smartwork system security has recently become an issue. Many researchers have proposed user authentication schemes for Smartwork systems. However, previous studies lacked the proper considerations for availability. In this paper, we present our development of cryptographic security requirements for Smartwork systems. We also propose an authentication scheme for these types of systems.*

**Keywords:** Smartwork, Authentication, ID-based, Protocols

## 1 Introduction

With the rapid development of Internet technology, we are now able to access any service from any place and at any time; this means that we are able to work from home or from the office. The systems used to make these connections are commonly called Smartwork systems. With user authentication of these systems, the study of system security has become an issue because Smartwork Service Connection are done in a varety of environments. This means that Smartwork systems are often exposed to attacks.

The Internet and its accessory services are convenient. However, this convenience comes with a threat to security. Additionally, Smartwork system providers should have a secure identification system to ensure that only legal users have access.

In order to prevent issues in this regard, several professional organizations have been researching the security of Smartwork systems and have been developing standards and reports. We will provide a brief overview of this research.

This thesis studies various approaches to enhance user authentication by focusing on protecting private data and guaranteeing user anonymity.

## 2 Related Work

### 2.1 Smartwork

Smartwork. It is quite common for a person to work at home, in front of his/her personal computer or mobile device; this person does not need to be physically present in an office. Smartwork is a far-reaching concept and a powerful tool that has given birth to new working conditions, organizational cultures, and social environments, resulting in huge benefits for all actors involved. Companies familiar with this non-traditional work concept usually enjoy a notable increase in productivity accompanied by a reduction in operating costs due to employee absences [1]. Various studies and surveys have been conducted to prove that Smartworkers benefit from “independence” in managing their work time and tasks and a reduction in the transportation costs involved in commuting.

### 2.2 Security Requirements

Smartwork systems have heterogeneous wired and wireless networks and a variety of protocols. Thus, Smartwork systems require more security than general systems. In this section, we review security requirements based on user authentication [2].

| Security Requirements | Description |
|---|---|
| Confidentiality | The data transmitted between nodes (Mobile-Office, Tele-Office, Office) should be protected by encryption. Furthermore, a secure user authentication is essential for connection. |
| Integrity | Smartwork systems should ensure the integrity of the transmitted message. |
| Availability | In Smartwork systems, availability is more important than confidentiality, because if the Smartwork system is down, important information may be inaccessible. Therefore, the Smartwork systems should be designed to always be on. |

This research was supported by a Basic Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (No.2011-0007755).

You-jnag Jang is with the Department of Information Security Engineering, Soonchunhyang University, Korea (e-mail: yjjang@sch.ac.kr).

Jin Kwak, is with Department of Information Security Engineering, Soonchunhyang University., Korea, (corresponding author to provide e-mail: jkwak@sch.ac.kr)

# 3 User Authentication Protocols

In this section, we propose a mobile user authentication protocol for secure and appropriate Smartwork system communication. Most Smartwork systems require message broadcasting and secure communications. As we described in the previous section, although the user authentication schemes for Smartwork systems provide secure unicast communications, these schemes cannot, at the same time, support multiple users or multiple environments. Therefore, we developed a user authentication protocol to support improved security [3].

Notations and definitions: The following notations are used in this paper:

- *: Each object (TO: Smartwork Office, TS: Smartwork Server)
- $R_U$: number of random users
- $R_S$: number of random servers
- ID*: Each ID
- T: Timestamp
- h(): Hash function
- H, S, N: Secure parameters

## 3.1 Registration Process

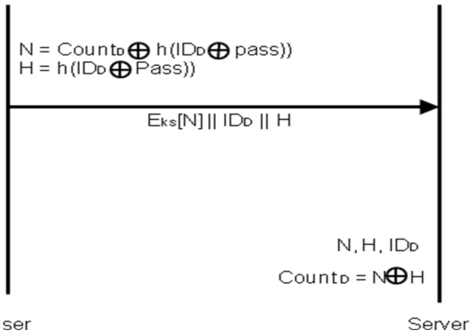


Figure 2. Registration Process

In the initial counter value step in the registration device and the ID of the device during initial setup, enter the password that is generated by the value of N. The value of N after the security parameters with the timestamp value will be used to authenticate the device.

The server user id, counter value, and the security parameter N value are stored. ID, and then check that the counter value N is used for the identification captured.

For each object value, store:

User: N, H, $ID_D$, $Count_D$
Server: N, H, $ID_D$

## 3.2 R Authentication Process

After the registration process, the steps listed in the user ID of the Server and N values, the counter values are stored. The security parameters are stored by the user through the N S value to create the security parameters and the ID, and the timestamp is sent with a value. This way, devices do not expose confidential information. Server confidentiality should also counter the existing value and the timestamp that is stored. A comparison of the values indicates that this system can be effective against man-in-the-middle attacks. In addition, because of the increased values of the counter-certified devices authenticated through the smart meter, mutual authentication takes place.

Comparison of Values

$$H(S`) = ?\ H(s)$$

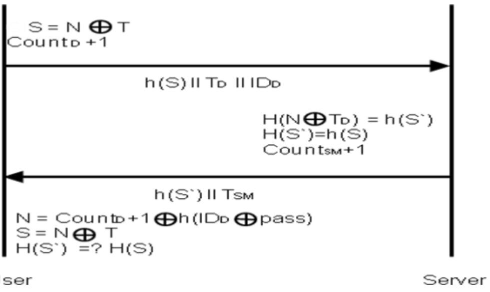


Figure 3. User Authentication

# 4 CONCLUSION

With the rapid development of Internet technology, we are now able to access any service from anywhere and at any time. As a result of this capability, workers can enjoy the popularity and convenience of Smartwork systems. However, Smartwork systems can be vulnerable to a variety of attacks—if such systems are attacked by unauthorized users, there can be devastating consequences. To prevent this type of damage, professional organizations have conducted research on the security of Smartwork systems; however, many security problems remain. In this paper, we suggest that user authentication schemes for Smartwork systems are among the security problems.

We proposed an improved user authentication scheme for secure Smartwork communications.

# 5 References

[1] Karen Scarfone, Paul Hoffman, and Murugiah Souppaya, “Guide to Enterprise Telework and Remote Access Security”, NIST SP 800-46 Rev 1, 2009

[2] Office of Information Security, “Telework and Remote Access Security Standard”, 2010

[3] Kiyomoto S, Tanaka T (2008) Anonymous Attribute Authentication Scheme Using Self-Blindable Certificates. In: IEEE International Conference Intelligence and Security Informatics, pp 215-217

# A Facial Recognition System based on User Authentication for Mobile Office Environments

**Yun-Sang Byun [1], SungYong Ryu[2], and Jin Kwak [1]**
[1]Dpt. of Information Security Engineering, Soonchunhyang University., Korea
[2`]Dept. Of Business Administration, Soonchunhyang University, Korea

**Abstract** - *Mobile office workspaces provide flexible and convenient work environments. Many countries have already introduced such mobile office environments.Most users of mobile offices have remote access to confidential data. With remote access from outside comes the possibility that unauthorized users might gain access to and leak that confidential information. In this paper, we propose a user-based facial recognition authentication scheme for mobile office environments*

**Keywords:** Mobile Office, User Authentication, Facial Recognition System

## 1 Introduction

With the rapid development of the Internet and electronic commerce technology, many services are conveniently provided via the Internet, such as distributed electronic medical records systems and mobile office workspaces.

Mobile office workspaces are non-traditional work environments without time or space constraints. However, mobile offices have remote access to internal data from the outside. With remote access from the outside comes the possibility that unauthorized users might gain access to the personal information of employees and leak confidential information. In this paper, we propose a mobile office environment using a smart device camera and a facial recognition authentication scheme..

## 2 Related work

Please use the styles contained in this document for: Title, Abstract, Keywords, Heading 1, Heading 2, Body Text, Equations, References, Figures, and Captions.

*Do not add any page numbers and do not use footers and headers (it is ok to have footnotes).*

### 2.1 Mobile office

Mobile offices are future-oriented work environments. Also Mobile office use the smart phone, tablet PC, laptop, etc. use the work environment that can be processed. Home offices have a similar form. TABLE I shows a comparison of a mobile and home office.

TABLE I. Comparison of a mobile and home office

| Division | Contents |
|---|---|
| Mobile office | - Mobile phone with Internet access, wireless Internet service<br>- Company operations in real time will enable quick processing |
| Home office | - Work at home<br>- No rush hour |

### 2.2 User authentication

User authentication is a means of identifying users and verifying they are allowed to access a restricted service. User authentication is necessary to ensure the IT system is secure. User authentication is an integral part of IT security.

Current user authentication technology consists of ID/password (PW), certificates, biometrics, etc. These can be classified as shown in TABLE II

TABLE II. User authentication technology

| Kind | Contents |
|---|---|
| ID/PW | - Using the Registered by the user with a unique ID/PW |
| Certificates | - Using the public key certificate, an electronic document used in cryptography |
| Biometrics | - The identification of humans by their characteristics or traits |

**Fig. 1. User authentication technology**

## 2.3 Facial recognition

Facial recognition is a type of biometrics that uses images of a person's face for recognition and identification purposes. It involves comparing an image or video of a person to one that is in a database. It does this by comparing the structure, shape, and proportions of the face; the distance between the eyes, nose, mouth, and jaw; the upper outlines of the eye sockets; the sides of the mouth; the location of the nose and eyes; and the area surrounding the cheek bones

# 3 Proposed scheme

In this paper, we propose a user authentication scheme for mobile office workspaces to connect to company networks via facial recognition technology-based password authentication. Therefore, the need for additional equipment is eliminated, resulting in cost savings.

The proposed scheme involves having smart device users download and run the application once they are connected to the company server. They then use a camera to capture an image of their face. In the next step, a 5 x 5 matrix is applied. The matrix is partitioned by applying user you select an image, and select your desired location along with the image moves.

### Registration steps

① The user's Information is sent to the server and legitimate users are authenticated by the server.

② To use the application, users employ the smart device cameras mounted on the front.

③ After registering, a 5 x 5 matrix is applied and the user selects the hoped images split and chooses where to move.

④ The images chosen to be moved to the desired location on the server are registered.

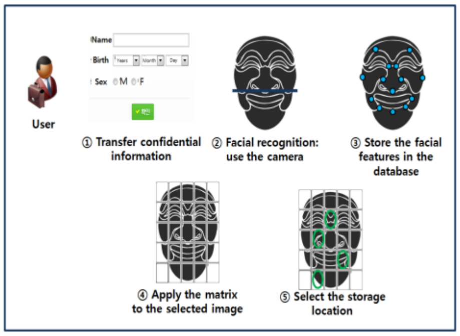

User
Name
Birth Years Month Day
Sex M F
확인
① Transfer confidential information
② Facial recognition: use the camera
③ Store the facial features in the database
④ Apply the matrix to the selected image
⑤ Select the storage location

**Fig. 2. Registration steps**

### Authentication steps

① User applications when connecting to companies server; users will require an ID and facial recognition.

② To take advantage of the application, users employ the smart device cameras mounted on the front.

③ The user's image is sent to the company's server and is compared to an image stored in the database. It is then authenticated.

④ Be rearranged randomly an image select to be registered in the output image and find the secret value to be move the location set.

⑤ The secret value, image number, and device serial number are transmitted to the authentication server.

⑥ The authentication server compares the data stored on the server to that provided by the user.

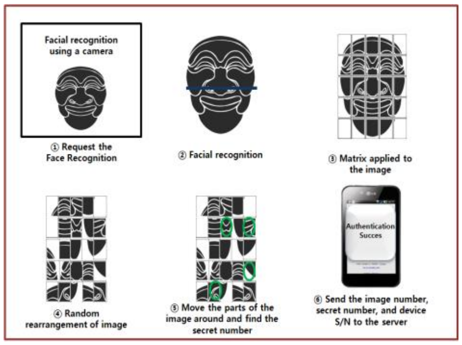

Facial recognition using a camera
① Request the Face Recognition
② Facial recognition
③ Matrix applied to the image
④ Random rearrangement of image
⑤ Move the parts of the image around and find the secret number
Authentication Succes
⑥ Send the image number, secret number, and device S/N to the server

**Fig. 3. Authentication Steps**

# 4 Efficacy and safety analysis

The proposed user authentication system is compared to the current authentication method in TABLE III.

**TABLE III. Efficacy and safety analysis**

| Division | Existing method | Proposed method |
|---|---|---|
| Recognition device | - Needs an additional reader device | - Does not need an additional reader device |
| Non-authentication user access | - Accessed to data using Picture | - Unaccessed to data using Picture |
| Used data user tracking | - Impossible to determine which users access data | - Possible to determine which users access data |

# 5 Conclusion

In this paper, we proposed a user authentication system for remote work environments that uses smart devices to distinguish users. We believe this system is both safe and efficient and expect it will be applied in the future.

References

[1] D. Maio, D. Maltoni, R. Cappelli, J.L. Wayman, and A.K. Jain, "FVC2004: Third Fingerprint Verification Competition," *Proc. Int. Conf. on Biometric Authentication (ICBA)*, Hong Kong, July 2004, pp.1–7.

[2] Guodong Guo, Stan Z. Li, and Kapluk Chan "Face Recognition by Support Vector Machines," *Image and Vision Computing*, vol.19, pp.631–638, 2001.

# Robust ECC-based three factor user authentication preserving biometric privacy

Tien-Ho Chen[1],Hsiu-Lien Yeh[2], Kuei-Jung Hu[3], Wei-Kuan Shih[4]
[1,3,4] Department of Computer Science, National Tsing Hua University, Taiwan
[2]Institute of Information System and Applications, National Tsing Hua University, Taiwan
d918325@oz.nthu.edu.tw,s9865805@m98.nthu.edu.tw,s9962629@m99.nthu.edu.tw,wshih@cs.nthu.edu.tw

***Abstract*— Recently, to achieve privacy protection using biometrics, Fan-Lin proposed a three-factor authentication scheme based on password, smart card and biometrics. However, we have found that Fan-Lin's proposed scheme has flaws in the design of biometrics privacy, fails to maintain a verification table, making it vulnerable to stolen-verifier attack and modification attack, and is vulnerable to insider attacks. Thus, we propose an ECC-based authentication scheme that is improved with regard to security requirements. Our proposed scheme overcomes the flaws of Fan-Lin's scheme and is secured from attacks. Furthermore, we present a security analysis of our scheme to show that ours is suitable for the biometric systems.**

***Keywords- Security; ECC; Biometrics; Authentication***

## I. INTRODUCTION

With the current advance of network services, proper user identification for remote user authentication over insecure communication channels is increasingly essential. Contrary to traditional password-based remote user authentication, biometrics-based authentication has greater security and is more reliable for remote user authentication [1]. In addition, some three-factor authentication schemes have been proposed in many publications [2-6]. Biometrics-based authentication systems are increasingly common for remote user identity authentication schemes. Due to its physiological or behavioral characteristics, remote authentication schemes can provide enhanced security using such techniques as fingerprint verification, iris analysis, facial analysis, and keystroke analysis [1,7].

Recently, Lee et al. [2] proposed a remote user authentication scheme based on smart card and fingerprint without a verification table to maintain records. In [3], we found that Lee et al.'s scheme is vulnerable to the masquerade attacks and replay attacks, and [4,5] showed that Lin-Lai's scheme is vulnerable to the server spoofing attack and does not provide proper mutual authentication. However, Li et al. [6] point out that Li and Hwang's scheme [5] fails to provide proper mutual authentication and is vulnerable to man-in-the-middle attacks. Unfortunately, the Li et al.'s scheme fails to securely update the new password and is also insecure.

The above mentioned schemes consider privacy protection using biometrics on the user's side without considering biometric characteristics on the server's side. For privacy protection using biometrics, the biometric data and settings have to be considered. Some methods, such as those based on error-correcting codes and fuzzy encryption, use biometric data to key encrypt a secret and then match the biometric template after extracting the secret. In Fan and Lin's scheme [7], user data is only stored on the user's side while still permitting the server to perform the authentication. Despite the benefits of Fan and Lin's scheme, it is still subject to privacy and security threats. It is obvious that Fan-Lin et al.'s schemes need to maintain a verification table in order to provide protection from inside attacks. In this paper, our authentication scheme employs a different approach. We improve the Fan et al.'s scheme and enhance the security and privacy protection. This leads to a robust three-factor remote authentication protocol based on the Elliptic Curves Cryptosystem (ECC).

The remainder of this paper is organized as follows. In section 2, we analyze the Fan and Lin′s scheme. In section 3, we propose a robust three factor biometric-based authentication scheme with ECC. Then, in section 4, we provide the security analysis and comparisons. Finally, we present some concluding remarks in section 5.

## II. CRYPTANALYSIS OF FAN-LIN'S SCHEME

In this section, we have analyzed the security flaws of Fan et al.'s scheme. First, we summarize the notations used throughout this paper as follows.

- $U_i$: The *i*th user
- $ID_i$: The identity of the user $U_i$
- $PW_i$: The password of $U_i$
- $h(\cdot)$: A public one-way hash function
- $||$: String concatenation operation
- $E(\cdot)$: A symmetric encryption function
- $\delta_k$ : The function of XOR operation with secret key $k$
- $S_i$: The iris template of the user $U_i$
- $\mathcal{E}_{Si}(\cdot)$: Encryption function with biometric template $S_i$
- r: A random string
- A: An extracting algorithm
- $\oplus$: A string XOR operation
- $\rightarrow$: A common channel
- $\Rightarrow$: A secure channel

### *A. Assumption 1*

If the adversary successfully manages the server under owning the right of authentication, the adversary can request to login procedure and pass the authentication. Furthermore, the $y_i = E_x(ID_i|| \; h(PW_i)|| \; SS_i)$ will be easily retrieved by the adversary due to the identity stored in a verification table. However, a verification table suffers easily from an adversary's attacks, and further is unable to resist the stolen-verifier attack and modification attack.

### B. Assumption 2

Assume that an adversary uses the $SID^*$ to impersonate $SID$ and replays messages to the remote server to encrypt $C_1^*$ with a random string $v$. Then, the remote server sends the messages to the user. Until the user's smart card accepts the pretended $SID^*$. Thus, a user will encrypt the function with adversary's random string $v$ and sends the encryption messages to an adversary. The remote server can be accepted by an adversary's login request because he/she owns the password and biometrics.

### C. Assumption 3

Registration phase, a user $U_i$ has an identity $ID_i$ to register the license for remote server. Additionally, the Fan-Lin's scheme must store $ID_i$ to a verification table insider remote server. And then the remote server can perform to check whether the $ID_i$ is legitimacy during the authentication phase. When $U_i$ want to register to more than one server with the same identity $ID_i$ and authentication key $h(PW_i)$, any server can impersonate the eligible user and access other servers to obtain a login request. Obviously, the insider attack is possible in the assumption.

## III. THE PROPOSED SCHEME

We propose a robust three-factor authentication scheme with Elliptic Curves Cryptosystem (ECC) for the network communication. A three-factor authentication scheme involves a client, a server, and consists of four phrases: initiation phase, registration phase, login phase and authentication phase.

### A. Initiation Phase

In the system initiation phase, the server sets up the following system parameters for session key generation:

*1)* The user and server choose an elliptic curve order n over $E_p(a, b)$ generated by $P$, where $n$ is a large prime number for the security considerations.

*2)* The eligible server randomly selects $q_s \in Z^*_P$ as its own private key, and then computes the point multiplication as user's authentication key. That is, the server computes the corresponding public key $Qs = q_s \times P$.

*3)* The server employs the MD5 one-way hash function $h(.)$.

*4)* The smart card is prestored with the secret parameters $\{W, h(.), P, Qs\}$ in user and server side respectively, and the encryption function $\mathcal{E}Si(r)$ is prestored in users' smart card.

### B. Registration Phase

The $U_i$ wants to register to the remote server and setup the secret codes into the smart card for the $U_i$.

*1) Step 1:* $U_i \rightarrow$ server : $\{ID_i, h(PW_i \oplus r), \delta_r(S_i)\}$

The $U_i$ enters his/her username $ID_i$ and password $PW_i$ for computing $h(PW_i \oplus r)$. Here, $U_i$ scans the biometric characteristic as a template $S_i$ and chooses a random string $r$ to encrypt as $\delta_r(S_i) = r \oplus S_i$ using an encryption key $S_i$. That is, the user submits his/her $ID_i$, $h(PW_i \oplus r)$, and $\delta_r(S_i)$ to remote server if the user wants to convert into a new eligible user.

*2) Step 2:* Server $\Rightarrow U_i$'s smart card: $\{W, h(.), P, Q_s\}$.

After receiving the message from $U_i$, the server computes $Q_S = q_s \times P$ and $W = h(P \oplus h(PW_i \oplus r))$. Finally, the server stores the secret parameters $\{W, h(.), P, Q_s\}$ to a smart card and issues the smart card to the user over a secure channel.

*3) Step 3:* The $U_i$ checks the $W$ in the smart card.

$U_i$ 's smart checks whether $W = h(P \oplus h(PW_i \oplus r))$ is correct. If the condition is truth, a user will accept the smart card via a secure channel. Otherwise, the smart card does not come from the server and rejects the smart card.

*4) Step4:* The sketch $\mathcal{E}_{Si}(r)$ is stored in the smart card using his/her biometric template $S_i$ is an encryption key.

### C. Login Phase

*1) Step 1:* $U_i$ submits a $PW_i^*$ and his/her own biometrics, $S_i^*$, and the random string $r_i = A(\mathcal{E}_{Si}(r))$ is decrypted by the sketch $\mathcal{E}_{Si}(r)$ function which using $S_i^*$ to retrieve. Then, the smart card will compute the value $SS_i^* = \delta_r(S_i^*) = r_i \oplus S_i^*$ .

*2) Step 2:* The server validates $W$. The server computes $W$ and validates whether $W = h(P \oplus h(PW_i^* \oplus r_i))$ is correct. If it holds true, the system accepts the login and proceeds the authentication phase. Otherwise, server rejects the login request and authentication is terminated.

### D. Authentication Phase

After receiving the login request from the user, the detail descriptions of the authentication phase are described in the following operations.

*1) Step 1:* $U_i \rightarrow$ Server : $m_1 = \{Q_1, Q_u, M_u\}$.

The $U_i$ randomly chooses a private key $q_u = r_i^*$ and computes $Q_u = q_u \times P$, where $Q_u$ is $U_i$'s public key (Here, let the random string $r_i$ convert to $r_i^* \in Z_p^*$, $r_i^* < n$). And then $U_i$ computes the following formulas for the authentication procedure. Recall that $Q_S$ is the server's public key in the system initiation phase. $Q_1 = q_u \times Q_S$, $M_u = N_u + Q_u + Q_1$, where $N_u$ is chosen by $SS_i^*$ which is provided by $U_i$. Then, $U_i$ sends the $m_1 = \{Q_1, Q_u, M_u\}$ to the server.

*2) Step 2:* Server verify whether the $m_1$ message come from $U_i$. After receiving the $m_1$ message, the server computes $Q_S = q_S \times P$ and $Q_1 = q_S \times Q_u$ and then checks whether the $N_u^* = M_u - Q_u - Q_1 = N_u$ is correct. If it holds true, the $m_1$ message definitely comes from the $U_i$, otherwise, the verification is failure.

*3) Step 3:* Server $\rightarrow U_i$ : $m_2 = \{T_S, M_S, Q_S^*\}$

The server computes $Q_S^* = q_S^* \times P$ and $T_S = N_u^* + Q_S + Q_1$ and $M_S = N_S + Q_S + Q_1 + N_u^*$ where the $N_S$ is chosen by $SID$ which is provided by the server. Then, the server sends the $m_2$ message $\{T_S, M_S, Q_S^*\}$ to $U_i$.

*4) Step 4:* $U_i \rightarrow$ Server :$m_3$ $\{L = N_S + Q_u + Q_1\}$

After receiving the $m_2$ message, $U_i$ computes $N_u^{**}$ and checks whether $N_u^{**} = T_S - Q_S^* - Q_1 = N_u$ is correct. If it holds true, the $m_2$ message surely comes from the server, otherwise, the verification is failure. $U_i$ computes $N_S^* = M_S -$

$Q_S$ - $Q_1 - N_u^*$ and $L= N_S^* + Q_u + Q_1$, and then sends the $m_3$ message $\{L= N_S+ Q_u + Q_1\}$ to the server.

5) *Step5:* Server checks $N_S$.

The remote server compares $N_S$ with computed $N_S^{**} = L - Q_u - Q_1$ and these two are the same. If it holds true, the server accepts the $U_i$'s login request. Otherwise, the server rejects the login request.

## IV. SECURITY ANALYSIS AND COMPARISONS

### A. Security Against the Diverse Attacks

*1) Proper mutual authentication:* Our authentication scheme is based on ECC and provides the proper mutual authentication between the user and the server. In login phase, the user's password can be verified by the server computing $W = h(P \oplus h(PW_i^* \oplus r_i))$. During authentication phase, the user $U_i$ sends the $m_1$ message to the remote server. The server first validates whether the $N_u^* = N_u$ is equal, then sends the $m_2$ message $\{T_S, M_S, Q_S^*\}$ to user $U_i$. Then the user $U_i$ checks the condition whether $N_u^{**} = N_u$. Finally, the server validates whether $N_S^{**}$ is equal to $N_S$.

*2) Resist insider attacks:* If an adversary masquerades the eligible user to login the system. Note that in our registered phase, a user $U_i$ has the different authentication key for each system or server with the same password $PW_i$. The user $U_i$ computes the authentication key $h(PW_i \oplus r)$ and access the remote server, where $PW_i$ is chosen by the user $U_i$. Therefore, our scheme can resist insider attacks.

*3) Not need of a verification table:* Our scheme is based ECC mechanism, and the remote server has no need to store the password or a verification table insider computer. That is, the remote server only maintains the secret parameters. Thus, the proposed scheme can resist the stolen-verifier attack and modification attack.

*4) Allow user securely to change or update password:* The $U_i$ can compute the new value $h(PW_i^* \oplus r)$ and sent the message $\{ID_i,\ h(PW_i^* \oplus r), \delta_r(S_i)\}$ to the remote server. After receiving the demand for password change, the remote server computes the new value to update $W^*= h(P \oplus h(PW_i^* \oplus r))$ into the smart card.

### B. Comparisons

Recall that the scheme of Fan-Li [7] and other [4-6, 8], we compare our scheme with other referenced schemes in security properties and computation cost. Table I summarizes the comparisons among our scheme and other referenced schemes. Obviously, our scheme can overcome the security flaws of Fan-Li and other schemes. In terms of the requirements for a remote user authentication scheme, our proposed scheme solves all listed table problems and achieves.

## V. CONCLUSIONS

Obviously, biometric-based authentication can assure more reliable authentication than traditional password-based authentication. Additionally, recent concerns in biometric-based authentication focus on the issues of security and privacy protection. In this paper, we propose a robust three-factor remote user authentication scheme based on the ECC. In our assumption analysis, Fan-Lin's scheme fails to resist insider attacks, stolen-verifier attacks and modification attacks, and has security pitfalls due to storage of a verification table inside the server. In addition, we also found the other referenced schemes to be unsafe. Our proposed scheme can overcome security pitfalls and strengthen the security and privacy protection. Our scheme is practical and suitable for biometrics-based remote authentication.

TABLE I. COMPARISON AMONG THE REFERENCED SCHEMES.

| *Item* | *Our scheme* | *Lin-Lai scheme[3]* | *Li-Hwang's scheme[5]* | *Fan-Li scheme[9]* |
|---|---|---|---|---|
| Proper mutual authentication | Yes | No | No | Yes |
| Resist insider attack | Yes | Yes | Yes | No |
| Resist stolen-verifier attack and modification attack | Yes | No | Yes | No |
| Without a verification table | Yes | Yes | Yes | No |
| Securely change /update password | Yes | Yes | Yes | Yes |

## REFERENCES

[1] V. J. Matyas, Z. Riha, "Toward reliable user authentication through biometrics", IEEE Security & Privacy Magazine, vol. 1, 2003, pp. 45-49.

[2] J. K. Lee, S.R. Ryu, K.Y. Yoo, "Fingerprint-based remote user authentication scheme using smart cards", Electronics Letters, vol. 38, 2002, pp. 554-555.

[3] C.H. Lin, Y.Y. Lai, "A flexible biometrics remote user authentication scheme", Computer Standards & Interfaces, vol. 27, 2004, pp. 19-23.

[4] M.K. Khan, J.S. Zhang, "Improving the security of a flexible biometrics remote user authentication scheme", Computer Standards & Interfaces, vol. 29, 2007, pp.82-85.

[5] C.T. Li, M.S. Hwang, "An efficient biometrics-based remote user authentication scheme using smart cards", Journal Network and Computer Applications, vol. 33, 2010, pp. 1-5.

[6] X. Li, J.W. Niu, J. Ma, W.D. Wang, C.L. Liu, "Cryptanalysis and improvement of a biometric-based remote authentication scheme using smart cards", Journal of Network and Computer Applications, vol. 34, 2011, pp. 73-79.

[7] C.I. Fan, Y.H. Lin, "Provably secure remote truly three-factor authentication scheme with privacy protection on biometrics", IEEE Transactions on Information Forensics and Security, vol. 4, 2009, pp. 933-945.

[8] H.L. Yeh, T.H., Chen, P.C. Liu, T.H. Kim, H.W. Wei, "A secured authentication protocol for wireless sensor networks using elliptic curves cryptography", Sensors, vol. 11, 2011, pp. 4767-4779.

# SESSION

# SECURITY AND MANAGEMENT: NOVEL APPLICATIONS AND ALGORITHMS

## Chair(s)

### Prof. Hamid R. Arabnia

# Utilization of Fingerprint System in The Mobile E-Government

KyoungYul Bae, Prof, Sanmyung Univ and Hyun Byun, M.S Sangmyung Univ

Abstract___ Information society and penetration of smart device has been raised. Leading countries have already offered various civil services and administration services on the internet or mobile applications through smart device. E-government can be exposed to the threat such as the leakage of private information, national confidential. So there are several countries now trying to protect information in many ways. in this paper, we explore a fingerprint technology under Mobile E-Government's environment and propose Utilization of fingerprint recognition system in the Mobile E-Government.

## Ⅰ. Introduction

Mobile Government is using wireless technology to perform a government business and provide service to the citizen and company.[1][2] Mobile Government is the one of forms of Government. E-Government is using ICT (Information and Communication Technology) to improve its service level and focuses on providing Government services by using ICT plus Smart phone and Smart device So Mobile Government has an advantage of wireless technologies. therefore Government can offer their services without restriction of time and place[3] To meet growing requirements, it is now under way to embark on mobilizing Government service.

When it comes to mobile banking and application, Korea is the one of the countries to meet the global standards. According to Korea Communications Commission at 2012 February a new member of smart phone was about 20 million. which accounts for 47.7% in the whole new mobile members. Considering that millions of people are sign up for a new smart phone member every month, It is anticipated that at April the number of members exceeded 50% of the whole new mobile members.[4] If the amount of smart phone user exceeds phone user at the first half 2012 at the 2016 globally the number of entire smart phone user overtakes that of phone user.

Security of internal affairs and mobile E-Government service now being provided is the major issue.

Specially it is not one way services but interactive services being offered by the Mobile E-Government. So Identification is important problem. In the first chapter, we examine finger recognition sensor trend and the second chapter, we cover algorism technology trend and In the third chapter we handle promoting example of Mobile E-Government and based on the third chapter we suggest direction of Mobile E-Government by dealing with using finger recognition.

## Ⅱ. Finger recognition sensor

The reason why Current using finger recognition smart phone, and device are not so popular is that there are many problem : not providing reasonable price, breaching human rights. in addition, finger recognition sensor is not providing reasonable price to customers but through continuous development of the smart phone and device, it will be more miniaturized and its price will fall. If Government starts public certification by finger recognition technology, it will be more widely used. Finger recognition sensor is composed of optical method which uses the light or non optical method which doesn't use the light. Finger recognition sensor can be more classified as capacitive array sensor and optical array sensor, Thermal Array Sensor, CCD/CMOS Sensor, Ultrasonic Sensor, Tactile Sensor.

### 1) Capacitive array sensor

finger size silicon chip captures finger image when it contacts chip. It can be produced massively and miniaturized while static electricity by human body can make it stop chip's function and sensor is so sensitive to humidity that person's humidity can affect image quality.

### 2) Optical array sensor

Optical array sensor can be easily maintained and fixed. It is easy to get a fingerprint image but if a big and dry finger contact the sensor, a low quality image is captured. in addition, children's finger size is so small that the sensor can't display that image.

### 3) Therml array sensor

When finger contacts sensor, it measures temperature of finger by using difference of temperature so it capture finger image. Despite the humidity, it can recognize finger and its size is smaller than any other sensor so can be adopted well while fever sensing film paralleled with finger's curve sensor can't recognize finger.

### 4) ETC Sensors

CCD/CMOS Sensor captures image when finger contacts CCD/CMOS sensor it is low powered electronics and it has a wide surface width and fingerprint image captured is good. Ultrasonic Sensor use the way that it projects ultrasonic waves to the finger which is on the surface of sensor and gets the reflected signal and then, turns that signal into electronic signal. It shows the stark contrast and has wide gray scale area. Ultrasonic Sensor can get a high quality of image so it can be used in investigating criminal but price is high and size of sensor is big.

Tactile Sensor use the way that when the finger contacts the sensor's surface it gain an image and the part contacted with sensor becomes a digitalized image.

### 5) Using mobile camera recognition of fingerprint

Quality of image that is captured by mobile camera has a distinctive feature compared to other images. First, fingerprint image captured is colorful. Second, background area of fingerprint can be different depending on acquisition place and time. third, fingerprint's ridge and valley's intensity of illumination difference is smaller than that of contact sensor so it can be easily influenced by noise[5]

## Ⅲ. Algorithm of fingerprint recognition

Fingerprint image is not consistent, when acquiring it. Fingerprint image is prone to distortion, change of the size and quality when acquired. Because of the change of fingerprint status, the external factor such as scratching , habit, abrasion, dryness and wetness can not be standardized So it is hard to decide how similar it is and compare between fingerprint images. So we should examine algorithm to deal with these problems.

| Sensor type | Advantage | Disadvantage | Durability | I/O |
|---|---|---|---|---|
| Optical Camera | Data reading speed is fast | In case of dry finger get a low quality of image.<br>In case of small size finger Sensor can't recognize fingerprint | Good | Parallel Port / USB |
| Capacitive Sensor | Massive production is possible and sensor is thin it can be miniaturized | image quality depends on the humidity | ESD | Parallel Port / USB |
| Thermal Chip | Regardless of humidity image can be obtained and the size of the sensor is small | In case fingerprint curve parallel with sensor it can't recognize fingerprint | good | USB |
| CCD Sensor | The size of sensor is small and low consuming electronics and wide surface width | Price of sensor is expensive and arounding circuit is complex | Fair | Parallel Port |
| Tactile Sensor | The size of sensor is small and prise is cheap | Quality of fingerprint image is low | Fair | PCMCA Card |

<Table 1> Compare with sensor's specification

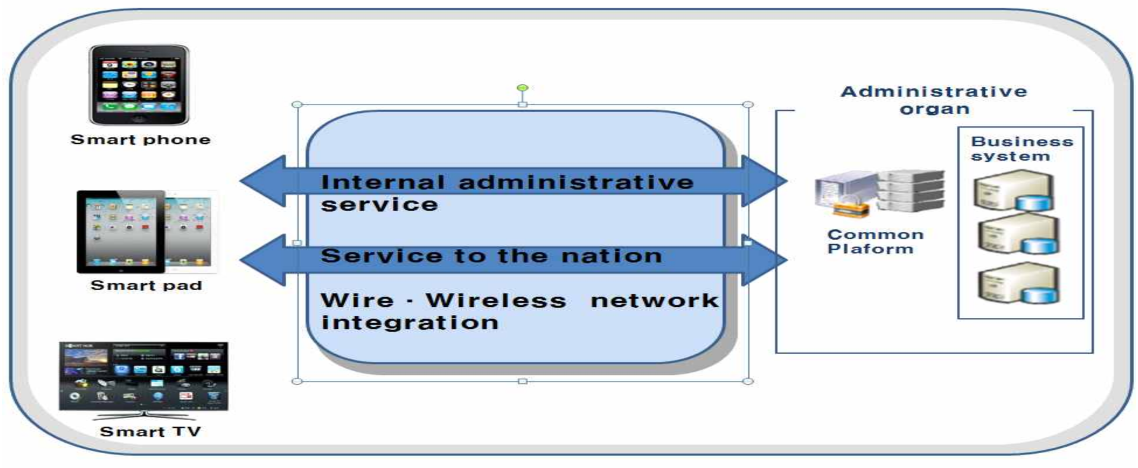
| Algorithm | Advantage | Disadvantage |
|---|---|---|
| String Aligment-based Matching [Jain, Hong, Bodle,1997] | By modeling Non-liniarity modification at the polar coordinates error estimation is convenient | Cause false Minutiae and incorrect arrangement Minutiae matching error occurs |
| Ridge Line Following [Main, Maltoni, 1997] | It reduce resource cost when preprocessing | Take the impact on the image quality into account |
| Image Enhancement [Hong, Wan, Jain 1998] | Considering quality of image it remove false Minutiae in advance | Cause using Gabor filter amount of calculation is enhanced |
| Structural Classification [Cappeli, Lumini, Maio 1999] | As a continuative categorization it is effective | It doesn't have a resistance to any amount of lines and character |
| FingerCode [Jain, et al 2000] | when fingerprint matching it reduce template that using comparison effectively | It takes long time when making a template |
| Triangular matching & Dynamic time warping [Kovacs-Vajna,2000] | It is more effective way about displacement and rotation image than matching method | Searching area is broad so there is a large amount of computation |

<Table 2> Comparison algorithm

## Ⅳ. Mobile E-Government

Recently the reason why all the countries in the world pushed forward to Mobile E-Government is that mobile technology is gradually developed and there is future plan for better government services. The background reason why Mobile E-Government now emerges is that first wireless ICT (information & communication technology) device has been increased rapidly not only in the Korea but other countries as well and the penertration rate of Smart device takes over penertration rate of PC. According to ABI research, anticipation of the amount of 2012 phone shipment is 1.8 billon and it is expected that it reaches 1.9 billion in 2016 while in 2012r smart phone shipment is 6 hundred bilion and it is anticipated that in 2016 11 billion smart phone will be shipped.[6] Second mobile ICT devices is portable, so Government can provide more additional services than cable service does. smart phone and most of smart devices are equipped with high quality of camera, so smart device can send multimedia contents directly that is obtained on the spot. Also Smart device is equipped with GPS, so Government can acquire user's location information. Those Smart device character such as interactive multimedia communication and location information acquisition enables Government to provide various services . Last Ubiquitous Government underlies Mobile E-Government and the latter is essential element as a final developmental form of Ubiquitous

Government. from the long term Viewpoint system for E-Government introduction is necessary.[7]



<Pic 1> Environment of Mobile E-Government[Ministry of Public Administration and Security]

Mobile E-Government service can be categorized as a <Table 3>

| service object | Service form | | Service content | Example |
|---|---|---|---|---|
| internal service | Business performance by using mobile device | | For processing Government's internal business by using Mobile device | Information inquire and enter by using Smaart device |
| Service to nation | one way | Provide information that form of massage | Using short message service or Multimedia Messaging Service offer information to the nation | Earthquake alert massage |
| | | Massage form report receipt | Using short message service or Multimedia Messaging Service report or receipt | Illegality reporting on the site |
| | inter active | Mobile web page service | On the Mobile web page information inquiry and settle a civil complaint | Bus information inquiry and file a civil complaint and issue a certification |
| | | Mobile application service | Provide applicaton that can be used at the Smart device | Using GPS iquiry closest public office application |

<Table 3> Categorize of Mobile E-Government service[9]

Table3 shows categorization of mobile E-Government services whose examples of Mobile E-Government services are arranged at table 4. As written on the table most of all the services are provided ,but a form of Mobile application service as a Mobile office is not offered yet.

### A. Implications of Mobile E-Government's certification

Considering that these days service provided by E-Government and diffusion rate of smart device, the usage rate and level of application development will rapidly increase in terms of Mobile administration. Each country interested in building Mobile Government has been increased and concerns about the security has been raised. Using Mobile E-Government means that important national classified information can be exposed on the internet [10]Example of each country's security infrastructure is as follows: America in 1998 October through GPEA("The Government Paper-work Elimination Act) has all of federal agencies using electronic signature, During an online transaction. In 1999 On behalf of federal agencies, private companies provide digital certification service and In 2000 June electronic signature and electronic documents get legal status equivalent with paper documents by the Electronic signature law. [11] Second in 2001 Japan pushes forward a legislation on private information protection and installing security equipment, Such as firewalls and intrusion detection systems to protect the main system through mutual authentication and data encryption to protect communications [12] Third

| service object | Service form | | Example |
|---|---|---|---|
| internal service | Business performance by using mobile device | | o Field Enforcement performance<br>- Enforcement parked car illegally<br>o Firefighting/Disaster prevention/Public peace<br>- e-911: Ministry of Public Administration and Security<br>- wireless Criminal record inquiry |
| Service to nation | one way | Provide information via massage | o Short message service<br>- Emergency be wanted, emergency call<br>- weather special report<br>- weather report, disease and pest information<br>- Inquiry of civil complaint process result |
| | Inter active | Mobile web page service | Ministry of Public Administration and Security: Inquiry of civil complaint process<br>Ministry of Justice :<br>Handling legal civil complaint via Mobile service<br>Korea Customs Service :<br>personal belongings clearance information<br>Military Manpower Administration :<br>militery service information |
| | | Mobile application service | o Legal Information application<br>the Office of Legislation :<br>Supreme Court precedent information service<br>o Weather information service<br>the Meteorological Administration : weather special report , current weather, real time weather information service<br>o Land information application<br>Ministry of Land :<br>Real estate information, address searching service |

<Table 4> Example of Mobile E-Government service[9]

singapore in 1948 personal ID card has been released so Government system ID card and PIN code are used to verify the identity [13]

## Conclusion

PKI such as existing certifications is usually stored on a smart phone or a smart device and used so there is a risk of identity theft but Fingerprint is convenient when it comes to portability there is no risk of losing or being stolen and imitating fingerprint is almost impossible so it has advantages of high reliability [14] but considering the problem of self authentication rate and fingerprint data theft, finger print needs to be linked with other security tools such as Smart card, password, other Biometrics recognition rather than to be used separately. If the fingerprint information is connected with private Interest, gender, address, Specialized government services are provided with individuals via smart devices. fingerprint recognition technology in this thesis will help Policy decisions and strategies related to corporate strategies.

## Reference

[1] Kesavarapu, Srikanth, and Choi, Mun-Kee, ATheoretical Framework of Knowledge Management in M-Government, International Journal of Computer and Communication Technology, Vol. 1, No 1, pp.1-18.

[2]Ovum: Mobile E-commerce - Market Strategies,Ovum Press Release, 2003.

[3]Lallana,E.C.,mGovernment: Mobile/WirelessApplicationinGovernment http://www.egov4dev.org/mgoverment/, 2004

[4]http://www.bloter.net/archives/103591

[5]Lee CheolHan, Lee Sanghun, Kim JaeHee, Kim SungJae "Fingerprint Segmentation and Ridge Orientation Estimation with a Mobile Camera for Fingerprint Recognition", Institute of Electronic Engineering Article 42 SP Part No. 6, 2005.11.

[6]http://www.bloter.net/archives/103591

[7]Jeong GukHwan, Korea's u-Gov Construction and future strategies, 1st U Government Forum, 2005.

[8]Ministry of Public Administration and Security

[9][Seo YongWon, Kim TaeHa, "Mobile Government Service Classification and Policy Implications", Korea Institute of Industrial cooperation Vol. 11, No. 4, pp. 1475-1482, 2010]

[10]Lee KyongHo, Kim SoJung, Lim JongIn, "E-government and privacy", Information and Communication Sciences Article 13, No. 3.

[11]Kim DaeHo, Oh IlSuk, "U.S. e-government information security law trends", Information and Communication Sciences Article 13, No. 3.

[12]An MunSeok, Park SungJin, Maeng BoHak, "An inquiry on E-Government Information Security Response System ", Information and Communication Sciences Article 13, No. 3.

[13]Ryu SeokSang, Lee YeonWoo, Park JungEun. "Comparative Analysis of the four major countries of e-government" NCA Information yisyubunseok,2003. 5.

[14]Yoon DukHyun, Kim KiUk, Kim ChangSu "Implementation of a fingerprint identification system integrity verification tools"2002 Proceedings of Korea Multimedia Society Spring

# Resource-Efficient Multi-Source Authentication Utilizing Split-Join One-Way Key Chain

**Seonho Choi**[1], **Kun Sun**[2], **Hyeonsang Eom**[3]
[1]Department of Computer Science, Bowie State University, Bowie, Maryland, U.S.A.
[2]Center for Secure Information Systems, George Mason University, Fairfax, Virginia, U.S.A.
[3]School of Computer Science and Engineering, Seoul National University, Seoul, Korea

**Abstract—***In wireless ad hoc networks, most of the authentication protocols assume a single source of trust. However, in the presence of multiple trust sources (called source group in this paper), it becomes difficult to design resource (or energy) efficient authentication protocols, especially for multicast/broadcast services, utilizing multiple trust sources at the same time. Some traditional authentication approaches may be extended and used for this purpose. However, the communication overhead, e.g,, may increase significantly in proportion to the number of trust sources.*

*In this paper, we propose a new scheme named as Multi-source Authentication with Split-Join One-Way Key Chain (SOKC). In this new technique, the communication overhead is small and constant, and the memory requirement at the verifier node is also minimal. The source group node needs to store n keys where n represents the key chain length, which may be a reasonable requirement considering that the trust sources usually have more resources compared to other regular node(s) in the network (e.g., base stations in wireless sensor networks).*

**Keywords—**authentication, security, protocol, wireless ad hoc networks

## 1 Introduction

In wireless ad hoc networks, most of the authentication protocols assume a single source of trust. For example, in a wireless sensor network (WSN), it is typically assumed there is one trustworthy base station and it is the only source of the trust. However, in the presence of multiple trust sources (called source group in this paper), it becomes difficult to design resource (or energy) efficient authentication protocols utilizing multiple trust sources at the same time. Some traditional authentication approaches may be extended and used for this purpose. However, the communication overhead, e.g,, may increase significantly proportional to the number of trust sources. In this paper, we propose a new scheme named as Multi-source Authentication with Split-Join One-Way Key Chain (SOKC). In this new technique, the communication overhead is small and constant, and the memory requirement at the verifier node is also small. The source node needs to store *n* keys where *n* represents the key chain length, which may be a reasonable requirement considering that the trust sources usually have more resources compared to other regular node(s) in the network (e.g., as in sensor network).

Our technique utilizes a delayed key disclosure mechanism as in TESLA and uTESLA approaches [8,9], and also extended the one-way key chain technique to achieve the goals of minimal communication overhead and minimal storage requirements at the verifier nodes.

Our SOKC scheme may be applied to both unicast and multicast/broadcast authentication services. But, the application of our scheme would be simpler for unicast cases, and for most of the cases broadcast authentication services are more important since conveying the information from the trustworthy source to other nodes may be more critical compared to the communication between non-trustworthy nodes. For example, several routing protocols were proposed based on periodic broadcasting (e.g., flooding) of routing (or beacon) messages. These include TinyOS beaconing [6], directed diffusion and its multi-path variants [10], etc. Also, several location discovery schemes have been proposed which utilize broadcasting capabilities to estimate node locations [8]. Even though more advanced broadcast techniques may be utilized in the network, simple flooding may be preferred or required due to the simplicity or instability of network connections. Our proposed approach may be applied in both cases. Hence, we will focus on developing and applying the SOKC scheme for the multicast/broadcast services.

This SOKC scheme may be applied for various authentication problems. However, to show its applicability we chose two authentication problems. For example, in wireless ad hoc networks, some attacks exploit the fact that it is hard to authenticate the actual path (or number of hops) data packets traversed - especially the attacks against the broadcast services. Sinkhole and wormhole attacks belong to this attack category [6]. With the multi-source authentication capabilities, each node would be able to detect and cope with such attacks. A new path authentication technique may be developed by utilizing our Multi-source SOKC scheme. For example, the source group keys may be duplicated and randomly distributed across the network, so that the verifier nodes may be able check whether a packet has really passed through a certain number of source group nodes along the routing path from the claimed origination point.

The SOKC scheme may also be applied to WSNs with multiple base stations. It is typically assumed that a WSN has only one base station. However, there may exist several drawbacks. Degraded reliability may be a problem due to a single point of failure. The latency may be an issue if the number of hops in the delivery paths may become large, which may cause the reduced lifetime of the sensor nodes and, thus, the entire sensor network. The deployment of multiple base stations were proposed to overcome these limitations [1,2,12]. However, in the presence of multiple base stations, it would be more difficult to provide robust authentication services since it would be required to tolerate compromise of multiple base stations as well as sensor nodes. If we assume that the base stations can communicate with each other directly using a separate channel, then our SOKC based approach may be used in providing multi-source broadcast authentication services. It is also assumed that all the base stations need to participate in authenticating the broadcast messages to provide increased security levels. If we consider the importance of the broadcast messages in WSNs, this would be a valid assumption.

## 2 Related Works

Several security mechanisms for authentication and secure routing protocols in wireless ad-hoc network are based on public key cryptography ([5], [14]). However, until now, the public key cryptography is still too expensive for the resource constrained mobile nodes. Secure routing protocols based on symmetric key cryptography have been proposed (e.g., [3], [4]). SEAD [3] is a distance vector routing protocol based on DSDV. The basic idea is to use one-way hash chains to authenticate the metrics and the sequence number of a routing table. The destination node can authenticate the source node; however, it cannot authenticate the intermediate nodes along the path from the source node to the destination node. Ariadne [4] uses per-hop hashing technique and source routing techniques to prevent route misbehaviors. However, it requires a precise time synchronization among all the nodes, which is usually difficult to be achieved in the mobile networks. Moreover, the communication overhead may increase significantly when including all the identifies and corresponding MACs for all the nodes along the path.

Authenticating broadcast (or multicast) traffic in wireless ad hoc networks is also a hard problem since the traditional approaches like digital signatures may not be adequate due to the heavy resource requirements. TESLA and μTESLA approaches [7,9] were proposed as viable solutions to the authentication problem in such networks. μTESLA utilizes the delayed key disclosure and one-way key chain techniques. First the packet is broadcast with a calculated keyed Message Authentication Codes attached along with the original data portion, and only after sufficient time is elapsed for all the nodes in the network to receive it, the corresponding key will be disclosed to the network nodes for authentication of the previously sent data and MAC. TESLA and μTESLA requires loose time synchronization among the network nodes.

Researchers have proposed several mechanisms to prevent the false data injection attacks. Przydatek, Song, and Perrig propose SIA [11], a secure information aggregation scheme for sensor networks that addresses the issue of false data injection using statistical techniques and interactive proofs, ensuring that the aggregated result reported by the aggregation node is a good approximation to the true value, even if a small number of sensor nodes and the aggregation node may have been compromised. SIA focuses on the accuracy of query results reported from the base station, whereas our scheme focuses on the authenticity of the reports from sensor nodes and provides a means to filter out any injected false data as early as possible. Both schemes can be combined to make the network more robust to false data injection attacks.

SEF [13] is a statistical en-route filtering mechanism to detect and drop false reports during the forwarding process. Authenticating event reports requires that nodes share certain security information, however, attackers can obtain such information by compromising a single node. To prevent any single compromised node from breaking down the entire system, SEF carefully limits the amount of security information assigned to each node, and relies on the collective decisions of multiple sensors for false report detection. First, SEF divides a global key pool into multiple partitions and carefully assigns a certain number of keys from one partition to individual node. Given that any single node knows only a limited amount of the system secret, compromising one or a small number of nodes cannot disable the overall network from detecting bogus reports. Second, by assuming that the same event can be detected by multiple sensors, in SEF each of the detecting sensors generates a keyed message authentication code (MAC) and multiple MAC are attached to the event report. As the report is forwarded, each node along the way verifies the correctness of the MAC's probabilistically and drops those with invalid MACs. Finally, the sink verifies the correctness of each MAC and eliminates remaining false reports that elude en-route filtering. Comparing to statistical solution provided by SEF, our solution can provide a more resource-efficient path authentication, and it cannot handle the broadcast authentication.

Zhu et al [15] present an interleaved hop-by-hop authentication scheme for addressing the false data injection attack launched by the compromised nodes. The scheme guarantees that the base station will detect any injected false data packets when no more than a certain number t nodes are compromised. To defend against false data injection attacks, at least $t + 1$ sensor nodes have to agree upon a report before it is sent to the base station. t is a security threshold based on the security requirements of the application under consideration and the network node density. Further, it provides an upper bound for the number of hops that a false data packet could be forwarded before it is detected and dropped, given that there are up to t colluding compromised nodes. In other words, it also attempts to filter out false data packets injected into the network by compromised nodes before they reach the base station, thus saving the energy for relaying them. [15] is the most similar work to our proposed scheme, but their approach cannot handle the broadcast authentication. Our solution approach utilizes a new SOKC technique along with the delayed key disclosure to achieve a much smaller communication overhead.

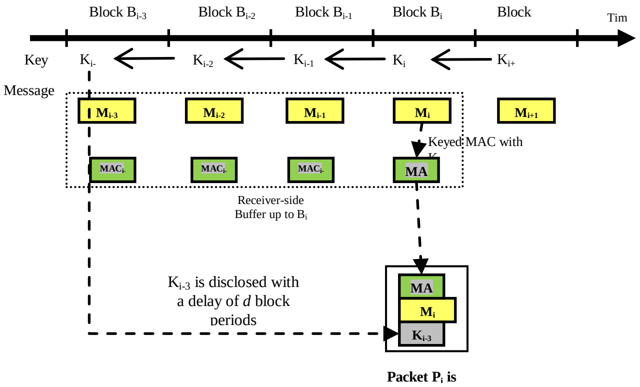


**Figure 1. µTESLA technique [7,9] is shown with a key disclosure delay of *d*=3 block periods. Note that keys are disclosed later while the message and MAC portions are disclose in the corresponding block.**

# 3 Methodology

## 3.1 Assumptions

We refer to the minimum number of hops necessary for a packet to reach from any node located at one extreme edge of the network to another node located at the opposite extreme, as the *diameter* of the ad hoc network. Packets may be lost or corrupted in transmission on the network. A node receiving a corrupted packet can detect the error and discard the packet. Nodes within the ad hoc network may move at any time without notice, and may even move continuously, but we assume that the speed with which nodes move is moderate with respect to the packet transmission latency and wireless transmission range of the particular underlying network hardware in use. We assume that nodes may be able to enable *promiscuous* receive mode on their wireless network interface hardware, causing the hardware to deliver every received packet to the network driver software without filtering based on link-layer destination address. Even though this feature is not required, by utilizing this feature, the performance of our scheme may be enhanced especially when the mobility level is high in the network.

The local clocks of the nodes are assumed to be (at least) loosely synchronized with a maximum time synchronization error Δ. Various time synchronization techniques were proposed for wireless ad hoc networks, and any of them may be utilized to achieve this requirement. Similar assumptions were made in the broadcast authentication schemes such as TESLA, µTESLA, etc. [7,9].

Also, the time line is assumed to be divided into block periods as in TESLA and µTESLA approaches [7,9]. In each block period one packet may be sent out for broadcasting by any valid broadcast originator (which may be determined by the application). Delayed key disclosure mechanism is adapted and incorporated in our scheme. Each broadcast packet contains the message, the authentication-related information, and the key information that is disclosed for the previously sent out message. The key disclosure delay is denoted as *d* block periods. These can be seen in Figure 1.

We assume that there is one source and one or more recipients that are involved in each session (one or more data packets are delivered in each session). That is, our authentication approach may be used for both unicast and multicast/broadcast communications. However, we will focus on developing protocols for broadcast services as mentioned before. The packets are transmitted along a multi-hop delivery path to the receiver(s). The delivery path will be determined by a routing protocol used in the network. Many routing protocols were proposed for wireless ad hoc networks, and any of them with reasonable route change rates (due to mobility) may be utilized in the network.

Finally, it is assumed that the number of different source group keys in one source group (which is denoted as *m*) is an odd number.

## 3.2 Overview of the Protocol

The protocol carries out the following three processes to provide the multi-source authentication. In this scheme it is assumed that the number of source group keys, *m*, is an odd number.

1. Offline SOKC generation (Figure 2): SOKC is generated offline by utilizing the source seed ($Z_0$), source keys ($a_i$), one-way hash operation, and the bitwise EXOR operation. Source nodes with a secret source key ai will be equipped with a chain of keys that are obtained from the intermediate keys, $Y_j$ ($1 \le j \le n-1$), by applying the EXOR operation with $a_i$. The keys generated in this process are denoted as $SOKC^i = \{K^i_{n-1}, ...K^i_2, K^i_1\}$.

The intermediate keys that are generated in this process are named as $Y_j$ and $Z_j$, which will be explain in more detail in a later section.

2. Semi-encrypted key pre-distribution (Figure 3): when the original sender node (this may be one of the source group nodes or may not be one of them) has some message to send, it will first send a packet that has the following field:

- random nonce $R_j$ of $k$ bits - this would be used to prevent the disclosure of the next key ($Y_j$) in the SOKC that is needed for the next round of validation.

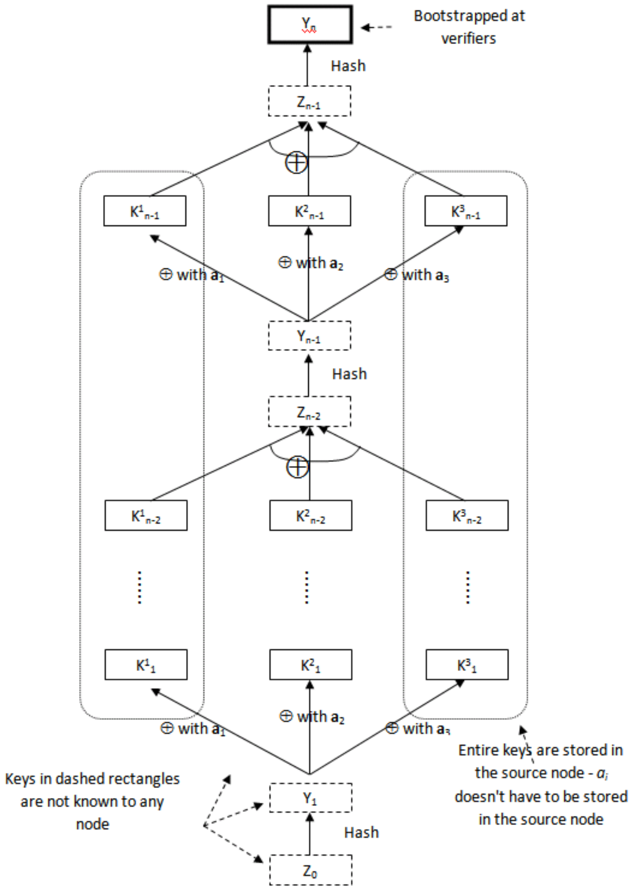

Y_n
Bootstrapped at verifiers
Hash
Z_{n-1}
⊕
K¹_{n-1}
K²_{n-1}
K³_{n-1}
⊕ with a₁
⊕ with a₂
⊕ with a₃
Y_{n-1}
Hash
Z_{n-2}
⊕
K¹_{n-2}
K²_{n-2}
K³_{n-2}
K¹₁
K²₁
K³₁
⊕ with a₁
⊕ with a₂
⊕ with a₃
Entire keys are stored in the source node - a_i doesn't have to be stored in the source node
Keys in dashed rectangles are not known to any node
Y₁
Hash
Z₀

**Figure 2. SOKC generation: the entire key chain is generated offline and only the keys in the solid rectangles are stored in the nodes. The keys, $K_j^i$, are stored in the *i*-th source node(s). $Y_n$ is bootstrapped in each verifier node. The intermediate keys and even the secret source keys, $a_i$, are not stored in any node in the network.**

Once this field is filled with a random number generated by the original source and the packet is sent out, the first source group node from $a_i$ (this may be the same as the original sender node) will apply EXOR operation with this random number to (its next key in the $SOKC^i$ ⊕ its Message), and forward the packet to the next node in the delivery path. The next source group nodes with different source keys will carry out the same process: get the value from this field and apply EXOR operation to it. But, this process is done only once for each source key $a_j$, $1 \leq j \leq m$. In other words, if there are multiple source group nodes with the same source key in the delivery path, only the first one will carry out this process. When the packet finished traversing all the source group nodes with *m* different source keys, then the verifier nodes in the remaining path will have the following value in the packet field:

$$\Phi_j = R_j \oplus a_1 \oplus a_2 \oplus a_3 \oplus Y_j \oplus M_j$$

$M_j$ stands for the message in the *j*-th packet. This value will be stored in the verifier nodes for later authentication purposes. Verifiers may store the other field values such as the actual message ($M_j$) depending upon the scheme).

3. delayed key disclosure with verification (Figure 3): After the key disclosure delay (*d* block periods), the original sender of $R_j$ will start the key disclosure process by including the following fields in the packet:

- disclose the actual random nonce used *d* block period before ($R_j$) - the original sender will initialize this field to all 0s.
- key disclosure field for accommodating the $SOKC^j$ keys from the *m* source group nodes (they will be EXORed and this field requires only *k* bits)

Each group source nodes will apply EXOR operation between the next key in $SOKC^i$ and the value from the key disclosure field mentioned above, and store back the result into the key disclosure field again. After the packet traverses all of the *m* source group nodes with different source keys, the packet will contain the following value in its key disclosure field:

$$a_1 \oplus a_2 \oplus a_3 \oplus Y_j = Z_j$$

Once the packet reaches a verifier node, and if the packet is claimed to have traversed *m* different source group nodes, then the verifier node will carry out a sequence of steps. First, it will extract the intermediate key, $Z_j$, from the key disclosure field, and checks whether this intermediate key is really from the authentic SOKC by applying the one-way hash operation and comparing the result to the already stored $Y_{j+1}$. If they don't match, then the key disclosure packet is discarded, and the already stored message $M_j$ will not be authenticated. If they match, then the verifier extracts the intermediate key, $Y_j$, by multiplying $a_1 \oplus a_2 \oplus a_3$ to $Z_j$, and stores it as $Y_j$ as a newly disclosed authentic SOKC key to be used in the next round of authentication.

Then, the verifier will check the following condition to compare to the already stored $\Phi_j$.

$$R_j \oplus a_1 \oplus a_2 \oplus a_3 \oplus Y_j \oplus M_j = \Phi_j$$

If the equality holds, then the previously (in TESLA) stored $R_j$ and $M_j$ are validated.

## 3.3 Basic Scheme

**Notations**

The following are defined for our authentication process:

- Source group: a group of nodes equipped with SOKCs generated from *m* different source keys, $a_i$ where $1 \leq i \leq m$, are distributed among $N_{src}$ number of source nodes ($m \leq N_{src}$). It is assumed that *m* is an odd number in this scheme. The source nodes may or may not be located in close proximity, and some source nodes may have the same source key if $m<N_{src}$. The source key size is denoted as $k$ bits.

- Verifier group: $N_{vrf}$ nodes (e.g., multicast group members or all the nodes in the broadcast case) are equipped with a verification information for authenticating a packet's traversing of at least $m$ source nodes with different $a_i$ in the routing path. That is, a verifier node has the ability to verify that the packet passed through all the source group nodes with $m$ different source keys.

The following are the information kept in the source group node and the verifier group node:

- Source node from $a_i$ keeps the following items:
  - ○Split-Join One-way Key Chain from $a_i$ : $SOKC^i = \{K_{n-1}^i, ... K_2^i, K_1^i\}$
  - ○public source key sum $a = a_1 \oplus a_2 \oplus \cdots \oplus a_m$
  - ○cryptographic one-way hash function
- Verifier node keeps the following items:
  - ○public source key sum $a_1 \oplus a_2 \oplus \cdots \oplus a_m$
  - ○last key, $Y_n$, in the SOKC
  - ○cryptographic one-way hash function

**SOKC generation with *m* source keys, *$a_i$*, *1≤i≤m* (Figure 2):**

This process is assumed to be carried out offline before the network launch time. The possible issues that may arise in developing online SOKC generation will be addressed in the proposed research section later. For the offline case, the detailed algorithm is shown in Figure 2 and the detailed steps are listed as follows:

1. Apply a cryptographic hash function to $Z_0$ to generate $Y_1$ which is also k-bit long.
   a. That is, $Y_1 = \mathrm{H}(Z_0)$.
2. Calculate a key $K_1^i$ in the chain by applying the EXOR operation with a secret source key $a_i$, $K_1^i = a_i \oplus Y_1$ .
3. Calculate $Z_1$ = $a \oplus Y_1 = (a_1 \oplus a_2 \oplus \cdots \oplus a_m) \oplus Y_1$.
4. Apply a cryptographic hash function to $Z_1$ to generate $Y_2$. That is $Y_2 = \mathrm{H}(Z_1)$.
5. Calculate the second key $K_2^i$ in the chain, $K_2^i = a_i \oplus Y_2$ .
6. Calculate $Z_2$ = $a \oplus Y_2 = (a_1 \oplus a_2 \oplus \cdots \oplus a_m) \oplus Y_2$.
7. Repeat steps 4 through 6 until key $K_{n-1}^i$ is obtained.

The one-way key chain at the source is now obtained as $SOKC^i = \{K_{n-1}^i, ... K_2^i, K_1^i\}$ . The keys in this $SOKC^i$ are bootstrapped in the source group nodes. These keys are used in reverse order starting from $K_{n-1}^i$. The last key $Y_n$ in the entire *SOKC* is assumed to be bootstrapped at each verifier node.

**Packet Format**

The *j*-th packet, *1≤j<n*, has the following packet format consisting of 5 fields:
(1) SRC index bits
(2) Semi-Encrypted Key ($\Phi_j$) pre-distribution– k bits
(3) Nonce ($R_{j-d}$) disclosure - k bits
(4) SOKC Key (for $a \oplus Y_{j-d}$) disclosure - k bits
(5) Message ($M_i$)

SRC index bits are used for showing which source group nodes the packet has been traversed. For example, it its *i*-th bit is set to 1, then it means that the packet is claiming that it has already traversed a node with the source key $a_i$, and, if its *i*-th bit is 0, then it means that the packet has not traversed any such node. If another source node with the same $a_i$ receives the packet whose *i*-th SRC index bit is set to 1, then it will forward the packet without modifying any of the fields, even though it can repeat the process without any adverse effect – but, it will waste resource if it does.

The other three fields following the SRC index bits field will be used in the semi-encrypted key pre-distribution and the delayed key disclosure with verification processes. Note that the nonce disclosure and SOKC key disclosure fields will disclose the values that were previously used in the (*j*-*d*)-th block period.

After the key disclosure delay (i.e., *d* block periods), another packet needs to be sent by the original sender and it would contain the following fields:

(1) SRC index bits
(2) Semi-Encrypted Key ($\Phi_{j+d}$) pre-distribution– k bits
(3) Nonce ($R_j$) disclosure - k bits
(4) SOKC Key (for $a \oplus Y_j$ ) disclosure - k bits
(5) Message ($M_i$)

**Semi-encrypted key pre-distribution at the source node with a source key *$a_i$* (*1≤i≤m*):**

Again, the purpose of this process is to let source group nodes to reveal their SOKC keys in a semi-encrypted form by applying the EXOR operation to the random nonce ($R_j$) sent out by the original sender of the packet. Let's assume that the a *j*-th packet is sent out by the original sender. The process is shown in Figure 3 and the detailed steps are described as follows:

1. Original sender generates a random nonce ($R_j$) and insert it into the semi-encrypted key pre-distribution field in the packet before sending it.
2. Each source group node from a source key, $a_i$, in the delivery path will carry out the following:
   a. if the SRC index bit (whose index value is *i*) is equal to 1, then go to step 3.
   b. extract the value in the semi-encrypted key pre-distribution field (let it be denoted as *x*).
   c. extract $M_j$ from the packet.
   d. calculate $M_j \oplus K_j^i \oplus x$ and store this as a new value in the semi-encrypted key pre-distribution field of the packet.
   e. set the SRC index bit (whose index value is *i*) as 1.
3. Each verifier node will perform the following steps:
   a. if all of the *m* bits are set to 1 in the SRC index bits field, then the node will extract the value of the semi-encrypted key pre-distribution field, and store it as $\Phi_j$ to be used at the future verification time (after *d* block periods).
4. Forward the packet if it is needed.

**Delayed key disclosure and verification after the key disclosure delay (Figure 3)**

Because the actual SOKC keys are disclosed after the delay (*d* block periods), the verifications of the keys and messages that were included in the *j*-th packet may be carried out when the nodes receive/process a packet in the (*j*+*d*)-th block period. So, we disclose the *j*-th keys from the SOKCs in the (*j*+*d*)-th packet, and the verifications will be carried out by the verifier nodes upon the receipt of the (*j*+*d*)-th packet. The process is depicted in Figure 3, and the detailed steps are described as follows:

1. Original sender discloses the nonce ($R_j$) by including in the (*j*+*d*)-th packet's nonce disclosure field.
2. Each source group node from a source key, $a_i$, in the delivery path will carry out the following:
   a. if the SRC index bit (whose index value is *i*) is equal to 1, then go to step 3.
   b. extract the value from the SOKC key disclosure field (let it be denoted as *y*).
   c. calculate $K_j^i \oplus y$ and store this as a new value into the SOKC key disclosure field of the packet.
   d. set the SRC index bit (whose index value is *i*) as 1.
3. Each verifier node will perform the following steps:
   a. if all of the *m* bits are set to 1 in the SRC index bits field, then the node will perform the following steps:
      i. extract the value of the SOKC key disclosure field (let it be denoted as *z*).
      ii. check whether $H(z)=Y_{j+1}$
         a) If not, then the key validation for SOKC fails, discards the packet and exit from the algorithm.
         b) If the equality holds, then store $a_1 \oplus \cdots \oplus a_m \oplus z$ as a valid $Y_j$ for future SOKC validation.
            A. calculate $R_j \oplus z \oplus M_j = R_j \oplus a_1 \oplus a_2 \oplus a_3 \oplus Y_j \oplus M_j$ and compare this to $\Phi_j$ that were extracted and stored in the previously received *j*-th packet.
               i. If they are the same, then the multi-source authentication succeeds for $M_j$.
               ii. If they don't match, then the multi-source authentication fails for $M_j$.
4. Forward the packet if it is needed.

**Resource Requirements (Figure 3)**

The resource requirements at a source group node from $a_i$ are:

- $n \times k$ bits are needed for storing the SOKC keys

The resource requirements at a verifier node are:

- *k* bits: for storing the SOKC validation key $Y_n$
- $d \times k$ bits: for storing $\Phi_j$ in at most d *d* consecutive block periods
- *2k* bits: for temporarily storing $Z_j$ and $a_1 \oplus a_2 \oplus a_3 \oplus Y_j = Z_j$

Hence, the total memory requirement at a verifier node is $(d+3) \times k$ bits.

The communication overhead at each packet (purely needed for our scheme) consists of the following:

- *m* bits: for SRC index bits
- *3k* bits: for semi-encrypted key pre-distribution field, nonce disclosure field, SOKC key disclosure field

Hence, the total overhead is *m+3k* bits for each packet.

## 4 Conclusion

We presented a new resource-efficient multi-source authentication scheme with Split-Join One-Way Key Chain (SOKC). In this new technique, the communication overhead is small and constant, and the memory requirement at the verifier node is also minimal. This technique may be effectively used for wireless ad hoc networks when there exist multiple trust sources to be utilized.

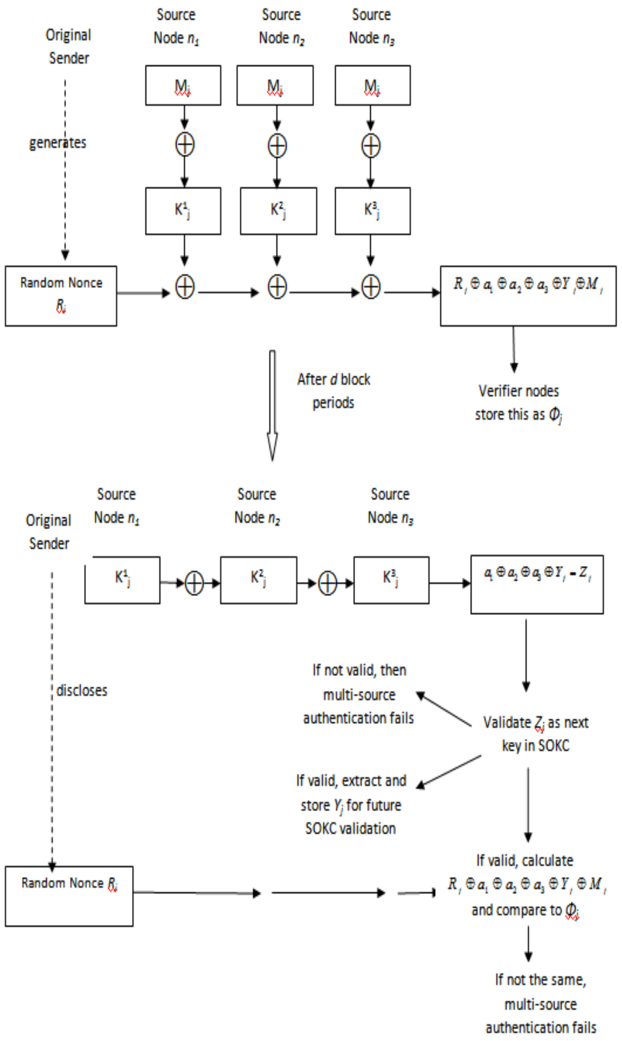

Original Sender
Source Node $n_1$
Source Node $n_2$
Source Node $n_3$
$M_j$
$M_j$
$M_j$
generates
$K^1_j$
$K^2_j$
$K^3_j$
Random Nonce $R_i$
$R_i \oplus a_1 \oplus a_2 \oplus a_3 \oplus Y_j \oplus M_i$
After $d$ block periods
Verifier nodes store this as $\Phi_j$
Original Sender
Source Node $n_1$
Source Node $n_2$
Source Node $n_3$
$K^1_j$
$K^2_j$
$K^3_j$
$a_1 \oplus a_2 \oplus a_3 \oplus Y_i = Z_i$
discloses
If not valid, then multi-source authentication fails
Validate $Z_j$ as next key in SOKC
If valid, extract and store $Y_j$ for future SOKC validation
Random Nonce $R_i$
If valid, calculate $R_i \oplus a_1 \oplus a_2 \oplus a_3 \oplus Y_i \oplus M_i$ and compare to $\Phi_j$
If not the same, multi-source authentication fails

**Figure 3. Semi-encrypted key pre-distribution and a delayed key disclosure along with the verifications at verifier nodes**

# 5 Acknowledgement

This work was supported in part by US Army Research Office (ARO) grant W911NF-12-1-0060, and in part by Seoul R&BD Program (No. JP110034).

# 6 References

[1] J. Deng, R. Han, and S. Mishra, "Enhancing Base Station Security in Wireless Sensor Networks", Technical Report CU-CS-951-03, University of Colorado, April 2003.

[2] S. Gandham, M. Dawande, R. Prakash, S. Venkatesan, "Energy efficient schemes for wireless sensor networks with multiple mobile base stations," IEEE Globecom '03, pp.377-381, January 2004.

[3] Y.-C. Hu, D. B. Johnson, and A. Perrig, “SEAD: Secure efficient distance vector routing for mobile wireless ad hoc networks,” in Proceedings of the 4th IEEE Workshop on Mobile Computing Systems and Applications (WMCSA 2002), June 2002, pp. 3–13.

[4] Y.-C. Hu, A. Perrig, and D. Johnson. Ariadne: A secure on-demand routing protocol for ad hoc networks. Wireless Networks Journal, 11(1), 2005.

[5] J. Hubaux, L. Buttyan, and S. Capkun, “The quest for security in mobile ad hoc networks,” in Proceedings of the ACM Symposium on Mobile Ad Hoc Networking and Computing (MobiHOC 2001), 2001.

[6] Chris Karlof and David Wagner, Secure Routing in Wireless Sensor Networks: Attacks and Countermeasures, In Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications, May 2003.

[7] Donggang Liu, Peng Ning, Multi-Level u-TESLA: A Broadcast Authentication System for Distributed Sensor Networks, Submitted for journal publication. Also available as Technical Report, TR-2003-08, North Carolina State University, Department of Computer Science, March 2003.

[8] S. Park, A. Bhatia, and J.-H. Youn (USA). Hop-Count based Location Discovery in Ad Hoc Sensor Networks. *In Proceeding* (424) Wireless Networks and Emerging Technologies - 2004

[9] A. Perrig, R. Canetti, D. Song, and D. Tygar, Efficient and secure source authentication for multicast. In Proceedings of Network and Distributed System Security Symposium, 2001.

[10] R. di Pietro, L. V. Mancini, Y. W. Law, S. Etalle and P. Havinga A directed diffusion-based secure multicast scheme for wirless sensor networks.First International Workshop on Wireless Security and Privacy (WiSPr'03)

[11] B. Przydatek, D. Song, and A. Perrig. SIA: Secure Information Aggregation in Sensor Networks. In Proc. of ACM SenSys 2003.

[12] Y. Ramamurthy, B. Xue, " A Key Management Protocol for Wireless Sensor Networks with Multiple Base Stations," Proceedings of the IEEE International Conference on Communications, 2008, ICC'08, pp. 1625-1629, Beijing, China, May 2008.

[13] Fan Ye, Haiyun Luo, Songwu Lu, Statistical En-Route Filtering of Injected False Data in Sensor Networks. IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 23, NO. 4, APRIL 2005

[14] L. Zhou and Z. Haas, “Securing ad hoc networks,” IEEE Network Magazine, vol. 13, no. 6, November/December 1999.

[15] Sencun Zhu, Sanjeev Setia1, and Sushil Jajodia. An Interleaved Hop-by-Hop Authentication Scheme for Filtering of Injected False Data in Sensor Network

# Forensic Acquisition and Analysis of VMware Virtual Hard Disks

Manish Hirwani, Yin Pan, Bill Stackpole and Daryl Johnson
Networking, Security and Systems Administration
Rochester Institute of Technology
mhirwani@gmail.com; {yin.pan; bill.stackpole; daryl.johnson}@it.rit.edu

**Abstract— With the advancement in virtualization technology, virtual machines (VMs) are becoming a common and integral part of datacenters. As the popularity and the use of VMs increases, incidents involving them are also on the rise. There is substantial research on using VMs and virtual appliances to aid forensic investigation, but research on the appropriate forensics procedures for collecting and analyzing evidence within a VM following is lacking.**

**This paper presents a forensically sound way to acquire and analyze VM hard disks. A forensics tool for analyzing VM snapshots and vmdk files is developed and has been proven to be forensically sound.**

**Keywords**

Digital forensics, Virtual Machines, virtual hard disk, Sleuthkit.

## 1. INTRODUCTION

Traditionally, computer systems such as desktops and servers have been considered physical devices. With the introduction of virtualization in the IT industry, this may no longer be the case.

With the benefits of virtualization, virtualization is becoming a widely adopted practice across organizations of various sizes [2, 13]. VMware is a popular provider of virtual machine (VM) software and holds a large share of the market. Products offered by VMware include VMware Workstation, VMware Server, and VMware ESXi among others. With the growing trend in virtualization, more and more production systems, workstations and desktops are being virtualized. Being a popular provider, VMware virtual environments are likely to be encountered by forensic investigators. To address the increase in exposure to VMs, this research will focus on acquisition and analysis of VMware products.

VMware VMs are implemented using virtual adapters for devices such as network cards, memory, etc. The VM, however, is stored in a set of files. VMware Workstation creates files with extension like virtual machine configuration (.vmx), virtual hard drive (.vmdk), snapshot of the virtual machines' memory (.vmem) etc [18]. The conventional methods for incident response and evidence acquisition - such as pulling the plug of a machine containing the VMs or suspending and resuming the VMs to gather evidence - may not be the best solution for forensic analysis of VMs since resuming a VM could potentially change both volatile and non-volatile evidence leaving the evidence to be NOT admissible to court.

In this paper, the authors propose a sound forensics methodology to acquire and analyze VMs hard disk evidence by utilizing the VMware artifact – the VM files. The rest of this paper is organized as follows. Section 2 reviews the existing work performed in the area of forensic analysis of VMs. A forensically sound procedure to acquire and analyze virtual hard disks is presented in Section 3. The authors also present a forensics tool for analyzing VM snapshots and vmdk files and prove it to be forensically-sound in section 3. In Section 4, the authors evaluate the proposed forensic analysis tool. Section 5 addresses the limitations of the proposed methodology. This paper is concluded in Section 6.

## 2. LITERATURE REVIEW

According to Kruse & Heiser [11], the conventional forensic process can be broadly classified into four main phases, namely Acquire, Preserve, Analyze and Report. The acquisition state of the process involves capturing as much volatile system data as possible, then powering down the system and creating a forensic image of all the remaining non-volatile storage devices that are found [5]. A forensic image of a device is a bit-by-bit copy of the drive. The bit stream copy can be either stored as a file on another device or can directly be copied to another drive of similar or greater capacity. The bit stream copy of the storage drive is generally acquired using a *dd* based tool [15]. This image is stored in a raw format supported by a *dd* or a propriety format which is typically based on *dd* [16]. The acquired image is either an identical copy of the storage device, for example dd image, or the data from the device which is stored in a format that can be used to evaluate the contents and presented in court as permissible evidence. Analysis can be conducted using any of the many open-source or propriety tools available such as Sleuthkit [17], Forensics ToolKit [1], EnCase [9] etc.

Most of the research conducted in the area of virtualization and forensics, makes use of VMs as forensics tools. VMs can be used to conduct analysis of evidence. VMs provide the examiner the ability to have a clean operating system without having to wipe a drive and install a fresh operating system on it for every new case. Helix [7] & Penguin Sleuth Kit Virtual Computer Forensics and Security Platforms [6] are popular Linux-based operating systems tailored for forensics acquisition and analysis. Both platforms are readily available as virtual appliances that can be used with VMware products. Similarly other virtual appliances are available which use virtualization to assist in conducting a forensic investigation.

Mrdovic et al. used a tool called Live View, which creates a VMware VM from a raw image of a drive or a physical drive [14]. Guo et al. use a similar process of using Live View to boot an image acquired by dd and use that to augment their static forensic methods [10]. This enables the investigator to boot up the disk in a virtual environment and gain an interactive, user-level perspective of the suspect’s environment. All this is done without modifying the image or the physical drive and is considered to be forensically sound.

As incidents related to VMs are on the rise, they have caught the attention of forensics experts. New methods to collect evidence from VMs are needed. Fiterman and Durick in their article titled “Ghost in the Machine: Forensic Evidence Collection in the Virtual Environment” point out that tools and options that enable an examiner to investigate virtual data are currently limited [8]. Similar concerns regarding the absence of forensics tools and procedures for VM analysis are raised and methodologies are proposed by Beek [4]. He also suggests a tool which compares the memory files (.vmem) of snapshots created by VMware products for any new files or processes.

Bares [2] in his research studied the amount of data that could be recovered from VMs in a NTFS partition. His research also showed that lesser amounts of data were recoverable if the VM was incorrectly shut down as compared to when it was shutdown gracefully. In this paper, we propose a solution that is able to acquire and analyze VM hard disk evidence on a running VM, without shutting down, in order to preserve the most evidence with the least number of modifications.

In conclusion, there is abundant research on using VMs and virtual appliances to aid forensic investigation, but research on collecting and analyzing evidence from VMs hard disk is lacking. The proposed forensics methodology provides forensics examiners a forensics sound procedure and tool to acquire and analyze VMs hard disk images using only the existing VM files.

## 3. FORENSICALLY-SOUND PROCEDURE FOR VIRTUAL DISK ACQUISITION AND ANALYSIS

The use of VMs in corporate and personal environments is rapidly increasing; 18% of servers were virtualized in 2009 and that grew to 25% in 2010. It is expected that by 2012, about half of the servers hosted will be virtualized and hosted virtual desktops will reach 49 million units by 2013 [13]. With the growing number of virtual systems it becomes imperative that a methodology to analyze virtual systems is developed. Many systems carry out critical tasks that cannot be stopped. If such systems are compromised, or are suspected of being compromised, they cannot be taken offline for analysis. In such a scenario it becomes important to conduct a live analysis of the system. However, when a live analysis is carried out, the investigator may change information that resides in memory and any remaining open network connections will be terminated. According to Kurse & Heiser [11], the common practice to conduct forensic analysis of a physical machine is to take the machine offline at some point. The machine’s hard disk is then imaged, and its data acquired for analysis.

When VMs are involved in an incident, the VM is usually suspended or a snapshot of the machine is created to preserve the processes and network status for forensics analysis. There are two paths that a forensics investigator can follow: a) resume the suspended VM and use the normal procedure to acquire and analyze the live machine, and b) analyze the VM files without resuming the suspended machine. However, the method of resuming a suspend VM before acquisition may potentially change the evidence, leading to the possibility that the evidence will NOT be legally admissible.

In this section, the authors describe a forensics solution to acquire, preserve, and analyze snapshots of disk images using VM suspend or snapshot of VM files created by the VMware utility without resuming or shutting down VMs.

### 3.1 Virtual Disk Acquisition

The aim of forensic image acquisition is to minimize contamination and ensure legally-admissible evidence. To accomplish this, the digital evidence acquisition process has to follow an appropriate procedure. Acquiring non-volatile data from a physical hard disk entails many steps [11]. A machine is first powered off by disconnecting the power supply from the machine (i.e. pulling the plug). The hard disk is then removed from the suspect machine and connected to a forensic analysis machine. The hard disk is then imaged using any of the many tools available for imaging a disk such as dd, FTK Imager, EnCase, etc. This image is then used by a forensics investigator to conduct an analysis of the events the machine may have experienced.

When working with suspended VMware images, there are two options for acquiring the virtual disks: resuming the suspended system, then use bit-by-bit copy or to directly work with the VMware .vmdk and snapshots files. The problem with resuming a VM is that during the resume process, many files stored on the hard disk are changed, which may destroy evidence. Another disadvantage of resuming the suspended VM is the loss of information stored in the memory as the state of the VM changes. Such information could be vital to the investigation being carried out.

To overcome the shortcomings of resuming the suspended VM, the better solution is to create a snapshot of the VM and then work directly with the VM files that are stored on the host system. Upon taking a snapshot, the state of the hard disk is preserved and any changes to the disk are stored in a separate file. The virtual memory of the VM is stored in a file and any state changes are written to another virtual memory file. Following this procedure, we ensure that the evidence is preserved and can be presented to court as forensically sound and admissible evidence.

Both EnCase and FTK support conversion of .vmdk files to raw (dd) format. When FTK is pointed to a snapshot for converting it to a raw image, it converts the snapshot along with any previous snapshots and the base vmdk files (see Figure 1).

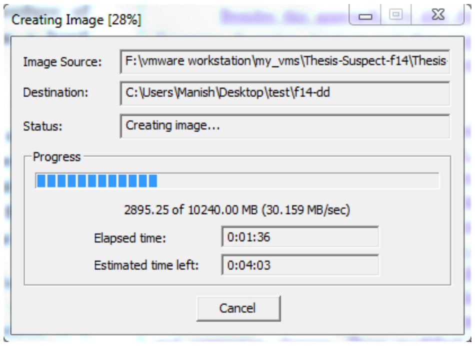

Creating Image [28%]

Image Source: F:\vmware workstation\my_vms\Thesis-Suspect-f14\Thesis

Destination: C:\Users\Manish\Desktop\test\f14-dd

Status: Creating image...

Progress

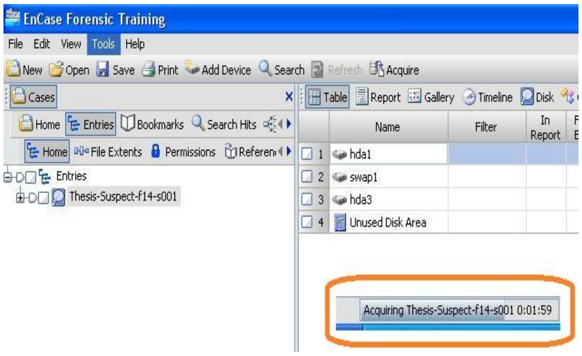
2895.25 of 10240.00 MB (30.159 MB/sec)

Elapsed time: 0:01:36

Estimated time left: 0:04:03

**Figure 1: FTK acquiring .vmdk in raw (dd) format.**

When EnCase is pointed to the base clean vmdk files, it successfully converts them to raw/dd format, but when EnCase is pointed to a later *snapshot*, it only converts the delta that is created after snapshoting a VM. As a result, to include the previous snapshots and the base images as well, one has to assimilate a flat vmdk file by using utilities that are packaged with VMware Workstation, namely vmware-vdiskmanager.exe. Vmware-vdiskmanager.exe creates one vmdk file that includes: the selected snapshot, any previous snapshots and the base vmdk files. The resulting single vmdk file can then be converted to raw/dd format using EnCase (see Figure 2).


EnCase Forensic Training

Entries
Thesis-Suspect-f14-s001

| | Name | Filter | In Report |
|---|---|---|---|
| 1 | hda1 | | |
| 2 | swap1 | | |
| 3 | hda3 | | |
| 4 | Unused Disk Area | | |

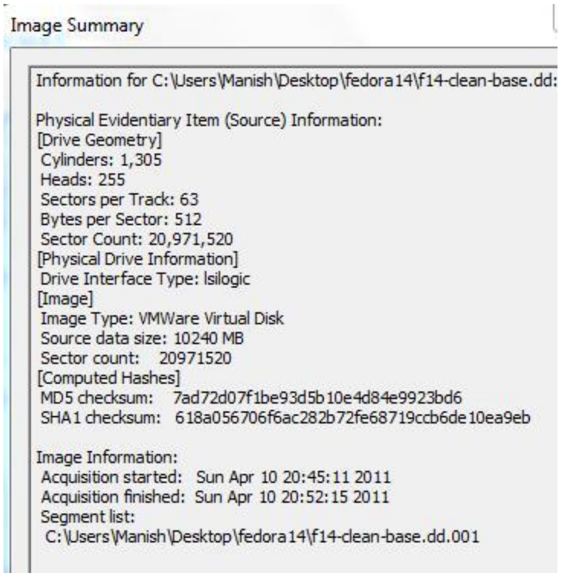
Acquiring Thesis-Suspect-f14-s001 0:01:59

**Figure 2: EnCase acquiring .vmdk in raw (dd) format.**

*3.1.1 Forensically Sound Virtual Disk Acquisition*

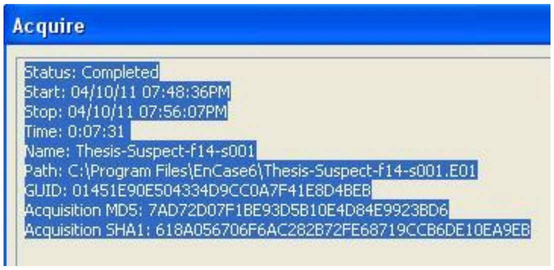
It is now clear that both FTK and EnCase can acquire the contents of the virtual hard disk without shutting down the VM if a snapshot of this machine is created at the time of incident response. The next question is whether this solution is forensically sound. In other words, do FTK Imager and EnCase change the original VM disk image?

In general, both FTK imager and EnCase require a write blocker device to image a live physical drive. However, since VMware virtual disks are implemented as files, can they be acquired without the use of a write blocker device using FTK and EnCase? In this experiment, the authors studied the functionality of both tools when applied to VM file conversion. In particular, the authors used both tools to create raw images for VM hard disks and calculated hashes of the raw images. We found that both tools produced the matching MD5 and SHA1 hashes, as can be seen from figures 3a & 3b. Therefore, we conclude that VM hard disk files can be safely converted to raw/dd images using either tool without relying on a write block device.



```
Image Summary

Information for C:\Users\Manish\Desktop\fedora14\f14-clean-base.dd:

Physical Evidentiary Item (Source) Information:
[Drive Geometry]
 Cylinders: 1,305
 Heads: 255
 Sectors per Track: 63
 Bytes per Sector: 512
 Sector Count: 20,971,520
[Physical Drive Information]
 Drive Interface Type: lsilogic
[Image]
 Image Type: VMWare Virtual Disk
 Source data size: 10240 MB
 Sector count:    20971520
[Computed Hashes]
 MD5 checksum:    7ad72d07f1be93d5b10e4d84e9923bd6
 SHA1 checksum:   618a056706f6ac282b72fe68719ccb6de10ea9eb

Image Information:
 Acquisition started:   Sun Apr 10 20:45:11 2011
 Acquisition finished:  Sun Apr 10 20:52:15 2011
 Segment list:
  C:\Users\Manish\Desktop\fedora14\f14-clean-base.dd.001
```

**Figure 3a: MD5 & SHA1 sums obtained from FTK Imager**



```
Acquire

Status: Completed
Start: 04/10/11 07:48:36PM
Stop: 04/10/11 07:56:07PM
Time: 0:07:31
Name: Thesis-Suspect-f14-s001
Path: C:\Program Files\EnCase6\Thesis-Suspect-f14-s001.E01
GUID: 01451E90E504334D9CC0A7F41E8D4BEB
Acquisition MD5: 7AD72D07F1BE93D5B10E4D84E9923BD6
Acquisition SHA1: 618A056706F6AC282B72FE68719CCB6DE10EA9EB
```

**Figure 3b: MD5 & SHA1 sums obtained from EnCase**

*3.1.2 Guest System Time Skew*

When the guest system (the VM) is being acquired it is critical for the incident response professional to record the time of the guest operating system as well as that of the host operating system. The time of the guest operating system could be skewed and if this is recorded the forensic examiner can make more definitive and accurate statements about activities that may have taken place. If the guest system clock is synchronized with the host system clock, the incident response professional should make sure to check if the host system time is correct else he should record the skew. If the guest is using an external source besides the host system clock to synchronize the time, any skew that exists should be recorded. The Forensics Snapshot Tool takes into account the time skew of the guest operating system to conduct analysis.

### 3.2 Forensics Snapshot Analysis Tool

With the acquired image from section 3.1, one can use open-source or commercial forensics tools, for example, EnCase, FTK, and Sleuthkit to conduct a forensics analysis. However, with the additional VM files created from the VM and features supported by VMware, are there other efficient techniques that could assist forensics investigation?

VMware offers the Snapshot feature that allows one to freeze the state of a VM at a given point of time [18]. In this research, the authors developed a *Forensics Snapshot Analysis* tool that compares the snapshot with a pre-staged (recorded at an earlier time) snapshot to identify possible malicious activities.

This tool is written as a Bash script and incorporates existing tools, such as Sleuthkit and md5sum, to verify the integrity of the evidence and also automate parts of the forensic analysis. The script extracts files from both the clean and the compromised image snapshots. Then a comparison is made to determine the changes detected - such as files created, changed or deleted. The tool is also capable of identifying MAC time changes, content changes and permission changes. These modified files are reported and can then be further investigated by a forensics examiner.

To prove that the Forensics Snapshot Analysis tool is a forensically sound, the authors computed the MD5 checksum of the image before and after applying this tool and found that the hash results are identical. This validates that the tool does not modify the evidence files or their contents.

## 4. EXPERIMENTS AND ANALYSIS

### 4.1 Experiment Setup

**Suspect virtual Machine:**

For our experiments, VMware Workstation 7.0.1 was installed on a Windows 7 Home Premium operating system with a New Technology File System (NTFS) partition. A Fedora 14 operating system was installed in VMware Workstation using an extended file system (ext), viz. ext3, with 10GB of disk space. Once the system was installed a snapshot was taken to establish a clean base system. Various packages were installed and changes to various files were made. Changes to file permissions such as execute bit, set user ID, etc. were also made to emulate a real world scenario where a malicious user might change file permissions to gain access to restricted parts of a system. Once these actions were performed, another snapshot of the VM was made. Since FTK Imager is a better solution to convert .vmdk and snapshot files to a completed raw image, we used FTK Imager to convert both the snapshots and create a raw disk images.

**Forensics Analysis Machine:**

Another Fedora 14 operating system was installed as a VM which was used for forensics analysis with the *Forensics Snapshot Analysis Tool* and its dependencies such as Sleuthkit and md5sum installed.

### 4.2 Analysis and Results

The experiment follows the processes shown in Figure 4.

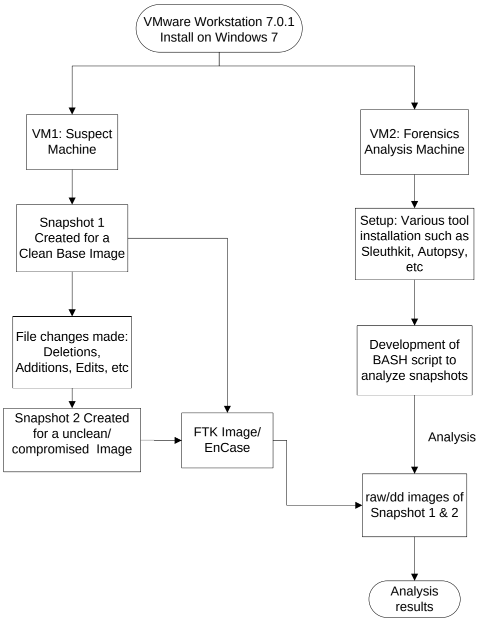
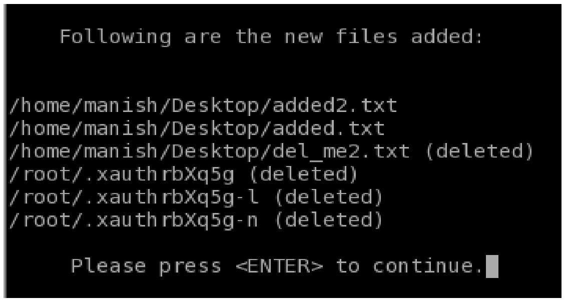

VMware Workstation 7.0.1 Install on Windows 7
VM1: Suspect Machine
VM2: Forensics Analysis Machine
Snapshot 1 Created for a Clean Base Image
Setup: Various tool installation such as Sleuthkit, Autopsy, etc
File changes made: Deletions, Additions, Edits, etc
Development of BASH script to analyze snapshots
Snapshot 2 Created for a unclean/ compromised Image
FTK Image/ EnCase
Analysis
raw/dd images of Snapshot 1 & 2
Analysis results

**Figure 4: Flow Chart of the processes**

As shown in Figure 5 below, the Forensics Snapshot Analysis tool successfully analyzes and compares snapshots of the same virtual machine taken at different points in time. Using the tool, a forensics examiner can generate a list of files that have been added, deleted, modified and changed by comparing the snapshots. The tool produces formatted reports of each analysis procedure it carries out. The forensics examiner can then decide on areas of further investigation based on the items of interest generated by the Forensics Snapshot Analysis tool.

The Forensics Snapshot Analysis tool is forensically sound and does not modify the raw files in any way. This can be proven by computing the hashes of the raw files after analysis is complete. The MD5 & SHA1 hashed computed for the raw files after the tool has analyzed the raw files match the hashes computed before analysis was run.



```
    Following are the new files added:


/home/manish/Desktop/added2.txt
/home/manish/Desktop/added.txt
/home/manish/Desktop/del_me2.txt (deleted)
/root/.xauthrbXq5g (deleted)
/root/.xauthrbXq5g-l (deleted)
/root/.xauthrbXq5g-n (deleted)

     Please press <ENTER> to continue.
```

**Figure 5: Output of *Forensics Snapshot Analysis* tool showing a list of recently deleted files**

### 4.3 Other usage of this tool

This tool can also be used to study OS behavior. One example would be inode reallocation. In some OSes, if a new file is created soon after deleting an existing file, the inode used by the deleted file is marked as "not in use" and could be assigned to a new file. This result can be seen from Figure 5 above when the file del_me.txt is deleted and added2.txt is created. The inode of file del_me2.txt is assigned to added2.txt.

## 5. LIMITATIONS: POSSIBLE METHODS OF OBFUSCATION

The Forensics Snapshot Analysis tool relies heavily on the MAC time of files to generate files of interest. A possible method for obfuscation could be the use of a tool that does not modify MAC times of files. TrueCrypt is one such tool. Depending on how the preferences are set, when a TrueCrypt volume is modified it can be configured to update only the change time, leaving the modification (mtime) and access (atime) times unchanged. The authors used TrueCrypt with their script to study this behavior of TrueCrypt and added features to the script which now checks for change time (ctime) difference between the files found in both snapshots.

The Forensics Snapshot Analysis tool cannot view the contents of files that have been encrypted. The tool will still list files which have differing modification & change times but the examiner will not be able to view the contents of the file, if it is encrypted. A method to analyze encrypted files and volumes can be developed and incorporated in the tool.

## 6. CONCLUSIONS AND FUTURE WORK

The authors studied the solutions for acquiring and analyzing live virtual machines based on VM files. A forensically sound procedure to acquire virtual disk images is provided. In addition, the snapshot analysis tool, Forensics Snapshot Analysis, was developed. It is a powerful tool for forensics investigators to analyze VM hard disk images and present pertinent evidence in court. This tool can also be useful in incident response to confirm a breach and to carry out further forensic analysis.

The Forensics Snapshot Analysis tool can also be used for academic and training purposes. Students are taught that booting a suspect machine or VM that has been shutdown or suspended can modify the contents, changing potential evidence. Snapshots of a shutdown or suspended VM can be acquired before and after booting and these snapshots can be analyzed using the tool developed which will practically demonstrate why booting is not a forensically sound procedure.

Several avenues can be pursued as an extension to this research. The effect of encrypted files and volumes on the analysis conducted by this tool can be studied further. Changes can be made to the developed tool to handle other file systems such as FAT, NTFS and other file systems.

## 7. REFERENCES

[1] Access Data Forensics ToolKit, Retrieved from http://www.accessdata.com/forensictoolkit.html, 2010

[2] Bares, R., “Hiding in a virtual world using unconventionally installed operating systems”, IEEE International Conference on Intelligence and Security Informatics. Dallas, TX, 2009.

[3] Barrett, D., and Kipper, G., “Investigating Dead Virtual Environments, Virtualization and Forensics”, Syngress, Boston, 2010, Pages 83-107.

[4] Beek, C., “Virtual Forensics”. Retrieved from: http://securitybananas.com/wp-content/uploads/2010/04/Virtual-Forensics_BlackHatEurope2010_CB.pdf, 2010.

[5] Brown, C. L. T., “Computer Evidence: Collection & Preservation”, Hingham, MA: Charles River Media, 2005

[6] Ebaca, “Penguin Sleuth Kit Virtual Computer Forensics and Security Platform”. Retrieved from http://www.vmware.com/appliances/directory/249, 2010.

[7] E-fense, Cyber Security & Computer Forensics Software Page. Retrieved from http://www.e-fense.com/helix/, 2010.

[8] Fiterman, E. M., & Durick, J. D., “Ghost in the machine: Forensic evidence collection in the virtual environment”, Digital Forensics Magazine, 2, 2010, pp. 73–77.

[9] Guidance Software, EnCase Forensic. Retrieved from: http://www.guidancesoftware.com/forensic.htm, 2010.

[10] Guo, H., Huang, D., & Zhang, Y. (2012). Implication of Virtualization Technologies in Computer Forensic. Energy Procedia, Volume 13. Pages 4133-4137, ISSN 1876-6102

[11] Kruse W. G., & Heiser, J. G., “Computer Forensics: Incident Response Essentials (1st ed.)”, Addison Wesley Professional, 2002.

[12] Metasploit, Metasploit Anti-Forensics Project. Retrieved from http://www.metasploit.com/research/projects/antiforensics/, 2010.

[13] Messmer, E., “Gartner predicts nearly half of server workloads will be virtualized”, Network World. Retrieved from http://www.networkworld.com/news/2009/102009-gartner-server-virtualization.html, 2010.

[14] Mrdovic S., Huseinovic, A., and Zajko, E., “Combining static and live digital forensic analysis in virtual environment. Information”, Communication and Automation Technologies, 2009. ICAT 2009. XXII International Symposium on, vol., no., pp.1-6, 29-31 Oct.2009.

[15] Nelson B., Phillips A., Enfinger F., and Steuart, C., “Guide to Computer Forensics and Investigations”,

Second Edition. Boston, MA: Thomson Course Technology, 2006

[16] Rude, T., "DD and Computer Forensics", Retrieved http://www.crazytrain.com/dd.html, 2010.

[17] sleuthkit.org, Sleuthkit. Retrieved from http://www.sleuthkit.org/sleuthkit/, 2010.

[18] VMware, "What Files Make Up a Virtual Machine?" http://www.vmware.com/support/ws55/doc/ws_learning_files_in_a_vm.html, 2010.

# ZLA: Towards Zero-Leak Private Information in Online E-commerce Transactions

Hiroshi Fujinoki, Christopher A. Chelmecki, and David M. Henry

*Department of Computer Science*
*Southern Illinois University Edwardsville*
*Edwardsville, Illinois 62026-1656, USA*
E-mails: *hfujino@siue.edu, chris.chelmecki@gmail.com, dahenry@siue.edu*

## Abstract

*In this paper, we propose new security architecture, called ZLA (Zero Leak Architecture), which guarantees zero possibility of customers' private information leak from online shopping web sites even in the worst case: the security administrator of the e-commerce web site becomes an attacker. The proposed security architecture assumes three parties of merchants, shipping carriers, and credit card companies. ZLA covers a whole procedure in an online shopping transaction from the time a customer browses an online shopping site, up to the time the customer confirms delivery of the correct products. Our performance studies suggested that the cost factor of running ZLA is 1.8 (1.8 times more computational power and faster network) to achieve the same queuing delay and transaction throughput compared to the existing architecture, where there is no protection against the customers' private information leaks in the worst case. The proposed architecture will contribute to flourish of e-commerce by eliminating the risk of private information leaks with relatively low cost of more powerful computations, especially as predicted by Moore's law.*

## 1. Introduction

The risk of private information leaks from servers has been a serious issue. Our previous survey revealed Internet users' conflict between their favor to online shopping and fear for their personal information leaks [1]. Many of online shopping sites are currently implemented based on the client server model, which practically forces their customers to provide their personal information to the servers owned by merchants. This implies that there is nothing customers can do to protect their privacy once they provide their personal information to the servers but to trust the server side security administrations.

The problem is that it is difficult for customers to know how serious the server side is for protecting customers' information. If some merchants do not look serious in protecting customers' privacy, all what we should do is just to avoid them, but how can we know for sure which merchants are good ones? In an extreme case, if the security administrator commits stealing the personal information from the merchant's server, there is nothing customers can do to proactively stop such crimes.

In designing a security architecture that guarantees no information leak, we categorized attackers in the two groups of those who have physical accesses to the server host computers where customers' information is stored and those who do not. We called the former "internal attackers" and the latter "external attackers". Internal attackers are server side system administrators, or someone (in the facility the administrators work) who manage to steal the access rights of the administrators by some means, such as duplicating the server room key or accidentally watching a memo that shows the administrator's password on it.

Handling customers' information in electronic commerce has another dimension. Any possible dispute regarding a transaction, such as a claim about delivery of a damaged product, must be resolved while any of the three parties should not have access to the complete information about each transaction. We defined "the complete information about a transaction" to mean "who ordered what product". Also from the view point of the legal issues, the ability to reconstruct the complete information for each transaction is mandatory, since the United States Money Laundering Act (U.S. Code 12, Section 1829) prohibits anonymous transactions valued over $100 [2]. The complete information about each such online transaction must be reconstructed and disclosed to a legislative authority up on a court's disclosure order, while none of the three parties involved in a transaction should have access to the complete information without a court order.

The recent security incidents caused serious damages to electronic commerce customers due to the leakages of their privacy information. A new solution is needed, which guarantees no leakages of customers' private information no matter what happen in the server side, while the solution satisfies the legal requirement. This paper proposes a new solution, called ZLA (zero leak architecture), which eliminates the possibility of leaking customers' private information in a sense that it is minimized to a level customers can assume that such leakages will not happen at the same time the solution satisfies the legal requirement.

The rest of this paper is organized as follow. Section 2 describes the existing related work. In Section 3, the zero leak security architecture for online shopping applications is described by applying the architecture to online transactions typically performed in online shopping web sites today. In Section 4, ZLA is analyzed for its protections against different types of unauthorized and malicious accesses to customers' private information. This section also describes our performance analyses of ZLA for its queuing delay and throughput. Section 5 summarizes the conclusions, followed by the references.

## 2. Existing related work

The risk of private information leaks from servers has been a serious issue. It becomes a significant issue especially when service vendors and data storage providers are separated in cloud computing. Many of the existing data protection methods, such as encrypting a whole disk [3] and separation of stored data from code execution [4], will be effective to external attackers, but neither of them will be an effective solution against threats by internal attackers.

To secure personal information from insiders' threats, encrypting personal information using customers' biometric information without storing decryption keys in the server side have been proposed by Uludag and Zhang [5, 6]. Although these solutions will be effective in hiding customers' information in e-commerce applications, none of the merchants, shipping carriers, or credit companies except the customers have access to the complete customer information, which fails to satisfy the legal requirement.

Mazières proposed a distributed file system, SUNDR, which prohibits unauthorized users from performing any read access to the files stored in their storage devices [7]. SUNDR chops the contents in each file and spread the chops to multiple segments, called "data blocks" in such a way that the mapping of the segments in a file is hidden by "virtual i-node".

Mattsson proposed column-level database encryption to protect users' information from both external and internal attackers [8]. The column-level database encryption is performed by encrypting security sensitive data, calculating the HMAC hash of the data, and attaching the hash to another column in the same row of the table. The owner of the data uses the hash to find target data by matching the hash value. The decryption key for the secure data is not stored in the database, but in the applications that use the database.

Although the above two solutions are effective for protecting customers' information from external attackers, the solutions will not be effective against information threats by the internal attackers. Since the solutions are applied to the system level, neither of them can protect customers' information after customer's information is produced by an application but before such information is passed to the system-level solutions. This implies that the true zero-leak protection should also cover the application level, especially in the client side of applications.

Secure Electronic Transaction (SET) protocol has been proposed to secure electronic payments using credit cards for online transactions such as purchasing goods through merchants' web sites [9]. Since its introduction, several extensions to SET have been proposed, some of which were proposed for realizing customers' anonymity in SET [10].

Rennhard proposed new network application architecture for online shopping using SET, which hides customers' network address from the merchants and credit card companies to protect customers' privacy [10]. Rennhard's solution consists of an overlay network called "Pseudonymity Network (PN)", which uses intermediate proxies and a security protocol (for key exchanges and encryptions) called "Pseudonymous Secure Electronic Transaction (PSET) protocol" for the messages exchanged by the involved parties to hide customers' network addresses and identities.

Although SET and its enhancements will be effective for protecting customer payments by hiding customers' credit card information from the merchants, none of them prevents customers' information from leaking at their servers if shipping goods is handled through merchants. As most of the online customers do not pick up the products at the online stores, supporting shipping the products while the merchants do not have access to customers' identity is another "must" to achieve true zero leak online shopping sites. Inclusion of product shipping as a part of zero leak design is essential along with the context of service integration in service oriented architectures (SOAs), where various services in online shopping, such as payments for the goods and shipping the goods to customers, should be integrated to automate businesses for better transactional turnarounds and for reducing running costs [11, 12].

To enhance customers' anonymity in online transactions, Tygar introduced the concept of "certified delivery" by enhancing "goods atomicity" [13]. The term "certified delivery" means that a customer will receive the product if and only if the money for the product is transferred from the customer to the merchant (this concept was called "goods atomicity") and that the delivered goods must be correct. The true challenge for realizing certified delivery is in the trade-off between disclosures of customer's information to all the involved parties and achieving anonymity. The more information about each transaction is disclosed to each party, the easier it will be for them to handle any dispute for each transaction, but it obviously sacrifices anonymity in each transaction.

Zhang proposed three security enhancements to SET [14]. They are protection of the customers' private key using Active-X, database encryption, and guaranteeing “certified delivery” in SET. Active-X is used to protect customers’ private key. When the private key of a customer is referenced by the customer’s local host process, active-X prompts the user to enter the user’s secret key to approve the use of private key.

Lekkas proposed a new approach for securing electronic payments, called “E-coins” [15]. E-coin is a new approach in that it performs payments by securely transferring virtual currency (i.e., E-coins) issued by an individual instead of transferring payment information for actual currency as SET does. One of its features is “untraceable”, meaning that it is not possible to identify who (as a real human) issued the currency and for what product it was used for, although the first who proposed the concept as “on-line virtual currency” was Chaum [16] in our best knowledge. E-coins is proposed as an off-line electronic payment method, meaning that payment transactions can be completed without contacting a central bank. An electronic payment system proposed by Eslami [17] is based on the same concept.

Despite their effectiveness in improving anonymity in online transactions, none of the above solutions guarantees no information leak when a shipping of the ordered goods is handled by merchants. Thus, a new security architecture that guarantees zero possibility for customers’ information leak, from the time a customer browses an online shopping site, up to the time the customer confirms delivery of the correct product(s) ordered by the customer, is needed.

## 3. Description of ZLA

The proposed architecture assumes the four parties of the customer, the merchant, the shipping carrier, and the credit card company, none of whom, except the customer, has access to the complete information regarding each online transaction. The customer is a user who visits an online shopping web site where the user makes orders to particular products. The merchant is the party who owns and maintains an online shopping web site to take orders from customers. The carrier is the party who sends the products the merchant sells to their customers. The credit card company is the party who makes payment to the merchant for the products a customer ordered.

ZLA must allow merchants and the other parties to handle any accident in a transaction, while no single party owns complete information about each transaction. For example, if a customer has a question about the payment amount charged by the credit card company for a particular order, the merchant and the credit card company need to collaborate on the investigation. If the products are lost during their shipping, the customer may request the merchant to provide its shipping and product information to the carrier for an insurance claim. For a disclosure order from a court, the three parties collaborate. To handle such situations, there must be a mechanism that logically integrates users’ information about a particular transaction while no single party accesses to the complete information about each transaction.

Each of the merchant, the carrier, and the credit card company is assumed to have their digital certificate that contains their digitally signed public key. Merchants have obtained the digital certificates of the shipping carriers and the credit card companies. The merchants make those certificates available to their customers through their online shopping sites. It is also assumed that the web server and browser used by the merchants and the customers support secure connections (e.g., SSL).

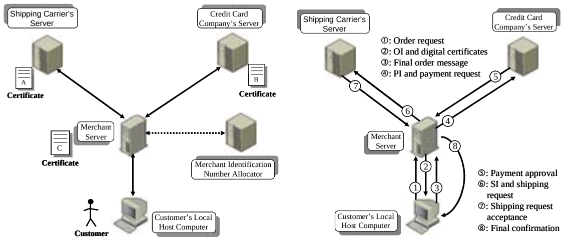
ZLA does not require customers to have their digital certificate. It may not be a reasonable requirement to mandate customers to have digital certificate, since requiring digital certificate to customers, as many existing secure transaction protocols do, costs customers and such solutions will not be widely adopted.

Table 1 lists the major cryptographic keys used in ZLA. The keys with a star are those provided in the digital certificate endorsed by a SA (security authority). “OT” in the key names means “one time”. For each transaction, eight one-time keys are created, but they are all “one-time”, meaning that once a transaction is completed (when the merchant server finalizes a transaction), they are discarded.

| Name of the key | Symbol | Purpose |
|---|---|---|
| Merchant public * | $M_P$ | To encrypt the messages to this merchant |
| Merchant private * | $M_S$ | To decrypt the messages to this merchant |
| Credit card company public * | $C_P$ | To encrypt the messages to this credit card company |
| Credit card company private * | $C_S$ | To decrypt the messages to this credit card company |
| Carrier public * | $S_P$ | To encrypt the messages to this carrier |
| Carrier private * | $S_S$ | To decrypt the messages to this carrier |
| Customer OT public for merchant | $U_{OTP\text{-}M}$ | To decrypt signature for the final order message |
| Customer OT private for merchant | $U_{OTS\text{-}M}$ | To encrypt signature for the final order message |
| Customer OT public for cc company | $U_{OTP\text{-}C}$ | To decrypt signature for the protected PI message |
| Customer OT private for cc company | $U_{OTS\text{-}C}$ | To encrypt signature for the protected PI message |
| Customer OT public 1 for carrier | $U_{OTP\text{-}S1}$ | To encrypt the tracking number from the carrier |
| Customer OT private 1 for carrier | $U_{OTS\text{-}S1}$ | To decrypt the tracking number from the carrier |
| Customer OT public 2 for carrier | $U_{OTP\text{-}S2}$ | To decrypt signature for the protected SI message |
| Customer OT private 2 for carrier | $U_{OTS\text{-}S2}$ | To encrypt signature for the protected SI message |

**Table 1** – List of the major cryptography keys used in the proposed zero leak architecture

ZLA assumes that each merchant is assigned a unique identification number. To guarantee the uniqueness in the numbers, a central organization that manages and assigns the merchant identification numbers is proposed. The organization of the architecture is shown in Figure 1 (a).


Shipping Carrier's Server
Credit Card Company's Server
A
Certificate
B
Certificate
Merchant Server
C
Certificate
Merchant Identification Number Allocator
Customer
Customer's Local Host Computer
Shipping Carrier's Server
Credit Card Company's Server
①: Order request
②: OI and digital certificates
③: Final order message
④: PI and payment request
⑤: Payment approval
⑥: SI and shipping request
⑦: Shipping request acceptance
⑧: Final confirmation
1
2
3
4
5
6
7
8
Merchant Server
Customer's Local Host Computer

(a) Organization of the zero-leak online shipping site

(b) Procedure in the message exchanges

**Figure 1** - The organization of ZLA (a) and its procedure (b)

The procedure of the zero leak transaction is as follow. First, a customer visits the online shopping site operated by a merchant. After the customer's browser and the merchant web server establishes an application level secure connection, the browser downloads the merchant's digital certificate to confirm the merchant's identity. Then, the customer specifies the product names, the number of the products, the carrier, and the credit card she wants to use for this transaction (message ① in Figure 1 (b)).

When the customer finishes entering the preliminary order information, the merchant server calculates the total payment amount and creates an "order token" for this order. The order token is the information that uniquely identifies each transaction in all the four parties. The order token consists of the unique merchant identification number, the date of the token's creation in year, month and day, followed by a unique transaction number issued by the merchant.

Once the order token is created, the merchant server creates "order information (OI)" that contains the details of the order specified by the customer, such as the product names, the number of the products, the carrier name, and the credit card name. The merchant then sends the OI, the order token, and the digital certificates of the shipping carrier and the credit card company in one message, called the "order confirmation" to the customer through a secure connection (message ② in Figure 1 (b)).

When the customer's browser receives the order confirmation, the client side process displays the contents in the order confirmation and asks the customer's approval for the order. Then, the public keys of the carrier ($S_P$) and the credit card company ($C_P$) are extracted from their digital certificates. The customer side process constructs the four pairs of one-time asymmetric keys followed by construction of the final order message.

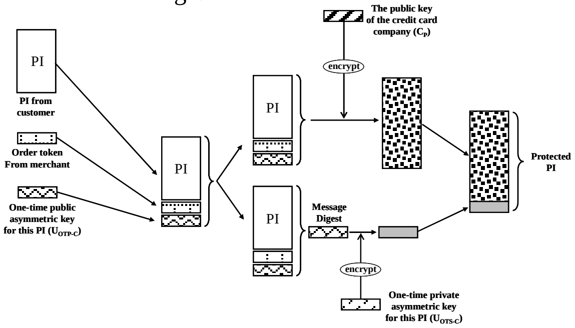


**Figure 2** – Structure and the construction of the protected SI segment for the selected carrier

The final order message is constructed as follow. First, the customer creates shipping information (SI) for the carrier. SI contains the mailing address of the customer, type of shipping (express, insured, etc), and other information needed for shipping the ordered products. Then, the customer side process constructs the protected SI, which contains the SI, the order token from the merchant, and the two one-time asymmetric public keys for the carrier ($U_{OTP-S1}$ and $U_{OTP-S2}$). The customer digitally signs the whole data structure using her first one-time asymmetric private key ($U_{OTS-S1}$), which is then encrypted by the carrier's public key ($S_P$). The digital signature is attached to the encrypted data structure to make it "protected SI" (Figure 2).

The payment information (PI) is prepared by the customer in a similar way to SI. PI contains the information needed to acquire payment approval from the credit card company, such as the credit card number, the name of the card holder, expiration date, and etc. The protected PI is created in the same way as SI, except that it contains only one one-time public key, $U_{OTP-C}$ (Figure 3).

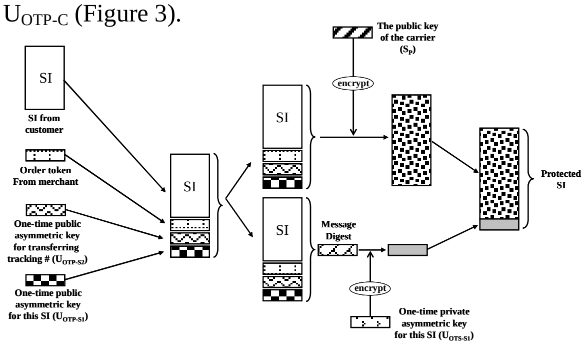


**Figure 3** – Construction of the protected PI

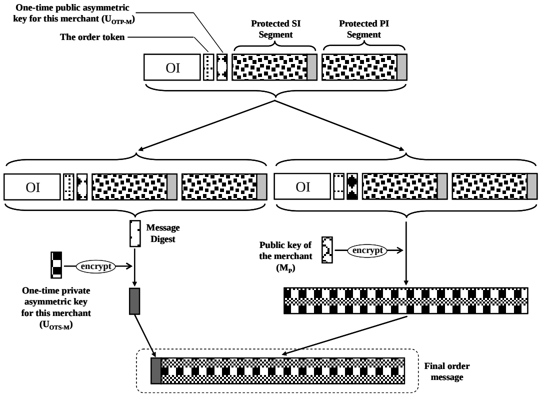


**Figure 4** – Construction of the final order message

After the customer's local process prepares the protected SI and the protected PI, it constructs the final order message by concatenating, the copy of the OI from the merchant, the order token from the merchant, the one-time public key ($U_{OTP-M}$), the protected SI, and the protected PI. Then, the local process calculates the message digest of the whole final order message, which is encrypted by the one-time private key of the merchant ($U_{OTS-M}$). After that the final order message is encrypted by the merchant's public key ($M_P$). The encrypted message digest is attached to the encrypted final order message. Finally the whole structure is sent to the merchant using a secure connection (message ③). The procedure of constructing the final order message is shown in Figure 4.

When the merchant receives the final order message, the merchant decrypts the whole message using its private key ($M_S$). The merchant extracts the OI, the order token, its one-time public asymmetric key ($U_{OTP-M}$), the protected SI, and the protected PI. Using his one-time public asymmetric key ($U_{OTP-M}$), the merchant decrypts the encrypted hash digest and tests the integrity of the whole message. Then, the merchant concatenates the protected PI, the order token, and the requested payment amount. The merchant sends the information to the credit card company (message ④). This message should be digitally signed by the merchant. Note that the merchant can not decrypt the protected PI or the protected SI, since their decryption requires the credit card company's private key ($C_S$) or the shipping carrier's private key ($S_S$).

When the credit card company receives the protected PI from the merchant, the credit card company decrypts the protected PI using its private key ($C_S$) and confirms the integrity of the protected PI using the one-time public key ($U_{OTP-C}$) in the protected PI. After the integrity is confirmed, the credit card company identifies the customer's account and compares if the order token and the payment amount extracted from the protected PI match those sent from the merchant. If they match, the credit card company examines the customer's payment credential to approve the payment request. The credit card company constructs the approval message that contains the same order token, and the approved payment amount. The message is digitally signed and encrypted by the credit company. Then, it is sent to the merchant with the digital signature (message ⑤).

After the merchant receives the payment approval from the credit card company, the merchant sends the protected SI and the order token together to the carrier specified in the OI (message ⑥). When the carrier receives the message, it decrypts the protected SI using its private key ($S_S$) and examines the signature using the second one-time public key from the customer ($U_{OTP-S2}$). The carrier issues a tracking number for this shipping. The tracking number and the order token are concatenated together and they are encrypted by the first one-time asymmetric public key from the customer ($U_{OTP-S1}$). Then the information is sent to the merchant as an acknowledgement for the shipping request (message ⑦). When the merchant receives the acknowledge from the carrier, the merchant sends the encrypted tracking number and the order token to the customer (message ⑧). The merchant does not have access to the tracking number, since it is encrypted by the one-time key issued by the customer ($U_{OTS-S1}$).

When an accident happens in a transaction, any party who is involved in each such accident can reconstruct the necessary information about the transaction using the order token. For example, if a customer has a question about payment amount charged by the credit card company for a particular order, the customer contact the merchant and the credit card company and show the order token as the only information needed to identify her transaction. Any accident that involves the carrier should be resolved in the same way. Each administrator can never provide the information about customers to the administrators in other parties without a disclosure order from a court.

## 4. Performance evaluations

In this section, the zero leak architecture is analyzed from the viewpoint of its effectiveness against the existing security threats. The five security threats are eavesdropping, replay, modification, masquerading, and traffic analysis. This section also describes the results of the performance evaluations for ZLA by implementing a prototype of ZLA.

### 4.1 Analyses on ZLA's coverage of different security risk types

In analyzing the resistances to different types of unauthorized or malicious accesses to customers' private information by external and internal attackers, we developed a model for analyses as shown in Figure 5. We expect the primary methods of unauthorized accesses by internal attackers are through direct accesses, while those for external attackers are performed by bug exploits and message interceptions. Message interceptions are typically performed in the forms of eavesdropping, replay, modifications (i.e., man-in-middle), masquerading, and traffic analysis.

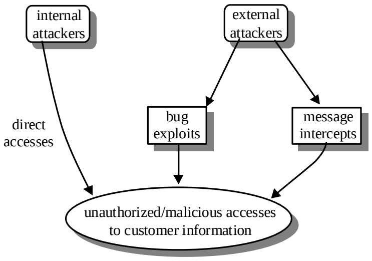


**Figure 5** – A model for unauthorized/malicious accesses to customer's private information

The followings are our analyses for the effectiveness of ALZ against unauthorized/malicious accesses by internal attackers through direct accesses:

(i) **Internal attackers at merchants**: The attackers have access only to the information about the products ordered by their customers, the order tokens for each transaction, and the network address of the customers. Any other information is not stored at the server, nor accessible by the attackers.

(ii) **Internal attackers at shipping carriers**: The attackers have access to the information that identifies each customer, such as the customer's full name and mailing address, but not the one about the products ordered by the customers or the amount of the payment for the products. Since the order tokens contain a unique merchant identification number, if the merchant's name provides any clue to the products, it may be possible for internal attackers at shipping carriers to guess the products ordered by each customer.

(iii) **Internal attackers at credit card companies**: The attackers have access to the information that identifies each customer, such as the customer's full name, mailing address, telephone number, credit card number, and the amount of the payment for each transaction. Similar to internal attackers at shipping carriers, the order tokens allow the attackers to identify the merchant in each transaction, which may allow the attackers to guess what products were purchased.

The followings are our analyses for unauthorized accesses by external attackers through message interceptions:

(i) **Eavesdropping**: Protection against this threat is provided by encryptions in multiple network layers. For example, SSL, IPSec, STS, WPA, and encryptions performed in the application process guarantee that the message contents will not be released to external attackers during their transmissions, assuming that the keys for encrypting messages are correctly exchanged between two parties.

(ii) **Replay**: Although the external attackers who have physical access to network equipments, such as routers, and transmission cables, can perform replays, they will not cause any effect to a transaction since none of the eight messages (those shown in Figure 1) is cumulative and each message is unique. The uniqueness of each message is guaranteed by the unique order token issued by a merchant. If two identical messages arrive, whichever arrives later will be dropped without any effect to the application (this assumes that "replay" causes duplicates of the initial messages and those duplicates always arrive after the initial message).

(iii) **Modification**: Any unauthorized modification of each of the eight messages will be detected by their message digest. However, it has been known that the attackers can perform man-in-the-middle attacks to intercept Diffie-Hellman key exchange algorithm. This problem can be prevented by confirming the identity of the merchants, shipping carriers, and the credit card companies using their digital certificates in the application level.

(iv) **Masquerading**: Masquerading merchants, shipping carriers or credit card companies by external attackers is impossible since the correctness of their public keys is guaranteed by their digital certificates in the application level. If an attacker tries to take over an ongoing transaction, by masquerading an existing customer, the attacker needs the order token, which is encrypted by a merchant's, a shipping carrier's, credit card company's or a customer's private key.

($v_1$) **Traffic analysis between a customer and a merchant**: We assumed that external attackers have access only to the packet headers, but not to their payload field when transport-layer secure connections are used. Since the merchant's network address may identify the merchant (and what business the merchant does), it is important to make sure that customer's network address does not identify the customer. If the above conditions are met, the traffic analyses will not be possible.

($v_2$) **Traffic analysis between a merchant and a carrier**: From the network addresses of a merchant and a shipping carrier, they can be identified easily from their network addresses. However, determining the merchant and the shipping carrier will not reveal who is the customer. This satisfies our zero-leak requirement because anonymity of the customer is still preserved.

($v_3$) **Traffic analysis between a merchant and a credit card company**: In the same was as $v_2$, anonymity of the customer will be preserved as long as the transport-layer connections are used for ZLA.

For bug exploits by the external attackers, since none of the three servers holds complete information for each transaction, an external attacker who manages to access the data stored at a server still can not obtain complete information about a transaction by the same reasons for the internal attackers.

### 4.2 Evaluations of ZLA on queuing delay and throughput

The online shopping web site shown in Figure 1 was implemented using ZLA. Its prototype was evaluated for the response time and the throughput when the workload from customers was changed. The workload was controlled by changing the time interval between two transactions generated at the customer computer. Each customer's request that consisted of the eight message exchanges was treated as a transaction.

The testbed consisted of an isolated local area network with one computer for each of the three servers and a customer. The merchant server consisted on a Intel Core 2 Quad CPU Q6600 at 2.4 GHz. The shipping carrier, the credit card company, and customer's computer were all Intel Pentium 4 CPU at 3.2 GHz each.

All three servers were available for SSL2/3 and TSL1 clients and all provided 4096-bit 3DES cipher block chaining (CBC) certificates to clients. In CBC mode, each block of plain text was XOR'd with the previous encrypted block before itself being encrypted. This ensured that blocks were interdependent and not subject to dictionary attacks. 3DES encrypted the data once using DES, performed decryption with a second arbitrary DES key, and re-encrypts the data using a third arbitrary DES key.

Using the testbed we built, we performed the following experiments:

**Experiment Definition**: The effect of workload to queuing delay and the throughput of the transactions were measured and analyzed. The merchant server logged the start time ($\boldsymbol{T_S}$: arrival of message ③) and the end time ($\boldsymbol{T_E}$: arrival of ⑦) of a transaction. We first measured "base transaction turnaround time ($\boldsymbol{T_{BASE}}$)" as $\boldsymbol{T_{BASE}} = \boldsymbol{T_E}$-$\boldsymbol{T_S}$ when only one transaction was requested by a customer. We then measured $\boldsymbol{T_E}$-$\boldsymbol{T_S}$ (= $\boldsymbol{T_\Delta}$ hereafter) for each transaction when multiple transactions were requested. We applied Poisson distribution to randomly generating the transactions. We tested 14 different transaction intervals: 0.1ms, 1ms, 10ms, 50ms, 100ms, 125ms, 150ms, 175ms, 200ms, 225ms, 235ms, 250ms, 300ms, and 500ms.

Each experiment consisted of ***n*** transactions (***n***=100) and we repeated the same experiment ***m*** times (***m***=20). We calculated the average queuing delay as:

$$\frac{\sum_{j=0}^{m}\left(\sum_{i=0}^{n}(T_{\Delta}, i, j)\right)}{m+n} \qquad (1)$$

where ($T_{\Delta, i, j}$) indicates $T_{\Delta}$ of the ***i***-th transaction in an experiment for its ***j***-th run. Finally, we calculated the average of the averages for the 20 runs of an experiment. For transaction throughput, we calculated how many transactions were completed at the merchant server (arrivals of message ⑦) per second and averaged them for the 20 runs of the same experiment.

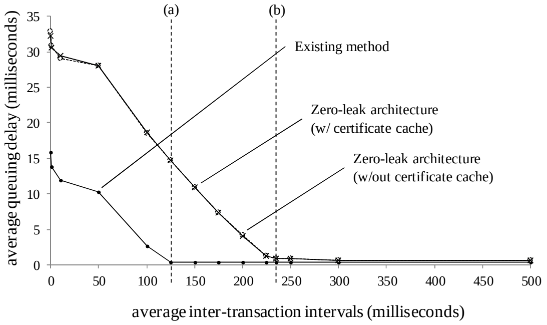


**Figure 6** – Results of the queuing delay analysis

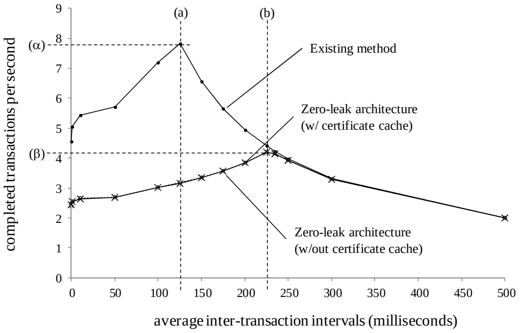


**Figure 7** – Results of the throughput analysis

Using the experiment testbed and configurations above, we measured $T_{\Delta}$ for transactions in the existing method, ZLA without certificate caching, and ZLA with certificate caching. In the existing method, none of the message contents was encrypted against the merchant. In "certificate caching", the client computer already had the digital certificates for all the three servers, while the certificate were first downloaded by the client at the beginning of an experiment in "without certificate caching".

**Analysis #1** ("queuing delay analysis"): Figure 6 shows the observed queuing delay in the experiment. We observed that "takeoff points" of the queuing delay were 125ms interval (i.e., 8 transaction per second) and 225ms (4.44 transactions per second) for existing method (indicated by "(a)" in the figure) and ZLA without certificate caching ("(b)" in the figure). Its ratio was 125:225 = 1.0:1.8. We did not observe meaningful effect from caching certificates. The average queuing delays at 0.1ms transaction interval were 15.8ms and 29.4ms for the existing method and ZLA without certificate caching (their ratio is 15.8:29.4 = 1.00:1.86).

**Analysis #2** ("throughput analysis): Figure 7 shows the observed throughput in the experiment. The peak throughput for the existing method was 7.8 transactions per second ("α" in the figure) and it was 4.2 for ZLA without certificate caching ("β" in the figure). Their ratio is 1:1.85.

## 5. Conclusions

The proposed ZLA is designed to protect the customers' private information in online e-commerce applications from malicious uses even in the worst cases, where the internal security administrators become attackers. This is a challenging goal since there is a regulation in most of the countries that obligates the e-commerce providers to investigate the details in transactions (such as U.S. Code 12, Section 1929).

To achieve the goal, we designed and developed a new architecture for e-commerce applications, called ZLA (Zero Leak Architecture). The ZLA's core concept is to split customers' transaction information to three different parties in such a way that none of them has complete information about each transaction. Complete information for a transaction can be reconstructed only if the three parties collaborate, which should happen only when the law enforcement organization requires the three parties to collaborate.

We described how the proposed ZLA provides protections against both external and internal attackers for three different categories of attacks: message interceptions, bug exploits and direct accesses. We evaluated the overhead of the ZLA from the viewpoints of the queuing delay and the throughputs of the transactions when workload to a merchant server was increased. The results of the performance analyses suggest that the ZLA's overhead in both queuing delay and throughput is a factor of around 1.8 (longer queuing delay by a factor of 1.8 and less throughput by a factor of 1.8). These results suggest that, with 1.8 times more computational power and faster networks, ZLA will provide equivalent operational performance of queuing delay and throughput to the existing unsecure systems, while the ZLA will significantly reduce the possibility of private information leaks from e-commerce servers. Along with the still-existing Moore's law, the hardware overhead will be most probably justifiable. The recovery cost, especially from large-scale customers' information leaks, will be more than the hardware cost in the future. In many such cases, they will also significantly damage the reputation of the companies, which adds to the cost of recoveries.

## References

[1] Hiroshi Fujinoki, Jacob W. Keister, Clinton W. Bandy, and Steven R. Lickenbrock, "SoKey: New Security Architecture for Zero-Possibility Private Information Leak in Social Networking Applications," *Proceedings of IEEE Communications Quality and Reliability Workshop*, pp. 1-6, May 2011.

[2] Retention of records by insured depository institutions, the U.S. Code Online via GPO Access, pp. 1084-1087, TITLE 12-BANKS AND BANKING, CHAPTER 16-FEDERAL DEPOSIT INSURANCE CORPORATION, Laws in effect as of January 3, 2007.
URL: http://*frwebgate.access.gpo.gov/cgi-bin/ getdoc.cgi?dbname=browse_usc&docid=Cite: +12USC1829b*

[3] Laszlo Hars, "Discryption: Internal Hard-Disk Encryption for Secure Storage," *Computer*, vol. 40, no. 6, pp. 103-105, 2007.

[4] Bryan Parno, Jonathan M. McCune, Dan Wendlandt, David G. Andersen, and Adrian Perrig, "CLAMP: Practical Prevention of Large-Scale Data Leaks," *Proceedings of IEEE Symposium on Security and Privacy*, pp. 154-169, May 2009.

[5] Umut Uludag, Sharath Pankanti, Salil Prabhakar, and Aniket K. Jain, "Biometric Cryptosystems: Issues and Challenges," *Proceedings of the IEEE*, vol. 92, no. 6, pp. 948-960, 2004.

[6] Ge Zhang and Chao Zhang, "A Biometric-based Framework for Digital Rights Protection," *Proceedings of the International Conference on Signal Processing*, vol. 3, pp. 2314-2317, 2004.

[7] David Mazières and Dennis Shasha, "Don't Trust Your File Server," *Proceedings of the Hot Topics in Operating Systems*, pp. 113-118, May 2001.

[8] Uif T. Mattsson, "A Practical Implementation of Transparent Encryption and Separation of Duties in Enterprise Databases: Protection against External and Internal Attacks on Databases," *Proceedings of the IEEE International Conference on E-Commerce Technology*, pp. 559-565, 2005.

[9] IBM Corporation, "An overview of the IBM SET and the IBM Commerce Point Products", http://*www.software.ibm.com/commerce/set/overview.html*, June 1998.

[10] Marc Rennhard, Sandro Rafaeli, Laurent Mathy, Bernhard Plattner, and David Hutchison, "Towards Pseudonymous e-Commerce," *Electronic Commerce Research*, vol. 4, pp. 83-111, 2004.

[11] Boualem Benatallah and Hamid R. Motahari Nezhad, "Service Oriented Architecture: Overview and Directions," *Advances in Software Engineering*, vol. 5316, pp. 116-130, 2008.

[12] Mike P. Papazoglou and Willem-Jan van den Heuvel, "Service Oriented Architectures: Approaches, Technologies and Research Issues," *The International Journal on Very Large Data Bases*, vol. 16, pp. 389-415, 2007.

[13] Doug Tygar, "Atomicity versus Anonymity: Distributed Transaction Electronic Commerce," *Proceedings of the International Conference on Very Large Databases*, pp. 1-12, August 1998.

[14] Xuan Zhang, Qinlong Huang, and Peng Peng, "Implementation of a Suggested E-commerce Model Based on SET Protocol," *Proceedings of ACIS International Conference on Software Engineering Research, Management and Applications*, pp. 67-73, May 2010.

[15] Dimitrios Lekkas and Diomidis Spinellis, "Implementing Regular Cash with Blind Fixed-Value Electronic Coins," *Computer Standards & Interfaces*, vol. 29, no. 3, pp. 277–288, March 2007.

[16] D. Chaum, A. Fiat, and M. Naor, "Untraceable Electronic Cash," *Proceedings of Advance in Cryptography*, pp. 200-212, 1990.

[17] Ziba Eslami and Mehdi Talebi, "A New Untraceable Off-line Cache System," *To appear in Electronic Commerce Research and Applications*, 2010.

# Clustering online social network communities using genetic algorithms

**Mustafa H. Hajeer*** **Alka Singh*** **Dipankar Dasgupta*** **Sugata Sanyal#**

*Center for Information Assurance
Department of Computer Science
The University of Memphis
Memphis, TN 38138, USA

#School of Technology & Computer Science
Tata Institute of Fundamental Research
Homi Bhabha Road,
Mumbai-400005, INDIA

## ABSTRACT

To analyze the activities in an Online Social Network (OSN), we introduce the concept of "*Node of Attraction*" (NoA) which represents the most active node in a network community. This NoA is identified as the origin/initiator of a post/communication which attracted other nodes and formed a cluster at any point in time. In this research, a genetic algorithm (GA) is used as a data mining method where the main objective is to determine clusters of network communities in a given OSN dataset. This approach is efficient in handling different type of discussion topics in our studied OSN –comments, emails, chat sessions, etc. and can form clusters according to one or more topics. We believe that this work can be useful in finding the source for spread of rumors/misinformation, help law enforcement in resource allocation in crowd management, etc. The paper presents this GA-based clustering of online interactions and reports some results of experiments with real-world data and demonstrates the performance of proposed approach.

**Keywords:** *Online Social Network, Node of Attraction, GA-Based clustering.*

## 1. INTRODUCTION

Blogging, twitting and online social networking have become an integral part of modern-day life. Studies have shown that people spend a lot of their time in online media such as Facebook, MySpace, Twitter, YouTube, Google+ and in other blogosphere. These web-based services/applications are designed to assist people and businesses alike to stay in touch, communicate and collaborate more effectively, and socialize across borders and in general, maintain a second avatar in cyberspace. The way blog sites (nodes) attract people for participating in discussion, share information and/or spread rumor is very interesting and fascinating to study. The dynamic interaction among these nodes and the period of increased interaction and manner in which such interaction get decreased can provide useful information about the current events, topics of discussion, and human behavior. In the analysis of online social network (OSN) data, the basic problem lies in the detection of groups of closely connected nodes which are called 'network communities'. In this work, we consider the relation among the nodes in OSN, and group them according to the strength of their relations/interactions. The work is motivated by the need to understand how the interactions in virtual-world can manifest real-world security concerns or law and order situations. Such a clustering of online social communities may also help to detect insider threats, employee behavior, or recognizing competitor interests in social activities as possible motivators.

In this research, a genetic algorithm (GA) is used as a data mining method where the main objective of clustering is to find network communities from a given OSN dataset. There exist some works which addresses different aspects of online social behavior and used un-weighted social network data for clustering [2], i.e. every edge in the network has equal value/weight and distributed community detection in delay tolerant networks [3, 9-10].

## 2. GA-BASED APPROACH

An OSN community can be defined as a set of users who are frequently interacting with each other and participating in some discussion e.g. group, subgroup, module, and cluster.
A Genetic algorithm (GA) is an optimization method which uses biological process as a model to find solutions in the search problem. We used a GA here for finding the clusters in OSN data having multiple values for each edge indicating individual node participating in more than one discussion groups or activities. The goal is to improve the processing and efficient interpretation of real-world OSN data, relevant knowledge, and subsequent information processing and representation. The underlying approach tries to find dense clusters using an edge removal strategy in an intelligent way based on the context of discussions.
The reasons for using GA-based approach are

- Huge search space and non-linearity in search space.
- The dynamic nature of network topology and the need of adaptive multi-fitness function.
- GAs perform global search in problem space
- GAs are easy to interface to existing simulations and models
- GAs are extensible and easy to hybridize
- GAs are remarkably noise tolerant

The implementation details of the GA approach for OSN data clustering is explained in the next section, the section 4 provides some experiments and results. The last section gives some concluding remarks.

### 2.1 Node of Attraction:

As we know, nodes in a social network are not physical nodes rather web sites, postings, user blogs, pages with some news, etc. For example, the documentary on

Invisible Children (Kony 2012) was posted at a video-sharing website, *YouTube.com* on 5th March 2012. In two weeks time, over 78 million viewers visited the site; another website *vimeo.com* also posted the same video and attracted over 16.6 million viewers during the same time. This indicates that the sites which hosted the video attracted significant number of viewers forming clusters with the posted page as the focal point.

We introduced the concept of *Node of Attraction* (NoA) which represents the most active node in a network community. This NoA is identified as the origin of a post/communication which attracts other nodes to form a cluster. The need of finding such nodes and to keep track of these nodes in social networks has many real-world applications such as detection of spreading news or rumor in the society.

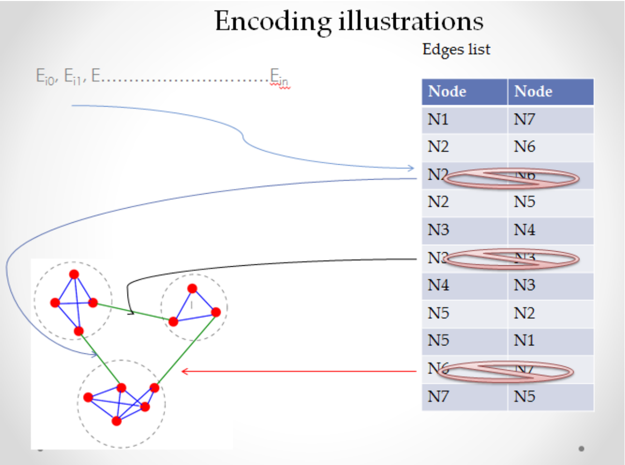
## 3. IMPLEMENTATION DETAILS

In this work, a genetic algorithm is used to find clusters/groups in online network data by removing minimum number of edges while the clusters have maximum number of ties.

### 3.1 Encoding schemes

We have used two encoding schemes where each chromosome represents groups or clusters.

In the first scheme, each chromosome is variable in size, composed of edge list as shown in Figure 1 (top left). These edges are derived from nodes in each row of the table (right). The genetic process will remove edges from the chromosome to form clusters as illustrated in Figure 1 (bottom left).

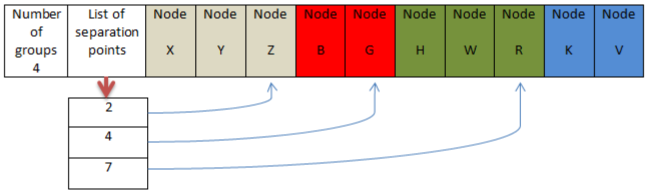

Encoding illustrations

Edges list

| Node | Node |
|---|---|
| N1 | N7 |
| N2 | N6 |
| ~~N2~~ | ~~N6~~ |
| N2 | N5 |
| N3 | N4 |
| ~~N3~~ | ~~N3~~ |
| N4 | N3 |
| N5 | N2 |
| N5 | N1 |
| ~~N6~~ | ~~N7~~ |
| N7 | N5 |

**Figure 1: First encoding scheme showing edge removal from the OSN data**

The second encoding scheme uses a real-valued representation, where the first number denotes the number of groups and the rest fields denote the points of separation between nodes to form groups as shown in Figure 2.



| Number of groups 4 | List of separation points | Node X | Node Y | Node Z | Node B | Node G | Node H | Node W | Node R | Node K | Node V |
|---|---|---|---|---|---|---|---|---|---|---|---|

| |
|---|
| 2 |
| 4 |
| 7 |

**Figure 2: The second encoding scheme**

### 3.2 The fitness function

Two properties of clustering such as high intra-cluster similarity and low inter-cluster similarity are considered for fitness measure of each chromosome. Accordingly, the number of edges connected to a node form a group are considered for fitness, where we calculate the clossness of groups/clusters and penalize the chromosome with smaller groups with same fitness. The fitness function uses the most updated edge values rather than old values in order to adapt to the changes in the interactions.

### 3.3 Setting of GA control parameters

To create diversity in the candidate solution list, the initial population is generated with random edges, and the fitness for each chromosome is calculated.

We run a steady-state GA with two different datasets. A single-point crossover with probability of 0.85, population size 100, maximum evaluations 10,000, and the binary tournament selection operator are used. The high rate of mutation is used so that the genetic search can adapt to dynamic user interactions (as reflected in data) over time. After crossover and mutations are performed, each chromosome is repaired (if necessary) in order to remove repeated edges in the chromosome.

For the second encoding scheme, we used a different type of crossover where each chromosome uses the swap operator where some genes are exchanged, and the mutations are restricted to the second and first parts of the chromosome.

## 4. EXPERIMENTS AND RESULTS

In order to determine the correct NoA for an attribute, the chromosome exhibiting the best fitness is searched and then it is decoded/ mapped to form groups. After the mapping operation, these groups are sent to a function which finds the NoA (the node with maximum number of edges) and is saved in a file to keep track of all NOAs over a period of time.

To demonstrate the performance of the GA-based clustering approach, we first experiment with a hand-crafted synthetic datasets and then we use some real-world OSN data.

### 4.1 Synthetic (small) dataset

Table 1 shows the small dataset containing 15 edges each with 3 attributes as different type of interactions between a pair of nodes. Such interactions represent multi-valued edges among the nodes. Here the interactions/edges represent number of emails, number of comments and the number of posts between nodes. Figure 3 is the corresponding network diagram (of Table 1) with some multi-valued edge. If there is no edge between any two nodes which indicates that attribute values between these two nodes are zero.

**Table 1: A small synthetic dataset**

| Node A | Node B | Number of emails | Number of posts | Number of comments |
|---|---|---|---|---|
| **1** | 2 | 4 | 4 | 4 |
| **1** | 3 | 3 | 3 | 3 |
| **1** | 4 | 4 | 4 | 4 |
| **1** | 5 | 4 | 5 | 5 |
| **5** | 4 | 3 | 4 | 4 |
| **5** | 3 | 4 | 3 | 3 |
| **2** | 3 | 3 | 3 | 3 |
| **2** | 4 | 3 | 3 | 3 |
| **4** | 7 | 1 | 29 | 1 |
| **8** | 14 | 2 | 1 | 30 |
| **5** | 6 | 1 | 39 | 1 |
| **6** | 7 | 5 | 4 | 3 |
| **6** | 8 | 3 | 5 | 4 |
| **6** | 9 | 4 | 4 | 4 |
| **7** | 8 | 4 | 3 | 5 |
| **7** | 9 | 3 | 4 | 4 |
| **8** | 9 | 4 | 5 | 5 |
| **6** | 10 | 1 | 1 | 35 |
| **10** | 11 | 4 | 4 | 3 |
| **10** | 12 | 3 | 3 | 3 |
| **10** | 14 | 4 | 3 | 4 |
| **11** | 12 | 2 | 4 | 5 |
| **11** | 13 | 3 | 3 | 4 |
| **11** | 14 | 5 | 4 | 5 |
| **12** | 13 | 4 | 5 | 4 |
| **13** | 14 | 3 | 4 | 5 |
| **12** | 14 | 4 | 3 | 4 |
| **14** | 15 | 3 | 4 | 5 |

After running the program, we observed that the GA is able to form clusters corresponding to each discussion groups (attributes) as shown in Figure 4.

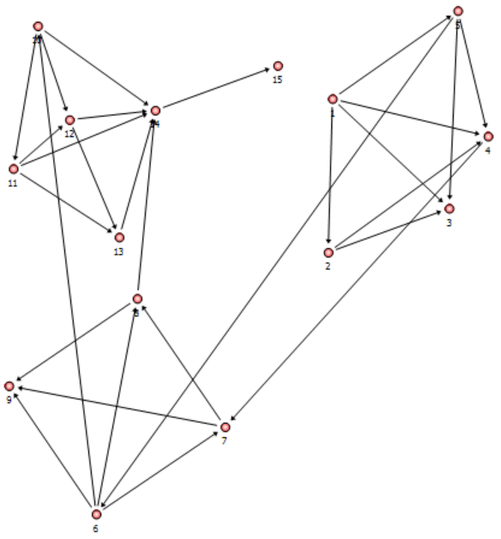


**Figure 3: A graphical representation of the small dataset.**

Figure 4(a) graphically represents results of a GA run which contains edges list forming network clusters according to the first attribute of the synthetic dataset (Table 1).

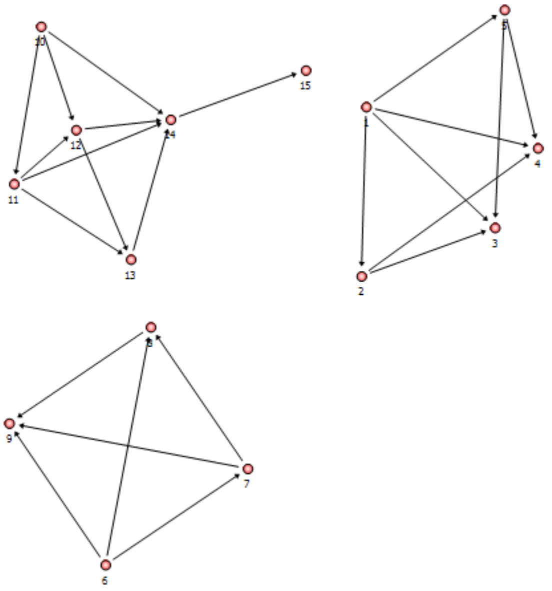


**Figure 4(a): GA-based clustering according to number of Emails.**

Similarly, Figure 4(b) and Figure 4(c) show the clusters formed based on other two attributes: number of comments and number of posts.

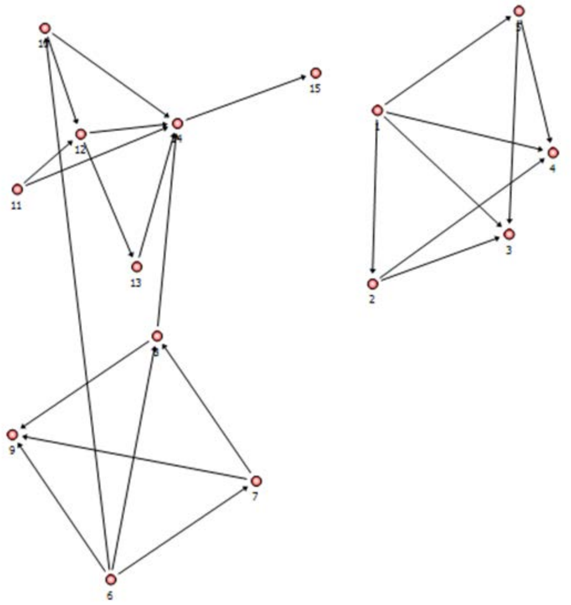


**Figure 4(b): Clustering according to number of comments.**

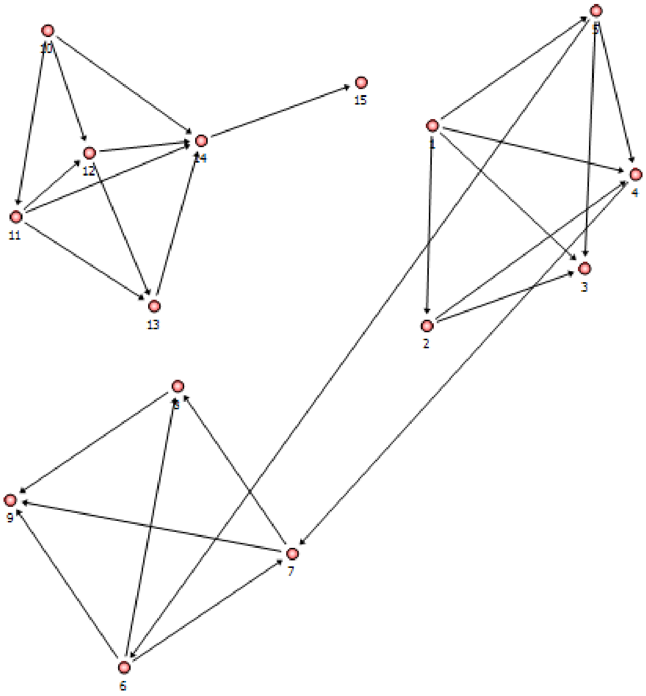


**Figure 4(c): Clustering according to number of posts.**

As per definition, the NoA for different attribute can be observed in each figure above. Accordingly, in figure 4(a) NoAs are {1, 9 and 14}, in figure 4(b) NoA nodes are {14 and 1} and in figure 4(c) these nodes are {6 and 14}. However, this provides a simplest analysis of real online social network interactions. Only one attribute is not sufficient to classify the entire social network. In particular, groups should not be clustered only on the basis of email exchanged among nodes. Other attributes should be selected in conjunction.

Now, let us take a look at figure 5(a) and Figure 5(b) below, where inter-group interactions and overlapping are demonstrated. If the OSN data are clustered according the first and the second attributes as per the dataset (Table 1), the overlapped groups are formed.

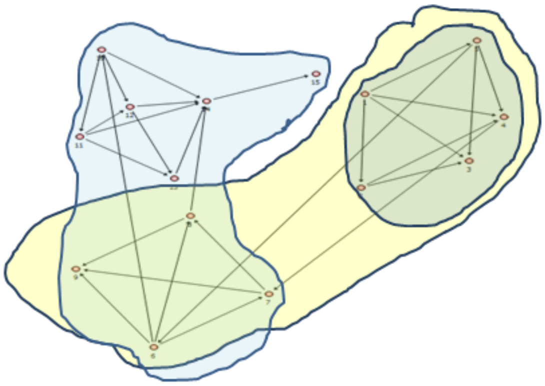


**Figure 5 (a): Groups formed according multi-valued edges**

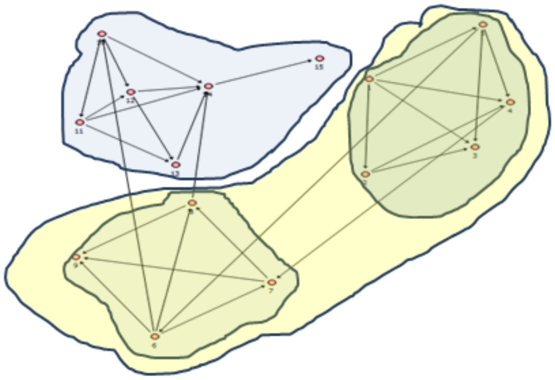


**Figure 5(b): Groups formed according multi-valued edges**

If we consider multiple attributes, it is observed that some nodes connect two or more separate clusters; such a node is identified as the "linkage node". These nodes help to analyze the interactions between two different discussion groups or topics (figure 5).

To illustrate the dynamic interactions in social networks and to demonstrate the performance of our GA-based clustering approach, we changed some attribute values during the run. At time $T_0$ during the execution of the algorithm, clusters started to form, based on the dataset as shown in figure 6, and after that some changes are made in the first attribute. This process is illustrated through the following example:

Here the second group is formed with the nodes {6, 7, 8, 9} before changes were made and NoA was the node number {9}. Sometime after $T_0$, node X is added to the second group and it was connected to all group members. Since it was connected with all group members, and it was observed that the system immediately considers X as the new node of attraction.

So after initial clustering, the groups' formation adapts to changes made to the network interaction. Table 2 shows the results of these dynamic changes, where the changes take place between time period $T_0$ and $T_1$ as follows:

- Node X was added to the data set.
- Edges were connected from X to nodes {6,7,8,9}

and the changes which take place between time $T_1$ and $T_2$ are:

- Node Y added to the dataset.
- Edge from X to node 6 and node 7 is removed to reduce its value.
- Edges from Y added to nodes {6, 7, 8, 9, X}. Hence node Y becomes the new node of attraction.

**Table2: second group dynamics**

| Time | Group members | Node of attraction |
|---|---|---|
| $T_0$ | 6,7,8,9 | 9 |
| $T_1$ | 6,7,8,9,X | X |
| $T_2$ | 6,7,8,9,X,Y | Y |

Figure 6 illustrates these changes where the blue edges represent the regular nodes which are added to the dataset and red edges show new Node of Attraction.

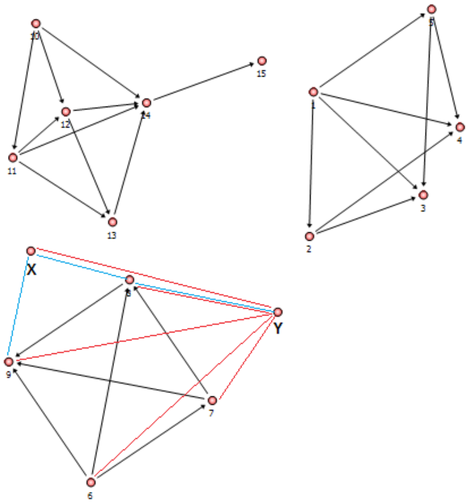


**Figure 6: Cluster results with dynamic changes in the synthetic dataset**

## 4.2 Experiments with real-world dataset

The real-world dataset considered here is taken from Stanford Large Network Dataset Collection [5], it is called Gnutella peer to peer network. This dataset contains 20,777 edges and random value is assigned to each edge (Figure 7),

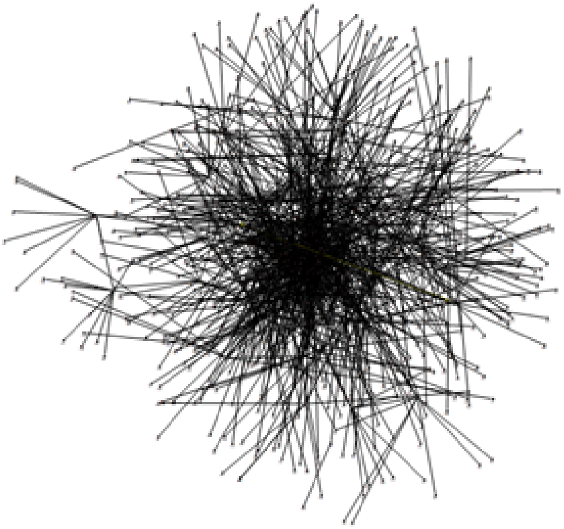


**Figure 7: An example network dataset with 1000 edges [5]**

Our solutions contain the groups which start to form slowly with smaller groups with strong ties. Hence after each (100) iterations we extracted the solution and observed more groups being formed in the solution set. The GA-based system responds slowly especially in the beginning (takeoff time) because of the huge solution space, but after forming the initial groups, it could quickly reflect changes to the groups, and produce better and reliable results as the time passes.

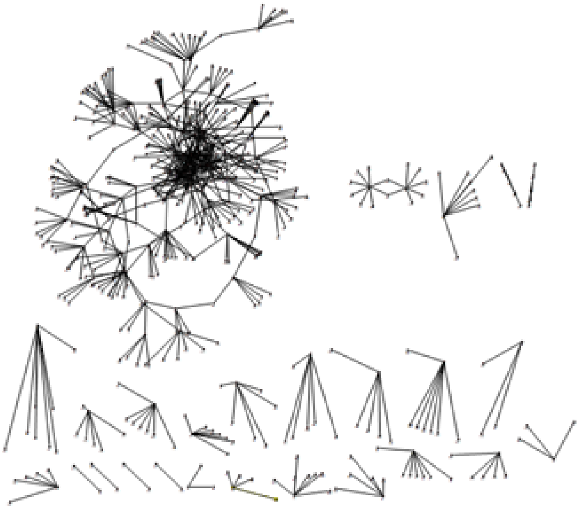


**Figure 8: Clusters formed in 1000 edges network data after running GA for 1000 iterations**

The clustering results obtained after 1000 iteration of a GA run is visualized in Figure 8. It shows that some small groups and a big group are formed. If we keep running the program, the big group would have further divided into smaller groups. It was observed that the groups which are already formed are adaptable to any change in the input dataset.

The differences between Figure 7 and Figure 8 are clearly visible; the original data representation in Figure 7 contains all nodes which are connected to each other in some manner (these figures are a presentation of clusters). The size of the groups varies based on the complexity of interactions.

Again, to illustrate the dynamism, some nodes are added to the dataset (Blue); the nodes with high value and high connectivity (shown in red) are the new nodes of attraction (Red) which are shown in Figure 9 below. It shows that updated clusters are formed in the social network interactions.

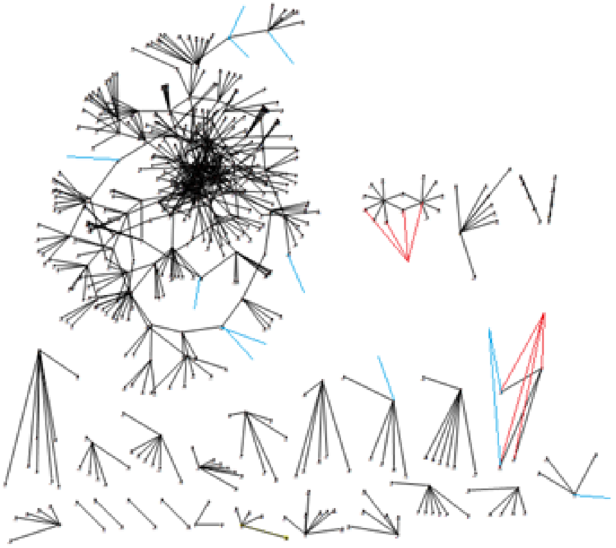


**Figure 9: Changes on the dataset during running the system causes restructuring of groups or clusters.**

Because of multi-attribute clustering technique used, this work provides the power to observe the points of connections nodes, indicating the nodes that connect more than one group, and it's considered as linkage nodes.

## 5. DATA AVAILABILITY

In the first test case (Table 1 -Synthetic dataset), we assumed that the user has access to data that have multiple weights for each connection, like weight for messages exchange and weights for discussions, but in real-world we may not have such information. So for the second set of experiments (real-world dataset), we considered equal values for all connections (i.e. either there is a connection or no connection) since we have only edge-list in the datasets [5].

## 6. CONCLUDING REMARKS

An OSN community can be defined as a set of users who are frequently interacting with each other and participate in some discussions. Determining such communities in an OSN has a wide verity of applications, such as

- Understanding the interactions among group of people
- Visualizing and navigating huge networks of NoAs
- Forming a basis for other tasks such as data mining
- Marketing, handling law and order situations

This approach allows us to form clusters based on interactions among group members in OSNs; we introduced the term NoA (Node of Attraction) which represents a node that attracts most of other nodes in the same group at a given time. The Node of Attraction in a social network, captures interaction dynamics “subsets of actors among whom there are relatively strong, direct, intense, frequent or positive ties”[1]. This NoA can help to predict the forming and merging of groups and subgroups. For example, if nodes from one sub group starts to interact with NoA of another sub group, there is a high probability that these sub groups may merge soon to form a larger group. Another benefit of NoA is that it helps to identify the source of a post/news and also may help in community detection in blogsphere.

.Because of the huge size of the solution space, we chose to use a GA approach to solve this problem, the main idea is to cluster the whole network into groups and at the same time keep track of NoAs in dynamic interactions.

The experimental run takes some time in producing reliable results but results seems to be persistent even with the dynamic changes. In case of a snapshot of the best individual (best so far) solution is desired from the population, the software is capable of producing such a result at any given time.

## 7. References:

[1] Stanley Wasserman and Katherine Faust, Social Network Analysis, Methods and Applications (Book), November 25, 1994.

[2] Santo Fortunato. Community detection in graphs, Complex Networks and Systems Lagrange Laboratory, ISI Foundation, Viale S. Severo 65, 10133, Torino, ITALY.

[3] Pan Hui, Eiko Yoneki, Shu-Yan Chan, Jon Crowcroft. Distributed Community Detection in Delay Tolerant Networks, University of Cambridge, Computer Laboratory Cambridge CB3 0FD United Kingdom

[4] Jorg Reichardt and Stefan Bornholdt. Statistical Mechanics of Community Detection, Institute for Theoretical Physics, University of Bremen, Otto-Hahn-Allee, D-28359 Bremen, Germany, February 3, 2008

[5] Stanford Large Network Dataset Collection http://snap.stanford.edu/data/ , Collaboration network of Arxiv General Relativity.

[6] M. E. J. Newman, 'Modularity and community structure in networks', PNAS103 (23), pages 8577-8582, June 6, 2006.

[7] Scott White and Padhraic Smyth, A spectral Clustering Approaches To Finding Communities in Graphs, in 'SDM', 2005.

[8] KONY 2012: Invisible Children http://www.youtube.com/watch?v=Y4MnpzG5Sqc

[9] Shihua Zhanga, Rui-Sheng Wangb, Xiang-Sun Zhang. Identification of overlapping community structure in complex networks using fuzzy *c*-means clustering. Physica A, 2007.

[10] Lei Tang and Huan Liu. Graph Mining Applications to social network analysis. Book chapter C.C. Aggarwal and H. Wang (eds.), *Managing and Mining Graph Data*, Advances in Database Systems 40, Springer Science+Business Media, LLC 2010.

# An Overview of Cloud Computing Challenges and Its Security Concerns.

**Krishnun Sansurooah**
SECAU Security Research Centre, School of Computer and Security Science
Edith Cowan University, Perth, Western Australia, Australia

***Abstract** - There has been an increasing advancement about Cloud computing during the past couple of years. Cloud computing has become a new computer model which aims to deliver reliable, customizable and scalable computing environment for end-users. Companies are choosing to move their data, applications and services to the Cloud. The advantages are significant ranging from increasing the availability, reliability, light weight, easily accessible applications, and low cost but so are the risks associated with. Companies that require application hosting could potentially benefit from the provisioning of computing infrastructure resources as a service. In addition to the economic advantages of an on-demand computing environment, businesses also enjoy the flexibility to scale up or down their services to accommodate the changing nature or the business requirement without having to invest in new equipment however, migrating data to the Cloud exposed the data to be an easy and vulnerable target for all the maliciously intended actors all over the world. This paper brings an introduction overview to Cloud computing, it's enabling technologies behind such a design, its evolution and finally the security concerns that is entails.*

**Keywords:** Cloud computing, infrastructure models, cloud security

## 1 Introduction

Cloud computing derives its name from the drawings normally used in the description of the Internet with a new retention and delivery of Information Technology (IT) services. It originated in the late 2007, but currently came forward as an enticing topic due to its abilities to offer and deliver adaptive dynamic IT infrastructures, quality of standards computing environment and softwares services that can be used by businesses having different needs.

The idea of Cloud computing definitely represents a change in the way of addressing things which the end user does not have to fully be knowledgeable of the details of a specific technology. According to Google trends [8], it has been reported that the Cloud computing has overtaken virtualization in popularity as depicted by figure 1 below with the blue line which is sitting above both the virtualization technology in yellow and the red line representing Grid computing [5].

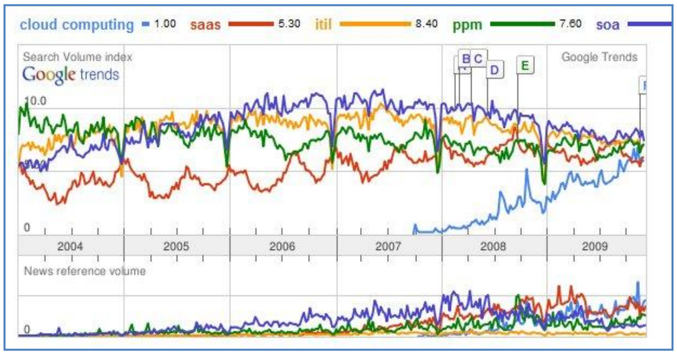


*Figure 1. Cloud computing surpassing virtualization in popularity* ***(Google Trends, 2009).***

Huge companies such as IBM Blue Cloud [10] and the European Union have joined forces to collaborate on a research programme for Cloud computing, Amazon Elastic Compute Cloud, [2] scientific projects such as Nimbus and Stratus which are a set of open source tools that offer an "Infrastructure-as-a-Service" (IaaS) cloud computing solution. [14].

The term Cloud computing is a common overused computer term that most computer professionals have used as "public internet" at one time or the other. There is still no such universal definition for the Cloud computing despite that Cloud computing practices have characterized much attention in the past years. Many arguments have lead to this setting:

- Cloud computing includes working in collaboration with researches, network engineers, from different experiences and environments;
- The technologies that permits clod computing to exists is still in evolution and progressing;
- Prevailing computing Clouds however comes short of acceptable measure of deployment and usage which would substantiate the conceptualization of Cloud computing.

This paper will consolidate what Cloud computing is using simple and commonly acceptable terms whilst elaborating on the other sections of this paper such as the evolution of Cloud computing, the various Cloud computing infrastructure models, its enabling technologies, what benefits does Cloud computing has to offer and an in-depth analysis of the security concerns with remedial actions to mitigate the threats that Cloud computing poses.

## 2 Definition of Cloud Computing

Gradually, businesses of all sizes are choosing to migrate their data, applications and services to the Cloud [6].Cloud computing is becoming one of the next IT trends where IT Administrators, managers and CEO are porting their data and applications to remote "Clouds" and accessing these data in a simple and ubiquitous manner from anywhere around the world. However not much considerations is given due to the vulnerability of the data which becomes a tantalizing target for hackers which will be discussed further in the overview of the security concerns of the Cloud computing. Cloud computing can be generally defined as numerous diverse ways to distribute information or services to customers who pay what they use. Obviously, the customer can either be an individual or a business purchasing for a service or information. Cloud computing is nothing less that a new approach of conducting business that seizes the advantage of housing various competencies into a particular structure that can then extended to host out services for multiple businesses.

Due to the ongoing evolution of the Cloud Computing, it is quite premature to come up with an exact definition of what is Cloud computing? However, at this early stage we believe that the Cloud computing definition could be summaries as:

- *"A range of network enabled services, having the ability to be expanded, quality of standards, usually customizable, with inexpensive computing infrastructure and environment on demand to the availability of the customers that can be accessed easily form anywhere."*

However according to National Institute of standards and Technology (NIST) in 2009, their definition of Cloud computing was defined as:

- *"Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This Cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models."*

## 3 Evolution of Cloud Computing

The growth of Cloud computing can be traced back to the Grid computing. The idea of "The Grid" gained in popularity after Foster and Kesselman [5] published their work *"The Grid: Blueprint for a new Computing Infrastructure"*. It was founded on the electric grid that provides electric power to your home and business. This time it would be oriented towards using the same concept in hardware and software which would be provided from the grid on-demand from a computing level [3].

One of the most interesting downfalls with Cloud computing is that the same issues that hit the grid did the same for Cloud computing which at present still suffers from the formation of standards, whether vendor have to be locked-in, etc.

Cloud computing is about moving services, computation and or data for an inexpensive cost which will benefit the business. It is location-transparent with a centralized facility. By making data available through the Cloud, it can be more easily accessed from everywhere at a much lower cost, increasing its value by allowing favourable circumstances for improved collaboration, integration and analysis on a shared common environment/platform which leads to have various choices of Cloud computing infrastructure models.

### 3.1 Cloud Computing Infrastructure Models

Prior to migrating the data, applications and services to the Cloud, there are several considerations that need to be taken into account as moving from a standard enterprise application deployment to a Cloud computing architecture could cause the entire business to collapse.

There are public and private Clouds that propose integrative advantages which IT companies can select to extend the use of applications on private, public or hybrid Clouds each of which have their gives-and-takes illustrated in figure 2.

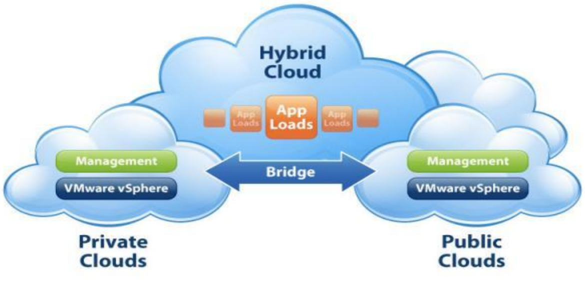


*Figure 2. Illustrates the different types of Clouds and their interactions, public, private and hybrids.*

### 3.1.1 Public Clouds

Public Clouds are usually run by third parties; and applications from various clients are probably merged together on the Cloud's server, storage systems and networks as depicted in figure 3. Public Clouds are normally hosted offsite from the client's localities which hence reduces the customers cost and risk by accommodating an adaptive extension to the enterprise infrastructure. It is also to be noted that if the public cloud introduce performance, security and data centralization, the existence of the other applications running on the Cloud will have to be transparent to both the Cloud architect and the end users. In this sense the benefit of public Clouds is that they can be much larger than a company's own private Cloud providing the ability to scale up or down on demand.

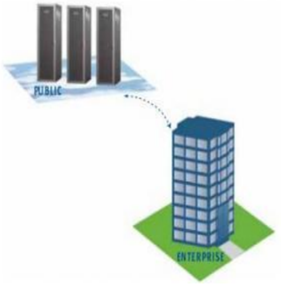


*Figure 3 depicts a public Cloud providing services to multiple customers which set up at a collocation facility [13].*

### 3.1.2 Private Clouds

Private clouds are built for the exclusive use of one customer, accommodating the absolute control over data, security, and quality of service as depicted in figure 4. The company is the owner of the infrastructure and has got total control over how applications are set up on it. Private clouds can either be built in an enterprise data centre or they can also be deployed at a shared facility.

Private Clouds can be put together and controlled by a business own IT department or by a Cloud provider. In this model a company can install, configure, and operate the infrastructure to support a private Cloud within a company's enterprise data centre. This infrastructure allows businesses with an utmost level of control over the use of Cloud resources while accommodating the necessary expertise to settle and manage the environment.

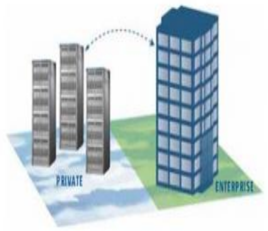


*Figure 4 shows how Private Clouds may be hosted at a collocation facility or in an enterprise data centre. They may be supported by the company, by a Cloud provider, or by a third party such as an outsourcing firm.*

### 3.1.3 Hybrid Clouds

Hybrid Clouds as it name say it all it is a combination of both public and private Cloud models as shown in figure 5. These Clouds assist with on-demand and externally provisioned scale. The ability to improve a private Cloud with the benefits of a public Cloud can be applied to preserve the service levels in the case of quick evolving workload variations. This is visible in making use of storage Clouds to assist Web 2.0 applications for example. Hybrid Clouds brought forward the intricacies of deciding how to allocate the applications across both a public and private Cloud. Among the different concerns that needed to be taken into account is the connection between data and processing resources. If the data is small and of not major importance, or the application is in a stateless condition, a hybrid Cloud would have more success rather than transferring a large amounts of data into a public Cloud for a small amount of data processing.

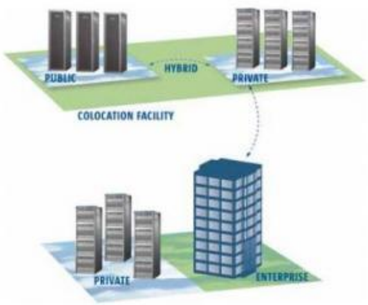


*Figure 5 shows a Hybrid Clouds combine both public and private Cloud models, and they can be particularly effective when both types of Cloud are located in the same facility.*

### 3.2 Architectural Layers of Cloud Computing

Ideally, users access computing platforms or IT resources from Cloud computing and then run their applications from the inside. Therefore, Cloud computing provide the users with services to access hardware, software, platform based, infrastructure and finally data resources, thereafter an integrated computing environment as a service in a evident way as illustrated in figure 6.

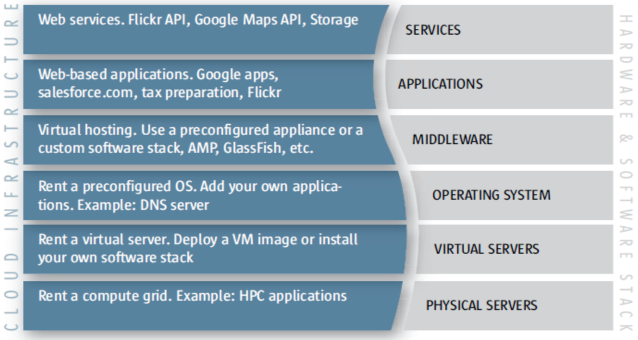

CLOUD INFRASTRUCTURE
Web services. Flickr API, Google Maps API, Storage
SERVICES
Web-based applications. Google apps, salesforce.com, tax preparation, Flickr
APPLICATIONS
Virtual hosting. Use a preconfigured appliance or a custom software stack, AMP, GlassFish, etc.
MIDDLEWARE
Rent a preconfigured OS. Add your own applications. Example: DNS server
OPERATING SYSTEM
Rent a virtual server. Deploy a VM image or install your own software stack
VIRTUAL SERVERS
Rent a compute grid. Example: HPC applications
PHYSICAL SERVERS
HARDWARE & SOFTWARE STACK

*Figure 6 depicts Cloud computing as one that can provide services at various layers from hardware to applications.*

#### *3.2.1 Hardware as a Service (HaaS).*

As a result of technology advancing rapidly, hardware virtualization has done the same hence allowing businesses to buy IT hardware, if not, a whole data centre, as a pay-as-you-go subscription facility [16]. The HaaS is flexible, scalable and manageable to meet your needs.

#### *3.2.2 Software as a Service (SaaS).*

This model is designed to be hosted as a service and provided to customers across the Internet. In a nutshell this model provides everything and simply rent out the software to the user through some sort of front end or web portal thus eliminating the burden of installing and running the application on the client's local machines. A single instance of the software runs on the Cloud and services multiple end users or client organizations. Another advantage of this SaaS is that it relieves the client's burden of software maintenance and reduces the expense of software purchases by on-demand pricing.

#### *3.2.3 Platform as a Service (PaaS).*

With this particular approach of Cloud computing, PaaS as a service covers a panel of software and delivers it as a service that can be adopted to develop or build higher-level services. This can be categorized into 2 groups. Firstly, PaaS, provides a platform for working while combining an operating systems, application software and a development environment that is then put at the disposition of the client as a service and secondly, as a customer making use of the PaaS as a service would see a tremendous advantage in a wrapped service that is proposed to them. Since PaaS model of Cloud computing includes all phases of the System Development Life Cycle (SDLC), the customers could interact with the platform through an application program interface (APIs), website portals or gateway software.

#### *3.2.4 Infrastructure as a Service (IaaS).*

This model encapsulates the delivery of basic storage and calculates competencies as standardized services over the network. As the name implies, the customer is buying the infrastructure. Servers, storage systems, switch, etc... are made available to administer workloads from computing application.

#### *3.2.5 Data as a Service (DaaS).*

Data in numerous formats and from different sources could be accessed through the services by customers on the network. User could therefore alter the remote data just like performing data changes on a local disk or accessing data in an acceptable manner. Therefore, the user pays only for the storage capacity that they require and the bandwidth for that particular storage. However the system is flexible and scalable as it allows businesses to add capacity as needed, in-demand.

# 4 Technologies Behind Cloud Computing

There are various enabling technologies that contributes to the successful achievement of Cloud computing. These precursors are identified as:

**Service Oriented Architecture (SOA)**

It is essentially a collection of services which communicate with each other by either simple data passing or it could involve two or more services organizing some activities. Since computing Clouds are normally opened to work with Web Services such as WSDL, SOAP and UDDI, the services provided by the SOA could be used to orchestrate and organize these services. Furthermore, a set of Cloud services could be used in a SOA application environment thus taking advantage of having them available of multiple distributed platforms accessed through the Internet.

**Virtualization technology.**

Virtualization is software technology which uses a physical resource such as a server and divides it up into virtual resources called virtual machines (VM's) which offers virtualized IT infrastructures on-demand. Virtualization permits users to strengthen physical resources, simplify arrangement and administration whilst abating power and cooling requirements. While virtualization technology is most popular in the server world, virtualization technology is also being used in data storage such as Storage Area Networks, and inside of operating systems. Virtual networks such as Virtual Private Networks (VPNs) assist users with customized network environments to access Cloud resources as they are the foundations of Cloud computing as they depict flexible and scalable hardware resources.

**Inexpensive Worldwide Distributed Storage.**

A network storage system which is supported by distributed storage providers also known as data centres (sometimes called server farm) which are at the disposition of the users for loan. Therefore, the need of having a centralized repository for the storage, management and dissemination of data and information. One key advantage of the data centre storage concept is that physical hard drive storage resources are combined into a collection of storage pools. Another benefit that contribute to the use of Cloud computing is the consolidation of all facilities encapsulating HVAC, electrical, network connections, wiring, hardware, software and personal. Hence a distributed data system which can support data sources to be accessed in a ubiquitous manner.

**Web 2.0**

Web 2.0 is a new technology which apparently helps in distinguishing the use of the World Wide Web technology and Web design pertaining to the creativity, information sharing, and collaboration to the Web. The main idea behind Web 2.0 is to uplift the interconnectivity and the interactivity of the Web applications enabling users to access the web more easily and efficiently. And since Cloud computing resources are based on web applications, it makes it natural that while evolving, Cloud computing will favour the use of Web 2.0.

**Service flow and workflow.**

Cloud computing provide an entire set of service templates on-demand, which could be comprised of services from the inside Cloud. Therefore, the Cloud should be able to naturally if not automatically coordinate services from various sources to form a flow of service which is transparent and dynamical to the users.

## 4.1 Benefits of Cloud Computing.

The primary aim of Cloud computing is to decrease the cost for computing resources whilst leaving if not increasing the flexibility and scalability of the system. This model also minimizes capital expenditures, since the concept of renting out the services from a different provider on a peruse fee the business only pays for the used resources. The sharing of resources and purchasing power of very large scale, data centres provides economic advantages are listed below:

- ***Reducing high fixed-capital cost to low variable expense***
  Creating an internal cloud within a company supports an efficient service platform while monitoring internal capital consumptions for the IT infrastructure. This could then be exploited with an external cloud service provider that could jump on board to supply overflow service capacity when demand increases above the internal capacity.
- ***Flexibility***
  As with large businesses, the ease of deploying a full service without having to organize the base environment to support it can be even more attractive that cost savings. There is the ability to update the hardware and software quickly to comply and suit customers' demands and updates in technology.
- ***Smoother scalability path.***
  Cloud computing allows for many singles services to scale over a wide demand range. It multiplies resources load balance peak load capacity and utilization across multiple hardware platforms in numerous locations.
- ***Drastically reduces cost (Disaster Recovery).***
  Especially with small to medium enterprises, make no investment in disaster recovery (DR) plans. But with enabling technologies such as the virtualization – VMWares (virtual machines) to be transferred to the Cloud for access whenever it is needed, it has evolved in a cost effective disaster recovery model since the DR plans cost twice the infrastructure whereas with a Cloud computing approach the DR is available for much cheaper hence making a significant difference in cost for the businesses.
- ***Self-service IT infrastructure.***
  Cloud computing service models are often self-service, even in internal models. As before the business had to partner with the IT to develop your applications, port it onto a platform and run it but with Cloud computing all this is not of the past. You are purchasing infrastructure.
- ***Increased automation and portability***
  Moving into the Cloud computing requires a much higher level of automation because moving off-premises terminates on-call systems administrators. It also increases the portability of the system whereby users can work from home, work or at client location thus increasing mobility that employee can access information anywhere they are.
- ***Centralization.***
  Centralization of critical data improves security by eliminating data from user's computer. Cloud computing providers have the necessary knowledge and resources to provide all the latest security features to guard the data.
- ***Maintenance.***
  With centralized applications, they are much easier to maintain than their distributed counterparts. All updates and changes are made in one centralized server rather that on each user's computer.

## 4.2 Security Concerns of Cloud Computing.

While cost and ease of use are among the great advantages of Cloud computing, there are alarming security concerns that need to be taken into consideration when being inclined to move critical applications and especially sensitive data to public, private or hybrid Cloud environments. To focus on these issues, the cloud provider must implement and develop acceptable measures to match the same if not extended layers of security than if the business would have if they were not to use the cloud. Listed below are the main concerns that should be taken into consideration prior to porting all their sensitive data and applications to the remote "Clouds" as these can very quickly become appealing targets for people of criminal intentions. Below is an overview of these issues that have been summarized into three different categories that should be thought prior to change to the Cloud computing infrastructure.

**Conventional Security:** These issues would be involving computer and networks intrusions or attacks which could have a potential of occurring since moving to the cloud.

**i)** **Authentication and Authorization**: the company's authentication and authorization structure does not extend into the cloud which if it does can cause problems with the cloud provider's schema provided that the cloud provider has got one in place. According to Jericho Forum (2010), it would be easier to lock and protect information if it's managed by a third party than it was in-house. So the question here is how the organization remodels its existing structure to accommodate the cloud computing infrastructure model. Furthermore, how does an organization combine cloud security within its own security policies?

**ii)** **Phishing:** What are the countermeasures that are set up to reduce the risks of being hit with phishing scams which have new attack vectors?

**iii)** **Virtual Machine attacks**: Potential attacks lies within the virtualization area – i.e. VM technology used by cloud vendors hence the introduction use of firewalls and intrusion detection systems should be part of the system to monitor system activities.

**iv)** **Network attacks**: The cloud user should ensure that the infrastructure used to connect and interact with the cloud is protected at all times which could be a gigantic task as the cloud is designed and set up outside the firewall in most of the cases.

**v)** **Forensics implications:** At present, the conventional methods of digitally forensically analysing equipment is to seize the devices and image them and perform an in-depth analysis of the media and then recover the data from. But with cloud computing the likelihood of conducting forensics analysis would be very complicated to perform due to the nature providers maintaining their own multi-server infrastructure and hence data being overwritten multiple times.

**Availability of data:** Based on critical applications and data being made available for the users to access anytime and from anywhere around the globe.

**i)** **Uptime:** most cloud providers dispute the fact that their server uptime is with reasonable parameters compared to the availability of the cloud users' own distributed centres.

**ii)** **Single point of failure:** Cloud computing services are meant to provide more availability which is not the case but instead are open to more single points of failure.

***iii)*** **Computational integrity:** Can an organization guarantee that the cloud provider is being trustworthy and faithful in running hosted applications and resources which are feeding back trustworthy outcomes for the smooth running of the organization.

**Data-control by external parties:** What are the legal ramifications and implications of data an applications being under the control of external parties? Because this area is not well understood and very complex, there are potential lack of control and transparency when a third party is involved in the process of containing all the data.

**i)** **Audit ability**: Is another issue of lack of control in the cloud which then bring up the question is there enough transparency in the way of conducting operations from the cloud provider's side? This is very important as it hence build up the trust between both the provider and the user.

**ii)** **Contractual Obligations:** One of the various issues that emerge when using a third party infrastructure is the legal implications. Hence the implementation of service level agreement (SLA) or what can and what cannot be done bounded by a written contract.

**iii)** **Data loss or leakage:** **T**here a numerous ways of compromising the data. Modification, fabrication, deletion of data without a proper backup of the original or updated data is a very common example. Loss of encryption key could lead to data not being disposed properly. The threat of compromising data is multiplied on the cloud due to the numerous interactions of users with the environment is dangerous due to the architectural design of the cloud environment.

**iv)** **Data lock-in:** How does a cloud user abstain from locking-in to a particular cloud vendor? Also there is the problem that the cloud users have no control over the regular changes that the cloud provider is effecting from his side.

# 5 Remedial Strategies for Addressing Cloud Computing Risks

The risks that an enterprise might recognize must be dealt with in an effective manner. A robust risk management program that is flexible enough to deal with continuously evolving information risks should be in place. In an environment where privacy has become the utmost priority to enterprise customers, unauthorized access to data in the cloud is of major concern. When agreeing to a service level agreement (SLA) with a cloud provider, the organization (user) should imperatively record all his information assets.

The organization should also make sure that data are correctly catalogued. This process should be implemented while drafting the SLA which will determine a legally contract between both the user and the provider. Any specific needs of the user such as data need to be encrypted before being transferred or transiting or stored from one location to the other, and additional controls for information that is sensitive or of high value to the organization should be clearly defined in the agreement.

The SLA is a legal document that characterizes the relationship between the business and the cloud provider, it also serves as a powerful tool to protect against information entrusted to the cloud. Listed below are some of the remedial actions that should be implemented when considering cloud computing environment:

- ***Abuse use of Cloud computing***
  - i) Effectively analysis of user network traffic whilst monitoring public blacklisted network blocks
  - ii) Tighter initial registration and validation process when it comes to using credit card. ( everyone can use a credit card to purchase and immediately start using the cloud services)
- ***Insecure Application Program Interfaces***
  - i) Scrutinize the security model of the cloud providers
  - ii) Making sure that strong authentication and authorisation is implemented with encrypted transmission.
- ***Threats from insiders***
  - i) Making sure that there is a reliable security breach notification in place.
  - ii) Requesting transparency in information security practices as well as compliance reporting
  - iii) Including human resource requirements as part of legal contracts
- ***Shared technology***
  - i) Monitor for unauthorized activities.
  - ii) Conduct vulnerability scanning, configuration auditing while ensuring that SLA mentioned the patching and upgrading of the environment
  - iii) Promoting strong use of authentication and access controls for any levels of operations.
- ***Data Loss or leakage***
  - i) Implementation of strong encryption to protect integrity of data in transmission
  - ii) Ensure that strong key generation, storage and management and destruction policies are in place
  - iii) Analyse and monitor data protection on both ends.
  - iv) Prohibit the sharing of account credentials between users and resources
  - v) Use strong 2-way authentication wherever it is possible.
  - vi) Understand cloud provider security policies and SLAs.

# 6 CONCLUSION

Cloud computing proposes valid advantages to businesses striving for competitive benefits. Many more providers are spreading into the concept of providing Cloud computing services and are therefore driving the competition to a higher level with even lower prices which is an absolute benefit for businesses. Since the Cloud computing provides appealing pricing, the ability to remodel the human resources towards other tasks, the ability to pay for services when needed will drive more businesses to adopt the Cloud computing infrastructure. But before any services are moved to cloud, the business managers should confirm that these actions are rational for the good running of the business. In this paper as we have seen there are the pros and cons of Cloud computing which need to be weighed carefully before any decisions to be made.

# 7 REFERENCES

[1] ACIS (2001) Status Project. Retrieved 9 August, 2010 from http://www.acis.ufl.edu/vws

[2] Amazon Elastic Compute Cloud (2007). Amazon Elastic Compute Cloud. Retrieved 24, July, 2010 from http://aws.amazon.com/ec2/

[3] Baca, S. (2010). Cloud Computing: What It Is and What It Can Do for You. Retrieved 27 June , 2010 from http://images.globalknowledge.com/wwwimages/whitepaperpdf/WP_VI_CloudComputing.pdf

[4] Craig Balding (2008). Assessing the Security Benefits of Cloud Computing. Retrieved 5 September, 2010 from http://cloudsecurity.org/blog/2008/07/21/assessing-the-security-benefits-of-cloud-computing.html

[5] Foster, I. & Kesselman.C (1998). *The Grid: blueprint for a new computing infrastructure*. Morgan Kaufmann

[6] FIERCIO (2009). Securing the cloud: Designing Security for a new Age. Retrieved 28 June, 2010 from http://i.zdnet.com/whitepapers/eflorida_Securing_Cloud_Designing_Security_New_Age.pdf

[7] Global Cloud computing test bed (n.d.). Global Cloud Computing Research Test Bed. Retrieved 21 May 2010 from http://www.hp.com/hpinfo/newsroom/press/2008/080729xa.html/

[8] Google Trends (n.d). Cloud Computing Surpasses Virtualization in Popularity Retrieved 02, August, 2010 from http://www.elasticvapor.com/2009/04/google-trends-cloud-computing-surpasses.html

[9] Here comes HaaS (2006). Here comes HaaS. Retrieved 14 May, 2010 from http://www.roughtype.com/archives/2006/03/here comes haas.php/

[10] IBM Blue (n.d) Cloud Initiative Advances Enterprise Cloud Computing. Retrieved 15, June , 2010 from http://www-03.ibm.com/press/us/en/pressrelease/26642.wss

[11] IBM (2009). Architectural manifesto: An introduction to the possibilities (and risks) of cloud computing. Retrieved 11 May, 2010 from http://www.ibm.com/developerworks/architecture/library/ar-archman10/index.html?S_TACT=105AGX20&S_CMP=EDU

[12] Mell, P & Grance T. (2009). The NIST Definition of Cloud Computing. Retrieved 21, July, 2010 from http://csrc.nist.gov/groups/SNS/cloud-computing/

[13] Sun’s View (2009). Overview of Cloud Computing. Retrieved 29, July , 2010 from http://hiencs.wordpress.com/2009/09/01/overvire-cloud-computing-architecture/

[14] Nimbus (2010) Nimbus Project. Retrieved 10 July, 2010 from http://workspace.globus.org/clouds/nimbus.html/
Understanding Cloud Computing (2009). A white paper for executives making decisions on computing resources. Retrieved 19, July, 2010 from https://ritdml.rit.edu/bitstream/handle/1850/7821/LWangConfProc11-162008.pdf;jsessionid=4D6164F12ABFD986C5891A8762A9B4A9?sequence=1

[15] Understanding Data Centers and Cloud Computing from http://images.globalknowledge.com/wwwimages/whitepaperpdf/WP_DC_DataCenterCloudComputing1.pdf

[16] Wang & Laszewski. (2008)Cloud Computing: a Perspective Study. Retrieved 20, August, 2010 from https://ritdml.rit.edu/bitstream/handle/1850/7821/LWangConfProc11-16-2008.pdf;jsessionid=4D6164F12ABFD986C5891A8762A9B4A9?sequence=1

[17] The Jericho Forum's demolition framework (2008). Don’t Cloud Your Vision. Retrieved 4 August, 2010 from http://www.ft.com/cms/s/0/303680a6-bf51-11dd-ae63-0000779fd18c.html?nclick_check=1

[18] Trapani, G (n.d). The Hidden risks of cloud computing. Retrieved 15 August, 2010 from http://lifehacker.com/5325169/the-hidden-risks-of-cloud-computing

# SCIENCE OR EXPERIENCE; what is more relevant?

**Ken Fowle [1,2] and Hadyn Green [2]**
[1] School of Computer and Security Science, Edith Cowan University, Perth, WA, Australia
[2] Centre for Forensic Science, University of Western Australia, Perth, WA, Australia

***Abstract -*** *It is very important that when we use science to determine the validity of evidence or information that it is done in a manner that is acceptable to the scientific community and the legal community, but what happens when "experience" is used. The use of forensic practitioners to provide 'expert' evidence and opinion must meet the Daubert/Frye and now Kumho tests. This paper will endeavour to demonstrate .what is best for a practitioner to have and what does the judiciary require for 'expert' evidence to be accepted? Science and/or Experience, what is more relevant? Evidence and the Courts depend upon the establishment of a reliable basis of fact. because at the end of a trial, a Judge or a Jury will be compelled to reduce a complex slice of human experience with all its subtlety, to what is, in essence, a one line answer: "I believe you, or I don't.".*

**Keywords:** Practitioner, Science, Experience, Knowledge

## 1 Introduction

It is very important that when we use science to determine the validity of evidence or information that it is done in a manner that is acceptable to the scientific community and the legal community, but what happens when "experience" is used. Is a scientific approach more valued than experience? This paper will look at the role of forensic practitioners using science and/or experience in supporting (or not) evidence and information being presented in the courts. This paper will look at how science and experience has been used, what has been the result and will endeavour to demonstrate using cases, as to why forensic practitioners need to keep evaluating themselves in relation to their forensic expertise.

In the United States, the National Research Council [1] of the National Academy of the Sciences concluded that; with the exception of nuclear DNA analysis, no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source.

The council also stated; "For a variety of reasons—including the rules governing the admissibility of forensic evidence, the applicable standards governing appellate review of trial court decisions, the limitations of the adversary process, and the common lack of scientific expertise among judges and lawyers who must try to comprehend and evaluate forensic evidence—the legal system is ill-equipped to correct the problems of the forensic science community. In short, judicial review, by itself, is not the answer."

In the same year Chief Justice Robert French of the High Court of Australia in a presentation to the Medico Legal Society of Victoria said "the more technically or scientifically complex the issue for determination, the greater the challenge for the courts whether in patent law or other fields. There are some areas, particularly those involving computer science and complex software that may test the limits of the capacity of the courts to answer the composite questions of science and law to which they give rise" [2].

We are living in world that is using complexity to resolve complexity. We expect advancement, we expect solutions and we expect it to be right. As forensic practitioners there is an expectation that we are experts in our field, we have qualifications, we have accreditation, we have practical experience and we have the under pining knowledge of how our speciality works, is used and accepted, but what happens for the practitioner who has qualifications but limited experience (in the field) and the practitioner who is experienced but has only limited or no qualifications?.

Gary Edmond [3] said that the failure to engage individuals with the requisite knowledge, training and experience can produce a variety of mistakes, faulty assumptions and risks, even if these are not appreciated during trial and appeal processes.

Were as James Robertson [4] made comment, that it is a worrying outcome if academic researchers were to be excluded from giving "relevant" evidence simply on the basis of not being practitioners. He continued by saying that he does value experience; it is an inescapable qualitative factor which is relevant. However practitioners should not hide behind experience as an excuse or substitute for appropriate research and academic rigour. There is differing opinion as to what a forensic practitioner should have, thus there is a need to consider whether the value of science outweighs experience, or vice versa, or are both equal given the circumstances..

### 1.1 Definitions

*as defined by 2005 NSW Law Reform Report 109 – Expert Witnesses.*

- ***expert***, in relation to any question, means a person who has such knowledge or experience of, or in connection with, that question, or questions of the character of that question, that his or her opinion on that question would be admissible in evidence.

- ***expert witness*** means an expert engaged for the purpose of:
  - providing a report as to his or her opinion for use as evidence in proceedings or proposed proceedings, or
  - giving opinion evidence in proceedings or proposed proceedings.
- ***expert's report*** means a written statement by an expert (whether or not an expert witness in the proceedings concerned) that sets out the expert's opinion, and the facts on which the opinion is formed, and contains the substance of the expert's evidence that the party serving the statement intends to adduce in chief at the trial.

.

# 2 Background

In what follows, the authors have used cases where the court has mentioned what is expected of experts who provide evidence and what is not acceptable.

.

## 2.1 The Frye vs Daubert Cases in the USA

In the USA, historically, scientific evidence, broadly defined, had to be generally accepted as reliable in the field in which it belongs, before courts would admit opinion testimony based on a particular technique or discipline. This was based upon the 1923 decision *Frye v. United States 293 D 1013 (DC Cir 1923)* and as such a "general acceptance" test was established by the testimony of experts in the particular field.

In 1993, *Daubert v Merrell Dow Pharmaceuticals Inc, 113 S Ct 2786,* the US Supreme Court supersede the *Fry* test and established as requirements for the admissibility of expert evidence that:

1. The expert must be qualified.
2. The methodology employed by the expert must be reliable.
3. The testimony must assist the trier of fact.

These requirements were reflected in an amended version of US Federal Rules 702:

1. Whether the theory or technique had been tested.
2. Whether it had been subjected to peer review.
3. The rates of error in the technique and any standards controlling the technique's operation.

Whether there is general acceptance of the theory or technique in the scientific community..

## 2.2 The Position of the Australian Courts

In contrast to the US Federal Rule 702 which states:

"*If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise*"

The Australian provision for expert opinion evidence is section 79 of the *Evidence Act* 1995 (Cth) which states:

*"If a person has specialised knowledge based on the person's training, study or experience, the opinion rule does not apply to evidence of an opinion of that person that is wholly or substantially based on that knowledge."*

In his paper Deflating Daubert: Gary Edmond [5] said that "*Daubert* and more recently *Kumho*, have provided judges with more (rhetorical) resources for excluding evidence. But it also illustrates how close *Frye* and *Daubert* really are." He continues "*Daubert* is an attempt to make sure the experts have actually employed the generally accepted theory. But if a qualified or experienced expert comes to court from a recognised field using a *generally* or *significantly* accepted technique, it is hard to conceive why the issue of faithfulness to the technique or particular approach could not be explored through cross examination".

Scott Mann [6] wrote that in Australia, the *Uniform Evidence Act* allows opinion based on specialised knowledge deriving from a person's training, study or experience, leaving specialised knowledge undefined. Under the common law it is accepted that expert opinion must derive from a 'field of expertise' and points out, 'Australian law has never clearly resolved the test for a "field of expertise".

All of this means that the onus remains on the legal representatives to take a very active role in the expert defence of their clients' interests; to prevent bias, bribery and untruth from winning the day through their own mastery of the crucial scientific issues, vigorous critical interrogation of expert witnesses for the other side, appropriate selection and use of their own witnesses and ongoing scientific education of judge and jury.

Justice Wood [7] from the Supreme Court of New South Wales in a presentation at the 2002, 16th International Symposium for Forensic Science said "that it was unresolved in Australia whether the appropriate test for the admissibility of expert evidence should be the *Frye* or the *Daubert* test".

From what has been presented to date from both practitioners and judiciary as to what is used to determine the acceptability of evidence, that it is unclear and that a resolution of this issue has real significance if is to be excluded from presentation as forensic evidence. The only clear thing and what we do know is that it is the role of a forensic practitioner to assist the court in understanding the facts presented in a trial and providing an opinion (if required).

## 2.3 Experience

As forensic practitioners we must be able to understand what we do, what we use (in our respective discipline) and how it works. We need to present our work and opinions in a scientific manner to the courts whilst being mindful that it needs to be understood by the triers of fact, often the jury.

For example: a Registered Professional Engineer and long-time State Traffic Safety, vehicular homicide expert was asked a question by the court in regards to a case based on a questionable application of critical speed formula [8].

The Court: Mr Godfrey, let's go back to some high school physics here just to complete the record. What is the scientific basis for the critical speed formula?

Mr Godfrey: Newton's Laws.

The Court: Which is?

Mr Godfrey: Well there are three of them, three different Laws

The Court: Put them on the record, please.

Mr Godfrey: You're pressing me, your Honor, here in my advanced senility.

The Court: I just want to complete the record.

Mr Godfrey: There's three Newton's Laws. For every force there is an opposing force.

The Court: An object in motion stays in motion?

Mr Godfery: An object in motion tends to stay in motion. If it's in a circular motion, it will tend to move to the outside.

The Court: And these are the basis of the mathematics of the formula?

Mr Godfery: These are the basics of the mathematics of the formula, yes, sir.

Clearly the evidence was erroneous because Newton's three (3) Laws are [9] :

1. Every object persist in its state of rest or uniform motion in a straight line unless it is compelled to change that state by forces impressed on it to it.
2. Force is equal to the change in momentum per change in time. For a constant mass, force equals mass times acceleration.
3. For every action there is an equal and opposite reaction.

From the above either the expert witness was so confident, that the court would accept his testimony because he was called an expert or he made a mistake that as an expert witness. What the scientific community was liable to draw was there was a lack of understanding of the basics. Worse still a jury was liable to be misled by the testimony.

As practitioners we need to know/understand the technology and methodologies that we use in our field of specialisation, and be prepared to apply them even to matters such as cold cases or when fresh information emerges in current cases.

## 2.4 Science

Most practitioners use some form of science to support their finding. But it is very rare to find a case where only science has been used to obtain a conviction.
In a paper by Wendy Abraham [10] she cites R v Rowe in dismissing an argument that a verdict was unsafe on the basis that DNA was the only evidence of identification, the three presiding judges Bleby J, Doyle CJ and Gray J all agreed, with Justice Bleby's conclusions: "The evidence was the subject of expert opinion. It was subjected to close scrutiny by the trial judge who directed the jury that they must be satisfied beyond reasonable doubt as to the reliability and accuracy of the DNA analysis". It probably founded a safer basis for a conviction that the frailty often attending the evidence of a single eye-witness who gives evidence of identification of the offender.

## 2.5 Science and No Experience

In the Court of Criminal Appeal in New South Wales [11], Mr Gordon Wood had his conviction for the murder of Carolyn Bryne on 7 June 1995 overturned. Wood had been convicted in 2008 of the murder of Byrne some 11 years earlier. The prosecution contended that Wood had thrown Byrne from a cliff. It had initially been assumed that she had committed suicide.

Associate Professor Rod Cross took an interest in how Byrne met her death: Originally it had been assumed that she committed suicide by throwing herself off a cliff and landing on rocks below; to test that theory Cross conducted a series of experiments. These involved strong men throwing women into swimming pools and throwing dead weights; further having fit and able-bodied young women jumping and diving into pools. It was concluded that a strong, fit man could have thrown a woman of Byrne's weight from near a bend in a safety fence to where her body was found

Once Cross had reached this conclusion, it was decided to prosecute the applicant for murder. The prosecution reasoned that this evidence, together with the evidence of another witness, was sufficient to exclude the possibility that Byrne committed suicide and to implicate Wood in her murder. Later Cross wrote a book *Evidence for Murder: How physics convicted a killer.* In his book he admits that he has never investigated a cliff fall but his experience was in the study the physics of sport, falling fatalities and accidents. (Note: His book about the matter was tendered and admitted as new evidence on the appeal).

Wood's appeal barrister Tim Game SC presented nine grounds of appeal, which included evidence that forensic material presented at the trial was flawed. .

Chief Justice McClellan in his finding stated "Experts who venture 'opinions', (sometimes merely their own inference of fact), outside their field of specialised knowledge may invest those opinions with a spurious appearance of authority, and legitimate processes of fact-finding may be subverted." He also mentioned the following regarding Cross' expertise in the matter;

- Cross was allowed, without objection, to express opinions outside his field of specialized knowledge.
- It was submitted to this Court that at the very least A/Prof Cross' lack of expertise in these areas diminished the weight that could reasonably be attributed to his evidence.
- Cross' qualifications are in physics and his primary area of expertise is in plasma physics.
- He has spent some time since his retirement assisting the police in the investigation of incidents of persons falling and has published alone, or with others, some papers concerned with the physics of sport.
- In the course of these tasks he has applied his knowledge of basic physics.
- He has no qualifications or experience in biomechanics.

## 2.6 Science and Experience

In the case of a *The State of WA v Marteniz* [12] before Justice Heenan the accused were charged with causing the death of Phillip Walsham who, it was said, was pushed from an overhead footbridge to the ground below in the early hours of 28 February 1998. Heenan was made comment on whether the evidence of an expert witness could be admitted and pointed out a number of flaws, for example:

1. There was no attempt made to standardise the results and there was no error analysis. Heenan observed that all of the measurements actually relied upon (height, velocity, weight) were fixed or precise and not within a range: There was no allowance for error. Therefore they produced precise results that could not impress even a lay observer as being particularly scientific.
2. The calculations as to time for the fall and distance covered were expressed in terms of absolute accuracy with no allowance for error. The Court observed that the situation was most unlikely given the subjective nature of much of the data and rendered questionable conclusions based on a difference between 3.7 metres and 5 metres over the short span of the fall.

Although Heenan J was critical of the evidence of the expert witness Heenan J was satisfied that the expert witness had training, experience and expertise in the field of physics, mechanics and trauma analysis and he presented his report and findings in a manner that was acceptable to the Court and allowed the evidence presented to be challenged and questioned.

## 2.7 Discussion

Our evidence is being tested by other experts, challenged by researchers and the law and it is the forensic practitioner who needs to keep abreast of what is happening. The case you used your science and/or knowledge to determine an opinion, may have changed. In a new case or due to the length of the legal process (the same case) your workings, finding and opinion may change due to new science or experience and this must be reflected in your work and findings as developments occur.

In the above we have seen how courts have accepted (or not) expert evidence. In the cases where an appeal has been accepted due to in adequate evidence, we must also consider that to get to an appeal, there must have been a conviction. It is not the intent of this paper to discuss the issue of why was it accepted.

We have seen how courts have accepted evidence from both the scientific and the experienced practitioner but Doyle CJ [12] stated that experience teaches us that witnesses can be "100 percent certain", yet wrong. So long as juries determine the issue of guilt, jurors will be entitled to reject the confident testimony of lay and scientific witnesses, especially if it does not fit other evidence that they do accept.

So what should the practitioner and the legal profession be looking for in the capability of a forensic practitioner to prepare and present forensic evidence?

Judge Richard Posner [13] declared that the continued rapid advance in science is going to make life difficult for judges (and the Courts) this was because of the breakneck technological changes that are thrusting many difficult technical and scientific issues on judges, for which very few of them are prepared because of the excessive rhetorical emphasis of legal education and the weak scientific background of most law students.

Justice Kirby [14] also supported the notion that technologies themselves have now gone beyond the understanding of ordinary citizens, even highly educated ones, and it is essential that society should be able to look to experts in the technology to help in defining, and responding to, the implications for society of the technological advances.

From what has been presented we know that the courts are the gatekeeper of what can be admitted as evidence but we are still not assured of ensuring the accuracy of the evidence. In an article by the Australian Law Reform Commission [15] they mention that human failure is more likely to cause science to fail on the courtroom and automated equipment and better methodologies are available.

So the question is raised again how can we ensure that 'expert 'evidence is of high quality?

In a speech to the Federal Court/Law Council Case Management Workshop Justice French [16] stated "The subject matters upon which courts are required to make decisions inevitably attract many different kinds of "expertise" which it is claimed will assist them in their determinations. Their varieties are distinguished by more than subject matter. Differences in conceptual foundations and methodology and the nature of the intellectual or other enterprises they represent raise a question about the proper construction to be given to such phrases as "specialised knowledge based on training, study or experience" which appears in s76 of the *Evidence Act 1995* (Cth).

Therefore a forensic practitioner (or expert) must be able to demonstrate their "specialised knowledge" and "expertise' to the satisfaction o a court and this is done by presenting their training, study and experience in their specialised field, as depicted in *The State of WA v Marteniz* [12].

In the UK in the cases of RvWeller [17], the appeals court judgement stated "that if one tries to question science purely by reference to published papers and without the practical day-to-day experience upon which others have reached a judgement that attack is likely to fail, as it did in this case". The three Justices continued that they do hope that the courts will not be troubled in future by attempts to rely on published work by people who have no practical experience in the field and therefore cannot contradict or bring any useful evidence to bear on issues that are not always contained in scientific journals.

The appeal was based on the proposition that the evidence (DNA) was not sufficiently reliable for experts to express an evaluation of the probabilities due to the lack of relevant publications. In the judgement it was stated "It is unrealistic to examine a field of science of this kind by reference to published sources. A court in determining whether there is sufficiently reliable scientific bases for expert evidence... will be entitled to take into account the experience of experts".

Even in the UK, courts are making comments and decisions on science and experience as to what is more relevant.

From what has been discussed to date, it is acknowledged that appropriate Science and Experience of the area of expertise that is being relied upon is required and as Abraham [10] states "Only then can the strength and limitations of evidence be properly assessed, and if required, presented to a jury in an accurate and comprehensible manner with its true significance being exposed".

# 3 Conclusion

Science and technical advancement is providing the Forensic Practitioner with better tools to work with to undertake work. This also means that the Forensic Practitioner is required to have a greater understanding of their particular area of expertise. The increasing complexity of some evidence demands that Forensic Practitioners assist the courts in understanding certain events; gone are the days where once the Forensic Practitioner could say "trust me I am a expert" without demonstrating to the satisfaction of their client and ultimately the court.

The Forensic Practitioner plays a decisive role in only a minority of cases that come before the courts; however, if required, they can have a crucial bearing on the outcome of the trial, as in Wood v R 2012 [11]. Of concern to the courts is that a sound judgement is reached that is based upon 'the facts'. To reach this conclusion it may be the acceptance of the 'expert' due to their scientific and/or experience on the subject matter.

Forensic Practitioners must demonstrate good understanding of their area of specialisation and this may include science, technology and law they use, whether it be old (but still accepted) or new and revised. Additionally their underpinning knowledge and experience is paramount to the case, client and court, as it compliments, the science.

We are not saying that you have to have both, but from the cases provided and the publication presented the two go hand in hand.

Science and Experience, what is more relevant? Evidence and the Courts depend upon the establishment of a reliable basis of fact. So why not both, because at the end of a trial, at the end of an appeal, a Judge or a Jury will be compelled to reduce a complex slice of human experience with all its subtlety, to what is, in essence, a one line answer: "I believe you, or I don't."

# 4 References

[1] NRC, - National Research Council. 'Strengthening the forensic sciences in the US: The path forward". Washington DC: National Academies Press; 2009, p. 5.

[2] French, (CJ) R, "Science and Judicial proceedings: seventy six years on". Chief, Medico Legal Society of Victoria, 2 May 2009.

[3] Edmond, G , "Actual innocents? Legal limitations and their implications for forensic science and medicine". Australian Journal of Forensic Sciences, 43: 2, 177 — 212

[4] Robertson, J. “Truth has many aspects”. Stuart Kind Memorial Lecture, Journal of Science and Justice No.52 (2012) pages 62-66).

[5] Edmond, G. "Deflating Daubert: Kumbo Tire Co v Carmichael and the Inevitability of General Acceptance (Frye)” University of New South Wales Law Journal 38, 2000

[6] Mann, S. "Science, corporations and the law" Alternative Law Journal 289, 2001.

[7] Wood (J). “Forensic Science from the Judicial Perspective”. 16th International Symposium on Forensic Science, Canberra, Australia 13-17 May, 2002.

[8] Bohan, T.L. “Changing Forensic Science from Within” Forensic Science for the 21st Century, Sabdr Day O’Connor School of Law, Arizona State University, April 3-4, 2009.

[9] NASA, (2010), Glen Research Center: http://www.grc.nasa.gov/WWW/K-12/airplane/newton.html Accessed 20 May, 2012.

[10] Abraham, W. “Science not sorcery”. DNA in the 21st Century, Expert Evidence Conference, Canberra Feb, 2011.

[11] Wood v R, Supreme Court of New South Wales , Court of Criminal Appeal Decisions, Wood V R [2012] NSWCCA 21, 24 February, 2012.

[12] WA v Martinez & Ors, WASC 98, 20 March, 2006.

[13] Posner, R "The Role of the Judge in the Twenty-First Century". 86 Boston University Law Review 1049, 2006.

[14] Kirby, (Justice), “Computer & Law: The First Quarter Century“, NSW Society for Computers and the Law, 9 October, 2007.

[15] ALRC, “Making Forensic Science Work: some ideas for reform”, Australian Law Reform Committee, Reform 69, n.d.

[16] French, (CJ), "Expert testimony, opinion argument and the rules of evidence". Federal Court/Law Council Case Management Workshop, Expert testimony, opinion argument and the rules of evidence , 14 - 15 March, 2008.

[17] R v Peter Weller, EWCA Crim 1085. Case No: 2008/4666/B3. Thursday 4 March 2010**.** URL: http://www.bailii.org/ew/cases/EWCA/Crim/2010/1085.html Accessed 25 May, 20102.

.

# A Security and Privacy Model for Electronic Health Records

**Xuezhen Huang [1], Jiqiang Liu [1], and Zhen Han [1]**
[1] School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China

**Abstract** - *The adoption of Information and Communication Technologies (ICT) in healthcare field makes the electronic-healthcare (e-Healthcare) environment and information from different medical organizations collected together, forming shared e-Healthcare information (EHI) named electronic health record (EHR), so that availability and completeness of medical information are improved. The owners of EHI provide them to organizations and drug stores for scientific research and analysis. But they pose threats to security or privacy of patients. Building a secure EHR sharing environment has attracted attention in both healthcare industry and academic community. We propose a privacy preserving security model of EHR for manage the security issues of personal information. Firstly, patients make their privacy policy of medical data according to their own concerns and preferences. Secondly, user can access medical data only if they are authorized. Finally, the adoption of proper technology makes the security model satisfying privacy requirements when data are released for other uses.*

**Keywords:** privacy preserving; security model; EHR; EMR

## 1 Introduction

Independent medical organizations originally generated and managed patients' medical information. EHRs are shared among healthcare personnel to provide complete and accurate information and to improve the quality and efficiency of healthcare diagnostics. The ultimate goal is to realize all EHI shared in the network, to end paper based medical records and reduce duplication of inspection. Patients to any hospital, just to provide a medical record number, will be able to achieve all medical records needed and find a variety of test results. EHR systems are run by community, regional emergence, or national wide emergence organizations. They may violate the security and privacy of EHR.

Uses of several purposes of EHR have given rise to long-term challenges for securing EHR. EHR is personal and its primary purpose of uses is to support decision-making of clinical care of an identified individual. Other uses of EHR are involved to as secondary uses. The secondary uses of healthcare information can be regarded as a balance between individual security and society's necessity to reduce healthcare costs improving quality and efficiency of the healthcare treatment. It is necessary to use the EHRs in medical research, assessment of care quality and healthcare treatment planning and management. Therefore, the secondary uses of EHI ultimately enhance patients' benefits through a well-managed healthcare treatment environment.

It is required to design a secure environment for operating EHI to protect the privacy and confidentiality of the individual who receives healthcare services that are delivered through e-Health. Advances in security technologies have so far not eliminated the challenge posed by the need to secure EHR.

In [14], a security system for healthcare application clouds was proposed. Their system has three components: EHR secure collection and integration component, EHR secure storage and management component, and EHR secure usage component. Patients' EHRs are controlled by organizations not considering the special privacy requirements of patient themselves. Different individuals have different degrees of privacy demand in the same privacy case. Even an individual may have different degrees of privacy demand in different privacy cases. So the personal interests and privacy should be reflected in a security system.

In [1], a system providing privacy for EHR is proposed. The system for EHR includes two subsystems- the patient service and the data service. All access control policies are developed by the patients themselves. It fully takes into account personal interests and privacy. However, the reality of the situation will be very complicated. Some solutions of EHI access are not the simple problem of whether or not to release medical data. Some technologies of restrictive release should be considered. What's more, in the medical organizations, patients aren't familiar with the rights and responsibilities of doctors. The facts that all privacy preserving policies are made and implemented by patients are infeasible.

The contribution of this study is that we propose a security and privacy model for EHR which is a hybrid model of patients, doctors and organizations co-management through using a combination of means of three methods. Among them, a privacy preserving security model of EHR that offers raw EHI self-determination to the patients including usage control with implicit possibility to trace data flows after sensitive data has been legitimately disclosed. This part is managed by patient so called patient-centered. Furthermore, our security model involves a sub-model named T-EMR service for access control which is developed by local medical organization where patients get medical treatment to prevent malicious access. This part is managed by doctors and organizations so called organization-centered. As a result, there is no need to interact with the policy service for access to patients' EHR every time. Third but not the least, we adopt privacy preserving technology based on restrictive release in a sub-model for exploiting to protect patients' privacy when data are published for secondary uses.

# 2 Related research

Now we introduce the concept of Electronic Medical Record (EMR), Electronic Health Record (EHR), and Personal Health Record (PHR). EMR is the legal record of what happened to the patient during their encounter at a Care Delivery Organization (CDO) across inpatient and outpatient environments and is owned by the CDO. EMR is created, used and maintained by healthcare practitioners to document, monitor, and manage health care delivery within a CDO [14]. The electronic records sharing between different EMR systems are called electronic health records (EHRs). The EHR systems are created, maintained by community, state, or regional emergence, or national wide emergence organizations. The electronic personal healthcare records (EPHRs) [15] are online web-based healthcare records that are created, maintained and managed by the owner of the information. They have arisen as free services managed by IT enterprises, such as Google Health and Microsoft Health Vault. In fact, there is no clear boundary between EHR and EPHR considering the widespread cooperation between them.

We present some privacy preserving technologies which can be carried out on the EHI field.

## 2.1 Data Anonymization

Data anonymization is the current hot research. Anonymity is used to remove or dilute a data set of all a patient's identity information, making compromise between the risk of privacy disclosure and data accuracy, which take availability, security and privacy of data into account.

EHR typically contain three types of attributes. Explicit identifier is able to uniquely identify a single individual attribute, such as name, ID number and phone number. Quasi-identifiers are multiple attributes which combine to uniquely identify a person, such as zip code, gender, birthday and other co-expression. Sensitive attribute contains sensitive data, particularly in relation to the details of individual privacy, such as disease, illness records, personal salary etc. If the data sheets are published with only simple identifiers removed, private information may still be obtained through co-locating quasi-identifiers.

The principle of k-anonymity [2] requires that the release of the data table for each record cannot be distinguished from other records. Equivalence class is known as k records which cannot be distinguished. Generally the larger the value k, the better the protection of privacy, but lost the more information. There are many anonymity principles based on k-anonymity. The l-anonymity [3] requires the release of data sheet for each equivalence class has at least l different sensitive attribute values. The t-closeness anonymity [4] based on the of l-anonymity consider the sensitive attribute distributions within an equivalence class, requiring differences in the sensitive attribute distributions within an equivalence class and their values no more than t. Identity-keeping anonymity [5] is another anonymity approach. It is aimed at individuals in a data sheet, rather than data records. This principle protects privacy better in case of a single individual corresponding to multiple records. The above principles can redefine based on identity-keeping anonymity.

The above anonymity principles are for static data release, the following anonymity principles are for dynamic data. Under dynamic condition, we should ensure that not only every release satisfies the data anonymization, but also even the collaborative attack to multiple data releases can still protect privacy. The m-invariance anonymity [6] requires that in dynamic data release, there must be m pieces of records having different sensitive attribute values, and in the repeated releases of a record, sets of sensitive attribute values of the record's equivalence class must be equal. The l-scarcity anonymity [7] requires that at any time after the release of anonymization data, disclosure risk of any sensitive attribution values of any individual is no more than l.

There are different anonymity algorithms depending on the anonymity principles. We should pay attention to weigh carefully the security, privacy and accuracy of release data to select an anonymity principle of and an anonymous algorithm.

## 2.2 Data Pseudonymization

Anonymization removes explicit of the individual from EHRs mainly because the individual identity is unnecessary for secondary uses. However, situations exist where it may be required to re-create the link between the EHR and the owner of EHR [12]. Such situations include handling follow-up data, individual's request to withdraw their information, further treatment of a patient in light of new discoveries and quality control. Neubauer and Riedl [13] define the concept of pseudonymization as: a technique where identification data is transformed into, and afterwards replaced by, a specifier, which cannot be associated with the identification data without knowing a certain secret. Iacono[12] identifies two pseudonymization schemes based on the reversibility. The first is the one-way pseudonymization scheme, which generates pseudonyms which are impossible to be used to re-identify the patients. This type of scheme requires the maintenance of a mapping database to store associations between pseudonyms and explicit identifiers. The second is the reversible pseudonymization scheme, which allows the patient to be re-identified through the use of cryptographic mechanisms applied to the pseudonyms. The latter does not require a mapping database. Neubauer et al. [16] provides an methodology that combines primary and secondary use in one system and guarantees data privacy. The security analysis showed that the methodology is secure and protected against common intruder scenarios.

# 3 Security and Privacy Requirements

In practice, EHRs exist in a dynamic cycle in some form. The dynamic cycle includes collection, access, creation and publication of EHRs. In every link, there are specific security and privacy challenges. We list requirements for a privacy preserving EHR:

1) A patient may have his or her EHI in different EMR systems. To increase efficiency in medical services distributed EHI should centralized to EHR accurately, securely and fast.

2) When a patient sees a doctor, the medical organization where the patient is served should get patient's EHR accurately, securely and fast.

3) The medical organization should assure that doctors who provide the medical serve can obtain patient's EHR and unauthorized individual who is even at the same medical organization can't get the patient's EHR.

4) We need to address the authenticity of EHR with respect to both content authentication and source verifiability.

5) When a patient finishes treatment at a hospital, after completing the EHI this time, doctors can't access the EHR any more, unless the patient allows. If it's necessary to get raw information of patients for any uses, it must be allowed by patients and trails and logs can be audit.

6) If some organizations apply for EHI for secondary uses, information provider can't disclose patients' sensitive information.

# 4 A Security Model of EHR

We propose a privacy preserving security model of EHR including four sub-models: the EHR service, the temporary EMR (T-EMR) service, the EMR service and the patient service. Fig.1 shows the schematic architecture of these components and their resulting data flows.

The EHR service integrates distributed EMRs and controls the storage and disclosure of EHRs. Other medical organizations use the EHR service for storing or retrieving EHI. The T-EMR service is managed by an organization. A medical organization obtains the EHRs of patients, only when patients in hospital here. After authorized medical organization reaching EHR of a registered patient, the T-EMR service implement locally access control and generate new EHI. At this point, the organization is responsible for security and privacy of EHRs. That is to say, the patient is out of control of her or his own EHR which is more efficient and feasible, considering the critical moments. The EMR service controls the storage and disclosure of EMRs which is generated locally, after patients discharged from organization. Patients, other organizations or enterprises can obtain some local statistical information about any doctors or diseases by the EMR service. The patient service offers administrative communication and is an interface for patients to the system where they can develop privacy policies on the release of their data and check whether their enforcements are run as intended.

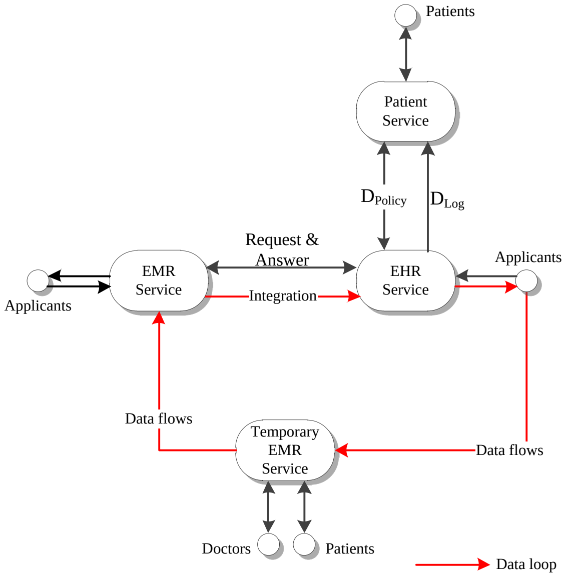


Figure 1. A security model of EHR.

# 5 EHR Service

The EHR service (Fig. 2) is one of the most important building blocks in the security model of EHR.

## 5.1 EHR Secure Collection and Integration

EHRs in the EHR Service are integrated from the distributed EMRs in the EMR Service. EMRs between the interaction entities meet cooperation and provide strong mutual authentication and responsibilities. Therefore, EHR system administrators verify authenticity, confidentiality, integrity and trust of EMRs from different medical institutions. EHR system administrators combine and integrate the successfully verified EMRs into a new composite EHR with a security certificate signed by the integrator. EHRs are encrypted and stored at EHR point.

## 5.2 Access Control of Raw EHRs

When there is an applicant request for raw EHRs, we implement following process. The policy enforcement point I (PEPI) enforces access control policies. Answer to an access request and storage of access control policies are done by the policy service, which combines a policy repository and the policy access point. When there is an applicant applies for

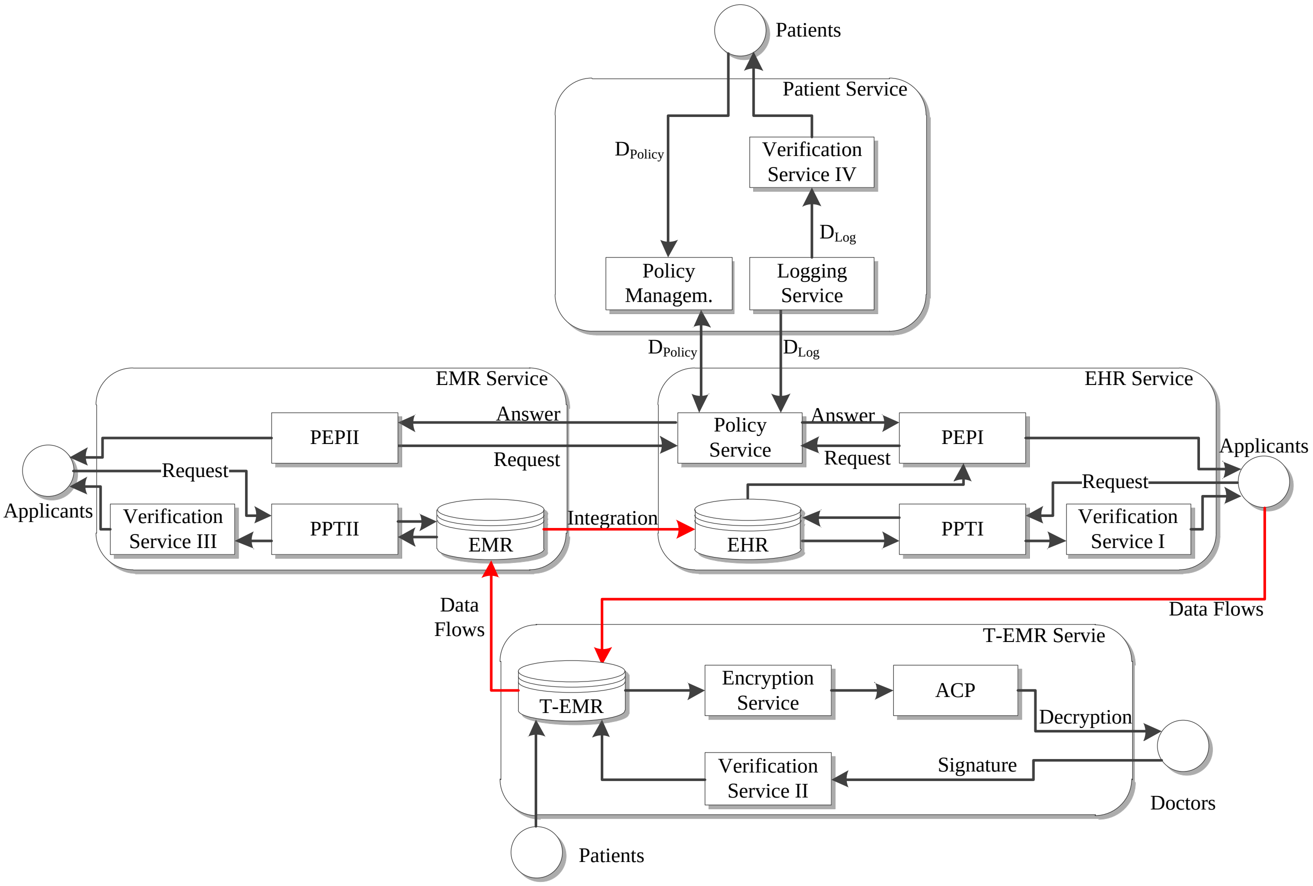


Figure 2. Architecture of the model

access EHR, the PEPI asks the policy service whether the request is admitted. The policy service decides according to the agreed-upon privacy policy $D_{Policy}$ made by the owner of EHR and returns the answer including relevant obligations. Then it sends $D_{Log}$ which records the access request and the resulting decision to the patient service.

Specially, when a patient registered in a hospital, the hospital will apply for raw EHRs of the patient with registration information signed by the patient. Of course, this application complies with privacy policy stated by the patient.

### 5.3 Access Control of EHRs after Handling of Privacy Preserving Technology

EHRs have an influential value for secondary uses. Centers for Disease Control need to collect EHRs for disease prevention and control. Research institutions for the good use of EHRs could benefit the patient. However, in this process, if the patient raw data release, which will inevitably expose sensitive data, the owners of sensitive data do not want to be exposed. So, whether it is data mining or data release, the raw data must take appropriate protective measures. On the one hand to protect personal privacy from being compromised, the other is to have sufficient information for analysis applies. Therefore, the privacy preserving technology I (PPTI) is necessary. PPTI tries to retain more validity of information while protecting the privacy of individual information. The verification service I verify the resulting data corresponding with the privacy preserving objectives and principles. Then pass EHR to the applicant. Privacy preserving technology based on restrictive release is to selectively release the raw data, not to release or to release of sensitive data of lower accuracy, in order to achieve privacy preserving.

Anonymization and pseudonymization especially anonymization are widely used in privacy technologies. Data anonymization and data pseudonymization attempt to address this problem by de-personalization of e-Healthcare information. However, these two methods ensure the privacy of the disclosure of the probability no more than a particular threshold. Data anonymization is a chronic problem. Therefore, before technologies taken into application, a thorough threat analysis is needed. Related methods and algorithms in anonymization and pseudonymization fields have be mentioned in section 2.

## 6 T-EMR Service

The temporary-EMR (T-EMR) service implements access control to protect the security and privacy of EHI locally. After a patient's registration in a medical organization, the organization applies for patient's raw EHRs. When the organization gets patient's EHRs, the encryption service encrypts them. The access control point (ACP) carries out access control according to the local doctor's roles and responsibilities based on the patient's treatment department and attending physician. The attending physician is probably in a medical stuff which is in charge of the patient. After the physician (or the stuff) completing the treatment and decision, the physician (or the stuff) generates new EMR with the signature of the physician (or the stuff). The signature of new EMRs generated locally need to be verified in verification service II. T-EMR point passes the successfully verified data to the EMR service.

## 7 EMR Service

The architecture of the EMR service is as Fig. 2. After the EMR point obtaining data of EMRs from the T-EMR service, it encrypts and storages them. The approach and the process of raw data on EHR access control and data privacy preserving technology handled on release process are both as same as the EHR service, except for the fact that supervisor of the EMR service and the data itself are different from the EHR service. The other difference is that the Policy Service is located in the EHR service. So every accessing to raw EMR, PEPII need to interactive with Policy Service in the EHR service in order to obtain the patient's security policy.

## 8 Patient Service

The patient service is the communication interface where patients can realize policy management and check data access log. It is divided into three components. The policy management is a point which patients use to view, present and modify access control policies $D_{Policy}$. Every access request creates an event that is delivered from the policy service of the EHR service to the logging service of the patient service. The verification service IV offers a list of access details to a verifier whether the system runs as intended and the agreed-upon policy has been enforced[1].

## 9 Conclusion

We have concluded security and privacy requirements for e-Healthcare application. Then we present an EHR security and privacy model for managing security issues in e-Healthcare information, which highlights four important core components. Patients are given right to decide on the usage and disclosure of personal raw EHRs. For the secondary use of EHRs, we adopt privacy preserving technology such as data anonymization to protect patients' privacy. Access control during patient treatment is managed by the medical organization where patient's treatment is implemented. The combination of the above three patterns of data management support security and privacy of e-Healthcare information. This hybrid model makes the management of EHI security, privacy, autonomy, quickness and efficiency.

## 10 Acknowledgment

This paper is based upon work partially supported by the National Natural Science Foundation of China (Grant No. 60973112) and Innovative Research Team in University and Science and Technology Foundation of Beijing Jiaotong University (No. 2007xm009).

## 11 References

[1] S. Haas, S. Wohlgemuth, I. Echizen, N. Sonehara, G. Müller, "Aspect of privacy for electronic health records," International journal of medical informatics, vol. 280, pp. 26-31, 2011.

[2] L. Sweeney. "k-anonymity: A model for protecting privacy," International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 5, pp. 557-570, 2002.

[3] A. Machanavajjhala, J. Gehrke, D Kifer, M. Venkita-Subramaniam. "l-diversity: Privacy beyond k-anonymity," Proceedings of the 22nd International Conference on Data Engineering (ICDE), Atlanta, Georgia, USA, pp. 24-35, 2006.

[4] N. Li, T. Li. "t-closeness：Privacy beyond k-anonymity and l-diversity," Proceedings of the 23rd International Conference on Data Engineering (ICDE), Istan Bottom-up k-anonymity, Turkey, pp. 106-115, 2007.

[5] Y. Hai, Y. Tao, S. Tang, D. Yang. "Identity-reserved anonymity in privacy preserving data publishing," Institute of software, No. 4, pp. 771-781(Ch) , 2010.

[6] X. Xiao, Y. Tao. "m-invariance: Towards privacy preserving republication of dynamic datasets," Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD), Beijing, China, pp. 689-700, 2007.

[7] Y. Bu, A. Fu, R. Wong, L. Chen, J. Li. "Privacy preserving serial data publishing by role composition," Proceedings of the 34th Very Large Data Bases (VLDB) Conference, Auckland, New Zealand, 2008.

[8] X. Xiao, Y. Tao. "Personalized privacy preservation," Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD), Atlanta, Georgia, USA, pp. 229-240, 2006.

[9] J. Xu, W. Wang, J. Pei, X. Wang, B. Shi, A. Fu. "Utility-based anonymization using local recoding," Proceedings of the 12th International Conference on Knowledge Discovery and Data Mining (SIGKDD), Philadelphia, PA, USA, pp. 785-790, 2006.

[10] T. Li, C. Tang, J. Wu, Q. Luo, S. Li, X. Lin, J. Zuo. "k-anonymity via clustering domain knowledge for privacy preservation," Proceedings of the 5th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Jinan, Shandong, China, No. 4 pp. 697-701, 2008.

[11] J. Pei, J. Xu, Z. Wang, W. Wang, K. Wang. "Maintaining k-anonymity against incremental updates," Proceedings of the 19th International Conference on Scientific and Statistical Database Management (SSDBM), Banff, Canada, pp. 5-14, 2007.

[12] L. Iacono. “Multi-centric universal pseudonymisation for secondary use of the EHR,” HealthGrid 2007, Geneva, Switzerland, 2007.

[13] T. Neubauer, B. Riedl. Improving Patients Privacy with Pseudonymization, IOS Press, 2008.

[14] R. Zhang, L. Liu, Security models and requirements for healthcare application clouds. Proceedings of the 3rd IEEE international conference on cloud computing (cloud2010), J Miami, Florida, USA, pp. 268-275, July5-10, 2010.

[15] D. B. Lafky and T. A. Horan. “Prospective personal health record use among different user groups,” Results of a multi-wave study. hicss, 0:233, 2008. ISSN 1530-1605. doi: http://doi.ieeecomputersociety.org/10.1109/HICSS. 2008.363.

[16] T. Neubauer, J. Heurix. A methodology for the pseudonymization of medical data. International journal of medical informatics, vol. 80, pp. 190-204, 2011.

# An Improvement over Chang and Lee's Electronic Voting Scheme

Hoda Ghavamipour
Computer Engineering & IT Department
Amir kabir University of Technology (Tehran Polytechnic)
Tehran, Iran
h.ghavamipour@aut.ac.ir

Maryam Shahpasand
Computer Engineering & IT Department
Amir kabir University of Technology (Tehran Polytechnic)
Tehran, Iran
m.shahpasand@aut.ac.ir

*Abstract*— In 2006, Chang and Lee [C. C. Chang, J. S. Lee, "An anonymous voting mechanism based on the key exchange protocol", Computers and Security vol. 25, 2006, pp. 307-314], proposed an e-voting protocol in which every voter must have a pair of public/private keys and some transmissions for setting session keys. In this paper, we propose an improvement over Chang and Lee's scheme such that the need of keys for voters is obviated and the number of servers and transmissions is decreased and it makes the proposed scheme more efficient in large scale electronic voting systems. Moreover, conspiracy between authorities can no longer violate privacy of voters. We further show This attack fails in our proposed scheme.

***Keywords: Electronic voting scheme; blind signature; key exchange protocol; Election.***

## I. INTRODUCTION

The great expansion of Internet use, allows the user to carry out a lot of tasks of all kinds such that: electronic trade, teleworking, databases enquiries, etc. Furthermore, the different governments and public administrations have been implied in this development and have put at citizen's disposal new services that have been called electronic government. In this approach, the society spreads to implant in the electronic environment all those performances that the citizens habitually develop and among them it can stand out the civic participation in the taking of decisions through what has been called the electronic voting. However, in order for electronic voting to replace conventional mechanisms, it must provide the whole range of features that conventional voting systems have. Further, due to the inherent lack of security in the Internet, electronic voting systems need to be carefully designed; otherwise these systems become more susceptible to fraud than conventional systems. Electronic voting has been intensively studied for over the last twenty years. Up to now, many electronic voting schemes have been proposed, and their security as well as their effectiveness has been improved. So far several electronic voting protocols have been appeared in the literature, [1-17].

In [8], Chang and Lee proposed a blind signature-based e-voting protocol which satisfies all security requirements mentioned in the literature. However, his scheme requires public/private keys for each voter and some transmissions between voters and register server for setting keys of their session. In this paper, we employ the ideas from [17] to present an improvement over Chang and Lee's scheme which decrease the number of servers and key exchanging between them and obviates the need of keys for voters but fullfiles the same security criteria as the scheme of Chang and Lee. Furthermore, our proposed scheme overcomes the corrupted servers attack which is reported for Chang and Lee.'s scheme in section 5.

The rest of this article is organized as follows: Section 2 and 3 introduces related techniques and notations used throughout the paper, respectively. In Section 4, we review Chang and Lee's electronic voting protocol and in section 5 we elaborate the security weakness. We propose a modified version of this scheme in section 6. Security analysis for our scheme is provided in section 7 and finally, comparison between the two schemes is outlined in the last Section.

## II. RELATED TECHNIQUES

Basically, the great majority is based on the use of three cryptographic primitives: mixnets, blind signatures and homomorphic encryption.

### A. Mixnets

Mixnets are similar to the anonymous channels that are used to distribute among the voters, in an anonymous and sure way, the credentials (digital certificate, etc.). In a more rigorous way, we can say that it is third trusted party that distributes messages among the voters in such a way that possible attackers are not able to determine the sender or receiver of a certain message. The use of the mixnets was proposed by Chaum [18].

### B. Blind sugniture

Blind signatures were initially used to design the first e-cash protocols. Later, they were used by Fujioka and Cuia. [10] to validate ballots in an electronic electoral outline. Roughly speaking, blind signatures schemes allow an authority to sign digitally some data (for example the ballot of a voter) without knowing the content of this data. As in the previous case, blind signatures were introduced by Chaum [19].

### C. Homomorphic encription

The homomorphic encryption was proposed by Cramer, Gennaro and Schoenmakers [16]. and it takes advantage of the characteristic properties of the homomorphic encryption to provide verifiability to the electronic vote schemes without contributing any information on the individual votes. In the homomorphic encryption model there are two operations: A sum ⊕, defined in the space of messages (votes), and a product ⊗, defined in the space of the cryptograms (ciphering votes), in such a way that the product ⊗ of two ciphering votes, $E(v_1)\otimes E(v_2)$, is the cryptogram of the sum of such votes: $E(v_1 \oplus v_2)$.

There are two other techniques which are essential in having secure communications. In the following we will briefly mention these two algorithms.

### D. RSA's operation

RSA's public operation (encrypting/verifying) is applied over the message x using m's public key $e_m$, i.e. x is encrypted (verified) as $x^{d_m}$(mod $n_m$). RSA's private operation (decrypting/signing) will be carried out using entity m's private key $d_m$ so that x is decrypted/signed as $x^{d_m}$ (mod $n_m$). Clearly we have: $x^{d_m e_m} = x^{e_m d_m}$=x (mod $n_m$).

### E. Diffie-Hellman key exchange protocol

In this subsection, we will introduce Diffie-Hellman key exchange protocol. The main concept of the protocol is that two participants, Alice and Bob, may negotiate a shared session key after some steps. We then define some notations used in the protocol. Let p be a large prime, and g be a primitive element in GF(p). Let $r_a$ and $r_b$ be two random numbers. Let $E_k()$ denote the encryption function with the encryption key k, and $D_k()$ denote the decryption function with the decryption key k. Let sk be the negotiated session key. The details of the protocol are introduced as follows.

Step 1: Alice generates a random number $r_a$ and computes $k' = g^{r_a}(mod\ p)$ . Then Alice sends the computation result to Bob.

Step 2: After Bob receives the message sent by Alice, he then generates a random number $r_b$ and computes the session key sk, where $= (g^{r_a})^{r_b}(mod\ p)$ . Next, Bob computes $E_{k'}$( $g^{r_b}$ (mod p))and sends it to Alice.

Step 3: Upon receiving the message sent by Bob, Alice computes: $D_{k'}(E_{k'}(g^{r_b}(\text{mod p})))$. to reveal $g^{r_b}$ (mod p). Then, Alice can computes the session key sk, where $sk = (g^{r_a})^{r_b}(mod\ p)$. Next, Alice computes $E_{sk}(g^{r_b}(\text{mod p}))$and sends it to Bob.

Step 4: While receiving the message sent by Alice, Bob computes: $D_{sk}(E_{sk}(g^{r_b}(\text{mod p})))$. to reveal $g^{r_b}$( mod p)and then checks if the computation result is equivalent to the one he calculated before. If it does hold, Bob can make sure that he is communicating with a legal user, and he sends $E_{sk}$( $g^{r_a}$( mod p))to Alice.

Step 5: When Alice receives the message sent by Bob, she then computes: $D_{sk}(E_{sk}(g^{r_a}(\text{mod p})))$. to reveal $g^{r_a}$( mod p) and then checks if the computation result is equal to the one she calculated before. If it does hold, Alice is convinced that she iscommunicating with a legal user. As a result, both Alice and Bob can use the session key sk to ensure the security of the following communications.

## III. NOTATIONS

Throughout the paper, we use the following notations.

| | |
|---|---|
| $(p_k, s_k)$ | the public key and the private key of CC, generated by RSA algorithm. |
| N | the product of $P_1$*$P_2$ , where $P_1$ and $P_2$ are two large primes, $P_1$ mod 4 = 3 and $P_2$ mod 4 = 3. |
| $((n_m , e_m ), d_m )$ | a set of RSA parameters for the entity m, where nm is the product of two large primes and $e_m$ . $d_m$ = 1(mod $n_m$ ). ($d_m$ ) is private while ($n_m$ , $e_m$ ) are public. |
| p | a large prime. |
| g | a primitive element in GF (p). |
| $V_i$ | The i'th voter. |
| $ID_i$ | the unique identification number of $V_i$. |
| $(x_r , y_r)$ | RC's public and private key, where $y_r = g^{x_r}$ (mod p). |
| $(x_m, y_m)$ | MC's public and private key, where $y_m = g^{x_m}$ (mod p). |
| $(x_v, y_v)$ | VC's public and private key, where $y_v = g^{x_v}$( mod p). |
| $(x_i, y_i)$ | $V_i$ 's public and private key, where $y_i = g^{x_i}$( mod p). |
| h() | a public one-way hash function. |
| $m_i$ | the marked ballot of $V_i$ . |
| $t_i$ | a time-stamp generated by $V_i$ . |
| $D_k$ (),$E_k$ () | the encryption and decryption function with the encryption key k. |
| f () | a one-way permutation function generated by the voting center. |
| TR/ TR' | the tally result of all votes. |
| A├ B(m) | A sends the message m to B. |
| RC | Registration Center(RC) is a trusted website at which all enrolled electors have to register in advance. |

CC Certification Center(CC) is a website which provides the blind signatures for all elector's ballots.

MC Monitor Center(MC) is responsible for supervising the whole procedure of electronic voting and announcing the final voting result.

VC Vote Counter (VC) is responsible for tallying the valid votes to obtain the final voting result.

## IV. A REVIWE OF CHANG AND LEE'S VOTING PROTOCOL

Here, we are going to review Chang and Lee's electronic voting protocol.

### A. Participants

In this scheme [8], there exist six participants listed as follows:

1. Registration Center (RC)
2. Certification Center(CC)
3. Monitor Center (MC)
4. Vote Counter (VC)
5. Proxy Server : The proxy server takes the responsibility for replacing the original network address of a ballot by another address such that the property, anonymity, can be achieved .
6. Voter : Voter denotes the person who has the right to vote.

Chang's electronic voting scheme consists of three phases, the initialization phase, the voting phase, and the publication phase.

### B. The initialization phase

At the beginning, MC and VC must negotiate a session key by Diffie-Hellman key exchange protocol.Then, RC also has to negotiate a common key with MC/VC.

Step 1: MC generates a nonce $N_1$, computes $k = y_V^{x_m} \bmod p = g^{x_V x_m} \pmod p$ and MC ├ VC($E_k(N_1)$).

Step 2: VC computes $k = y_m^{x_V} (mod\, p) = g^{x_m x_V} (mod\, p)$ and $D_k(E_k(N_1))$ to reveal $N_1$. Then, VC checks if $N_1$ is in the decryption result for freshness checking. If it does hold, VC ├MC($E_k(N_1 + 1)$).

Step 3: MC computes $D_k(E_k(N_1 + 1))$ and checks if $N_1$ +1 is in the decryption result for freshness checking. If it does hold, MC can make sure that it is communicating with the legal VC. Therefore, both MC and VC will use k as a shared key for secure latter communication.

Step 4: Both MC and VC use k to generate a common public key $k'$ by computing: $k' = g^k \, mod\, p$.

Step 5: MC and VC then use $k'$ to negotiate another common key $k^*$ with RC, respectively. The same steps between MC and VC will happen between MC/VC and RC to get a shared $k^*$ key for secure latter communications.

### C. The voting phase

Step 1: At first, the voter $V_i$ chooses two random numbers, $RD_1$ and $RD_2$, and computes $M_i = RD_1 \oplus RD_2 \oplus m_i$. Then, Vicomputes and $\hat{k} = y_r^{x_i} (mod\, p)$ and Vi ⊢ RC($E_{\hat{k}}(M_i, personal\, information,\ N_3)$).

Step2: RC computes $\hat{k} = y_r^{x_i} (mod\, p)$ and $D_{\hat{k}}\left(E_{\hat{k}}(M_i,\ personal\, information,\ N_3)\right)$ and checks the identification of $V_i$. If $V_i$ is a legal voter, it then computes $B_i = E_{k^*}(M_i \parallel t_i)$and generates a unique serial number SN-$V_i$ for $V_i$. Next, RC ├ $V_i$($E_{\hat{k}}(B_i, SN\_V_i, N_3)$).

Step 3: Vi computes$D_{\hat{k}}(E_{\hat{k}}(B_i, SN_V_i, N_3))$ to reveal $N_3$ and checks if $N_3$ is in the decryption result for freshness checking. If it does hold, $V_i$ then $C_i = h(B_i)RM^{pk} (mod\, N)$ where RM is a random number and RM ∈ ZN. Next, $V_i$ ├ CC ($C_i$).

Step4: CC computes $BG_i = (C_i)^{sk}\, mod\, N = h(B_i)^{sk} RM\, (mod\, N)$ and CC ├ $V_i$($BG_i$).

Step 5: $V_i$ computes$SG_i = BG_i RM^{-1} = h(B_i)^{sk}\ (mod\, N)$ to get the signature of h($B_i$).

Step 6: $V_i$ computes $h(SN - V_i)$and $V_i$ ├ MC(h(SN-$V_i$), $SG_i$, $B_i$, $RD_1$) and $V_i$ ├ VC(h(SN-$V_i$), $SG_i$, $B_i$, $RD_2$) through a trusted proxy server which can conceal the original network address of $V_i$.

Step 7: MC and VC execute the following procedure to confirm the validity of $V_i$.

h($B_i$) =?($SG_i$)$^{pk}$ (mod N)

If the above equation does hold, MC and VC will store (h(SN-Vi), SGi, Bi, $RD_1$) and (h(SN-Vi), SGi, Bi, $RD_2$) in their own databases, respectively. Besides, MC and VC have to make sure that h(SN-$V_i$) is stored in its database only once, and the duplications should be canceled.

### D. The publishing phase

Step 1: After the deadline of voting, VC├MC( $E_k(RD_2)$). At the same time, MC├VC($E_k(RD_1)$). Next, VC computes $D_k(E_k(RD_1))$ to reveal $RD_1$. while MC computes $D_k(E_k(RD_2))$ to reveal $RD_2$.

Step 2: MC and VC execute Steps 2-1 and 2-2, respectively, at the same time to tally the vote.

Step 2-1: VC computes $D_{k^*}(E_{k^*}(M_i \parallel t_i))$ and checks if $t_i$ is in the computation result for freshness checking. If it does hold, VC computes $m_i = RD_1 \oplus RD_2 \oplus M_i$. Finally, VC ├MC($E_k$(TR)), where TR is the tally result.

Step 2-2: MC computes $D_{k^*}(E_{k^*}(M_i \parallel t_i))$ and checks if $t_i$ is in the computation result for freshness checking. If it does hold, MC computes $m_i = RD_1 \oplus RD_2 \oplus M_i$. and then calculates the tally result TR'.

Step 3: MC computes $D_k(E_k(TR))$ to retrieve TR. Next, MC compares TR with TR'. If they are not equivalent, the tally result cannot be confirmed; otherwise, MC announces the final result TR.

## V. SECURITY WEAKNESS OF CHANG'S SCHEME

In this section, we show that if the authorities collude, they can identify voters of published ballots and violate privacy of voters as follows:

In the voting phase of this scheme, the registration authority RC must have the personal information (such as ID number) of legal $V_i$ to generate a unique serial number (SN-$V_i$) for him/her. Hence RC knows the link between each voter and his/her serial number. On the other hand, MC and VC, in the publishing phase should store h(SN-$V_i$) of the voter in their databases to avoid double voting. Note that the hash function h() is public and everyone who knows (SN-$V_i$) can compute h(SN-$V_i$). Therefore, if RC and one of MC or VC collude, they can build a table VDB with entries (Vi's personal information, h(SN-$V_i$), $M_i$) corresponding to each voter. So at the end of the voting period, the corrupted authorities can link any ballot ($m_i$) to the voter who has cast it. Therefore, they can violate privacy of voters.

## VI. IMPROVEMENT TO CHANG AND LEE'S SCHEME

In this section, we propose a modified version of Chang and Lee's scheme which obviates the need of keys for voters and decrease the number of voter's transmisions.

The main assumption of this improvement is that the CC has a database including all eligible voters'ID.

### A. Participants

In our proposed electronic voting scheme, there exist five participants listed as follows:

1. Certification Center(CC)
2. Monitor Center(MC)
3. Vote Counter(VC)
4. Proxy Server : The proxy server takes the responsibility for replacing the original network address of a ballot by another address such that the property, anonymity, can be achieved .
5. Voter : Voter denotes the person who has the right to vote.

As the chang's scheme, our improved scheme consists of three phases, the initialization phase, the voting phase, and the publication phase.

### B. The initialization phase

At the beginning, MC and VC must negotiate a session key by Diffie-Hellman key exchange protocol.

Step 1: MC generates a nonce $N_1$, computes $k = y_V^{x_m} \bmod p = g^{x_v x_m} \pmod p$ and MC $\vdash$ VC($E_k(N_1)$).

Step 2: VC computes $k = y_m^{x_V} modp = g^{x_m x_V} modp$ and$D_k(E_k(N_1))$ to reveal $N_1$. Then, VC checks if $N_1$ is in the decryption result for freshness checking. If it does hold, VC $\vdash$MC($E_k(N_1 + 1)$).

Step 3: MC computes $D_k(E_k(N_1 + 1))$ and checks if $N_1$ +1 is in the decryption result for freshness checking. If it does hold, MC can make sure that it is communicating with the legal VC. Therefore, both MC and VC will use K as a shared key for secure latter communication.

### C. The voting phase

Step 1: At first, the voter $V_i$ chooses seven random numbers, $R_1$, $R_2$, $R_3$, $R_4$, RM, $RD_1$ and $RD_2$, and computes:

hID = f(ID,RM);<br>
$BhID = R_1^{pk}hID$;<br>
$M_i = RD_1 \oplus RD_2 \oplus m_i$;<br>
$BM_i = R_2^{pk}M_i$;<br>
$BRD_1 = R_3^{pk}(RD_1 \| hID)$;<br>
$BRD_2 = R_4^{pk}(RD_2 \| hID)$.

Then, Vi $\vdash$ CC (ID, BhID, $BM_i$, $BRD_1$, $BRD_2$).

Step 2: CC checks the identification of $V_i$. If $V_i$ is a legal voter, then it computes:

$SBhID = BhID^{sk}$;<br>
$SBMi = BM_i^{sk}$;<br>
$SBRD_1 = BRD_1^{sk}$;<br>
$SBRD_2 = BRD_2^{sk}$;

Then, CC $\vdash$ $V_i$ (SBhID, $SBM_i$, $SBRD_1$, $SBRD_2$).

Step 3: $V_i$ then computes:

$ShID = SBhID(R_1^{-1}) = hID^{sk} \pmod N$;<br>
$SM_i = SBMi(R_2^{-1}) = M_i^{sk} \pmod N$;<br>
$SRD_1 = SBRD_1(R_3^{-1}) = (RD_1 \| hID)^{sk} \pmod N$;<br>
$SRD_2 = SBRD\ (R_4^{-1}) = (RD_2 \| hID)^{sk} \pmod N$;<br>
$EnSRD_1 = SRD_1^{e_{MC}} (\text{mod } n_{MC})$;<br>
$EnSRD_2 = SRD_2^{e_{VC}} (\text{mod } n_{VC})$.

Then $V_i$ $\vdash$MC (ShID, hID, $SM_i$, $EnSRD_1$) and $V_i$ $\vdash$ VC (ShID, hID, $SM_i$, $EnSRD_2$) , through a trusted proxy server which can conceal the original network address of $V_i$.

Step 4: MC and VC execute the following procedure to confirm the validity of $V_i$.

hID =?$ShID^{pk}$ (mod N)

If the above equation does hold, MC and VC will store (hID, $SM_i$, $EnSRD_1$) and (hID, $SM_i$, $EnSRD_2$) in their own databases, respectively. Besides, MC and VC have to make sure that hID is stored in their databases only once, and the duplications should be canceled.

### D. The publishing phase

Step 1: After the deadline of voting, VC computes $(EnSRD_2)^{d_{VC}}$ (mod $n_{VC}$))$^{pk}$ (mod N) to retrieve $RD_2$ and then VC $\vdash$MC($E_k(RD_2)$). At the same time, MC computes $((EnSRD_1)^{d_{MC}}$ (mod $n_{MC}$))$^{pk}$ (mod N) to retrieve $RD_1$ and then MC $\vdash$VC($E_k(RD_1)$). Next, VC computes $D_k(E_k(RD_1))$ to reveal $RD_1$ while MC computes $D_k(E_k(RD_2))$ to reveal $RD_2$.

Step 2: MC and VC execute Steps 2-1 and 2-2, respectively, at the same time to tally the vote.

Step 2-1: VC computes $M_i = SM^{pk_i}$ (mod N) and then computes $m_i = RD_1 \oplus RD_2 \oplus M_i$. Finally, VC $\vdash$MC($E_k(TR)$), where TR is the tally result.

Step 2-2: MC computes $M_i = SM^{pk_i}$ (mod N) and then computes $m_i = RD_1 \oplus RD_2 \oplus M_i$. Then calculates the tally result TR'.

Step 3: Upon receiving the message sent by VC, MC computes $D_k(E_k(TR))$ to retrieve TR. Next, MC compares TR with TR'. If they are not equivalent, the tally result cannot be confirmed; otherwise, MC announces the final result TR.

## VII. Security Analysis of the IMPROVED scheme

In this section, we analyze the security of the improved scheme and prove that the scheme satisfies the requirements mentioned in [8].

*Lemma* 1. (Completeness). All ballots are counted correctly in the proposed scheme.

*Proof.* As it is assumed that the databases of MC and VC are public all the time, any single voter can personally check that his/her vote is correctly recorded in these databases all the time and if not, he/she can publish the receipt. Therefore, the ballot of a registered voter which does not collide with other ballots is accepted and correctly counted. However, if a ballot collision occurs, only one of the collided ballots is accepted and correctly counted. We show that this does not happen in our scheme. Let vote1 = (hID, . . .) and vote2 = (hID, . . .) be two collided ballots corresponding to voters with identification $ID_1$ and $ID_2$, respectively. This means that hID = $f(ID_1,RM_1)$ = $f(ID_2,RM_2)$. Since f is a one-way permutation function, $ID_1$ = $ID_2$ and that $RM_i$'s are random numbers generated by different voters, this is impossible. To ensure the completeness of voting, all ballots cannot be removed, duplicated or altered. In our scheme, both MC and VC can only cooperate to reveal $m_i$ by computing $m_i = RD_1 \oplus RD_2 \oplus M_i$. Besides, both MC and VC will count all votes to have the voting result doubly checked such that miscounting votes will be eliminated from our scheme. Therefore, all ballots are counted correctly, i.e., the proposed scheme is complete.

*Lemma* 2. (Robustness) . No voter can disrupt the voting in the proposed scheme.

*Proof.* In order for a dishonest voter to disrupt the election, he/she must create a fake instance of vote where hID = $ShID_{pk}$ (mod N). However, the security of RSA makes this impossible. Furthermore, since no global computation is necessary our scheme, a single voting failure can not disrupt the whole election. Hence, the proposed scheme achieves robustness.

*Lemma* 3. (Uniqueness) . Every voter votes exactly one time.

*Proof.* In the voting phase, each eligible voter can receive blind signature from CC which he/she uses to create a ballot just one time. Otherwise, both MC and VC will detect the duplications in their databases and delete them. Hence, this requirement is confirmed in our scheme.

*Lemma* 4. (Fairness) . In the proposed scheme, no one can know the intermediate results of the voting.

*Proof.* No one can learn any information about tallied votes until the voting result is announced. It is due to that MC and VC cannot decrypt the vote result without the knowledge of $RD_1$ and $RD_2$. Before announcing the tally result, MC and VC have to exchange the knowledge of $RD_1$ and $RD_2$ such that they can compute $m_i = RD_1 \oplus RD_2 \oplus M_i$ to retrieve $m_i$. Consequently, this requirement is also confirmed in our scheme.

*Lemma* 5. (Anonymity) . all ballots must be secret, i.e., it should be impossible to link a ballot to the voter who has cast it.

*Proof.* During the voting phase, the information that CC receives, contains only random multiples of the hidden information of the voter (hID) plus ($M_i$) and its random numbers $RD_1$ and $RD_2$, which CC blindly signs. Hence, it is guaranteed that CC does not obtain any information about the parameters of the vote. In publishing phase, hID directly appears in the vote sent through an anonymous channel (that suppresses the origin of the message) to VC and MC. No one except the voter himself has seen hID so far and by the properties of one-way permutation function f, no link can be derived between hID and ID. So the scheme achieves anonymity.

*Lemma* 6. (Uncoercibility) . (a) A voter cannot prove to a coercer, how he has voted. As a result, verifiable vote-buying is impossible. (b) Only a voter can decide his/her intention.

*Proof.* We have to provefirst that verifiable vote -buying is impossible in our scheme. Since the intention of a voter $m_i$ is concealed as $RD_1 \oplus RD_2 \oplus m_i$, furthermore two random numbers $RD_1$ and $RD_2$ are concealed as $\{(RD_1 \| hID)^{sk} \pmod N\}^{e_{MC}} \pmod{n_{MC}}$ and $\{(RD_2 \| hID)^{sk} \pmod N\}^{e_{VC}} \pmod{n_{VC}}$, so voter can not prove (without sk and $d_{MC}$ and $d_{VC}$) the contents of his/her vote to a coercer. However, in addition to this, the exact definition of uncoercibility requires that only a voter can decide his/her intention. In [7], it is pointed out that to achieve this goal, we must sacrifice mobility and set voting booths for example. Hence our scheme achieves partial uncoercibility and can not solve the problem of a coercer participating in a voter's voting behavior

*Lemma* 7. (Convenience). Voters should be able to cast their ballots quickly, in one session, and with minimal equipment or special skills.

*Proof.* Our scheme does not require any special device and the whole voting procedure is carried out in one session. In particular voters do not need to learn any special skills. Hence convenience is assured.

*Lemma* 8. (Efficiency) . The whole election should be held in a timely manner and all computations done in a reasonable amount of time and voters are not required to wait for other voters to complete the process.

*Proof.* The computations performed by voters in our scheme, consist solely of modular exponentiation for which there are efficient algorithms. Further, voters do not need to wait for other voters to complete the process. hence efficiency is assured.

*Lemma* 9. (Mobility) . Voters are not restricted by physical location from which they can cast their votes.

*Proof.* As for mobility, with our scheme voters can vote through the Internet and voters are not restricted by physical location.

*Lemma* 10. (General election) . The intentions of voters are not just in ”yes” or ”no”, voters can choose someone from among several candidates.

*Proof.* This is obvious as the scheme imposes no restriction on voter’s intention.

## VIII. COMPARISON BETWEEN THE TWO SCHEMES

Both the scheme proposed in this paper and the scheme of Chang and Lee satisfy most of the security requirements mentioned in the literature. The main contributions of this paper was the decent of the transmissions in initialization and voting phasesand the obviation of keys for voters, with respect of all the security requirements.

also, if the authorities are corrupted, they can identify voters of published tickets at will and violate voters’ anonymity in Chang and Lee’s scheme. This attack fails in our improved scheme. From the viewpoint of efficiency, the numbers of transmissions are almost decreased, therefore achieve higher conveniencein large scale electronic voting systems.

The number of transmissions need for both schemes can be seen in table I.

TABLE I. COMPARISION BETWEEN THE NUMBER OF TRANSMISSIONS IN TWO SCHEMES

| Sum | phases | | | |
|---|---|---|---|---|
| | The initialization phase | The voting phase | The publication phase | |
| (9*N)+5 | 5 | 6*N[a] | 3*N | Chang and Lee’s scheme |
| (7*N)+2 | 2 | 4*N | 3*N | Our proposed scheme |

a. N is the number of voters.

## REFERENCES

[1] A. Fujioka, T. Okamoto, K. Ohta, “A practical secret voting scheme for large scale elections”, ASIACRYPT 92: Proceedings of the Workshop on the Theory and Application of Cryptographic Techniques, Springer-Verlag, London, UK, 1993, pp. 244-251.

[2] J. Karro and J. Wang, Towards a practical, secure, and very large scale online election, ACSAC 99: Proceedings of the 15th Annual Computer SecurityApplications Conference, IEEE Computer Society,Washington, DC, USA, 1999, p. 161.

[3] W. C. Ku, S. D. Wang, “A secure and practical electronic voting scheme”, Computer Communications vol.22, 1999.

[4] B. Schoenmakers, “A simple publicly verifiable secret sharing scheme and its application to electronic”, CRYPTO 99:Proceedings of the 19th Annual International Cryptology Conference on Advances in Cryptology,Springer-Verlag, London, UK, 1999, pp. 148-164.

[5] I. Ray, N. Narasimhamurthi, “An anonymous electronic voting protocol for voting over the Internet”, WECWIS 01: Proceedings of the Third International Workshop on Advanced Issues of E-Commerce and Web-Based Information Systems (WECWIS 01) IEEE ComputerSociety,Washington, DC, USA,2001,P.188.

[6] R. Joaquim, A. Zuqueteand, P. Ferreira, and Revs, “A robust electronic voting system”, Proceedings of IADIS International Conference e-Society, 2003, pp. 95103.

[7] H. T. Liaw, “A secure electronic voting protocol for general elections”, Computers and Security, vol. 23, 2004.

[8] C. C. Chang, J. S. Lee, “An anonymous voting mechanism based on the key exchange protocol”, Computers and Security, vol. 25, 2006.

[9] X. Chen, Q. Wu, F. Zhang, H. Tian, B. Wei, B. Lee, H. Lee, K. Kim, ”New receipt-free voting scheme using double-trapdoor commitment”, Information Sciences, vol. 181, 2011, pp. 1493–1502.

[10] F.Song, Z.Cuia, “Electronic voting scheme about elgamal blind-signatures based on XML”, International Workshop on Information and Electronics Engineering (IWIEE), 2012, pp.2721-2725.

[11] Y.F. Chung, Z.Y. Wu, ”Approach to designing bribery-free and coercion-free electronic voting scheme”, The Journal of Systems and Software, vol. 82, 2009, pp. 2081–2090.

[12] O. Cetinkaya, “Analysis of security requirements for cryptographic voting protocols”, The Third International Conference on Availability, Reliability and Security, 2008.

[13] H. Haiyan, S. Chang, “Study and implementation of the electronic voting plan on blind signature”, International Conference on Mechatronic Science, Electric Engineering and Computer August 19-22, 2011, Jilin, China.

[14] W. M. dos Santos, R. J. G. B. de Queiroz, “Preserving vote secrecy in end-to-end verifiable voting systems”, IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, 2011.

[15] T. Kobayashi, “An evaluation of the secret voting method using a one-way homomorphic function for network meetings”, International Symposium onCommunications and Information Technologies (ISCIT), 2010 , pp. 539 – 544.

[16] R. Cramer, R. Gennaro, B. Schoenmakers, “A secure and optimally efficient multi-authority election scheme”, Advances in Cryptology-Eurocrypt vol. 97, 1997, pp. 103-118.

[17] Z. Eslami, H. Ghavamipoor, “an improvement over liaw’s electronic voting scheme”, IADIS e-Society, 2010.

[18] D. Chaum, “Untraceable electronic mail, return addresses,and digital pseudonyms”, Communications of the ACM, vol. 24(2), 1981.

[19] D. Chaum, “Blind signatures for untraceable payments”, CRYPTO, 1982, pp. 199203.

# A Covert Channel Over Transport Layer Source Ports

**James R. F. Gimbi, Daryl Johnson, Peter Lutz, Bo Yuan**
B. Thomas Golisano College of Computing & Information Sciences
Rochester Institute of Technology, Rochester, NY, USA

**Abstract –** *Covert communication is a rapidly expanding field of research with significant impact on the security theater. These communication methods, or "covert channels", can be applied in a number of ways, including as a mechanism for an attacker to leak data from a monitored system or network. This paper sets out to contribute to this field by introducing a new covert channel which operates over transport layer protocols. The mechanism is flexible, covert, and has the potential to operate at relatively high bandwidth. In addition, this paper proposes a number of encoding schemes which can be used in conjunction with this channel to improve its bandwidth and covertness.*

**Keywords:** Network Covert Channels, Information Hiding, Network Security

## 1 Introduction

A covert channel can be defined as any communication method where both the data being transmitted and the existence of the channel itself are hidden from network and system authority figures. The field has generated much interest because of its applications on both sides of the information security industry; while covert channels are useful for defensive security applications and collaboration between legitimate security teams, they can also be used by attackers to covertly leak data from a secure environment.

This paper presents a novel method for leveraging transport layer source ports as a medium for covert communication. The technique is flexible and can be applied in a wide variety of environments. The paper then discusses a number of possible implementations of this channel. It will also introduce a collection of encoding mechanisms to use in conjunction with the channel and will review their utility. Some of these encoding techniques provide data integrity and obfuscate the channel from a would-be investigator.

## 2 Related Work

Covert channels have been the subject of research for some time. They were originally defined in [1] as any communication medium not designed or intended for data transfer but could be used as such. Multiple types of covert channels have been defined, including storage channels, timing channels and behavioral channels. This topic was explored in depth in [2]. A storage channel is essentially any channel where a shared storage medium is used to encode and transmit information. A timing channel is any communication which relies on the time between particular events to encode and transmit information, instead of shared storage media. Behavioral channels are broadly defined as any channel where the mechanism is non-stored and time independent.

The channels covered in [1] are exclusively single system process-to-process examples. Since then, the definition of a covert channel has gradually expanded to include channels between processes on two separate machines over a computer network. [3] provided solid groundwork for creating TCP/IP timing channels. The approach encoded data in the amount of time between to the arrival of two packets. TCP/IP storage channels were thoroughly examined in [4]. In this work, data is transmitted in header fields of TCP/IP packets. Two known works have identified transport layer source ports as a potential channel medium in passing but did not discuss how it might be accomplished [5] [6].

## 3 Covert Channel Over Source Ports

This section defines and outlines a method for using transport layer source ports as a covert channel. A technical background will be provided in subsection 3.1. The method itself is introduced in 3.2. The final subsection will propose a number of different technical implementations.

### 3.1 Technical Background

Communication between two computers over modern network protocols requires the use of what is known as a network socket. A socket is a tuple of data used to identify each unique and active connection on a particular machine, and a socket pair is a tuple containing information for both the local and remote sockets [7]. While the exact contents of this tuple will vary depending on which transport protocol is being used, the TCP socket pair includes the IP addresses of both machines as well as the port numbers each machine has committed to the session. The IP addresses help the computer keep track of which remote machine it is communication with, while the port numbers help keep track of individual sessions for that machine. This 4-tuple allows two computers to manage thousands of unique conversations between them without risk of data loss on any one session. For example, a web client can make two distinct GET requests to a particular the web server for different page elements at port 80. Because they are two separate requests, different source port numbers are selected by the client (i.e. port 1,111 for one socket pair, and port 2,222 for the second socket pair) so the server knows which remote socket to feed the appropriate HTML response.

While most services are given a dedicated port number with which to manage all connections, client ports are not static and will not always be the same under normal operation. Instead, the source port is generally a pseudo-random number selected from a given range. Ports selected like this are known as ephemeral or temporary ports. This gives the client operating system flexibility when establishing new connections. There is technically nothing to prevent a client from using any port within the 16 bit port range (ports 0 through 65,535) but there are suggested standard ranges which most transport layer protocol implementations observe. A commonly observed range is maintained by the Internet Assigned Numbers Authority (IANA) which mandates that the two most significant bits must be registered as ones, giving a usable ephemeral range of 49,152 through 65,535 for a total of 16,384 available ports [8]. While Microsoft operating systems now follow the recommended IANA range, legacy Microsoft systems, such as Windows XP and 2000, use the range 1,025 through 5,000 for a total of 3,976 available ports [9]. Linux systems tend to vary from distribution to distribution but most use either the IANA mandate or the range 32,768 through 61,000 for a total of 28,233 available ports.

## 3.2 Transport Layer Covert Channel

The following method takes advantage of the flexibility provided by layer four protocols in source port selection and can be applied to any type of network environment that uses a layer four protocol, such as TCP and UDP. It consists of a sender, which will transmit data over the channel, and a receiver, which will collects the transmitted data. The sender and receiver must be able to communicate in some legitimate fashion without being flagged by a security appliance. For example, this pair might be a web server and client (TCP), a streaming media server and client (UDP), or some proprietary protocol.

Each time a new source port is needed there is an opportunity to transmit up to sixteen bits of information from the client to the server in that source port field. A one-way channel is established when a user or process manipulates the source port to send data. Because all that is modified for this channel is the contents of a mandatory static length field, it can easily be piggybacked on top of legitimate traffic. The channel lends itself to a large number of different encoding mechanisms, two of which will be outlined in the next section.

While it is possible to use these bits to transmit absolute data (i.e. sending an ASCII ‘A’ by using port 65), the channel is made more covert and robust by using the delta between two consecutive source ports. Using a delta scheme, no data is actually stored in a given source port; an analyst could investigate the totality of a guilty packet and find no leaked data. By contrast, absolute data transmission can easily be detected by an analyst reviewing a packet. Further, delta schemes lower the likelihood of colliding consecutive source ports because repeated characters will not use the same port number. These collisions could cause problems if the channel is run over legitimate traffic [10].

Bandwidth for this channel might appear to be limited because of the tendency of most transport layer protocols, particularly TCP, to use one socket per session. However, it is not atypical for multiple sessions to be generated per task. For example, when a typical web browser retrieves the HTML document, style sheet, images and other elements from a single web page it will frequently establish several sessions with the remote server so that it can make many requests at a time, enhancing protocol performance. Each of those requests uses a different ephemeral source port, meaning that simply accessing a lone web site with many elements can provide adequate cover for this channel at high bandwidths. Similarly, any protocol that takes advantage of parallel network sessions could support high bandwidths with this channel.

If the sender and receiver communicate on a regular basis the channel does not need to generate any new traffic. If they do not normally communicate, there is much flexibility in the traffic that can be used because of the application-neutral nature of the channel. Virtually any protocol can be selected for packet generation. This makes the channel simple to customize for any number of environments without raising the suspicions of common security appliances or analysts.

The channel does have a number of inherent weaknesses. For instance, the prolific use of network address translation technology (NAT) stands to limit the utility of the channel as described. This is because many NAT implementations modify the socket pair so that the source port received by the receiver cannot be reliably controlled. As such, if the sender lays behind a NAT box this channel is limited to communicating with other machines behind the NAT box. Similarly, proxy servers typically change the socket pair, again limiting the applicable scope of the channel. Sometimes the use of proxy servers is enforced even within a LAN, potentially crippling the channel. Legitimate traffic from the sender can possibly interfere with the channel in two separate ways. First, if another unrelated process makes use of an ephemeral port, that port will be locked from other processes until the TIME_WAIT timer expires. This timer, built into TCP with RFC793, is designed to ensure that the socket can still properly handle traffic arriving late from a closed connection [10]. If the source port required for the next data transmission is still in TIME_WAIT, a poorly written or light implementation might crash. While it is possible to work around this issue, higher level permissions are generally required. Second, if the sender and receiver communicate for some legitimate reason outside of the channel process it is possible that the receiver misinterprets the source port used in that exchange as part of the message, corrupting the data and calling to question the integrity of data received over this channel. This last problem can be effectively eliminated by using a robust encoding mechanism like the one discussed in subsection 4.2.

## 3.3 Potential Implementations

This method can be implemented in any number of ways ranging from the very clumsy to the very elegant. A simple implementation might generate false traffic with no real purpose other than providing a medium for the channel. Such an implementation would have high bandwidth but would be easy to identify as it would carry no changing data except for the source port. A more sophisticated version might act as a local wrapper for applications to use which would replace source port addresses for packets it receives and map it back to the original address, not unlike the basic functionality of NAT. Legitimate client applications could be a modified to take advantage of the channel. For example, a web browser can be modified to use the proper ephemeral port unless it is communicating with the intended receiving server, in which case it would use encoded delta ports instead of standard ephemeral ports.

A much more elegant approach than these might include a kernel level modification on the sender. For instance, every time any application communicates with the intended receiver, the sender kernel selects encoded delta ports. An implementation like this would eliminate the need to manage redundancy checking (discussed in section 4.2), greatly improving bandwidth while only using legitimate user traffic to transmit data.

# 4 Encoding Mechanisms

## 4.1 Simple Encoding Schemes

One example of a simple delta encoding scheme for this channel is to use the difference between two raw consecutive port numbers as the value to be transmitted. For example, if a user wanted to transmit the message “ABC” over the channel, they might first start a session with the source port 50,000, followed by 50,065, then 50,131, and finally 50,198. The differences between each port are 65, 66, and 67 respectively, which are the values of the ASCII decimal representations of the above message. This is represented visually in table 1 where the non-italicized bits carry the encoded data.

Table 1: Basic 8-bit Encoding of “ABC”

| Port | Binary Representation |
|---|---|
| **50,000** | *1100 0011* 0101 0000 |
| **50,065** | *1100 0011* 1001 0001 |
| **50,131** | *1100 0011* 1101 0011 |
| **50,198** | *1100 0100* 0001 0110 |

As mentioned above, IANA recommends that the first two bits be set to one for ephemeral ports and, although the range is not a technical limit, traffic coming from any port not adhering to this rule may trigger a signature in an intrusion detection system or fail to pass through an internal firewall [11]. For that reason this encoding scheme should comply with IANA recommendations, giving the scheme a port range between 49,152 and 65,535. Once the upper limit of this range has been reached, the numbers can loop around to the bottom range picking up where they left off. This function is described in Equation 1 where $R_{min}$ and $R_{max}$ are the range limits, $V$ is the value to be transmitted, $P_1$ is the current port and $P_2$ is the next port to be used. For example, if the last port used was 65,500 ($P_1$) and the next value to be transmitted is 65 ($V$), it is clear that the port number is going to need to loop as the port 65,565 is beyond the upper limit. The difference of the 65,535 ($R_{max}$) and the last port used should be subtracted from the value to be transmitted. The sum of that difference and 49,152 ($R_{min}$) minus 1 is the next port to be used. In this case, the next port would be 49,181 ($P_2$).

$$P_2 = R_{min} + \big(V - (R_{max} - P_1)\big) - 1 \qquad (1)$$

This encoding scheme is somewhat inefficient. The problem is that no more than eight of the sixteen bits are ever being used at a time as the difference will never exceed 256. To increase efficiency while staying within the guidelines set forward by IANA, twelve or fourteen bits could be used on a rolling basis. Table 2 illustrates how a twelve bit implementation might encode the ASCII message “ABC”. Note that the first four bits, shown in italics, are ignored. The remaining bits are concatenated with the other port bits and interpreted as a single binary string. A twelve bit encoding mechanism such as this would enjoy 50% better throughput that the eight bit counterpart outlined earlier, and a 14 fourteen bit representation would have 75% better throughput.

Table 2: Basic 12-bit Encoding of “ABC”

| Port | Binary Representation |
|---|---|
| **49,156** | *1100* 0000 0000 0100 |
| **49,539** | *1100* 0001 1000 0011 |
| **52,320** | *1100* 1100 0110 0000 |

In some cases an implementation might not need to worry about the IANA port standard and would be free to use all sixteen bits. It may seem logical to simply divide the port bits in half and use the difference between them, but this method would forfeit the major benefit of delta encoding because the data would be completely contained in a single port number, making it easier for an analyst or security appliance to identify the channel and discover the data being transmitted. If a full sixteen bit scheme is selected, a better solution would be to use the delta between the first byte from two ports, followed by the delta between the first byte from the second port and the second byte of the first port. Finally, the delta between the second byte of the first port and the second byte of the second port is considered. At that point the pattern can be reversed and the cycle can continue. This is demonstrated in table 3.

Table 3: Simple 16-bit Encoding of “ABC”

| Port | Binary Representation |
|---|---|
| **131** | 0000 0000 1000 0011 |
| **19,654** | 0100 1100 1100 0110 |

## 4.2 Advanced Encoding Schemes

While functional, the above basic encoding methods can be problematic. The first major issue with these schemes, especially the eight bit scheme in particular, is that they are easy to identify. Second, they are all prone to data corruption. As discussed above, there is a risk that legitimate, unrelated communication between the sender and received could interfere with the channel by using source port numbers within the next delta range. There are 16,384 available ephemeral ports in the IANA suggested range, meaning the above eight bit implementation of the channel could be disrupted by an ephemeral selection of anywhere between 256 and 512 ports. This translates to a chance of data corruption between 1.56% and 3.13% for every unrelated source port number. While some practical implementations might be willing to call this acceptable loss in exchange for simplicity and bandwidth, there may be cases where a more robust approach is needed. In these cases improvements can be made to the encoding mechanism. One such improvement is defined below.

This more advanced encoding method uses the available bits left over from the data encoding scheme to help verify the contents of the next packet. In the previously discussed eight-bit scheme there remain eight bits in the sixteen bit port number which is further cut to six bits due to the IANA ephemeral port definition discussed above. The ones in the following bit string represent the bits in question: 0011 1111 0000 0000.

These bits will be used as a redundancy check (RC) to verify that the next source port received is, indeed, part of the message. To achieve this, an "exclusive or" (XOR) operation is run between the data bits of the current source port and the data bits of the previous source port. The resulting bit string is truncated to fit the available RC bits, depending on implementation. This method leaves the very first source port in the chain without data to XOR. To address this problem, both machines will share a key the same length as the RC bits. The RC bits in the first source port sent will be the result of an XOR between that key and the data bits to be transmitted. When these port numbers are considered in context, it is very easy to identify and ignore ports that are not a part of the message, greatly increasing data integrity. This process is illustrated in figure 1 and an 8 bit example is given in table 4. Note that the two leading IANA bits are in italics and ignored. The six bold bits for a given port are the result of an XOR operation between the data bits in that port and the data bits of the previous port or the initialization key.

Table 4: Advanced 8-bit Encoding of "ABC"

| Port | Binary Representation |
|---|---|
| **Key** | 10101010 |
| **49,643** | *11* **000001** 11101011 |
| **50,478** | *11* **000101** 00101110 |
| **57,201** | *11* **011111** 01110001 |

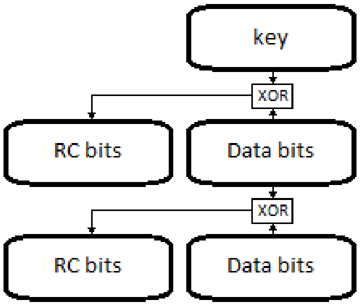


Figure 1: Advanced Encoding Process

This method allows only two bits worth of remaining offending ephemeral ports, or four ports total. This lowers the chance of data corruption to 0.02%. It should also be noted that this modified encoding scheme has the additional advantage of being more difficult for an analyst or security appliance to detect as it maintains the advantage of being present only in the delta while making the delta harder to discover and making the raw source ports jump around.

Once again, this encoding scheme stands functional but imperfect. If an implementation has no need to adhere to the IANA port standard, a much improved scheme can be developed. While maintaining eight data bits a full set of eight RC bits could be committed to the channel, leaving no chance of data corruption by unrelated traffic due to a perfect XOR. If the next legitimate port happens to be selected by an unrelated program, the real source port would be ignored due to incorrect RC bits, leaving no corruption. An example of this can be seen in table 5.

Table 5: Advanced 8-bit Encoding of "ABC"; not IANA Compliant

| Port | Binary Representation |
|---|---|
| **Key** | 10101010 |
| **16,875** | **01000001** 11101011 |
| **50,478** | **11000101** 00101110 |
| **24,433** | **01011111** 01110001 |

Even if IANA standards must be adhered to, an improved implementation is possible with an encoding implementation which uses seven data bits instead of eight. There would be seven RC bits remaining to ensure integrity, leaving no chance of data corruption by unrelated traffic for the same reason outlined above. An example of this can be seen in table 6.

Table 6: Advanced 7-bit Encoding of "ABC"

| Port | Binary Representation |
|---|---|
| **Key** | 10101010 |
| **61,429** | *11* **1011111** 1110101 |
| **57,163** | *11* **0111110** 1001011 |
| **49,870** | *11* **0000101** 1001110 |
| **61,200** | *11* **1011110** 0010000 |

# 5 Conclusion and Future Work

This work presented a new method for leveraging transport layer source ports as a covert channel. A number of implementation models were discussed, including an efficient and covert kernel modification. Additionally, a wide variety of encoding schemes were proposed and reviewed on their merit. These contributions open the door to a number of new methods that warrant further work.

One method that may be worth exploring for future work is a channel which duplicates regular network traffic while changing the source port in the duplicate packet. Instead of using live packets as the medium for the channel, a local listener on the sender could wait for outgoing traffic destined for the receiver. Once the traffic is identified, the sender will create duplicate packets of the legitimate traffic, changing the source port in each packet to encode the leaking data. In this instance, the encoded delta is between the legitimate packet and the modified packet, as opposed to the delta between two modified packets. This approach has some advantages. First, legitimate traffic will not have any effect on the channel; whatever ephemeral source port is selected, the modified duplicate packet will be able to use whatever port it needs for the encoding as it will not actually open the socket advertised locally. Similarly, there is no need to worry about TIME_WAIT status of the sockets because the socket is never actually opened. Finally, this approach will allow a much higher bandwidth in a TCP environment as it will not need to establish a connection for each delta. The primary disadvantage to this technique is that it dissolves the features that make source port delta channels appealing from the perspective of covertness. There would be a high amount of unusual traffic over the network, making it easy to tell that some sort of communication is going on. Further, the new packets are exact duplicated of legitimate traffic except for the source port, making it easy for an analyst to identify the source ports as suspicious and possibly leading to the discovery of the transmitted data.

Another promising method involves using destination ports in UDP as a way to transmit data. On many UDP protocols, when a server receives a connection from a client it replies back with a new port listed for this particular client to use. This method allows UDP protocols to keep track of different “connections” without the benefit of TCP connectivity facilities. However, there is no limit to this port switching technique and it may be feasible to leverage rapid port switching deltas as a covert channel. There would be some distinct advantages to this method, including that the channel would survive NAT and proxy interference. Further, since the role of the sender and receiver is swapped, this method shows promise as a medium for covert command and control. A disadvantage associated with this method is that it would be inherently lossy.

Detection of this channel has yet to be researched. One approach could be comparing the rapidly changing port numbers to ordinary network traffic patterns. It may be possible to identify or prevent this channel by noting source port selection outside a standard variance.

# 6 References

[1] B. W. Lampson, “A note on the confinement problem,” *Communications of the ACM*, vol. 16, no. 10, pp. 613–615, 1973.

[2] Johnson, D., Lutz, P. and Yuan, B., “Behavior-based covert channel in Cyberspace,” In: Vanhoof, K., et al (eds) *Intelligent Decision Making Systems*. World Scientific, New Jersey, pp. 311-318, 2009.

[3] S. Cabuk, C. E. Brodley and C. Shields, “IP covert timing channels: Design and detection”, in *Proceedings of the 11th ACM Conference on Computer and Communication Security*, Washington DC, USA, 2004.

[4] S. J. Murdoch and S. Lewis, “Embedding Covert Channels into TCP/IP”, in *Information Hiding Workshop Proceedings*, Berkeley CA, USA, 2005

[5] R. Bidou, F. Raynal, “Covert Channels,” (2009)

[6] J. Thyer, “Covert Channels Using IP Packet Headers”, presented at *DerbyCon*, 2011, Louisville, Kentucky.

[7] Stevens, W. R., B. Fenner, and A. M. Ruddof. *UNIX Network Programming: The sockets networking API*. Boston, MA: Pearson Education, 2004.

[8] Cotton, M., et. al. Internet Engineering Task Force , "Request for Comments: 6335 ." Last modified 2011. Accessed May 1, 2012. http://tools.ietf.org/html/rfc6335.

[9] Microsoft Corporation, "Important notice for users of Windows XP (SP3)." Last modified 2009 .http://support.microsoft.com/default.aspx?scid=kb;en-us;196271.

[10] Rfc 793: Transmission control protocol. (1981, September). Retrieved from http://tools.ietf.org/html/rfc793

[11] Firewall intrusion detection system signature enhancements. (n.d.). Retrieved from http://www.cisco.com/en/US/docs/ios/12_2t/12_2t15/feature/guide/ft_fwIDS.pdf

# A Physical Channel in a Digital World

Michael D. Deffenbaugh, Daryl Johnson, Bo Yuan, Peter Lutz
Rochester Institute of Technology
Rochester, NY
mddeff@zeroent.net,{daryl.johnson,bo.yuan,peter.lutz}@rit.edu

***Abstract* -- Gaming has always been a very prevalent part of our society. There is a long history of using various board games as methods of covert communication, from what we could call "differential" communication (i.e. the movement of pieces constitutes the "language") to the idea of "game state" communication (where the current state of all the manipulatable objects in the game world represents a message). For many years we have seen covert channels in games like chess, checkers and other simple board games and even evolution to more "complex" computer games like magnetron. Enter the 21st century and the advent of the high performance personal computer.**

*Keywords – Covert Channel; Computer Game; Garry's Mod; Valve; Gaming*

## I. Introduction

In the 80's, 90's and early 2000's we see amazing advances in computer games both in the aesthetic/graphical aspects of computer games but also in the types and complexities of game play, including the associated physics engines [1]. One of the biggest advances we see is in the client/server model where more of the interaction is handled by the server, but the rendering is done client side. The development of this technology happens around the same time when we start to see more and more powerful personal workstations (likely not by coincidence). The gaming companies utilize this and migrate to a client-render model and this opens up a litany of opportunities for covert channels.

As Lampson defined back in 1972, a covert channel is one that is not intended for communication [2]. Covert channels are nothing new however there hasn't been a lot of development into game-based channels. This is not the first occurrence of using video games as a method for covert channels. Sebastian Zander, Grenville Armitage, and Philip Branch authored a paper [3] in 2008 titled "Covert Channels in Multiplayer First Person Shooter Online Games". In this paper, they detailed the use of the Quake 3 engine as a means of implementing a covert channel. Their channel relied on making minute modifications to player's movements (imperceptible to the "unwitting players" but detectable by client side programs).

## II. Channel Overview

This covert channel utilizes a computer game called "Garry's Mod" (leveraging the Valve Source physics and graphics engine). Garry's mod is a "sandbox" game that allows users to spawn props and construct various inventions. An example is that a player can construct a car that they can then drive around the map. The Garry's Mod "paradigm" even allows users to create functioning electronics (using an add-on called "WireMod") that allows them to affect various in-game objects.

Covert channels span from physical channels such as leaving classified documents in a trash bag under a bridge to digital ones like embedding information in TCP checksums. This channel merges the two worlds. It is a digital covert channel in that it occurs in the digital world utilizing computer programs and network protocols however in the game it occurs as a "physical" channel (in that the player/sender creating the messages is manipulating physical game pieces).

This paper will discuss in detail the idea of using Garry's Mod in particular as a covert channel. We will provide an example language but it should be noted that this is one of many that can be used. Garry's Mod has a lot of different ways one could utilize various game mechanics to create a channel.

## III. Garry's Mod In-depth

Before we delve into the nitty-gritty of how Garry's Mod is to be used as a covert channel, we first must understand how it is to be used as a game.

Each player has the game client installed on a local system and when they start up the game they can browse for servers hosting various instances of the game on the internet. Both companies as well as private individuals host servers for various purposes/game types. A large portion of the servers setup for the Garry's Mod community are "build" servers, meaning they are just there for people to come and build various contraptions, in the true spirit of Garry's Mod. There is a large variety of other servers created for specific purposes such as Role Playing Game servers (where players take on various rolls such as shop keeper, and have inventions, such as a shop, to augment the "reality" created) or space-build servers (where players must build contraptions that will provide life support, propulsion, etc. in a simulated space environment). For the purposes of this discussion, we will focus on general build servers which are not only arguably the most prevalent type of server, but also the one that places the fewest restrictions on player actions.

Once a player enters a server they can start spawning props to construct whatever they want, whether it be some invention they thought of, or in our case, a covert message. They have the ability to spawn whatever prop or props they want and then are provided with a set of tools to either constrain props together (using methods such as "welding" two props together or binding them together with "rope") or attach forms of propulsion to them (adding motors and wheels, or adding vector-based thrusters). This leaves the player limited by almost nothing but their imagination.

While there is an entire ecosystem of modifications (mods) that can be added to Garry's Mod, the most prevalent is "WireMod" [4]. As described above, WireMod is a set of props that can have values or do simple (or complex) computations. The makers of WireMod have even developed a language called Expression2 [5]. A Perl/Gnu-C like language, Expression2 allows players to program some of the more complex chips to run their inventions.

An example of WireMod is where a player could build a car that, when it detects it's about to flip over, it fires a thruster to correct itself (via a programmed chip linked to a thruster on-board the car). This has even gone to the extent where we have people who have made their own operating system atop computers that interact in a client-server model with-in Garry's Mod (termed "ConOS" [6]) completely constructed out of WireMod components.

## IV. The Covert Channel

The proof-of-concept was developed to demonstrate that this game could in fact be used as a means for a covert channel. The "example language" that was developed utilizes the "color" tool within Garry's Mod to alter the RBGA(red, blue, green, alpha) values of a specified prop; in this case, a 55-gallon drum. The encoding method is accomplished by encoding a character into its ASCII decimal value (0-255) as one of the RGB values of the prop. As such each prop can hold 3 characters. So for example if you had one prop with RGB values of 87, 105, and 110 respectively, this would be translated to "Win".

Ordering is handled simply by the order in which they were spawned (i.e. the first prop spawned is the first "packet", second prop is the second packet, and so on). The advantage to using this ordering method is that it is something that the message author doesn't have to worry about, as they spawn containers, they are in proper order. The main disadvantage is that if one needs to change part of the message, the message effectively needs to be re-entered/re-encoded.

The final part is channel noise, determining what is in the channel and what is not. The alpha value is used to determine if a prop is in the channel or not. The alpha value of the prop (for game purposes) dictates how translucent the object is (an alpha value of 0 is completely transparent; an alpha value of 255 is completely opaque). Essentially you specify a "key" to be used for a specific alpha value, and any prop with said value is in one's channel; however, this method comes with some inherent downsides.

If a random player (not participating in the covert communication) happens to set one of their props to the specific alpha value used as the key for the covert channel, their prop (and its RGB values) will appear as in the channel and as such be decoded by the client and added to the message. This is less likely to happen with sufficiently random values for alpha as alpha values tend to either be exactly or close to 0/255. Choosing something relatively random (and not on a round number, i.e. not ending in 0 or 5) should eliminate most if not all of the noise from that source.

The other source of noise is a little more difficult to work around. Through our testing we noticed an interesting behavior in the Source Engine. It seems that when you have a gib (a piece of scrap from a larger prop that was destroyed) that is on fire (say from a wooden crate or flammable barrel), as the client turns to face away from said gib, the client side renderer has the gibs alpha fade from 255 to 0, thus causing the gib to temporarily "enter the channel". As the client for this language is constantly evaluating the current game server and all of the props inside and displaying its results live to the client's screen, a reader of the channel would notice some artifacts in the message. It should be noted that when the client turns to face the gib that is on fire again (to bring it back into view), the effect is reversed and it fades (quickly) from 0 to 255. The flame effects themselves can also produce random alpha values (thus sometimes falling into the channel).

In the demo video, available at http://file.zeroent.net/mddeff/pub/gmcc/, we see that the user takes 2:18 seconds to create the message "Covert Channel Example", encode it into ASCII decimal values, and then affecting the properties of the 55-gallon drums. That roughly translates to 0.159 Bps or 1.275 bps. Garry's mod allows a player to make a contraption off-line then effectively "copy and paste" it into the current server. It would allow us to create a message off line and then paste it into the server. With preloading the message, we were able to load and paste a 53 byte stream in 2 seconds, this works out to be 26.5 Bps. The downside to this is that there is more preparation required off-line.

### A. Advantages

As with every covert channel or covert channel, we must evaluate its ability to function effectively.

Noise in a covert channel is defined as the amount of data the falls into the parameters of the channel that was not intended to be part of the covert communication. In the case of the Garry's Mod covert channel, channel noise ends up being other players and other interactions in the game. However it all depends on what the language is defined as. For example, if your language is if a player spawns a certain prop at a certain time, that means a certain message; noise in that channel would be every time any player spawns any prop that wasn't intended to be part of the channel.

However with the proof-of-concept language discussed in this paper, the channel noise is any time a player modifies the default color properties (RBGA values) of a prop and happens to hit the specific alpha value, thus putting it in the channel.

Due to the sheer amount of Garry's Mod servers running at any time, it is not uncommon for a single player to be in a server "playing" by themselves. Normally it is the mark if an ineffective covert channel if it causes highly observable traffic, however in this case, even though the traffic is observable, it still appears normal.

Furthermore, the concern of having the channel be conspicuous to other non-covert channel users is to some extent a non-issue. Even if you have a player in the game who is doing nothing more than spawning barrels and changing the color (like the proof-of-concept language), other players or server administrators won't typically question what they are doing. They may not be viewed as terribly competent players, but no ulterior motives will be assumed. This allows us to operate almost any type of channel within Garry's Mod. As Harry Chriss (see acknowledgements) said, "Nothing is too weird for Garry's Mod."

While there are literally hundreds of Garry's mod Servers running at any time (see table 1), any average user has the ability to create a server at any time that other users can join. This allows the message sender to setup servers with specific settings that may aid in the particular language they have developed.

Table 1 - Survey of Garry's Mod Servers [7]

| Date/Time | Servers | Players | Slots |
|---|---|---|---|
| 5/22/2012 04:59 | 288 | 219 | 5283 |
| 5/23/2012 00:01 | 271 | 205 | 4978 |
| 5/24/2012 00:48 | 274 | 237 | 5068 |

Another property of this covert channel is that as long as the server stays up (and depending on the server configuration), any storage base channel can turn into a dead-drop style message. Some of the limitations on the dead-drop concept include the fact that many servers employ "prop cleanup" scripts where after a certain amount of time of a user disconnecting from the server, their props are automatically removed. That being said, if the message sender is running their own server, they can remove this limitation.

Depending on the language chosen, a Garry's Mod covert channel can either act as a timing based channel or a storage-based channel. For the purposes of this paper, we are utilizing the definitions as described by the Trusted Computer Security Evaluation Criteria (TCSEC). The proof-of-concept channel is a storage-based channel as it a form of communication by "modifying a stored object", in this case, a 55-gallon drum prop. An example of a timing based channel is where you have a player who sets off a series of explosions in the game to represent "S.O.S." in Morris code.

Observability is another advantage in using the Valve Source graphics/physics engine. The game client and server communicate in such a way that the client knows about every event that transpires on the server at any given instant. This means that if Player A spawns a barrel and colors it a certain color, all other players **in that server** know that just happened. Even though the actual user might not be able to see my barrel on their screen from across the map, each gaming client has the knowledge that it exists and in what state it exists. This is advantageous in that it doesn't require the receiver of the message to be in physical proximity (in the game) to the sender. This can even be more advantageous as when one does a "video capture", a client is essentially taking a recording of every event that is transpiring in the game (a player jumping, another one firing a gun, and yet another spawning a prop). When this "video" is played back, the client simply re-renders the set of instructions just as if it were getting them from the server.

## B. Limitations

Unfortunately there are some significant limitations to the practicality of this channel. The biggest limitation of the channel is the large client-side requirements (for those sending and receiving the message). Given that both the sender and receiver of the message must have the Garry's Mod client installed. The requirements of which include a relatively high end graphics card and a decent CPU, this limits the types of systems that can be communicating in the covert channel [8].

One of the other limitations is that as of now we have not developed a way to programmatically create or transmit a message. This means that in order to send a message, a user must manually log into the game and create the message. This currently is the biggest limitation in the bitrate of most of the channels created for the GMCCF, in that the speed of the channel is based largely on the speed of the player's ability to manually encode the message. That being said, the reading/decoding of the message has been automated and example LUA script can be found at http://file.zeroent.net/mddeff/pub/gmcc/init.lua.

Another limitation that we run into is that there must be a hosted dedicated server for us to log into and transmit the message on. And if one doesn't exist that suits the needs of the particular language, then one must be created. Another note to mention about server requirements is that there is a server-side variable set called "sv_pure". This variable determines if clients are allowed to run client side scripts (in this case, our reader lua script), if set to 0, we can run client side scripts, if set to 1, we can't. Almost all of the Garry's Mod servers have sv_pure set to 0 however it technically isn't guaranteed and as such must be acknowledged as a limitation of the channel.

The largest limitation that should be noted is one that is more about the practicality of using this channel. This channel has the potential for transmitting a lot of data at once, that being said, given the requirement of a custom gaming client using custom TCP ports, it requires that the transmitting party have the ability to either have access to a system with the Garry’s Mod client already installed (not very likely in a secured environment where information would need to be exfiltrated from) or the sending party would need to have the ability to install it to one of the local computers (would be rather non-covert).

Furthermore, the transmitting party would need to be able to actually have their client’s game traffic leave the network and hit the server being hosted on the internet. A lot of companies block common gaming ports inbound and outbound simply for loss of productivity reasons, and as such this presents another limitation. In summary, any covert channel leveraging Garry’s Mod requires that both the sender and the receiver have some modicum of control of the systems as well as the network in between their client computers and the server hosting the game instance.

## V. Additional language concepts

There are a few other ideas that are worth more research to determine the best method for communicating using Garry’s Mod.

One method that was tested was having a time sensitive storage-channel where the message was time sensitive in that it was only viewable for a certain amount of time before it self-destructed, using the explosive barrels as the prop. It was insinuated that if the reader was too close to the message when it self-destructed (exploded), it would be considered a “volatile” message, as it would likely kill the player.

A variation of the proof-of-concept language was where each of the props “z-axis” value, or “elevation” determined which prop or “packet” came in which order. As the Garry’s Mod physics engine allows the player to arbitrarily freeze a prop in any x, y, z, yaw, pitch, roll orientation, one could develop a language where the highest elevated prop is the first “packet” in the message, the second highest prop is the second packet, and so on.

Another covert channel could be stenography within the maps themselves. Some advantages include the fact that it does not require the actual existence of a server. The sender in this case would author a map that contains the message within the map, and then the reader would download it (either by joining a server running the map) or by downloading it out-of-band and then running it locally. The most prominent disadvantage includes the fact that it requires a different map to be created for each message that is sent. Map creation is no simple task and the time it would take to properly make a map and encode a message in it would be very lengthy, however could store upwards of gigabytes of data. It should be noted that computer game map stenography is not limited to Garry’s Mod and could be accomplished with any reasonably current video game.

## VI. Conclusion and Future work

The development of the Garry’s Mod Covert Channel (GMCC) shows us that new types of covert channels are emerging very rapidly. We can expect to see a lot more research done not only in the traditional gaming arena but in anywhere where there is a content-rich user experience. Possibilities for future work include the creation of an API to programmatically manipulate in-game objects as well as the development of other languages and performing a differential analysis as to which language is the most effective as a covert channel within the GMCC.

## VII. Acknowledgements

We would like to give special thanks to a personal friend, Harry Ethan Chriss. He was instrumental in implementing the proof-of-concept language discussed in this paper and was always there to bounce ideas off of.

## References

[1] Valve, "Valve Source Engine," [Online]. Available: http://source.valvesoftware.com/. [Accessed 24 5 2012].

[2] B. W. Lampson, "A Note on the Confinement Problem," *Communications of the ACM,* vol. 16, no. 10, pp. 613-615, October 1973.

[3] S. Zander, G. Armitage and P. Branch, "Covert Channels in Multiplayer First Person," in *Local Computer Networks*, Montreal, Canada, 2008.

[4] Wiremod Community, "WireMod Wiki," [Online]. Available: http://wiki.wiremod.com/wiki/Main_Page. [Accessed 24 05 2012].

[5] "Syranide", "Expression2 - WireMod Wiki," WireMod Community, 5 4 2012. [Online]. Available: http://wiki.wiremod.com/wiki/Expression_2. [Accessed 24 5 2012].

[6] "LooperNor", "ConOS Wire OS," Youtube, 5 7 2010. [Online]. Available: http://youtu.be/9ET8jqOfIYM. [Accessed 24 5 2012].

[7] M. Deffenbaugh, "Garry's Mod Server Stats," 24 5 2012. [Online]. Available: http://file.zeroent.net/mddeff/pub/gmcc/server_stats.xlsx. [Accessed 24 5 2012].

[8] Valve, "Garry's Mod," [Online]. Available: http://store.steampowered.com/app/4000/ . [Accessed 24 5 2012].

# An Advanced Data Loss Prevention System Being Able to Respond Data-Leaking Incidents Using e-Discovery Primitives

**Youngsoo Kim[1], Namje Park[2], and Sung Kyong Un[1]**
[1]Cyber Security-Convergence Research Laboratory, ETRI, Daejeon, Korea
[2]Department of Computer Education, Jeju National University, Jeju, Korea

**Abstract** - *Most enterprises or organizations have IDSs or firewalls to prevent or detect intrusions from the outside, and recently prepare DLP(Data Loss Prevention) systems to prevent insiders from leaking important data. However, these systems do not block all attacks. Diverse accidents of leaking privacy information requires that conventional DLP products should have additional functions being able to restore routine workflows or services provided immediately by finding the causes of data-leaking incidents besides the main function like preventing data leakage. In this paper, we propose an advanced DLP system providing this function using primitives of e-Discovery processes.*

**Keywords:** Data Loss Prevention, Data Leakage, e-Discovery, Forensic Readiness, Incident Response

## 1 Introduction

Most enterprises or organizations have IDSs or firewalls to prevent and detect intrusions from the outside, and recently prepare DLP(Data Loss Prevention) systems to prevent insiders from leaking important data. However, these systems do not block all attacks[1]. Diverse accidents of leaking privacy information causes them to require that conventional DLP products have additional functions being able to restore routine workflows or services provided immediately by finding the cause of data-leaking incidents as well as the main function of preventing data leakage. Their needs are caused by recent experiences that business workflows do not work, customer services are delayed and also recovery costs are increasing greatly, in case of data-leaking incidents. We use primitives of e-Discovery processes such as strategy establishment, collection, preservation, processing, review, or analysis to provide above function for conventional DLP system.

The discovery at Civil Procedures means that a litigant asks opponent parties to release related potential evidences and information. By requesting each other to open the opposing parties' evidences, documents, or witnesses, litigants can receive a fair trial under the same condition. Litigants should open all evidences they have by themselves prior to trial and can request the other party or the third party to make public theirs at the same time[2]. The purpose of this requesting procedure for opening evidences is to make clear a point at issue of suit and secure all evidences which might be hidden purposely on trial, and there are a lot of cases that compromise is achieved prior to trial because each party knows about the other party's evidences in detail. As ESI(Electronically Stored Information) is included in extent of evidence that become discovery's target in FRCP taken effect on December 1, 2006, terminology named e-Discovery was appeared[3].

Enterprises been always vexing in several litigations are hurrying to adopt systematic ESI administration and confrontation system to prevent a lawsuit from losing owing to failure in duty of presenting related evidences and to maintain their confidences[4]. To select only litigation-related data from huge amount of data which enterprises have, some procedures such as identification, preservation, collection, processing, review, analysis, production, or presentation are required. Since it is impossible to proceed above procedures by hand for ESI increasing exponentially, some tools and solutions being able to shorten litigation costs and time radically having various automated functions that is necessary in e-Discovery process are released[5].

In this paper, we propose an advanced DLP system being able to analyze incidents to find reasons and restore business workflows or services quickly, when data-leaking incidents occurs, using primitives of e-Discovery processes. We survey fundamental functions of conventional DLP system in Section 2 and describe processes of e-Discovery processes in section 3. Finally, the proposed system is introduced in Section 4.

This work was supported by the IT R&D program of MKE/KEIT[10035157, Development of Digital Forensic Technologies for Real-Time Analysis]. Corresponding Author: Namje Park (namjepark@jejunu.ac.kr)

## 2 Conventional DLP system

Figure 1 depicts functional blocks of conventional DLP system. It consists of integrated PC security block, physical device security block, network control block, and database security block.

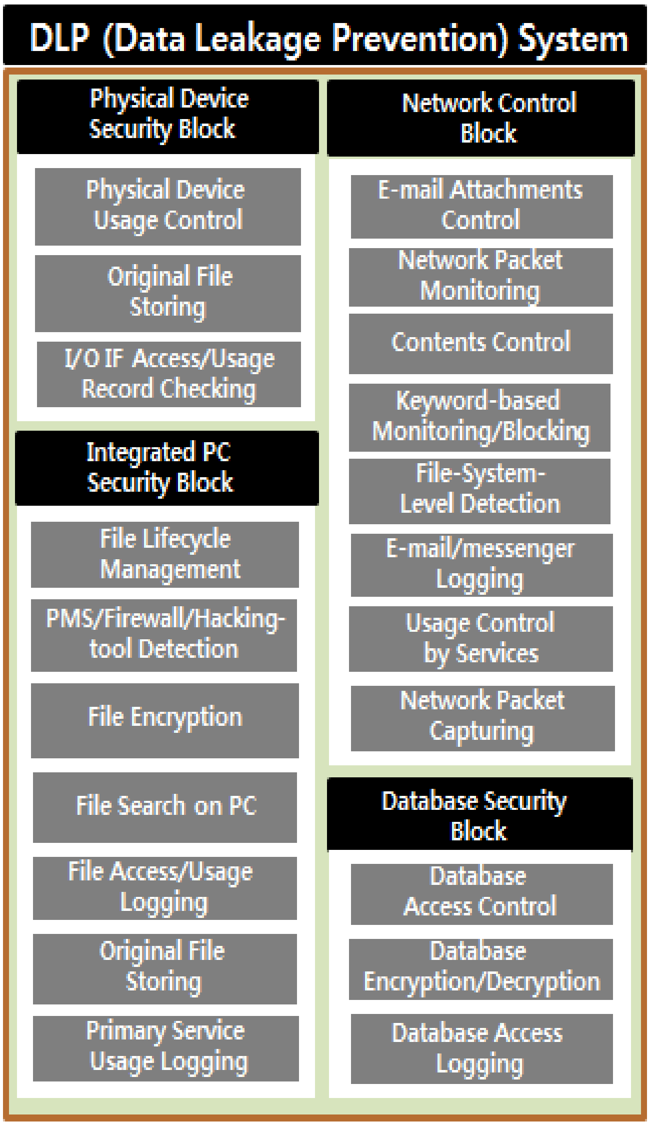

DLP (Data Leakage Prevention) System
Physical Device Security Block
Physical Device Usage Control
Original File Storing
I/O IF Access/Usage Record Checking
Integrated PC Security Block
File Lifecycle Management
PMS/Firewall/Hacking-tool Detection
File Encryption
File Search on PC
File Access/Usage Logging
Original File Storing
Primary Service Usage Logging
Network Control Block
E-mail Attachments Control
Network Packet Monitoring
Contents Control
Keyword-based Monitoring/Blocking
File-System-Level Detection
E-mail/messenger Logging
Usage Control by Services
Network Packet Capturing
Database Security Block
Database Access Control
Database Encryption/Decryption
Database Access Logging

Figure 1. DLP Blocks & Detailed Functions

Integrated PC security block plays a role of protecting data included at employees' personal computers or notebooks and comprises various functions such as file lifecycle management, PMS/Firewall/Hacking-tool detection, file search on PCs, file encryption, file access/usage logging, original file storing, primary service usage logging, etc.

- File lifecycle management: It manages all files created, modified, deleted, moved, or compromised.
- PMS/firewall/hacking-tool detection: It detects PMS(Patch Management System), private firewalls, or hacking tools for security reason.
- File search on PCs: It can search files could be required to monitor/block or important for security.
- File encryption: It prevents users having lower security-level from seeing contents of classified files.
- File access/usage logging: It stores and reports access/usage records of files in order to use for future investigation.
- Original file storing: It stores critical original files should not be modified or compromised safely.
- Primary service usage logging: It can help detecting users' behaviour for later use.

Physical devices security block which prevents information leakage from physical devices like printers includes an authority control function for using physical devices, a storing function of original printing copies, and a checking function of access/usage records for I/O interface.

- Authority control for using physical devices: It prevents users from printing and leaking out some information. Only authorized user can access and use these devices.
- Storing original printing copies: It stores hard copies of original ones being prominent.
- Checking access/usage records for I/O interface: It monitors access logs for I/O interfaces to detect unusual status.

Network control block which stops the outflow of information using networks comprises many functions such as e-mail attachments control, network packet monitoring, contents control, keyword-based monitoring/blocking, file system level detection, e-mail/messenger logging, usage control by services, or network packet capturing.

- E-mail attachments control: It looks into sending e-mails which attach classified files.
- Network packet monitoring: It monitors network traffics and decides if unusual symptoms are occurred or not.
- Keyword-based monitoring/blocking: It can input some specific keywords to finds related packets.
- E-mail/messenger logging: For future usage, it stores some logs of e-mails or network messengers.
- Usage control by services: It can control network services provided case by case.
- Network packet capturing: It captures important network packets for use of investigation.

Database security block which protects information stored at databases includes functions of database access control, database access logging, and database encryption/decryption.

# 3 E-Discovery Processes and Functions

E-Discovery related tools or solutions are designed and made referring EDRM. This reference model standardizes proceedings and defines each step's functional specification to effectively follow guidelines and recommendations described in FRCP[6]. EDRM offers general, scalable, and flexible frameworks to develop e-Discovery related products and services and evaluate them. This is used by general standard about e-Discovery that is authorized, because it was developed by co-works of various related organizations.

**Information Management** - It manages documents of enterprises or organizations with documents control and preservation policy. **Identification** - As step of deciding a scope of discovery-related documents, it prepares documents can be used in discovery potentially and decides documents should be collected and preserved. **Preservation** - It secures that documents do not change or destroy. **Collection** - It collects ESI from various media such as tapes, drives, portable storage, networks, etc. **Processing** - It filters duplicated or unrelated documents and changes format of ESI to be able to review them more effectively. **Analysis** - As step of making related summaries (Related subjects, persons, or documents) by analyzing ESI, it should be done to enhance productivity prior to detailed review step. **Review** - As step of establishing strategy on court, it evaluates collected ESI via relations and privileges and selects sensitive documents. **Production** - It stores ESI to various media and submits it to a court and opposing litigants. **Presentation** - It considers methods that can be seen effectively in trial.

We divide technologies for e-Discovery into 8 steps. These steps are assigned referencing EDRM and the Sedona Conference, yet another e-Discovery project[7]. The Sedona Conference is a non-partisan think tank on law and policy. A number of working groups consisting of lawyers, jurists, and consultants that work to provide best practices and commentary on e-Discovery issues address various issues of ESI and related topics. EDRM focuses on managing ESI and developing related solutions, while the Sedona Conference focuses on responses and workflows of enterprises or organizations in case of lawsuit.

**Information Management** – It is the step of preparing law suits and then is prior to occurrence of a civil suit. Some functions such as Policy Establishment & Adaptation, Policy Compliance Monitoring, Employees' Relationship Definition, ESI Category Definition, ESI Automatic Classification, and ESI Lifecycle Management are required. Document managers of an enterprise set-up various policies for managing information generated, modified or deleted. After setting-up policies, ESI could be managed through the function of ESI lifecycle management. Furthermore, a function of monitoring whether all employees keep them well is required, too. In this step, automatic classification techniques for documents are very useful for finding some evidences could be used at the court later. Therefore, functions of categorizing and classifying ESI automatically using those categories and some classifying algorithms are needed[8][9].

**Strategy Establishment** – When a law suit starts, litigant parties need this step, first. It requires some functions of Lawsuit Issues Examination, ESI Search, Early Case Assessment, E-Discovery Planning, Related ESI Identification, Litigation Hold Execution, and Data-map Creation & Management. At first, all the people concerned should understand issues of that suit, so they need a searching function of ESI and ECA-related functions. ECA (Early Case Assessment) refers to estimating risk (cost of time and money) to prosecute or defend a legal case[10]. ECA lifecycle will typically include the followings: A risk-benefit analysis, information preservation, gathering relevant information, process potentially relevant information for filtering, search term, or data analytics, reuse information in future case, etc. Based on ECA, litigant parties establish e-Discovery plan. They identify related ESI and execute Litigation-hold for preserving them. Creating and managing data-maps can help do above things better.

**Collection** – After establishing strategy, parties concerned should collect identified information. This step includes Collecting-method Choice, Identified ESI Backups, and Creating Copies (Imaging). They can collect only related data or disks including related data. Integrity should be considered to choose the way of collecting. Usually, parties concerned process or review a copied version of data, not original one, in order to prevent information being changed. The only copying way with integrity is imaging. The copying images from the original data or disks can be used. Furthermore, Identified ESI backups are also one of the prominent functions of this step.

**Preservation** – This step preserves candidate data which could be used as evidence information at the court and includes functions such as Policy Establishment & adaptation function and Litigation Hold Management. At first, parties concerned setup several policies for preservation like extension types, creation time, recent modification time, employees' name, IP address, MAC address, preservation starting time, preservation period, scope of preservation, preservation method, preservation type, etc. Additionally, parties concerned should monitor whether the Litigation Hold, started at strategy establishment step, is being executed well or not.

**Processing** – This is a step of processing data to review or analyze. It includes ESI Evaluation & Data Recovery, Data Format Transformation, Container-File Extraction, Metadata Acquisition, Similarity-based Hash Analysis, De-Duplication, Near-Duplication Analysis, Target ESI Indexing, and Condition-based Filtering. Each company stores ESI on several types of media. When the ESI is being created, received, or processed, or when it must be quickly and frequently accessed, it is stored at online storage like hard drives. Some ESI is stored at removable media such as DVDs, CDs, or flash drives. In this case, the files are available in a short period, such as a few minutes. Usually old ESI is stored at offline storage or backup tapes. Offline storage and archives is magnetic tapes or optical disks. It differs from online or removable storages in that the storage media are labelled, organized in shelves or racks, and accessed manually. Backup tapes, commonly using data compression, are sequential access media. The data is not organized for retrieving individual files. Retrieval typically requires restoring contents of the entire tape. In processing step, parties concerned evaluate and recover ESI from above types of media. This step requires also a function of transforming data format in order to review or analysis and extracting function for Container-files[11]. Additionally, it needs to acquire metadata showing file's information and filter well-known files like operating system files not being analysed using hash analysis. De-duplication and near-duplication are essential functions of this step. Through these functions, candidate data to review or analyse can be reduced prominently. Finally, target ESI indexing and condition-based filtering functions are also required in this step. Even though indexing takes long time to complete the index, it is very useful for searching a specific data. Filtering some ESI which mean nothing to reviewers can also reduce a respectable amount of reviewing data.

**Review and Analysis** – This is a step of extracting evidence data from processed one. Various functions are included such as Review Strategies & Planning, Review Format Transformation, Redaction, ESI Re-Search through Review Plan, Visualization of Integrated ESI, Privilege Log Creation, Tagging or Annotation, Grouping, Reviewing Result Reporting, Context-based Analysis, Relation Analysis between ESI and Suit, etc. After reviewing plans are made, ESI are re-searched through reviewing plans. Parties concerned use redaction, tagging, annotation, or grouping for providing convenience for reviewing and analysis. For high-level analysis, several functions such as visualization of integrated ESI, Context-based analysis, or relation analysis between ESI and suit are used. A function of privilege log analysis is also essential. Confidential conversations and communications that are protected by law from being used as evidence or revealed to others are referred to as privileged. Unless there's an exception, privileged ESI is not discoverable. Therefore, privileged ESI should be handled carefully using this function.

**Production** – This step transforms reviewed data to specific format files to present to the court. It needs several functions like Evidence-Producing Format Analysis, Specific File Format Production, Production Log Creation/Management, Load File Creation, Chain of Custody Log Creation, etc. After analysing evidence-producing format, parties concerned select a specific file format and create load files. Functions of loggings like production log or chain of custody log are also needed.

**Presentation and Destruction** – This is the Final step of e-Discovery. Functions like Evidence Visualization, Unprofessional Report/Diagram Creation, or Policy-based ESI Destruction are needed. To understand presented files in the court well, a visualization function is useful. Presentation report should be made unprofessionally and it is a better way to add several easy diagrams. After presenting, litigant parties destruct all ESI through some policies.

## 4 An Additional Response Module for Data-Leaking Incidents

Figure 2 depicts functions of responding module for data-leaking incident using component technologies of e-Discovery processes.

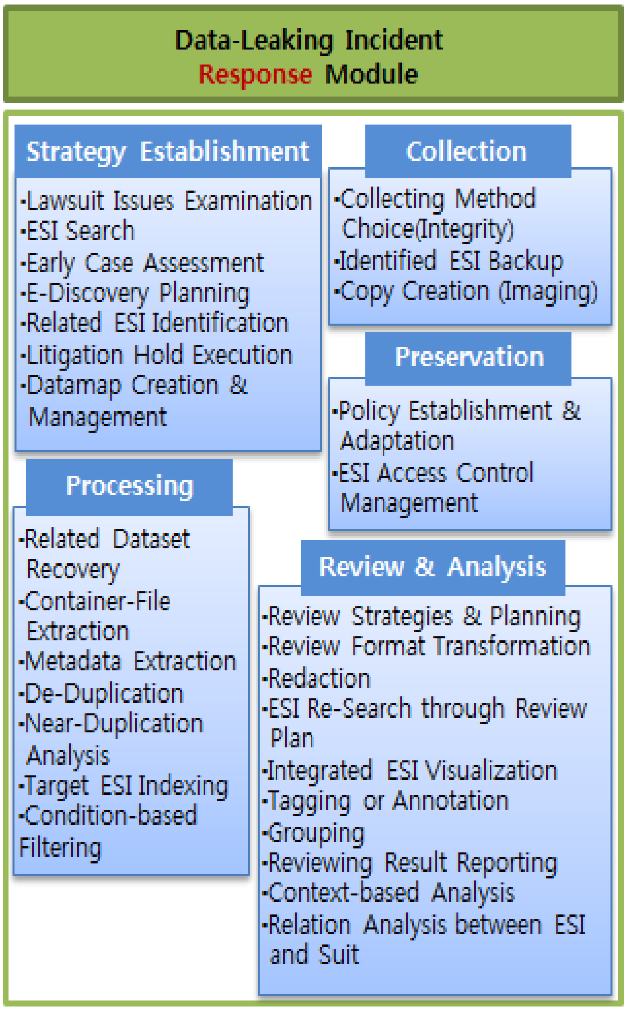

Data-Leaking Incident
Response Module
Strategy Establishment
·Lawsuit Issues Examination
·ESI Search
·Early Case Assessment
·E-Discovery Planning
·Related ESI Identification
·Litigation Hold Execution
·Datamap Creation & Management
Collection
·Collecting Method Choice(Integrity)
·Identified ESI Backup
·Copy Creation (Imaging)
Preservation
·Policy Establishment & Adaptation
·ESI Access Control Management
Processing
·Related Dataset Recovery
·Container-File Extraction
·Metadata Extraction
·De-Duplication
·Near-Duplication Analysis
·Target ESI Indexing
·Condition-based Filtering
Review & Analysis
·Review Strategies & Planning
·Review Format Transformation
·Redaction
·ESI Re-Search through Review Plan
·Integrated ESI Visualization
·Tagging or Annotation
·Grouping
·Reviewing Result Reporting
·Context-based Analysis
·Relation Analysis between ESI and Suit

Figure 2. Data-Leaking Incident Response Module

This module consists of 4 main functions such as strategy establishment, collection, preservation, processing, and review and analysis.

**Strategy Establishment** figures out issues of data leaking incident and makes some plans for solving problems and returning to normality. ESI search is essential function to need in this step, and also early assessment for incidents should be included to understand details and damages of incidents roughly. After figuring out incidents briefly, investigators establish some plans and countermeasures and identify related ESI. Since, related data should not be changed after the time of incidents for investigation, access-control for related ESI is also one of the prominent functions. Both systematic approaches and policy-based approach monitoring after announcing for every employees can be used. Finally, a function of creating and managing data-maps can be included to record essential details of data and find relationships among them.

**Collection** chooses a gathering way of data with data-integrity, backups identified ESI, and creates images. Preservation establishes preserving policies for ESI and manages access-control functions for related ESI, which are included at strategy establishment step.

**Processing** prepares data to find causes of data leaking incidents. Generally data is divided into active online data, near online data, offline data, data stored at backup tapes, deleted, fragmented, or compromised data, according to accessibility for data storage. Investigators decide what types of data to be prepared. It also extracts container files like attachment files of e-mails or compressed files and gets metadata of files. Furthermore, it requires functions of de-duplication getting rid of duplicated data to reduce amount of reviewing candidate files and near duplication removing previous version of documents and preserving only the final version of document. Target ESI indexing and condition-based filtering that chooses related ESI by condition are required at this step.

**Review/Analysis** step includes functions of real investigations for data-leaking incidents. It makes some strategies and plans for reviewing, transforms related ESI into review/analysis format, and sometimes requires a function of redaction. It has a function of ESI re-search through review plan and can visualize integrated results of analyzing ESI to find relationships among them. Furthermore, it gives some tags or annotations to ESI and makes some groups to analyze easily.

## 5 Conclusion

In this paper, an advanced DLP system is proposed. Besides it has conventional functions of preventing leakage of important data like personal information, privacy information, or customer information, it can find reasons of data-leaking incidents and restore routine workflows and services after it occurred.

If the duty of doing e-Discovery things becomes one of significant factors of compliances of enterprises or organizations in near future, this proposed system can be used as a prototype of developing automatic tools for e-Discovery.

## 6 References

[1] W.J.Blanke, "Data Loss Prevention using an Ephemeral Key," *HPCS 2011*, High Performance Computing and Simulation, 2011

[2] L.Volonino and I.Redpath, *E-Discovery for Dummies, First Edition*, Wiley Publishing, Inc, 2010.

[3] Y.Kim, S.Shin, and D.Hong, “ESI and e-Discovery,” *Weekly Technology Trends*. 2010.

[4] A.Cohen and E.Kalbaugh, *ESI Handbook: Sources, Technology and Process,* Aspen Publisher, 2010.

[5] Y.Kim, D.Hong, and S.Shin, “FRCP and e-Discovery,” *Weekly Technology Trends*, 2010.

[6] EDRM, Electronic Discovery Reference Model, http://edrm.net.

[7] The Sedona Conference, http://www.thesedona conference.org.

[8] Y.Yang and J.Peterson, “A Comparative Study on Feature Selection in Text Categorization,” *Proceedings of the 14th International Conference on Machine Learning*, 1997

[9] T.Joachims, “A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization,” *Proceedings of the 14th International Conference on Machine Learning*, 1997

[10] ECA, Early Case Assessment, http://en.wikipedia. org/wiki/Early_case_assessment.

[11] Container Format, http:// en.wikipedia.org/wiki/Con tainer_format_(digital).

# The 2011 IDN Homograph Attack Mitigation Survey

**P. Hannay**[1] **and G. Baatard**[1]
[1]ECUSRI, Edith Cowan University, Perth, WA, Australia

**Abstract -** *The advent of internationalized domain names (IDNs) has introduced a new threat, with the non-English character sets allowing for visual mimicry of domain names. Whilst this potential for this form of attack has been well recognized, many applications such as Internet browsers and e-mail clients have been slow to adopt successful mitigation strategies and countermeasures. This research examines those strategies and countermeasures, identifying areas of weakness that allow for homograph attacks. As well as examining the presentation of IDNs in e-mail clients and Internet browser URL bars, this year's study examines the presentation of IDNs in browser-based security certificates and requests for locational data access.*

**Keywords:** IDN, homograph, network security, internationalized domain names, computer security

## 1 Introduction

Internationalized Domain Names (IDN) allow for non-latin characters to be present in domain names. There are a number of security issues associated with this. Primarily this comes from the potential impersonation of domains by an attacker. This attack is achieved through the use of non-latin characters, which are visually indistinguishable from their latin counterparts. The aforementioned attack is known as an IDN homograph attack. This paper aims to investigate the strategies utilized by current web browsers to mitigate the impact of these attacks.

### 1.1 Instructions for authors

Domain names have been with us for a long time, first introduced in 1983 they provided a centralized means of abstraction for IP addresses (Mockapetris 1983, Mockapetris 1983). Since their inception domain names have become a key player in the information security arena, a known domain name inspires trust on behalf of the average user and as such is a high value item for would be attackers. Internationalized Domain Name Homograph attacks represent one attack vector that such an attacker could leverage for his/her advantage.

The initial implementation of domain names allowed only for alphanumeric characters and hyphens encoded as ASCII (Mockapetris 1983). In subsequent years it become apparent that this was an unacceptable limitation as audiences that make use of non-latin character sets were not able to have domains in their respective languages. In 1998 the initial work on Internationalized Domain Names (IDN) began. This work and subsequent work cumulated in 2003 with the publication of RFC3454, RFC3490, RFC3491 and RFC3492, a set of documents outlining the function and proposed implementation of IDN (Bell-ATL 2011).

The proposed IDN solution made use of UTF-8 character encoding to allow for non-latin characters to be displayed. In order to enable existing DNS infrastructure to handle UTF-8 domains a system known as Punycode was developed (Faltstrom, Hoffman et al. 2003). Punycode provides facility to represent IDNs as regular ASCII domain names, as such no changes are required for the majority of infrastructure (Costello 2003). An example of an IDN would be the domain name ☃.com, which would be represented as xn--n3h.com when converted to punycode.

### 1.2 Attacks

A number of visually indistinguishable glyhps (known as homoglyphs) exist within the Unicode character space. An example pair of glyphs are Unicode 0067 "Latin Small Letter G" and its counterpart Unicode 0261 "Latin Small Letter Script G" which are visually indistinguishable from one another. The aforementioned glyphs can be seen below in Figure 1.

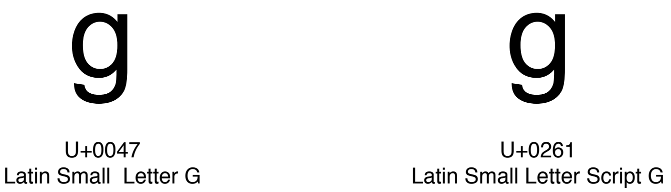


*Figure 1 – Example of Homoglyph for "g"*

Homoglyphs can be combined with characters from other scripts to form a series of glyphs, which as a whole are visually indistinguishable from their English counterpart. When a client or server interprets these homographs however they are treated in a distinct manner. Through the use of this trait attackers are able to craft domain names, which look familiar but are hostile in intent. These attacks can be deployed in the same manner as regular email phishing attacks, aiming to entice a user into accessing a hostile website in the belief that it is the genuine site being imitated. These attacks have been employed to steal financial data, passwords and corporate information.

Traditionally phishing attacks are mitigated through user education, encouraging users to check the legitimacy of links before clicking them, looking for unrelated URLs, not replying to emails asking for information if they are from an external domain name, etc. However when phishing campaigns are modified to make use of homograph domain names the ability for user education to provide mitigate is eliminated, as there is no way to make a visual identification of a fraudulent domain name. Figure 2 shows two domain names, both visually identical, however they lead to separate websites, with the one to the right making use of U+0261 rather than U+0047 for the second G.

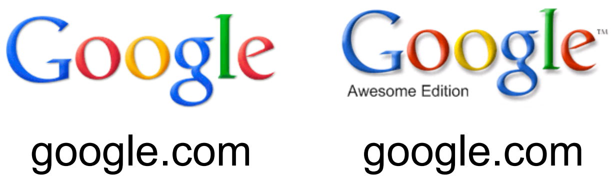


*Figure 2 – A pair of Homograph Domains*

### 1.3 Mitigation

A number of countermeasures have been implemented in order to mitigate the effectiveness of this attack. The majority of these involve displaying punycode in place of the actual UTF-8 text. Punycode is an ASCII representation of a Unicode domain name, originally implemented as the domain name service infrastructure did not support Unicode (Costello 2003). The punycode alternative is commonly displayed in both the address bar and the status bar on hover for a particular link.

When identifying domain names to display in punycode, there are two main methods used. The first (used by internet explorer 7 and above) is to use punycode only when a domain using mixed-script is detected (Fu, Deng et al. 2006). The implications of this are that any domain, which is intended to be spoofed via the replacement of one or more characters, will be detected, however in the event that the entire domain name is made from a single script it will be presented as intended by the attacker.

The other method employed by Mozilla Firefox and Safari both utilizes a whitelist in which all IDNs are presented as punycode unless they belong to a top level domain (TLD) that has policy in place preventing the spoofing of domain names in this manner. The policies employed via TLDs to prevent this attack often require that prior to registering a domain name containing homoglyphs, the registerer must own the domain name containing the western variant of those homoglyphs. In implementing this policy the IDN homograph attack is eliminated, however a number of TLDs have failed to implement this policy (Mozilla 2005).

A final strategy involves the color coding of various scripts in URLs (Krammer 2006). In this method Cryllic scripts are highlighted one color, while western scripts are left uncolored. In this situation mixed script URLs become immediately visible to the user, even though the characters themselves are visibly identical.

## 2 Testing strategy

For testing purposes we developed a virtual environment comprised of Windows 7 installation, which at the time of writing at all current updates applied. A snapshot was taken prior to the installation of web browsers or email clients.

Four primary attack vectors were identified with regards to IDN homograph attacks in web browsers, corresponding to the four most prominent locations in which an IDN may be shown to the user. If an IDN is shown in Unicode, a homograph attack could result in the user being tricked into believing that a URL is that of a legitimate website. The four attack vectors, in order of prevalence, are:

- The text shown in the browser's address bar, after the "Go" (or equivalent) button has been pressed.
- The text shown in the browser's status bar while the mouse is over a hyperlink.
- The text shown when viewing prominent information about a website's SSL certificate. As most users do not examine the details of a certificate, this attack vector relies upon the presentation of IDNs in immediately visible or accessible information.
- The text shown when the user is prompted to share their location using geolocation services.

In order to summarize the findings, an overall "Mitigation Rating" was calculated for each version of each browser tested. A value of zero is awarded if the browser does not support a particular attack vector, for example a lack of support for IDNs or geolocation services. A value of negative one is awarded if the browser supports an attack vector without mitigation against IDN homograph attacks. A value of positive one is awarded if the browser supports an attack vector and does mitigate against IDN homograph attacks, for example by presenting IDNs in Punycode. As the presentation of IDNs in the browser's address bar is by far the most prominent and influential vector of attack, values of positive and negative *two* are awarded for this vector. These values are shown in Figure 3 below.

| Address Bar | Status Bar | SSL Certificate | Location Request |
|---|---|---|---|
| -2 (Unmitigated) | -1 (Unmitigated) | -1 (Unmitigated) | -1 (Unmitigated) |
| 0 (No Support) | 0 (No Support) | 0 (No Support) | 0 (No Support) |
| +2 (Mitigated) | +1 (Mitigated) | +1 (Mitigated) | +1 (Mitigated) |

*Figure 3 – Mitigation Ratings*

By applying this metric, each browser version tested can be awarded a Mitigation Rating between positive five and negative five, representing a browser that supports and mitigates all attack vectors and a browser that supports but does not mitigate any of the attack vectors respectively.

Numerous versions of five web browsers were tested, based on averaged current market share data from a number of sources (Clicky 2011). Tested browsers were Internet Explorer (Microsoft), Firefox (Mozilla), Chrome (Google), Opera (Opera Software) and Safari (Apple). The authors attempted to test the initial release of each major version of the browsers since 2003, when RFC3454, RFC3490, RFC3491 and RFC3492 and ICANN's "Guidelines for the Implementation of Internationalized Domain Names" were published. All browsers were tested in a Windows 7 environment. The results of the testing are presented below.

# 3 Results

| Internet Explorer | | | | | |
|---|---|---|---|---|---|
| **Version & Release Date** | **Address Bar Mitigation** | **Status Bar Mitigation** | **SSL Certificate Mitigation** | **Location Request Mitigation** | **Mitigation Rating** |
| 7.0 (2006-10) | Punycode | Punycode | No Mitigation | No Support | +2 |
| 8.0 (2009-03) | Punycode | Punycode | No Mitigation | No Support | +2 |
| 9.0.8 (2011-03) | Punycode | Punycode | No Mitigation | No Mitigation | +1 |

*Figure 4 – Results for Internet Explorer*

Support for IDNs was added to Microsoft Internet Explorer in version 7, released in late 2006. IDNs in the address and status bars were shown in Punycode, and an icon providing further information about IDNs appears next to the address bar when one is used. Support for geolocation services was implemented in the latest major version of the browser, version 9, released in March of 2011. Internet Explorer currently offers no mitigation against IDNs in SSL certificate information or geolocation requests, showing them in Unicode. While Internet Explorer has protected itself against the most significant vector of IDN homograph attacks since support for IDNs was implemented, SSL certificate information and geolocation requests are presented without any mitigating features.

| Firefox | | | | | |
|---|---|---|---|---|---|
| **Version & Release Date** | **Address Bar Mitigation** | **Status Bar Mitigation** | **SSL Certificate Mitigation** | **Location Request Mitigation** | **Mitigation Rating** |
| 1.0 (2004-11) | None | None | None | No support | -4 |
| 1.5 (2005-11) | Punycode | Punycode | None | No support | +2 |
| 2.0 (2006-10) | Punycode | Punycode | None | No support | +2 |
| 3.0 (2008-06) | Punycode | Punycode | Punycode | No support | +4 |
| 3.5 (2009-06) | Punycode | Punycode | Punycode | Punycode | +5 |
| 3.6 (2010-01) | Punycode | Punycode | Punycode | Punycode | +5 |
| 4.0 (2011-03) | Punycode | Punycode | Punycode | Punycode | +5 |
| 5.0 (2011-06) | Punycode | Punycode | Punycode | Punycode | +5 |
| 6.0 (2011-08) | Punycode | Punycode | Punycode | Punycode | +5 |
| 7.0 (2011-09) | Punycode | Punycode | Punycode | Punycode | +5 |

*Figure 5 – Results for Firefox*

The first version of Mozilla Firefox was released in late 2004, and supported IDNs without any features to mitigate against homograph attacks. From version 1.5, released approximately a year later, IDNs in the address and status bars were shown in Punycode. From version 3.0, released in mid 2008, IDNs were shown in Punycode for SSL certificate information and were also placed more prominently in the interface. When support for geolocation services was implemented in version 3.5, mid 2009, requests were shown in Punycode. Firefox incorporated features that mitigate IDN homograph attacks fairly quickly, limiting its exposure in the two main vectors to a single major release.

| Google Chrome | | | | | |
|---|---|---|---|---|---|
| **Version & Release Date** | **Address Bar Mitigation** | **Status Bar Mitigation** | **SSL Certificate Mitigation** | **Location Request Mitigation** | **Mitigation Rating** |
| 1.0.154.59 (2009-04) | Punycode | Punycode | Punycode | No support | +4 |
| 2.0.172.27 (2009-05) | Punycode | Punycode | Punycode | Punycode | +5 |
| 3.0.197.11 (2009-08) | Punycode | Punycode | Punycode | Punycode | +5 |
| 4.0.302.3 (2010-01) | Punycode | Punycode | Punycode | Punycode | +5 |
| 5.0.396.0 (2010-05) | Punycode | Punycode | Punycode | Punycode | +5 |
| 6.0.495.0 (2010-08) | Punycode | Punycode | Punycode | Punycode | +5 |
| 7.0.544.0 (2010-10) | Punycode | Punycode | Punycode | Punycode | +5 |
| 8.0.552.224 (2010-12) | Punycode | Punycode | Punycode | Punycode | +5 |
| 9.0.597.16 (2011-02) | Punycode | Punycode | Punycode | Punycode | +5 |
| 10.0.648.205 (2011-03) | Punycode | Punycode | Punycode | Punycode | +5 |
| 11.0.696.77 (2011-04) | Punycode | Punycode | Punycode | Punycode | +5 |
| 12.0.742.112 (2011-06) | Punycode | Punycode | Punycode | Punycode | +5 |
| 13.0.782.218 (2011-08) | Punycode | Punycode | Punycode | Punycode | +5 |
| 14.0.835.202 (2011-09) | Punycode | Punycode | Punycode | Punycode | +5 |
| 15.0.874.21 (2011-09) | Punycode | Punycode | Punycode | Punycode | +5 |
| 16.0.904.0 (2011-10) | Punycode | Punycode | Punycode | Punycode | +5 |

*Figure 6 – Results for Google Chrome*

Despite only being released a few years ago, Google has released sixteen versions of the Chrome web browser. As the browser is in beta, the release cycle and version numbers are not as predictable as other browsers. All versions present IDNs in the address bar, status bar, SSL certificate information and geolocation requests in Punycode. Support for geolocation services was added in version 2 of the browser, released in mid 2009. All versions of Chrome have included defences against IDN homograph attacks, however the fact that it was first released much later than any of the other major browsers must be taken into account.

| Opera | | | | | |
|---|---|---|---|---|---|
| **Version & Release Date** | **Address Bar Mitigation** | **Status Bar Mitigation** | **SSL Certificate Mitigation** | **Location Request Mitigation** | **Mitigation Rating** |
| 7.00 (2003-01) | No support | No support | No support | No Support | 0 |
| 8.00 (2005-04) | No Mitigation | No Mitigation | No Mitigation | No Support | -4 |
| 9.00 (2006-06) | No Mitigation | No Mitigation | No Mitigation | No Support | -4 |
| 10.00 (2009-09) | No Mitigation | No Mitigation | No Mitigation | No Support | -4 |
| 11.00 (2010-12) | Punycode | Punycode | Punycode | Punycode | +5 |
| 11.51 (2011-08) | Punycode | Punycode | Punycode | Punycode | +5 |

*Figure 7 – Results for Opera*

Version 7 of the Opera web browser, released early in 2003, did not support IDNs. The next three major releases (in 2005, 2006 and 2009) supported IDNs but offered no mitigation to IDN homograph attacks. Support for geolocation services was added in version 11, late 2010, at which point IDNs in all attack vectors started to be shown in Punycode. While all vectors are not mitigated against IDN homograph attacks, the browser was vulnerable to the attacks for approximately five years.

| Safari | | | | | |
|---|---|---|---|---|---|
| **Version & Release Date** | **Address Bar Mitigation** | **Status Bar Mitigation** | **SSL Certificate Mitigation** | **Location Request Mitigation** | **Mitigation Rating** |
| 3.1 (2008-03) | No Mitigation | No Mitigation | No Mitigation | No Support | -4 |
| 3.2 (2008-11) | No Mitigation | No Mitigation | No Mitigation | No Support | -4 |
| 4.0 (2009-06) | No Mitigation | No Mitigation | No Mitigation | No Support | -4 |
| 5.0.1 (2010-07) | Punycode | Punycode | No Mitigation | Punycode | +3 |
| 5.1 (2011-07) | Punycode | Punycode | No Mitigation | Punycode | +3 |

*Figure 8 – Results for Safari*

Apple's Safari browser began showing IDNs in Punycode in the address bar, status bar and geolocation requests from version 5, released in mid 2010. Prior to that version, IDNs in the address and status bar were shown in Unicode and geolocation services were unsupported. It is worthwhile noting that the default settings for Safari hide the status bar, nullifying the mitigation possible when hovering over a hyperlink. No mitigation exists for SSL certificate information. The authors also noted that the first HTTP URL to be entered into the address bar upon launching the latest version of the browser was shown in Unicode. Successive URLs were presented in Punycode. Safari was vulnerable to IDN homograph attacks for a number of years, and remains vulnerable in small areas.

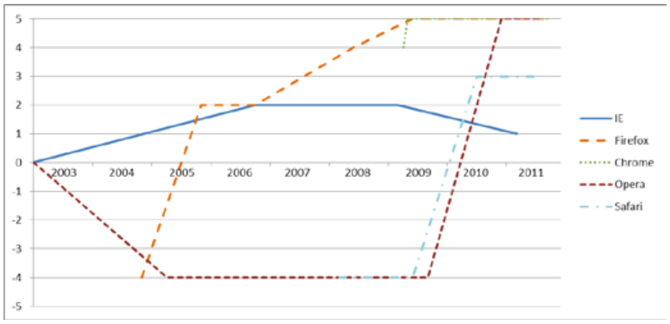


*Figure 9 – Results Summary*

# 4 Conclusion

The longitudinal data present from the nine years of software releases sampled provides interesting data. We can see that there is a strong trend towards effectively mitigating IDN homograph attacks in all products tested. However there still exists a need to ensure that location services and other potential areas of web browsers are secured in the same manner as the rest of the URL parsers & display mechanisms in the software. The lack of mitigation in some areas but not others in the same browsers suggests significant duplication of functionality in code, which is resulting in an increased attack surface. In order to better mitigate this issue it would be advantageous to consolidate these functions into single libraries which perform URL parsing, display and IDN homograph attack mitigation.

# 5 References

[1] Bell-ATL (2011). "Timeline of IDN." Retrieved 9th June, 2011, from http://www.bell-atl.net/articles/53357/Timeline-of-IDN.

[2] Clicky (2011, June). "Web browsers (Global marketshare)." Retrieved 8th June, 2011, from http://www.getclicky.com/marketshare/global/web-browsers/.

[3] Costello, A. (2003, March). "RFC3492 - Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)." Retrieved July 1st, 2011, from http://www.ietf.org/rfc/rfc3492.txt.

[4] Faltstrom, P., et al. (2003, March). "RFC3490 - Internationalizing Domain Names in Applications (IDNA)." Retrieved July 1st, 2011, from http://www.rfc-editor.org/rfc/rfc3490.txt.

[5] Fu, A. Y., et al. (2006). The methodology and an application to fight against unicode attacks, ACM.

[6] Krammer, V. (2006). Phishing defense against IDN address spoofing attacks, ACM.

[7] Mockapetris, P. (1983, November). "RFC882 - DOMAIN NAMES - CONCEPTS and FACILITIES."

Retrieved 9th June, 2011, from http://tools.ietf.org/html/rfc882.

[8] Mockapetris, P. (1983, November). "RFC883 - DOMAIN NAMES - IMPLEMENTATION and SPECIFICATION." Retrieved 9th June, 2011, from http://tools.ietf.org/html/rfc883.

[9] Mozilla (2005). "MFSA 2005-29: Internationalized Domain Name (IDN) homograph spoofing." from http://www.mozilla.org/security/announce/2005/mfsa2005-29.html.

# Covert Channel in the BitTorrent Tracker Protocol

Joseph Desimone, Daryl Johnson, Bo Yuan, Peter Lutz
B. Thomas Golisano College of Computing & Information Sciences
Rochester Institute of Technology, Rochester NY
{jwd1063, daryl.johnson, bo.yuan, peter.lutz}@rit.edu

***Abstract*— Covert channels have the unique quality of masking evidence that a communication has ever occurred between two parties. For spies and terrorist cells, this quality can be the difference between life and death. However, even the detection of communications in a botnet could be troublesome for its creators. To evade detection and prevent insights into the size and members of a botnet, covert channels can be used. A botnet should rely on covert channels built on ubiquitous protocols to blend in with legitimate traffic. In this paper, we propose a covert channel built on the BitTorrent peer-to-peer protocol. In a simple application, this covert channel can be used to discretely and covertly send messages between two parties. However, this covert channel can also be used to stealthily distribute commands or the location of a command and control server for use in a botnet.**

***Keywords: Computer Security; Covert Channels; BitTorrent; Botnets; Information Hiding***

## I. INTRODUCTION

Cryptography is useful in providing message confidentiality, or preventing a third party from uncovering the content of a message. However, cryptography is not designed to hide evidence that the communication has occurred. For some applications, even the existence of a communication between two parties could have disastrous consequences. These applications must rely on covert channels to prevent a third party from uncovering evidence that a communication has occurred. The term “covert channels” was first coined by Lampson nearly 30 years ago [1]. These original covert channels operated on a single machine to send information from a high security level process to that of a low security level process. Today, the majority of computing devices are attached to a network. Network protocols can be used to create covert channels were messages can be sent to remote machines in a stealthy manner [2]. In most cases, covert channels make use of unintended characteristics of network protocols that can be used to store information. The number of network protocols and their complexities allows for the creation of an almost unlimited number of covert channels.

Botnets are one possible application for covert channels. Botnets are groups of compromised machines that attackers control remotely for a variety of mostly malicious purposes. This includes conducting distributed denial of service attacks, sending spam messages, stealing account information, and conducting identity theft. Traditional botnets were controlled from an IRC server where the attackers could send commands to nodes of the botnet. However, this method is rarely used today as it presents a single point of failure for the botnet and they are more easily dismantled. As a result, malware writers have tried to devise better ways to control their army of compromised machines. Command and control over HTTP or HTTPS is common nowadays due to the prevalence of these protocols on the Internet. This makes identifying botnet traffic on the network more difficult. However, locating the command and control server is still a common failure point for botnets. For example, if they have static domain addresses programmed into the malware these can be null-routed and the botnet will go offline. This has led some malware writers to rely on DNS generation algorithms where DNS names are generated in a pseudo-random fashion. While these are more difficult to take down, it is not impossible. Another even more resilient method for botnet command and control relies on peer-to-peer network protocols [3]. This design completely decentralizes the botnet and makes it very difficult for defenders to track and dismantle. Malware authors will continue to explore resilient and stealthy methods to control botnets. Utilizing covert channels are another method that botnets use to elude detection from security researchers.

BitTorrent is one example of a peer-to-peer protocol. In recent years it has exploded in popularity and is responsible for the majority of the Internet traffic in most regions of the world [4]. The high volume and common use of the BitTorrent protocol would make an ideal target for malware writers that wish to blend in with legitimate traffic. Also, most BitTorrent networks are open and require no authentication. Therefore, a covert channel in the BitTorrent protocol would be an ideal candidate for covert channels, especially in the case of botnets.

## II. LITERATURE REVIEW

Much research has been done developing new types of covert channels [2]. Some of these covert channels rely on the TCP protocol to hide information. For example, Rowland proposed hiding messages in the Initial Sequence Number (ISN) field of a TCP SYN packet [5]. This sequence number is used to synchronize TCP packets in a communication and is normally randomly generated for the first packet of a TCP connection. However, a covert message can be inserted into the ISN field instead before being sent to the receiver. Other covert channels utilize the DNS protocol to send hidden information. In one such channel, information is sent over DNS lookup requests to a fake DNS server [2]. The message is encoded in the hostname field of the lookup request.

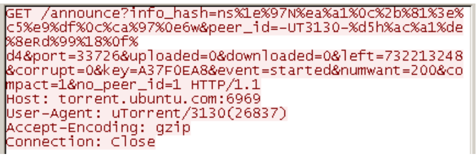
Research has also been done on categorizing types of covert channels and evaluating them based on common criteria. These criteria include the type, throughput, robustness, and probability of detection [6]. The most common types of covert channels are storage, timing, and behavioral based. The covert channel proposed in this paper can be considered a storage channel. Throughput measures the amount of information that can be sent over a given time interval. Covert channels can range from very low throughput rates of less than 10 bits per hour to high rates of megabytes per second. The robustness measures how resilient the covert channel as it proceeds through networking devices or other layers in the networking stack. Finally, detection measures how susceptible the covert channel is to detection from active listeners along the data path. Usually these criteria conflict with each other and a designer must select the most appropriate qualities for a given application. For example, generally as the throughput increases a covert channel's probability of being detected also increases.

The use of covert channels in botnet networks is not new. Johnson, Bo, and Lutz stated that malware authors might begin utilizing covert channels as a means to evade detection [7]. Also, Butler et al. proposed using a covert channel in DNS as a method for botnet command and control [8]. However, no published paper has documented the use of the BitTorrent network protocol directly. On the other hand, Li et al. proposed using torrent files to store covert messages [9]. Torrent files contain all of the necessary information needed to download a certain file, or collection of files, using the BitTorrent protocol.

## III. COVERT CHANNEL IN BITTORRENT

### A. Background

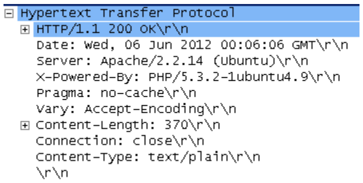
BitTorrent is a peer-to-peer protocol used for file sharing. Each user downloading a file also simultaneous uploads pieces that they have already received to other users. BitTorrent trackers are used so peers can locate other users that are downloading the same file. Users that are actively downloading files are known as leechers, while users that have completed downloading a file but remain uploading to other users are known as seeders. The BitTorrent tracker protocol operates over HTTP. Each torrent uses a unique SHA1 hash to identify the files or group of files that can be downloaded. After a user downloads and opens a torrent file, their BitTorrent Client will perform a GET request to the tracker. This GET request contains the info hash (unique SHA1) from the torrent file, the peer id of the client, an IP and port number of the client, an event message, a numwant field, among others [10]. At the start of the download the event message is "started." The peer id field is 20 bytes of randomly generated characters unique to each client. The numwant field is the number of peers the client wishes to receive. Figure 1 shows an example announce request to torrent.ubuntu.com. As you can see, the info_hash and peer_id fields are URL encoded.



```
GET /announce?info_hash=ns%1e%97N%ea%a1%0c%2b%81%3e%
c5%e9%df%0c%ca%97%0e6w&peer_id=-UT3130-%d5h%ac%a1%de
%8eRd%99%18%0f%
d4&port=33726&uploaded=0&downloaded=0&left=732213248
&corrupt=0&key=A37F0EA8&event=started&numwant=200&co
mpact=1&no_peer_id=1 HTTP/1.1
Host: torrent.ubuntu.com:6969
User-Agent: uTorrent/3130(26837)
Accept-Encoding: gzip
Connection: Close
```

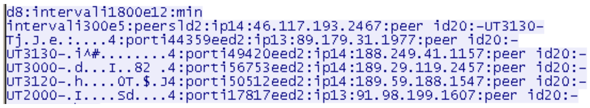
Figure 1. Example announce request

The tracker will respond to the client's request in a standard HTTP format. Figure 2 shows the packet header of the tracker response. Typically, trackers will return a maximum of 50 peers per request. Trackers will respond with a list of peers that are currently seeding or leeching the requested file. If more peers are present than the amount requested, peers are selected by the tracker randomly. This response can be in two formats. Some trackers allow the client to choose the response type, while others force a specific response. The first response format is known as a dictionary response. In a dictionary response, the server sends a list of IP, port, and peer ids of clients currently seeding or leeching the file. This response is structured in the Bencode format and contains the IP and port numbers in decimal notation. Figure 2 shows an example dictionary response from the BackTrack Linux tracker. The second response type is known as a binary or compact response. In a binary response, the server will respond with a list of IP and port numbers corresponding with other peers in network (big endian) notation. This response is typically the default as it requires less bandwidth and is also encoded with Bencode. Figure 3 shows a binary response from the BackTrack Linux tracker. As you can see from the figure, the response is not human readable. Peers are listed in the response with 4 bytes for their IP address and 2 bytes for their port number. No delimiter separates each IP/port combination in the list. The binary response omits the peer ID field entirely.



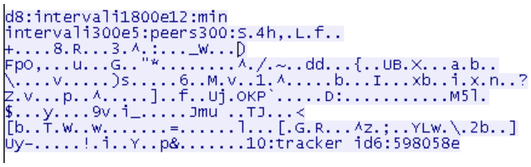
```
⊟ Hypertext Transfer Protocol
  ⊞ HTTP/1.1 200 OK\r\n
    Date: Wed, 06 Jun 2012 00:06:06 GMT\r\n
    Server: Apache/2.2.14 (Ubuntu)\r\n
    X-Powered-By: PHP/5.3.2-1ubuntu4.9\r\n
    Pragma: no-cache\r\n
    Vary: Accept-Encoding\r\n
  ⊞ Content-Length: 370\r\n
    Connection: close\r\n
    Content-Type: text/plain\r\n
    \r\n
```

Figure 2. Tracker response header



```
d8:intervali1800e12:min
intervali300e5:peersld2:ip14:46.117.193.2467:peer id20:-UT3130-
Tj.J.e.:....4:porti44359eed2:ip13:89.179.31.1977:peer id20:-
UT3130-.i^#........4:porti49420eed2:ip14:188.249.41.1157:peer id20:-
UT3000-.d...I..82 .4:porti56753eed2:ip14:189.29.119.2457:peer id20:-
UT3120-.h....OT.$.J4:porti50512eed2:ip14:189.59.188.1547:peer id20:-
UT2000-.I....Sd....4:porti17817eed2:ip13:91.98.199.1607:peer id20:-
```

Figure 3. Tracker dictionary response



```
d8:intervali1800e12:min
intervali300e5:peers300:S.4h,.L.f..
+....8.R...3.^.:..._W...D
FpO,...u...G.."*........^./.~..dd...{..UB.X...a.b..
\....v.....)s.....6..M.v..1.^.....b...I...xb..i.x.n..?
Z.v...p..^.....]..f..Uj.OKP`.....D:...........M5l.
$...y....9v.i_.....Jmu ..TJ...<
[b..T.W..w........=......l...[.G.R...^z.;..YLw.\.2b..]
Uy-.....!.i..Y..p&.......10:tracker id6:598058e
```

Figure 4. Tracker binary response

### B. Proposed Covert Channel

To covertly send messages using BitTorrent trackers, a client could hide information in the peer id field during an announce request. To receive the message, one could contact the same tracker with the same info hash used as the sender and perform another announce request. The target info hash and tracker must be established prior to sending/receiving the message. The receiver's request would need to specify a dictionary type response. The server will respond with peers downloading the file which will include the peer ID of the sender containing the covert message. Figure 2 shows an example topology for this situation. The sender connects to the tracker's web interface and sends an announce request with a peer id containing a covert message. This message is then stored in the tracker's database. To retrieve this message, the receiver performs an announce request to the tracker in the same fashion as legitimate P2P clients. The server replies to the receiver with a list of clients active in the specified torrent. This reply will contain the covert message. In this covert channel, 20 bytes of information can be sent at a time. However, for a more legitimate looking covert channel, this could be reduced to 12 bytes. Most BitTorrent clients reserve a portion of the peer id field for an identifier of the client name and version number. For example, the uTorrent client begins its peer id with the string "-UT3130-" where 3130 corresponds with version 3.1.3 [11]. The remaining 12 bytes are randomly chosen characters.

Unfortunately, many trackers do not support the dictionary response format and therefore the peer id field will not be visible to the receiver. An alternative covert channel could utilize the IP field to send messages. The BitTorrent protocol allows clients to specify IPs other than that which the tracker sees in the connection. This feature allows clients to connect to the tracker through proxy servers or from the same side of a NAT device. An IPv4 address is 4 bytes which allows 4 bytes of information to be stored in each announce request. For messages longer than 4 bytes, the message can be split across multiple requests to the tracker. For each request, the peer id field must be different. Otherwise, previous bytes will be overwritten in the tracker's database. The port field could be used as a sequence number for each message so the receiver can properly reorder the message chunks upon receipt. In order for the message receiver to differentiate between valid peer IP addresses or pieces of a covert message an XOR scheme was used to encode messages. The receiver could simply reverse the encoding mechanism and verify if the resulting message was composed purely of ASCII characters. However, this requires the original message only contains ASCII characters.

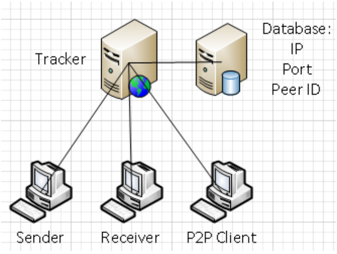


Figure 5. BitTorrent tracker topology

In an implementation of the covert channel using the IP field, we achieved a throughput rate of 20 bytes/second. A higher rate was obtained with the peer id field due to it allowing 5 times as many bytes than the IP field. Also, the peer id field covert channel has a higher degree of covertness. Hiding encrypted messages in the peer ID field would be impossible to distinguish from normal peer IDs due to their random nature. However, hiding messages in the IP field could be detected as these would resolve to IPs that are not actively participating in the download. Fortunately, most torrents will contain peers that cannot be contacted due to firewalls or peers that have closed their download client. We collected data across several popular ubuntu torrents to confirm this theory. Out of 1654 peers reported by the tracker, only 1517 could be contacted (91.7%). Additionally, some trackers are known to mix random IPs into their tracker responses to provide plausible deniability in file sharing lawsuits [12].

## IV. Discussion

We created proof of concept code to test the validity of the proposed covert channels. In both cases, we were successful in sending messages over the proposed covert channels. The backtrack-linux.org tracker supports the dictionary response and was used to send messages over the peer id field. To send messages over the IP and port field, the tracker from etree.org was used. An interesting issue occurred if more than 50 peers were downloading the target file. It could not be guaranteed that the receiver would see the sender's message with a single request. As a result, multiple requests were made to the server until the message was received. The selection of which torrent to rendezvous at is important for this reason. If the torrent has too many seeders and leechers then the message receiver is required to perform many requests to locate the message. However, if the torrent has too few seeders and leechers then plausible deniability is decreased. For most applications of this covert channel, I would expect a low number of total peers (<30) to be ideal. Also, torrents of copyrighted material would be less than ideal due to the higher possibility of monitoring (From the MPAA or other organizations).

The tracker will provide a min update interval to the client as part of an announce request. This is the minimum amount of time a client should wait before performing another announce request. Typically, trackers will remove clients that have not announced for twice the min update interval. As a result, the messages need to be reposted within this time frame to ensure they remain on the tracker. Typical min update intervals are 1 hour.

Both of these covert channels could be used in botnets. The BitTorrent trackers could be used as a rendezvous point for botnet clients and their controller. As an alternative to a domain name used to locate the IP of a command and control server, the IP address could send in the covert channel. This method would be more resilient to takedown than a domain name. One possible weakness that could be exploited to enumerate all nodes of the botnet would be to send the IP address of a monitoring server to the target tracker. This would be possible to anyone who knows the rendezvous locations and algorithm. To prevent insights into the size of the botnet and to prevent a possible hijack, public key cryptography could be used to sign the IP address. Unfortunately, utilizing 1024 bit RSA key would require almost as many bits for message signing and would be much too large for the covert channel. Elliptical curve cryptography (ECC) has the advantage of having an equivalent level of security with much smaller key sizes. For example, ECC with a 160 bit key offers the same level of security as RSA with 1024 bits [13]. Starnberger, Kruegel, and Kirda proposed using ECC in their botnet protocol with a 112 bit key [13]. This allows for 40 bit messages to be securely sent if a maximum cipher text length of 20 bytes is desired. Thus, a botnet controller could encode the command and control server using a single message in the peer id covert channel.

## V. Future Work

The BitTorrent specification should be analyzed for other potential covert channels. Also, mitigation or detection procedures should be developed for the proposed covert channels in this paper. For example, trackers should disallow the dictionary model response. Also, trackers should not allow peers to announce from arbitrary IPs. The info hash field should also be investigated as the potential for a covert channel. Many trackers will begin tracking any info hash that is announced to it. Therefore, arbitrary information could be inserted in this field to store up to 20 bytes of a message.

## VI. Conclusion

BitTorrent is one of the most commonly used protocols on the internet. Covert channels exist in the BitTorrent protocol that can be used to send messages in a stealthy and resilient manner. These covert channels could be used by a botnet to distribute commands or send the location of a command and control server.

## References

[1] B. W. Lampson, “A note on the confinement problem,” *Communications of the A.C.M.*,vol. 16, no. 10, pp. 613–615, Oct. 1973.

[2] S. Zander, G. Armitage, and P. Branch, “A survey of covert channels and countermeasures in computer network protocols,” *IEEE Communications Surveys & Tutorials*, vol. 9, no. 3, pp. 44-57, Oct. 2007.

[3] P. Wang, S. Sparks, and C. Zou, “An advanced hybrid peer-to-peer botnet,” in *First Workshop on Hot Topics in Understanding Botnets*, Cambridge, MA, April 2007.

[4] H. Schulze and K. Mochalski, “Internet Study 2008/2009,” 2009. [Online]. Available: http://www.ipoque.com/resources/internet-studies/

[5] C. H. Rowland, “Covert Channels in the TCP/IP Protocol Suite,” Technical Report, *First Monday: Peer Reviewed Journal on the Internet*, Jul. 1997.

[6] D. Johnson, B. Yuan, P. Lutz, and E. Brown, “Covert channels in the HTTP network protocol: Channel characterization and detecting man-in-the-middle attacks,” 2010. [Online]. Available: https://ritdml.rit.edu/ha ndle/1850/14797.

[7] D. Johnson, B. Yuan, and P. Lutz, “Behavior-based covert channel in cyberspace,” 2009. [Online]. Available: https://ritdml.rit.edu/handle/18 50/14795.

[8] P. Butler, K. Xu, and D. Yao, “Quantitatively analyzing stealthy communication channels,” in *Applied Cryptography and Network Security*, vol. 6715, J. Lopez and G. Tsudik, Eds. Springer Berlin / Heidelberg, 2011, pp. 238–254.

[9] Z. Li, X. Sun, B. Wang, and X. Wang, “A steganography scheme in P2P network,” in *IIHMSP ’08 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Harbin, China, 2008, pp. 20 –24.

[10] B. Cohen, “The BitTorrent Protocol Specification,” 2008. [Online]. Available: http://www.bittorrent.org/beps/bep_0003.html.

[11] “Bittorrent Protocol Specification v1.0,” 2011. [Online]. Available: http://wiki.theory.org/BitTorrentSpecification.

[12] “Perfect Deniability,” 2007. [Online]. Available: http://opentracker.blog .h3q.com/2007/02/12/perfect-deniability/.

[13] G. Starnberger, C. Kruegel, and E. Kirda, “Overbot: a botnet protocol based on Kademlia,” in *Proceedings of the 4th International Conference on Security and Privacy in Communication Networks*, New York, NY, USA, 2008, pp. 13:1–13:9.

# Audio Steganography Using High Frequency Noise Introduction

David Wheeler, Daryl Johnson, Bo Yuan, Peter Lutz
B. Thomas Golisano College of Computing & Information Sciences
Rochester Institute of Technology, Rochester NY
{dbw3113,daryl.johnson,bo.yuan,peter.lutz}@rit.edu

*Abstract*—**This paper presents a new method of audio steganography that allows character data to be encoded into audio in a way that is indiscernible to prying third-parties. Unlike typical methods of audio steganography that propose storage by modifying the least significant bits or the phase of the audio data, this approach makes use of frequency ranges that are undetectable to the human ear. The method proposed in this paper provides for a reasonably high-bandwidth and is resistant to common detection and prevention techniques.**

## I. Introduction

The act of being able to hide imperceptible information within digital media has become an area of increasing interest in the computing world. Data hiding techniques have many potential applications such as covert communication, hiding executable data, watermarking and digital rights management [2]. The method used to conceal data for all of the above situations is called steganography. Steganography, which literally means ”concealed writing”, is a method of covert communication that has existed for thousands of years. Today steganography has been adapted to the digital era and can be implemented in pictures, audio, text and even other forms of digital multimedia as well [3].

Digital steganography involves the use of two entities that make up the transfer file. The first entity is the cover object, which is the overt data being sent, and the second entity is the stego object, which is the secret or covert message embedded in the cover object. In addition, digital steganography has two rudimentary requirements that must be fulfilled. The first requirement is that the stego object is virtually imperceptible to any third-parties who may obtain the file or files, whereas the second requirement is that there must be a reasonably high bandwidth for the stego-data.

In this internet era, digital media is commonly transferred over the internet both through individual file transfer and streaming of raw data. Given these new developments there has been an increasing focus on embedding hidden information into audio. There are several classic ways that are used to hide information in audio. Least Significant bit (LSB) encoding, echo hiding, phase coding, and spread spectrum coding are among the most common techniques used [3]. This paper proposes a novel concept of audio steganography that encodes binary information into high-frequency signals. The proposed steganography implementation is a sort of hybrid mix of several of the more common methods, and combines a known technique with a new one. The maximum range that a human ear can hear is between the frequencies of 20 Hz and 20KHz. However, natural high-frequency hearing loss over time and the lack of speaker fidelity makes this perceptible frequency much lower [9]. This means that while audio files carry digital information all the way up to 20 KHz, some of the highest frequencies will be completely imperceptible to the human ear. The method illustrated in this paper will be both imperceptible to humans and reasonably resistant to preventative software techniques.

## II. Related Work

As mentioned above there are four main categories of audio steganography that are commonly used to hide information into auditory data: least significant bit coding, echo coding, phase coding and spread spectrum coding. Each method varies in implementation, bandwidth, and covertness. They all have advantages and disadvantages and are typically used for differing applications.

### *A. Least Significant Bit Encoding*

As the name implies least significant bit coding (LSB encoding) deals with modifying the least significant bit of each audio frame in order to encode binary information. This is an inherently simple task and has the advantage of high bandwidth but is unfortunately easy to prevent. Small format changes that occur during file conversion, compression, or through preventative techniques, can easily contaminate the hidden data [3]. There have been proposed LSB coding methods, however, that utilize higher level bits in same fashion as the LSB method that are robust against some issues that are present in this category of audio steganography [2].

### *B. Echo Hiding*

The process of echo hiding involves inserting echoes with varying characteristics into discrete audio signals. Three echo parameters, amplitude, decay rate and offset (essentially the delay time of the echo) are applied in varying ways in order to successfully encode binary information. This method is very covert as each echo occurs below the audible limit of the human ear. The disadvantage of this technique is that the process can sometimes yield a noticeable mix of echoes which increases the risk for detection [3].

### C. Phase Coding

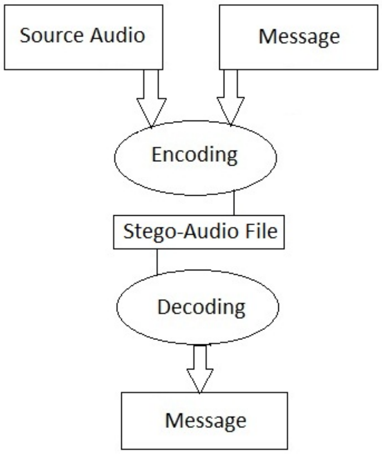
Phase Coding addresses the issue of covertness and detectability as components of sound (modified phase) are much more difficult for the human ear to perceive than the addition or subtraction of noise. To implement this method the audio file is broken down into discrete chunks, separated into phase groups and then shifted according to the binary data being encoded. The main problem with this process is that modifications of audio phase allows for a relatively low quantity of stego-data. Thus this technique is relatively low bandwidth and is typically used for applications such as watermarking or copyrighting of audio files [3].

### D. Spread Spectrum Coding

The final type of audio steganography worth mentioning is Spectrum Coding. Spectrum Coding takes bits of information and randomly spreads them over the entire frequency spectrum [5]. It is similar to LSB but is more robust against steganalysis techniques. This procedure, however, is still somewhat vulnerable to detection as it can introduce noise into the audio file [3].

## III. Method

Unlike some of the common steganography techniques listed above that try to mask the audio signals this approach uses more of a hidden-in-plain-sight kind of approach. While all wave files hold information between the ranges of 20 Hz and 20 KHz, not only are frequencies at the uppermost range rarely used, but they are also nearly impossible to perceive by the human ear. Given these characteristics of wave files, high frequencies can be injected into the cover audio file in order to produce a concealed binary stego signal. The basic approach to the process of embedding and reading binary information is presented below.



The method of encoding binary information in this manner is relatively simple. First, an audio file must be broken into discrete chunks known as frames. The character data to be hidden in the audio is then converted into binary information and mixed into the audio frames. The pseudo code algorithm used for mixing binary into audio is given below:

```
frame_buffer = frames.get(bufferSize)
for each binary_bit
if binary_bit == 1
   then addHighPowerHFT into frame_buffer
if binary_bit == 0r
   then noHFTInjection into frame_buffer
```

Essentially a group of frames is obtained from the audio file and then a HFT (high-frequency tone) is added to that buffer of frames depending on whether the next bit to be encoded is a 0 or a 1. One important point to note is that the size of the buffer must stay constant throughout the encoding process for the purposes of decoding. The buffer size can range anywhere from fifty to thousands of frames, which equates to between one and a hundred milliseconds. The disadvantage to increasing the buffer size is that the audio data will have a lower potential bandwidth.
The process used to encode audio is given below:

```
frames = Read_WaveFile(fileName)
characters = Read_Character_Data()
binary_bits = Convert(Characters)
foreach bit in binary_bits
   add_HFT(bit)
output_file = Write_WaveFile
```

In Fig1 a short one hundred and sixty millisecond clip of the cover audio file is shown. The graph measures the power levels at various frequencies over a span of time (160 MS). In Fig2 the same audio clip is shown once it has undergone the proposed HFT encoding. In Fig2 the levels of the HFTs would appear to be significant enough to be detected by human perception but are not, due to the fact they are located just past the peak of a humans auditory perception.

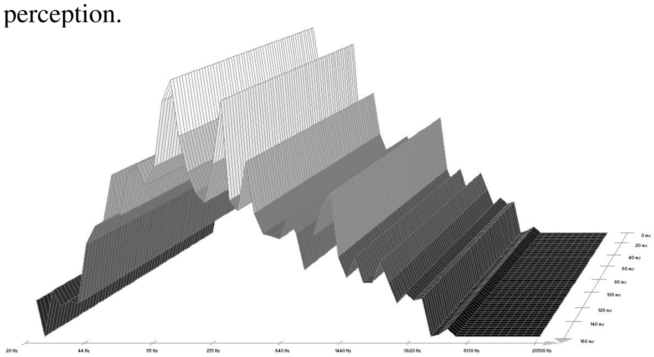


**Fig1: 160 ms of cover audio**

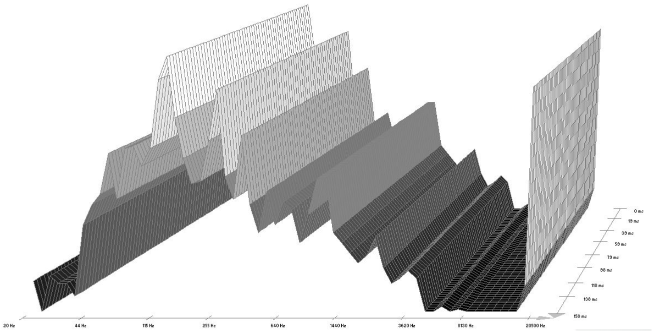


**Fig2: 160 ms of audio with encoded HFTs**

Decoding the stego object from the encoded wave file simply requires reversing the encoding process and applying a Fourier transform to the encoded frames. The decoding algorithm is given below:

```
frames = Read_WaveFile(stegoFileName)
fftData = FFT(frames)
data = HighPass(fftData
foreach buffer in data
   bits += get_Bit(buffer)
characters = Convert(bits)
```

In these steps the wave file with the hidden stego object is initially read into frames. In order to get the frequency breakdown of the frames a fast Fourier transform algorithm is applied to the data. A high-pass filter is applied to the frequency data which will only allow the HFT stego data to be read into the data array. Depending on the buffer size the data is read and each buffer is interpreted either as a 0 or 1 value. The binary information is then converted back into its character information and the hidden message is then fully decoded.

## IV. Experimental results

The proposed method described above was implemented in a java program called AudioStego which allowed for encoding of character data into wav-forFmat audio files. The program made use of a WavFile java class for basic input and output operations of wave files [1] and was otherwise implemented with standard java libraries. A multitude of different songs and audio information were tested. Each audio clip featured different dynamic and spectral ranges. All audio was sampled at 44.1 KHz with 16 bits of depth and the length of various audio files ranged from ten seconds up to five minutes in length.

Subjective tests were done on each of the tested audio files by several individuals in order to determine how discernable the introduction of HFTs were in the cover audio. The tests compared an original cover audio file with an audio-stego version that implemented varying buffer sizes and power levels for the encoded HFTs. Table 1 illustrates the different characteristics of HFTs used and their corresponding inaudibility as reported by the listeners.

TABLE I
Subjective Listening Results

| Buffer Size | Power Level | inaudibility |
|---|---|---|
| 100 | High | Sometimes |
| 1000 | High | Sometimes |
| 100 | Mid | Yes |
| 1000 | Mid | Yes |
| 100 | Low | Yes |
| 1000 | Low | Yes |

In all low power and mid-power cases it was reported that the additional HFTs in the audio signal were indiscernible. In both high-power cases some subtle white noise and crackling was discernable in the audio-stego file to some of the listeners. When inserting very low-power HFTs, data could be occasionally be decoded incorrectly due to noise in the audio signal. It also appeared that size of frame buffers played no noticeable role in detection. Given this data, the best apparent characteristics for indiscernible HFT injection would occur at the mid-power level and with buffers of 100 frames (this maximizes the bandwidth of the file).

## V. Detection and Prevention

There is a somewhat limited amount of research in the field of audio steganalysis (the analysis and detection of steganography). This has to do with the fact that many audio steganography schemes are quite advanced and the nature of high-bandwidth audio streams makes it difficult to produce consistent analyzing tools [4].

There are two main categories to talk about when referring to steganalysis. The first is detection and the second is prevention. Detection is particularly difficult in this case as HFT insertion is essentially in a category of its own for audio steganography methods. Prevention is much easier as any attack against the stego file that offsets the HFTs length, frequency, or power level could possibly disrupt the hidden data.

Along the lines of steganalysis, there are several major detection methods used that are worth mentioning. The first is the use of statistical distance measurements for steganalysis. This idea essentially measures the difference between cover audio and their stego-audio signal equivalents with the use of common techniques such as echo and phase coding. The statistical data that is generated is then compared against the stego-audio in question [7]. Another similar technique called audio steganalysis based on Hausdorff Distance finds the Hausdorff distance measure between a cover audio signal and its stego-audio equivalent. It decomposes the audio into wavelet coefficients and the Hausdorff distances between the wavelets are used to train a classifier about the difference between cover audio and stego-audio signals [3]. This type of steganalysis would prove to be ineffective against HFT addition since the algorithms would be gathering statistics from unrelated steganography algorithms.

While HFT addition is seemingly resistant to detective methods, it is much less robust against preventative methods. Since it is computationally difficult to analyze a high-bandwidth audio stream, recent research has been conducted to attack steganography by slightly distorting audio data. Some of the techniques that are being implemented are frequency shifting, white noise addition, and variable time delay [6]. All three of these techniques can potentially put the HFT method at risk. The main issue with attacking a stego-audio signal is that the attacker wishes to maintain the integrity of the audio as much as possible. This works out in favor of HFT as smaller changes to power levels, sample time adjustments, and frequency domains are less likely to distort the encoded binary information. In the case of variable time delay a common delay is no more than 10 milliseconds per second of audio data [6]. That level of change is unlikely to distort the stego-audio enough to distort the encoding. These preventative techniques pose a potential threat to the HFT method but do not completely invalidate it in many cases.

## VI. Conclusion

There are a number a number of proven methods for applying steganography to hide information within audio data. In this paper a new and simple approach was investigated that made use of rarely used and difficult to detect frequency ranges in raw audio files. It was shown through implementation and subjective experimentation that this novel method can effectively transmit binary information, by way of frequency injection, unnoticed to an end user. While the proposed method suffers from several drawbacks as far as robustness, however, it also serves the user with a channel for high bandwidth data transmission.

## VII. Future Work

There are several areas in which this proposed method could potentially be expanded on in the future. One particular addition that could be made to obfuscate binary data would be to apply an encryption algorithm on top of the binary data before encoding it in the audio file. A potential technique that could be applied would be to use a common encryption scheme such as AES and to apply it across the data before encoding and after decoding [8]. This would place a level of security on top of the hidden information so that any third-party that discovers the encoded bits would be unable to discern any kind of meaning from them.

Currently the proof of concept program, AudioStego, only encodes binary data to the raw wave file format. Due to the unwieldlieness of uncompressed audio data most auditory information is transmitted in compressed formats. Formats such as MPEG Audio Layer III(mp3), Windows Media Audio formats(wma) and Vorbis(ogg) are prime examples. As with many lossy formats MPEG Audio Layer III ignores potentially repetetive or unimportant frames and frequency bands [10]. Encoding data using HFTs would be harderto implement since the signal would have to be prominent enough so that it is not lost when the raw file undergoes lossy compression. There would either be a higher level of detectability or a higher error rate when decoding bits. To further complicate the issue of lossy-file conversion, formats such as MP3 significantly reduce signals present at very high frequency ranges. A quick experiment was conducted where a stego-audio file was converted into its lossy MP3 equivalent, converted back to a wav format and then was decoded. The HFTs were diminished to the point where the binary data was unable to be read. In order to avoid this conversion issue the AudioStego program would need to directly modify the converted format in order to inject HFT's in a manner in which they could be decoded.

## References

[1] Dr. Andrew Greensted, http://www.labbookpages.co.uk/audio/wavFiles.html

[2] Cvejic, N., and T. Seppanen.*Increasing Robustness of LSB Audio Steganography Using a Novel Embedding Method.* In Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference On, 2:533 –537 Vol.2, 2004.

[3] Ishaque, M. Qudus Khan, F. Abdul Sattar, S. *Investigation of Steganalysis Algorithms for Multiple Cover Media.* Ubiquitous Computing and Communication Journal. Vol 6, no 5, October 2011.

[4] C. Kraetzer and J. Dittmann, *Pros and Cons of Mel-cepstrum based Audio Steganalysis using SVM Classification.* Lecture Notes in Computer Science, vol. 4567, pp. 359 – 377, January 2008.

[5] H. Matsuoka, *Spread Spectrum Audio Steganography Using Sub-band Phase Shifting.* in Intelligent Information Hiding and Multimedia Signal Processing, 2006. IIH-MSP 06. International Conference on, 2006, pp. 3 – 6.

[6] M. Nutzinger, *Real-time Attacks on Audio Steganograhy.* Journal of Information Hiding and Multimedia Signal Processing. Vol. 3, no. 1, January 2012.

[7] H. Ozer, I. Avcibas, B. Sankur and N. D. Memon, *Steganalysis of Audio based on Audio Quality Metrics.* Proceedings of the Conference on Security, Steganography and Watermarking of Multimedia, Contents V, vol. 5020, SPIE, pp. 55 – 66, January 2003

[8] Sridevi, R. Damodaram, A. Narasimham, S. *Efficient Method of Audio Steganography by Modified LSB Algorithm and Strong Encryption Key with Enhanced Security.*Journal of Theoretical and Applied Information Technology. vol. 5, no. 6, pp. 768 – 771, June 2009.

[9] hypertextbook.com/facts/2003/ChrisDAmbrose.shtml

[10] MP3 Audio Format: http://wiki.hydrogenaudio.org/index.php?title=MP3

# User reaction towards End User License Agreements on Android Smartphones

**H. Cotton**[1] **and C. Bolan**[1,2]
[1]School of Computer and Security Science, Edith Cowan University, Perth, Western Australia
[2]secau – Security Research Centre, Perth, Western Australia

***Abstract*** *– Smartphones are increasingly recognized as the most popular computing platform, forming an integral part of the way users interact with the online world. Accompanied with the advent of user-installed content, End User License Agreements have surfaced mirroring issues previously arising on more traditional platforms. This survey conducted in Perth, Western Australia looked at user behavior when viewing and accepting EULAs on smartphone devices. The results show that a majority of users do not read such agreements citing issues of readability and length.*

**Keywords:** Information Security, Cellular Phones, Software Protection

## 1 Introduction

The evolution of computing from the more traditional personal computer to the rapidly establishing mobile plaforms has been accompanied by an increase in the range and type of applications [1][2]. Directly linked to such growth is the multiplication of vulnerability and the related increase in opportunity for exploitation [3]. Studies have found that as with other environments, that in the Smartphone platform vulnerabilities often play upon the lack of security knowledge of the user [4].

The primary sources of information about the actions taken by an application are the EULA, and in the permissions an application requests during installation [5][6]. Such importance is only magnified by the view that EULA's form a legally binding electronic contract between the vendor and the end-user [7]. It would appear that without a detailed and significant investigation into their applicability on the new platform that the implementation of EULAs has become an unclear issue in both the cyber and legislative domains [8]. From a cyber security perspective, the assent of a user to such a contract implies approval for the actions that a particular software application may take while providing mitigation to the vendors against any user led legal challenges [3][8].

Within the legal domain, several studies have sought to gauge the readership of these legally binding contracts [7][9]. However, such works often focus on unrepresentative sample populations and focus on platforms other than the smartphone area. Thus their findings whilst broadly applicable may have little bearing when viewed in the context of mobile platforms.

## 2 EULAs & Android

End User License Agreements (EULA) have been utilized previously on the desktop and online mediums to facilitate a supposed legally binding contract between the parties of the user and manufacturer [10].

In the context of the Android based Smartphones, a EULA is typically required when a user attempts to install a third party application. Upon presentation to the user the EULA asks the user to agree to the terms and conditions within by an action such as checking a box or a single button press. Yet, often the EULAs contain significant amounts of written text containing a high amount of legal terms, which are often difficult for the general user to understand [11].

Despite the seeming ubiquity of the EULA, its legal status remains somewhat uncertain. Such agreements (referred to as Clickwrap) would typically be seen to follow established contract law principles [12]. This is not universally accepted however as there are many disagreements as to their status under law [10].

If EULAs are to be considered a traditional contract then according to Australian legal precedence, deceptive conduct on behalf of either party is unacceptable [13]. Such precedent that exist seem to demonstrate that EULAs which contain deceptive language or do not allow for informed consent are unenforceable, however, when an EULA sets out the conduct of an application and assent is required the contract becomes binding. Such precedent means that EULAs may be used to accept behavior that would otherwise seem to be malicious [14].

However, central to any acceptance is the issue of informed consent. Satisfying the informed consent component requires the opportunity to read the contract, which in an Android environment is provided at the installation of the application. Some legal scholars assert that this reliance on notice and informed assent is outdated and somewhat insufficient to protect the user [8]. The question remains as to how such consent is established on the mobile platform and to what effect the length, readability and time based factors effect such consent.

Although a high degree of anecdotal evidence suggests users do not read EULA's only a limited number of studies have attempted to quantify these assumptions[8][9][11]. Of these studies few if any have focused on such agreements on Smartphone platforms.

# 3 The Study

## 3.1 Demographics & Setup

In an attempt to quantify these issues on a Smartphone platform a study was devised which incorporated the installation of an application with an accompanying EULA on an android smartphone. The study focused on participants aged over 18 with 107 participants representing 57% between 18-30 years old, 27% 31-45 years old and 15% between the ages of 46-65 with 15% of participants choosing not to provide this information. The gender mix represented 53% male and 35 % female with 12% undisclosed.

Each participant was supplied with a standardized device and asked to install a specific application upon the phone. To ensure that no contamination occurred participants were not given forewarning as to the nature of the research.

## 3.2 Time spent considering agreement

The EULA used in this study consisted of a total of 2406 words, available to the reader by scrolling through 13 screens. Of the 103 Participants that progressed toward the EULA stage, 5 of the 6 participants that scrolled through more than the opening page had an average readership of 147.8 words (SD=3.62). When the intention of the 1 participant who intended to read the agreement in its entirety the mean increases to 498.83 words (SD=238.98).

Individual participant reading speed has not been assessed due to the quantitative nature of this study, however; it is clear that of the participants that scrolled past the first screen, 5 participants merely "skim read" the agreement. This result was reflected anecdotally by participants during the debriefing exercise with a number of participants expressing the view they "did sometimes "skim read" the EULA".

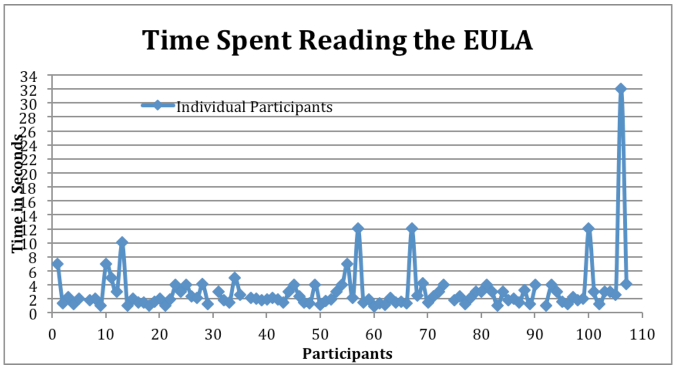


*FIGURE 1 – Time Spent Reading EULA*

After the conduct of the experiment the users were then issued with a survey to verify the results and provide further insight into the findings. The survey found that a small number of participants (14%) expressed the view "*I read EULAs when installing apps on my Smartphone*", which concurs with 11% of participants describing themselves as "readers" in the Bartlett and Plaut study [7]. Although the professed readership is relatively consistent the experimental results clearly demonstrated that participants did not read EULAs in practice. As only 5.8% of participants (n=102) attempted to read more than the opening screen of the EULA.

The reasons for this low level of readership may be explained in part through the survey responses relating to the complexity, enforceability, and readability of EULAs. Overwhelmingly participants reflected the belief that EULAs are too long and time consuming with 75.56% of participants in agreement. When participants were surveyed on their agreement to the statement EULAs were incomprehensible and hard to read, 55.13% agreed. This shows that although most participants felt they are too long and time consuming a reasonable percentage (17.75%) did not express concern over the complexity, rather made a conscious choice not to read. Although this research does not attempt to understand the individual heuristic factors behind non readership, it does show the willingness of participants to actively avoid reading EULAs regardless of the perceived readability of the document.

The "sameness" of EULAs was also examined during the survey with mixed results. A small majority of participants (42.98%) agreed with the statement that EULAs all say the same thing. However of note is the high number of participants taking a neutral position (23.36%) or answering "don't know" (14.95%). This may suggest that a high degree of confusion among participants of the content of this form of legal contract.

Overall, the study found that only 5% of participants took the effort to scroll past the opening screen. This was a lower figure than suggested in the survey responses where 14% of participants indicated they read EULAs (14%). The combined results seem to confirm the view of previous works which suggest EULAs are an ineffective mechanism of disclosure due to non readership.

# 4 Conclusions

This research has presented a practical and quantitative approach to assessing the readership of EULAs among Android Smartphone users. The results illustrate that, in such an environment very few users attempt to read the EULA, and, those that do spent a very short time “skim reading” the EULA.

The description of behavior expressed by participants during the survey concurred with the experimental results of comparable studies. During the experimental process a significant disparity was found to exist between the expressed views in the survey and the demonstrated behavior in the experiment. Although a number of participants expressed the view they read EULAs they then went on to spend less than 3 seconds on the screen.

The research demonstrates the degree of difference between traditional computing and the Smartphone domain raises new questions privacy and security. Further questions regarding the appropriateness of traditional EULAs and the effectiveness of permissions have been raised. Regardless of the opportunity to read and legal stature of the EULAs, users are left uninformed, and vulnerable to information attack. Some authors have put forward the idea of simplifying contracts to allow for greater readership and understanding. Moving forward all solutions must be explored as the issue of EULA is unlikely to abate in an ever more litigious society.

A study encompassing the various contexts in which applications are installed might garner a better understanding of normal user behavior. This may be accomplished by generation of application software which monitored user behavior “in the background” without the users knowledge. The major issues with such an approach would be the ethical considerations of installing such software on user’s devices without their prior knowledge.

# 5 References

[1] Androlib (2011). "App Stats - Statistics." Retrieved 27/02/2012, 2012, from http://www.androlib.com/appstats.aspx.

[2] Podonik, v., & Jezic, G. (2011). Mobile network fundamentals and evolution. Modern Communications: Linking people everywhere, Faculty of Electrical Engineering. University of Zargreb. .

[3] Oberheide, J., & Jahanian, F. (2010). When mobile is harder than fixed (and vice versa): demystifying security challenges in mobile environments. Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications,, Annapolis, Maryland.

[4] Coursen, S. (2007). "The future of mobile malware." Network Security 8: 7-11.

[5] Android Developer (2011). "Security and Permissions Developer Guide." Retrieved 03/05/2012, 2012, from http://developer.android.com/guide/topics/security/security.html.

[6] Good, N. S., et al. (2007). Noticing notice: a large-scale experiment on the timing of software license agreements. SIGCHI conference on Human factors in computing systems, San Jose, California, USA.

[7] Bartlett, R. and V. Plaut (2009). Blind Consent? A Social Psychological Investigation of Non-Readership of Click-Through Agreements. Law and Economics Workshop, UC Berkeley.

[8] Ben-Shahar, O. and C. E. Schneider (2010). The Failure of Mandated Disclosure, University of Michigan Law & Econ, Empirical Legal Studies Center

[9] Marotta-Wurgler, F. (2010). "Will Increased Disclosure Help: Evaluating the Recommendations of the ALI’s ‘Principles of the Law of Software Contracts." University of Chicago Law Review.

[10] Clarke, L. (2010). "Performance Risk, Form Contracts and UCITA." Michigan Telecommunication. & Technology Review 7(1).

[11] Clapperton, D. M. and S. G. Corones (2007). "Unfair terms in 'clickwrap' and other electronic contracts." Australian Buisness Law Review 35: 152-180.

[12] Carter, J., et al. (2007). Contract law in Australia. Australia, Butterworths.

[13] Gindin, S. E. (2009). "Nobody Reads Your Privacy Policy or Online Contract? Lessons Learned and Questions Raised by the FTC's Action Against Sears." New Journal of Technology and Intellectual Property 8(1).

[14] Desautels, E. (2005). "Software license agreements: Ignore at your own risk. US-CERT." Retrieved 13/4/2012, 2012, from https://www.uscert.gov/reading_room/EULA.pdf.

# UPnP Port Manipulation as a Covert Channel

Steven Monette, Daryl Johnson, Peter Lutz, and Bo Yuan
B. Thomas Golisano College of Computing & Information Sciences
Rochester Institute of Technology, Rochester, NY, United States

**Abstract**— *Port knocking traditionally has been a technique used from external connections to convey information to or request services from an internal private network [1]. UPnP as a standard allows for devices and services to open ports on network devices in order to enable functionality [2]. By combining these two techniques it is possible to port knock internally, opening ports for an intended viewer on an external network device. This paper proposes a covert channel using this technique to exfiltrate data or broadcast messages from a system behind a UPnP device to any Internet connected system.*

## 1. Introduction

Port knocking has primarily been a technique used to send a signal to a machine that exists within a private network or behind a firewall [1]. This has been used for many different applications such as requesting services or signaling internal machines. With the advent of technologies such as UPnP there is now a reliable method for an internal machine to signal or obtain a connection to an outside system discretely. UPnP allows for systems to request external routers to open ports for use with software and services that require open ports in order to function. This protocol however has no method for authentication, which makes it ideal for use as a covert channel. A machine can request ports to be opened in certain sequences/combinations that would appear innocuous to other systems on the LAN and invisible to systems on the Internet. Using a SYN scan the intended receiver can monitor for these ports opening and closing and interpret the message. SYN scans on external routers are common on the Internet and would not appear out of the ordinary [3]. Not only does this mask the receiver but UPnP traffic itself is innocuous and is used by many different services to open ephemeral ports on external routers.

## 2. Terms

The following abbreviations will be utilized throughout the paper:

UPnP
: A network extension of the concept of Plug & Play. Regulated by the UPnP Forum it seeks to deliver a set of robust network protocols to allow for dynamic configuration and setup of connected devices.

IGD
: Internet Gateway Devices are a sub scheme within the UPnP protocol. This category is restricted to network devices such as routers/NAT boxes employing the UPnP protocol [4]. Devices within this scheme support operations such as remote port mapping and remote configuration.

CTR
: For this channel the Clear to Read will be defined as a pre-determined port, between the sender and receiver, on the target UPnP IGD. This port when opened signals to the receiver that a message has been placed successfully.

Wait Timer
: The period of time before the sender/receiver will place a message/initiate a SYN scan. This allows for proper syncing between the sender and receiver.

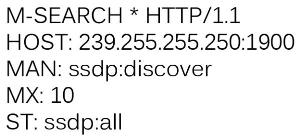
## 3. UPnP

UPnP is a network configuration technology designed to allow devices to auto-configure or communicate without prior configuration. To discover other services on the network a device will multicast a SSDP (Simple Service Discovery Protocol) packet requesting any UPnP enabled devices to respond with a list of their services. SSDP utilizes HTTPU (HTTP over UDP) to transmit this request which resembles a standard HTTP request. Once this request is received each device will send a packet back describing which UPnP service it has available for use. If a device has multiple services available it will send back one packet per service in order to advertise them to the requesting device. This packet provides a URL to the XML schema for a particular service containing information on what commands can be issued and more detailed information about the service. Once a service has been discovered the two devices can communicate via SOAP using the XML URLs returned during the discovery process.

The different types of devices defined under the UPnP protocol allows for a wide range of applications. Media servers can populate network attached media players song/video libraries automatically, network devices can configure a router to enable connectivity, or network devices can pull their own configuration from a server to enable connectivity.

## 4. Design of the Covert Channel

This covert channel seeks to exploit a weakness in how UPnP functions; primarily that it is an un-authenticated protocol [5]. By issuing commands to an IGD device we can open



```
M-SEARCH * HTTP/1.1
HOST: 239.255.255.250:1900
MAN: ssdp:discover
MX: 10
ST: ssdp:all
```

Fig. 1: SSDP Request Packet

up forwarded ports on the external side of a private network [6]. Using a large enough port range encoded messages can be made available to outside observers without making a direct external connection. Observers would scan the external NAT wall of the network within a pre-determined range and record which ports are open or closed. This sequence of open and closed ports would be a binary representation of some encoded message; for the purposes of this paper we will be using an ASCII string but 8-bit binary data is accommodated. The sender and receiver will use the CTR port number as the start of the port range for encoding the message. Next, following the CTR port will be three ports for a packet pre-amble, followed by nine ports for the encoded data and checksum, and ending with a three port post-amble to complete the packet. The packet will be a binary string which the sender will use to open the appropriate ports on the external side of the NAT. For each 1 in the string a port will be opened while a 0 indicates a port should not be opened. The receiver will scan this port range using a SYN port scan [7] and record the open and closed sequence. Using this information the binary string can be validated and the original message decoded. In addition to the CTR the sender and receiver must agree on a wait timer. This timer allows for the client and server to sync correctly and ensure transmission of the message.

## 4.1 Message Encoding

The message will be encoded in a packet format prior to being placed on the IGD. As ports can be interfered with during transmission this allows for error checking on the client side when decoding.

CTR

1 bit in length and is toggled last during packet transmission.

Pre-amble

3 bits in length it is very similar to the Ethernet frame pre-amble [8] as it alternates 1/0's meaning it should always be defined as 101.

Data

The payload of any packet in the channel will be 8 bits in length representing an ASCII formatted character.

Checksum

The checksum will be an odd bit parity check against the payload of the packet.

Post-amble

Identical to the pre-amble encoding this will always be 101 to signify the end of the packet.

## 4.2 Operation

The server sending the message operates in the following order:

1) Close the CTR port on the IGD and wait WAIT_TIMER
2) Encode the Packet
   - Assemble pre-amble
   - Encode payload
   - Create checksum
   - Assemble post-amble
   - Combine into packet
3) Process the Packet
   - Clear all ports needed for transmission from the IGD
   - Open the ports necessary for transmission
   - Set CTR and wait WAIT_TIMER
4) Wait the predetermined amount of time before resuming
5) Repeat for the next of the message

This method allows a client (the receiver) time to jump into the broadcast during the process and sync with the server. Since there is no communication between the sender and receiver the sender may wish to broadcast the message more than once in order to give the receiver a larger window of opportunity. The receiver at any point can begin scanning for the intended message. During the wait period is where the client will be able to sync up with the sender. The receiver samples the CTR port and waits for a transition from open to closed ignoring the rest of the ports until the transition. This will ensure that the parties are in sync and reading does not take place during the write phase. The necessary processing order for the client begins as follows waiting for a transmission from closed to open on the CTR port:

1) Scan CTR port on the IGD
   - If closed go to step 1
   - If open proceed to step 2
2) Process the remaining ports into a binary string
3) Check if the pre-amble and post-amble are intact
   - If not discard packet and return to 1
4) Validate Checksum
   - If invalid discard packet and return to 1
5) Process payload and convert back into an ASCII character
6) Return to step 1

Having the receiver wait for the CTR to transition from open to closed and then back to open allows reliable synchronization of the parties. Should the sender or receiver become delayed or interrupted resynchronization will occur at the next available opportunity. Some data loss may occur

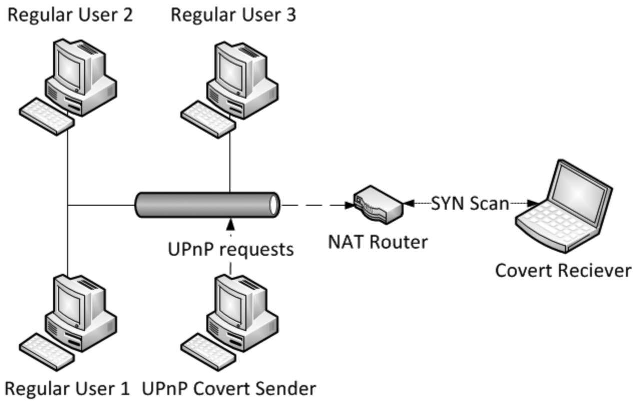


Fig. 2: Diagram of the testing environment

and would have to be detected and compensated for at the message level.

### 4.3 Testing

The testing environment for this covert channel was a simple NAT based environment using a Linksys WRT54-G as the IGD. The UPnP sender and receiver were coded using Perl, miniupnpc for port mapping, and nmap for scanning the IGD. The network layout used is diagrammed in figure 2.

Prior to communicating the sender and receiver agreed upon a CTR of 9000 and a wait timer of 20 seconds. The client's internal timer was set to 5 seconds. Using these values the initial test of the channel was:

1) Sender places the packet and enters the wait period
2) Client is initiated and scans for the packet
3) Repeat

This resulted with the intended message being received by the client. Following that the client was tested against the sender in various different states to ensure proper syncing would occur. In all of these instances the client was able to sync properly and received the message intact. In addition a scenario in which the client was first initiated and the server started some time later was tested. This also resulted in a successful message being transmitted to the receiver.

## 5. Properties of the Channel

### 5.1 Covertness

The channel if used in small bursts transmits packets on the network briefly, leaving only a small window for the traffic to be spotted. That being said it is possible to trace who the sender and receiver are when transmitting the message. The trace is limited though to the message sender and not the receiver. The sender would still be linked to the scans however and the possibility exists that someone could figure out the encoding scheme given enough time. With the frequency and persistence of port scans from a multitude of attackers and compromised host all around the world, if the receiver includes scans of the other ports it will appear little different than the rest of the background noise of the Internet.

### 5.2 Data Rate

The initial results conclude that our implementation of the channel could not reliably function under a 20 second wait timer between processing the message. The time it takes to process and place the message accounts for roughly 1 to 2 seconds. To scan the packet and display its contents also takes about 1 second. However these times vary depending on the amount of ports open, more so on the senders side than the receiver's. Given this the data rate is roughly 15 bits per 22 seconds or 5+ characters/minute.

### 5.3 Robustness

When implemented with a 20 second timer we did not see any errors generated during normal operating conditions. However there were instances in which the IGD took longer than average to process the UPnP requests sent to it. This resulted in longer send times than usual however these slow-downs did not occur consistently. Including these problems the client did not display incorrect data and was able to sync up again when the slowdown concluded. When dropping the timer down below 20 seconds some errors were generated. The most common error was a displaying of incorrect characters by the receiver. Other than that malformed and duplicate packets were dropped as designed. The dropping of duplicate packets based on the channels implementation may result in data loss if resynchronization needs to occur between the sender and receiver.

## 6. Applications

The channel's low bandwidth, while not suitable for long messages, can make it ideal for broadcasting short phrases. One example would be to control a bot-net using the IGD as a centralized control point. The attacker would, using the channel above, broadcast the command on the IGD where an infected bot would be able to find it. An infected machine using the same scanning technique demonstrated above would look for the instructions and execute them. As the implementation makes use of broadcasting the message over and over, as long as any machine listened long enough, it would receive the command in its entirety. This means that each machine does not have to begin listening at the same time for the message. Each bot scanning at different times reduces the footprint created preventing abnormally high spikes in traffic to a single machine (the AP).

There are many benefits to using this design. The first being that a single update to the controller would be heard by all. Infected machines do not have to be tracked in order to update them, they will do so automatically when they scan the target IGD. The attacker does not have to communicate directly with an infected machine, eliminating what would be a direct association with an infected bot. Any such connection could alert a user or administrator to the attack in progress. Last at no point in time are the commands being issued encoded in the traffic generated during the scan. A single packet contains no portion of the command which could give the attack away. Additionally no group of packets could be assembled to decode the command being issued.

## 7. Conclusions & Future Work

Our conclusion is that UPnP using ports for ex-filtration of data is a viable channel for communication. Our implementation was able to clearly send a message to a receiver and do so using existing protocols on the network. Unless someone was explicitly looking for an encoded message on the device no one would be aware of our communications. Future work that can be done to improve the channel would be to increase the speed at which the channel operates. Currently our implementation uses Perl and two external programs in order to place and check for the message. It is possible to encode these natively eliminating some of the overhead associated with the external calls. This would improve the functionality of the channel greatly.

## References

[1] E. Y. Vasserman, N. Hopper, and J. Tyra, “SilentKnock: practical, provably undetectable authentication,” *International Journal of Information Security*, vol. 8, no. 2, pp. 121–135, Nov. 2008. [Online]. Available: http://www.springerlink.com/index/10.1007/s10207-008-0070-1

[2] S. Son, B. Allcock, and M. Livny, “Codo: Firewall traversal by cooperative on-demand opening,” in *High Performance Distributed Computing, 2005. HPDC-14. Proceedings. 14th IEEE International Symposium on*, 2005, p. 233âĂŞ242.

[3] C. B. Lee, C. Roedel, and E. Silenok, “Detection and characterization of port scan attacks,” *Univeristy of California, Department of Computer Science and Engineering*, 2003.

[4] U. Forum, “InternetGatewayDevice:1 device template version 1.01,” Nov. 2001. [Online]. Available: http://upnp.org/specs/gw/UPnP-gw-InternetGatewayDevice-v1-Device.pdf

[5] C. Heffner and D. Yap, *Security Vulnerabilities in SOHO Routers*. Retrieved September, 2009.

[6] T. Maki, “Explicit Mechanisms for Controlling NAT/Firewall Systems Dynamically.”

[7] M. De Vivo, E. Carrasco, G. Isern, and G. O. de Vivo, “A review of port scanning techniques,” *ACM SIGCOMM Computer Communication Review*, vol. 29, no. 2, p. 41âĂŞ48, 1999.

[8] IEEE, “Part 3: Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications,” 2008. [Online]. Available: \url{http://standards.ieee.org/getieee802/download/802.3-2008_section3.pdf}

# Webpage Source Based Covert Channel

Tarun Madiraju, Daryl Johnson, Bo Yuan, Peter Lutz
B. Thomas Golisano College of Computing & Information Sciences
Rochester Institute of Technology, Rochester, NY USA

**Abstract** - *Covert Channels can be used for enabling hidden communication mechanisms that can facilitate secret message transfer. This paper presents a new covert channel based on the HTML source of a webpage. The new covert channel while featuring high bandwidth also demonstrates high imperceptibility as it doesn't involve any modifications to the source or the visibility of the webpage and is independent of timing of page requests. The availability of page source for a webpage on the Internet makes this covert channel easy to implement and effective.*

**Keywords:** Network Security, Covert Channels, Information Hiding, Webpage

## 1 Introduction

Covert channel were first defined by Lampson as a communication channel that was not intended for information transfer [1]. This covert channel, defined in 1973, was based on trusted host systems but can be applied to any shared environment. Covert channel research has migrated into many other environments since then such as networked services.

The DoD's Trusted Computer System Evaluation Criteria (TCSEC) [2] describes a covert channel as a communication channel that facilitates transfer of information between entities in such a manner that the system's security policy is violated. They also mention that the presence of a communication channel cannot be supported without the complete knowledge of the channel. A relatively recent definition states that a covert channel can be established by hiding and transmitting data over a legitimate communication channel [3].

Covert channels can be classified into three types: Storage Channels, Timing Channels and Behavioral Channels [4]. Covert storage channels include communicating entities that are involved in a direct or indirect form of writing to/reading from a storage location. Covert timing channels are described as scenarios where a process signals information to another process by modulating its own use of system resources in such a manner that the real response time observed by the second process is affected [2]. Behavioral type of channels are generally application specific wherein the behavior of an application is used to define and communicate information between the participating entities of the covert channel [4].

A typical HTML page on the Internet consists of HTML tags, regular text and URLs linking to other pages. In order to implement our covert channel, we utilize the text and the URLs available in the HTML source of a webpage. The fact that an existing webpage's source can be used to build a message dictionary without modifying any content illustrates the advantage this covert channel possesses. The purpose of this project was to demonstrate a practical covert channel that is difficult to be detected by an active administrator or a regular observer. Further, the administrator or the observer should not be able to understand the covert message if discovered.

## 2 Paper Organization

The remainder of this paper is organized as follows. Section 3 discusses the literature review which mentions existing work in HTML based covert channels. Section 4 details the design, methodology and experiments. Section 5 presents the characteristics of the proposed covert channel. Section 6 discusses the future work. Finally, Section 7 presents the summary of this new covert channel.

## 3 Literature Review

Previous works have discussed and implemented covert channels based on HTML. These works include embedding invisible characters into the html source, using uppercase and lowercase of letters in html tags, using attributes of html elements, etc [5], [6], [7], [8]. The method that involves embedding invisible characters into html source denotes white space as 0 and tab as 1. The invisible characters are embedded into the webpage by adding white spaces and tabs at the end of sentences or lines [6]. This method is simple to implement however, issues associated with this methods include: high perceptibility as a result of increase in webpage size, need to modify the web page source, and it is easier to destroy the steganographic covert channel by deleting additional white spaces and tabs. The second method is implemented by switching the case of the letters in html elements called tags [5]. The secret message was embedded by utilizing the letter case where uppercase letters were used to represent the binary value 1 and lowercase letters were referred to as 0. This method doesn't lengthen the size of a webpage but the switched case of the letters can be used to decode the embedded message. This covert channel can be broken by converting all the tag letters to uppercase or lowercase format. The third method takes advantage of the attributes of the HTML elements. The attributes of an element in HTML can be used in any order without effecting the webpage content and visibility. The authors Dongsheng Shen and Hong Zhao used a mapping scheme between permutations of attributes and binary strings to implement an embedding and extracting process thereby achieving information hiding [7]. This method doesn't increase the webpage size and is relatively robust. Another attribute based webpage information hiding scheme described in [8] makes use of the property "equal attribute object has identical function" to define a set of rules for embedding and extracting secret information.

The following covert channel mechanisms closely relate to our proposed covert channel. William Huba et al. briefly discuss a covert channel based on the webpage browsing pattern [9] which is currently unimplemented. Erik Brown et al. presented three techniques to implement covert channels using webpage browsing patterns [10]. In the first technique, a value 0 or 1 is assigned for a page request on the basis of its distance from the root node. In the second mechanism, every page of the web site is designated as 0 or 1. The third mechanism denotes values 0 if the pattern of the page requests were in breadth-first search format while a depth-first search would indicate the value 1.

The discussed mechanisms offer low bandwidth and hence do not facilitate high capacity channels. Our approach demonstrates the characteristics of imperceptibility and robustness without compromising on the channel capacity. Unlike the first three studies mentioned above which implement unidirectional channel from the web server to the client, our work focuses on establishing a unidirectional communication channel from client to web server. In addition it focuses on using words available in the webpage as message components and URLs available in the webpage for enabling covert transfer of the message without any modifications to the existing source of web pages.

# 4 Proposed Covert Channel

## 4.1 Design

The proposed covert channel establishes a unidirectional communication scheme between the communicating entities. Typically, the communicating entities of this covert channel are a web server, as receiver, and a web client, as the sender. For any covert channel to be implemented successfully, certain information must be shared between the communicating parties prior to transferring of covert messages. In this scenario, the sending entity and the receiving entity share three things: a webpage, sender's (web client) IP and the encoding mechanism. It is important to understand that in the current proposal the receiver is implemented at the web server and hence the receiving entity must own the web server or must have compromised the web server. But the covert channel can be implemented beyond the mentioned scenario. For example, the covert channel can be implemented for scenarios where the receiver exists between the web client and the web server. This man-in-the-middle receiver captures HTTP requests from the sender in order to decode the messages based on the requested pages. An advantage of this approach is that the receiver need not have control over a web server and any webpage could be selected rather than a webpage from the owned or compromised server. This overcomes the limitation of selecting a webpage from a web server that is either owned or compromised by the receiving entity. However, in this paper we restrict our study to owned/compromised web server as the receiver.

## 4.2 Methodology and Experiments

The methodology section details how the encoding and decoding operations are performed by the sender and the receiver respectively. In order to present the feasibility and practicality of our webpage source based covert channel, a proof of concept was developed in python. The implementation includes two scripts wherein one of the scripts is used to carry out encoding on the web client while the other is responsible for decoding the message at the web server.

*1) Sender Side:* The sender side implementation of this covert channel can be described as the following phases.

Phase 1: URL Dictionary: The first task is to obtain and read and parse the webpage source. We build the URL dictionary, as illustrated in Fig. 1, by assigning numbers (termed as URL numbers) to a uniquely sorted list of URLs identified from the webpage source. The URL numbers act as keys and the URLs represent their values. If there are n URLs retrieved from the webpage, every URL is mapped to a number starting from 0 to n-1. We reserve the first three URLs, termed as Control URLs, to define activities such as indication of start or end of a covert message transfer. The purpose of each Control URL is presented in Table 1.

TABLE I
CONTROL URLS

| URL Number | Indication |
|---|---|
| 0 | Start of communication |
| 1 | Wait for next URL |
| 2 | End of communication |
| 3 to N-1 | A word or a shift |

```
URL Number : URL

3 : http://www.rit.edu/alumni.html

4 : http://www.rit.edu/alumni_photos.html

5 : http://www.rit.edu/campuslife.html

6 : http://www.rit.edu/centers.html

7 : http://www.rit.edu/co-op.html

8 : http://www.rit.edu/colleges.html
```

Figure 1. A sample from URL Dictionary built using RIT's homepage

Phase 2: Word List & Mapping: In addition to the retrieval of URLs, we also retrieve words from the webpage source and uniquely sort the list of words in alphabetical order. The word dictionary is obtained by mapping words from this word list to the URL number(s) as demonstrated in Fig. 2 and Fig. 3.

It is likely that in a webpage, the number of URLs would be less than the number of words in most cases, so we use the following approach to map the words and URLs. The purpose of linking URLs to words is to send or receive a word of the message based on the URL requests. Tthe URL numbers {0}, {1} and {2} are reserved for representing control URLs so we begin mapping the words from URL number {3}.. We make use of Single URL mapping and Multiple URL mapping techniques to link URLs and words. In the Single URL mapping technique, a word is linked directly to only one URL while in the Multiple URL Mapping technique a word is linked to a sequence of URLs. Starting alphabetically, the words are assigned single URLs from {3} through {n-1} to indicate the word desired. Then we utilize the URLs {3} through {n-1} again as shift indicators followed by the URL {1} to indicate the the preceding URL was a shift and not a word. This sequence is then followed by another URLfrom {3} to {n-1}to indicate the word desired. So, some words in the word dictionary (like "about" shown in Fig. 2) are mapped to a single URL and hence it takes only one URL request to convey such words. Other words of the word dictionary are mapped to multiple URLs (like "research in Fig.3). These words are conveyed by requesting the shift URL and a Wait URL followed by the word URL. The wait URL is used to indicate that for this particular word there exists multiple URLs mapped to it. Hence the receiver program must examine the URL following each initial URL to determine if it is a single or sequence of URLs. If a URL is followed immediately by a Wait URL then it is a mutliple URL or sequence.. In the example presented, "about" takes only one URL request which is URL number {3} while "research" requires three URL requests which are URL number {12} followed by the URL number {1}, which is the Wait URL (to indicate there is more than one URL mapped to the word research), and URL number {48}. The URL {12} being followed by the URL {1} indicates that it is a multiple URL or sequence and that the next URL indicates the specific word

The Wait URL allows this scheme to represent (n-3) single URL encodings and (n-3)*(n-3) multiple URL encoded words. Therefore if we have n URLs in the webpage we can encode a maximum number of words of

$$.\{\text{max words}\} = (n-3) + (n-3)^2$$

Phase 3: User Input: The final phase at the sender's side would include accepting URL numbers (mapped to words) from the sender in order to invoke respective URL requests.

```
Word : URL Number

about : (3)

abroad : (4)

academic : (5)

academicaffairs : (6)
```

Figure 2. Single URL Mapping Sample

```
Word : URL Numbers

requirements : (12, 1, 47)

research : (12, 1, 48)

researcher : (12, 1, 49)

reserved : (12, 1, 50)
```

Figure 3. Multiple URL Mapping Sample

This way for every word of the covert message, mapped URL(s) are accessed. With reference to Fig. 2 and Fig. 3, for sending the covert message "about research", we must input the URLs {3} for about and {12, 48} for research. The same is illustrated in Fig. 4.

```
Words Available: 715
Total Available URLs: 51
Please enter the number of words you want to send: 2
Do you want to send a covert message now: [yes/no]? Yes
Start URL Sent
Instructions:
 1.enter the URL numbers mapped to the WORDS
 2.use space if multiple URL numbers are mapped to a WORD
word 1 – URL Number(s): 3
word 2 – URL Number(s): 12 48
END URL Sent
Covert message sent!!!
```

Figure 4. Send Covert Message

*2) Receiver Side:* The implementation on the receiver side is identical to the server side implementation except that the key-value pairs built for the dictionaries are reversed to facilitate decoding of covert message.

Phase 1: URL Dictionary: The first phase for both the sender and receiver are the same as it involves building a dictionary of URLs from the webpage that was decided prior to the communication. However, structurally the key, value pair of dictionary is reversed. In this implementation, the URL acts as the key while the numbers act as their corresponding values.

Phase 2: Word List & Mapping: After building the URL dictionary as mentioned above, the receiver builds a word list and then performs word-URL number mapping but in this case the URL number acts as key and its corresponding values are the words parsed from the webpage source.

Phase 3: Retrieval (PCAP/Apache Log): The final phase of the program on the receiver side is responsible for retrieving the message and displaying it to the receiver. This phase requires an additional input which is the packet capture file (.pcap) in our case. An apache access.log file is an alternative option. The reason for implementing using a packet capture is to imply that such a mechanism can be incorporated for other scenarios where a man-in-the-middle receiver, which is a system between the web client and web server, is the covert message receiver and captures the HTTP requests to decode the covert message. On receiving the packet capture file and sending entity's IP address as input, the receiver program retrieves the URLs requested by the sender from the packet capture and utilizes these URLs to decode the message sent as shown in Fig. 5.

```
H:\>python read.py

Enter your web page: http://www.rit.edu

Enter PCAP path: c:\captures\ritcovert.pcap

Enter Sender's IP: 192.168.0.6

Covert Message: about research
```

Figure 5. Receive Covert Message

# 5 Covert Channel Characteristics

## 5.1 Type

The proposed covert channel falls under the category of covert storage channel. This is because the sender's URL requests are captured in pcap (packet capture) file or the web server log and the receiver reads the packet capture or web server log to decode the message obtained.

## 5.2 Imperceptibility

It is very difficult for a regular observer or a capable administrator to detect the proposed covert channel as existing content of the webpage source and legitimate URL requests are utilized for its implementation and no alterations are made to the source code.

## 5.3 Robustness

The robustness of this channel is high considering the channel's ability of being persistent in the presence of firewalls, proxy servers, routers, etc. The message transfer is carried out by accessing URLs which are basically mapped to words using the encoding scheme. Since these URLs are legitimate HTTP requests they are not deterred by devices mentioned above.

## 5.4 Throughput

With reference to the word-URL mapping example described in the previous section, sending each word of the message takes either one or three URL requests. Every word either takes one or three URL requests as some words are mapped to a single URL and would hence require only one URL request while other words mapped to two URLs are implemented by requesting first URL followed by Wait URL and then the second URL. Considering this, a message having 10 words would take 10-30 URL requests excluding the Start URL and the End URL. This clearly indicates that the proposed covert channel demonstrates much higher bandwidth unlike the existing HTML based covert channels. Table II presents the number of URLs and words parsed from some pages at the time of writing this paper.

TABLE II
SOME WORD-URLSTATISTICS

| Webpage | Words Available | URLs Available |
|---|---|---|
| www.rit.edu | 730 | 51 |
| www.cnn.com | 1753 | 119 |
| www.premierleague.com | 931 | 174 |

## 5.5 Prevention

The imperceptibility factor plays a very crucial role in this covert channel. Though it is quite difficult to detect this covert channel, on identifying the existence of this covert channel, the channel can be disrupted by injecting HTTP requests for URLs existing in the URL dictionary with source IP address spoofed to that of the covert message sender.

# 6 Future Work

Though the proposed covert channel fares high on various characteristics, certain modifications and implementations can further improve the effectiveness of this channel. The message (word) dictionary part of the encoding/decoding mechanism can be improved by implementing an algorithm that would build the message dictionary from the webpage source on the basis of word occurrence frequency in the English language. This approach will help in ignoring words that are less likely to be used. Also, the implementation must be improved to ignore redirection of a URL to another URL that already exists in the URL dictionary as it might result in the transfer of incorrect messages. The widely varying sequence of URL requests would not be suspicious as it might look like a user browsing through different pages of a website however it might result in suspicion when these widely varying sequence of URL requests are made in a very short span of time by the same user. To prevent any suspicion or identification of the existence of this channel as a result of multiple URL requests from the same source in a short duration of time, the code can be modified to implement random wait times between multiple URL requests.

# 7 Conclusion

A communication channel, used to transfer secret message within the same system or across a network, where there is seemingly no communication taking place can be termed as covert channels. A new covert channel based on webpage source was exhibited in this paper. The covert channel presented, used words and URLs available in a webpage source to enable unidirectional covert communication from a web client to a web server. This covert channel does not modify the webpage source while demonstrating high imperceptibility and channel capacity.

# 8 References

[1] B.W. Lampson, “A note on the confinement problem,”*Commun. ACM*, vol. 16, no. 10, pp. 613–615, Oct. 1973. [Online]. Available: http://doi.acm.org/10.1145/362375.362389

[2] V. D. Gligor, “A guide to understanding covert channel analysis of trusted systems,” *ser. Rainbow Series. National Computer Security Center*, Nov. 1993. [Online]. Available: http://www.fas.org/irp/nsa/rainbow/tg030.htm

[3] R. C. Newman, “Covert computer and network communications,” in *Proceedings of the 4th annual conference on Information security curriculum development*, ser. InfoSecCD ’07. New York, NY, USA: ACM, 2007, pp. 12:1–12:8. [Online]. Available: http://doi.acm.org/10.1145/1409908.1409922

[4] D. Johnson, B. Yuan, and P. Lutz, “Behavior-Based covert channel in cyberspace,” in *4th International ISKE Conference on Intelligent Systems and Knowledge Engineering*, 2009, pp. 311–318.

[5] X.-G. Sui and H. Luo, “A new steganography method based on hypertext,” in *Radio Science Conference, 2004. Proceedings. 2004 Asia-Pacific*, Aug. 2004, pp. 181 – 184.

[6] H. Huang, X. Sun, Z. Li, and G. Sun, “Detection of hidden information in webpage,” in *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007.Fourth International Conference on*, vol. 4, Aug. 2007, pp. 317 – 321.

[7] D. Shen and H. Zhao, “A novel scheme of webpage information hiding based on attributes,” in *Information Theory and Information Security (ICITIS), 2010 IEEE International Conference on*, Dec. 2010, pp. 1147 –1150.

[8] Y. Yang and Y. Yang, “An efficient webpage information hiding method based on tag attributes,” in *Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on*, vol. 3, Aug. 2010, pp. 1181 –1184.

[9] W. Huba, B. Yuan, D. Johnson, and P. Lutz, “A HTTP cookie covert channel,” in *Proceedings of the 4th international conference on Security of Information and Networks*, ser. SIN ’11. New York, NY, USA: ACM, 2011, pp. 133–136. [Online]. Available: http://doi.acm.org/10.1145/2070425.2070447

[10] E. Brown, D. Johnson, B. Yuan, and P. Lutz, “Covert channels in the HTTP network protocol: Channel characterization and detecting Man-in-the-Middle attacks,” *The journal of Information Warfare*, vol. 9, no. 3, Dec. 2010.

# A Covert Channel in TTL Field of DNS Packets

Christopher Hoffman, Daryl Johnson, Bo Yuan, Peter Lutz
Rochester Institute of Technology
{cwh4129,daryl.johnson,bo.yuan,peter.lutz}@rit.edu

*Abstract*—**Covert channels are used as a means of secretly transferring information when there is a need to hide the fact that communication is taking place. With the vast amount of traffic on the internet, network protocols have become a common vehicle for covert channels, typically hiding information in the header fields of packets. Domain name service (DNS) packets contain a 32-bit time to live (TTL) fields for each response record. This is the number of seconds the entry is valid for before caching servers remove the entry. There is no prescribed value for this field making it an ideal covert carrier.**

## I. Introduction

The most common technique for transferring secure information is cryptography, using algorithms to mask what is being communicated [1]. Although this makes the information unreadable by third-parties, it does not hide the fact that communication is taking place. Covert channels are utilized to attempt to hide the existence of a communication channel. The original application of covert channels was to solve the Prisoners Problem, where two parties wanted to communicate but communication was mediated by a warden who was able to read the messages and determine if they were allowed. They had to devise a way to hide their secret conversation in a way that would not look suspicious.

Covert channels can be categorized into three categories: storage, timing, and behavior[2], [3]. Storage channels use a shared storage medium between two parties where one party writes and one reads. These include networking protocols or disk storage. Timing channels use relative timing of events to convey information using patterns. This is normally done by purposely altering resources such as CPU or other resource utilization. Behavior channels function more at the application layer and information is conveyed by selecting certain actions. An example of this is selecting certain moves in a board game.

With the vast amount of traffic on the internet, network protocols have become a common vehicle for covert channels, typically hiding information in the header fields of packets. DNS has a time-to-live(TTL) field, a 32-bit unsigned integer, which denotes how long the (domain name, address) pair is valid for, measured in seconds. This is different from the IP TTL field which is the number of hops a packet can make before being removed from the network. This keeps a packet from traveling in a continuous loop.

The DNS protocol does not outline what the TTL value should be so it is possible to assign it any desired value. Although most authoritative servers use discrete increments such as hours and days, due to caching, it is not uncommon to see values with minutes and seconds. This amount of uncertainty makes the TTL field ideal for a covert channel. Different methods for encoding information can be used to alter the bandwidth and covertness of the channel.

## II. Related Work

There has not been much work using DNS to transfer information other than encoding information in the domain name. Typically it is more common to use DNS as a device for tunneling TCP to restricted areas [4]. This is due to the robustness of DNS. In systems that require payment or authentication, DNS traffic is typically still allowed to ensure users do not cache incorrect information.

Even though multiple answers can be returned for each question, a long packet length can appear suspicious. Omar, Ahmedy, and Ngadi examined this issue while they were working with a DNS covert channel for tunneling IP [5]. Their work involved using an alphabet of domain names, each domain name corresponding to a set of characters. To increase bandwidth, more question/answer segments could be included per packet but this increased the size of the packet. The packet would become suspicious if it was either much shorter or longer than normal DNS packets. By keeping the packets near a normal size, the channel would be less suspicious.

Zander, Armitage, and Branch devised a covert channel using the IP TTL field [6]. The effectiveness of this field is dependent on the natural variation in the carrier network and the ability to encode information in this natural carrier. Since the hop distance between two hosts can change over time, it is not possible to assign a static value to their distance. They proposed using an alphabet of low and high values to signal 0 and 1 respectfully. To allow the channel to be more robust, natural distance between the two hosts was reacquired as it changed to provide a channel with less noise.

Zander, Armitage, and Branch went on to evaluate IP covert channels in [7]. They observed three basic techniques: direct encoding, mapped encoding, and differential encoding. These techniques fall short due to the problem of not being able to add this covert channel to a preexisting carrier. Instead of a channel where the endpoints are the covert sender and receiver, a preexisting channel can be utilized and the covert operators sit in the middle and either actively alter traffic or passively listen for the channel passing by. Zander *et al* developed a method for transmitting a covert channel over IP TTL where the covert sender and receiver are not the overt sender and receiver. The first method they proposed assumes that the distance between the covert parties is already known. The sender generates a TTL related to the message, the receiver then adds the known distance before decoding the message. They also proposed another scheme where the sender alters the TTL value, sending bits by repeating or changing the TTL value of subsequent packets.

Qu *et al* examined another TTL covert channel in the IP header protocol [8]. They used this carrier because it is common to see variance in the values it sends. This channel is also persistent from IPv4 to IPv6, even though the field name changes to Hop-Limit. In IP, the TTL field will decrement as it passes through each device on its path and remove itself from circulation when it hits zero. Since the hop distance between two hosts can change dynamically, care must be taken so the message doesnt deteriorate along its path. They devised a method utilizing Galois Fields to decrease the error rate in the channel.

## III. TTL in DNS

### A. Background

DNS is used for translating qualified domain names to addresses that can be used for communicating to remote servers. The DNS request message consists of a list of one or more server names it wants to communicate with. The DNS server generates resource records for the questions, returning address information. The reply can also contain additional resource records with information the server thinks the client may want to know. Generally authoritative name servers for the domain are also provided.

Part of the DNS reply is a time to live (TTL) field attached to each resource record. This is used to communicate how long that DNS entry is valid for caching purposes. When the entry reaches its life span it is removed from the caching table. Before the lifespan is reached, the caching server can generate a reply as opposed to forwarding the request to a higher level.

### B. Covert Channel

For this covert channel, communication will appear to take place normally. It will require a DNS server either be operated by the sender or a server that has been compromised by the sender. Clients require no special utilities to use the channel. When a client sends a request to the defined domain, the server will generate a valid DNS reply, answering all the questions. An analysis of the packet will show an answer for every question as to not raise suspicion of an observer or IDS system. It does not include additional answers to protect the channels presence.

For each question, a small portion of the message is inserted into the TTL field of the answer. To provide a higher level of covertness, messages can be encrypted instead of plain text. The proof of concept server and client are configured so that an encryption scheme can be placed into each of them easily to provide more security or covertness as needed.

A DNS request can contain more than one question statement and each question can have multiple resource records for clustered services such as the Google search engine. This can be leveraged to increase bandwidth by having more than one TTL field in the reply.

To examine TTL fields of large data sets, packet captures were obtained from the Cooperative Association for Internet Data Analysis(CAIDA) [9]. These packet captures were analyzed with Wireshark and scripting tools to examine the TTL fields. TTL values varied between a few minutes and a couple days. A large percentage of values were a discrete number of hours or days but values of minutes and seconds were not uncommon. Observing the DNS traffic, the average number of answer records, authority records, and additional records per response averaged a little above one with the high end of the range well above twenty, even reaching above 100. For testing, one question was used in the request and one answer resource record was returned along with one of each authoritative and additional resource records, a situation that would not look suspicious. Since some DNS servers strip authoritative and additional resource records, these TTLs will not be used for data transfer.

While testing, the TTL values were also examined to find their behavior. The data was graphed to observe the distribution. There was a constant distribution of all TTL values with the most prevalent spikes at 1 and 2 days. There was also more prevalent values in the low range. Statistical tests were also carried out. The mean TTL was found to be 98742 seconds,  27.4 hours, and a standard deviation of 81766,  22.7 hours. This gives a range of 4.7 to 50 hours that most traffic will fall into. Removing the 10 most used TTL values (5, 10, 15 minutes, 1, 2, 3, 4, 12 hours, and 1, 2 days), the mean and standard deviation become 38.44 hours and 42.86 hours respectively for a range of 0 to 81.30 hours. By removing the extreme outliers, more information can be gained about the behavior of traffic that isn't a highly common value. To be covert, an alphabet should be chosen that falls into this range to blend into the rest of the non-covert DNS traffic. The graph of TTL values with the top 10 removed is shown in Figure 1.

If the message is placed into the TTL field as the ASCII value, the packet capture will show the message. To increase the covertness of the channel, a layer of encryption can be used to mask the value. Many of the encryption schemes used for IP TTL field need to handle the decremental nature of the field. Since DNS TTL doesnt change between endpoints, except during caching, any reversible obfuscation algorithm can be used. This could incorporate encryption such as AES or it can be a simple addition based cipher. The client and server are configured to be able to replace the obfuscation algorithm.

Since the TTL field is 4 bytes long, it would be possible to encode 4 bytes directly into the TTL but this would create a TTL over 27 weeks which would be suspicious. Instead, one of the current implementation uses a cipher that takes the value of 2 ASCII bytes and multiplies it by 2 to generate larger TTL values. Calculation is shown in Equation 1

$$TTL = (byte_1 * 256 + byte_2) * 2. \tag{1}$$

This is sufficient to mask the hidden value from a packet capture. This scheme has a one standard deviation range of range of 13.04-17.02 hours so there is a high probability that

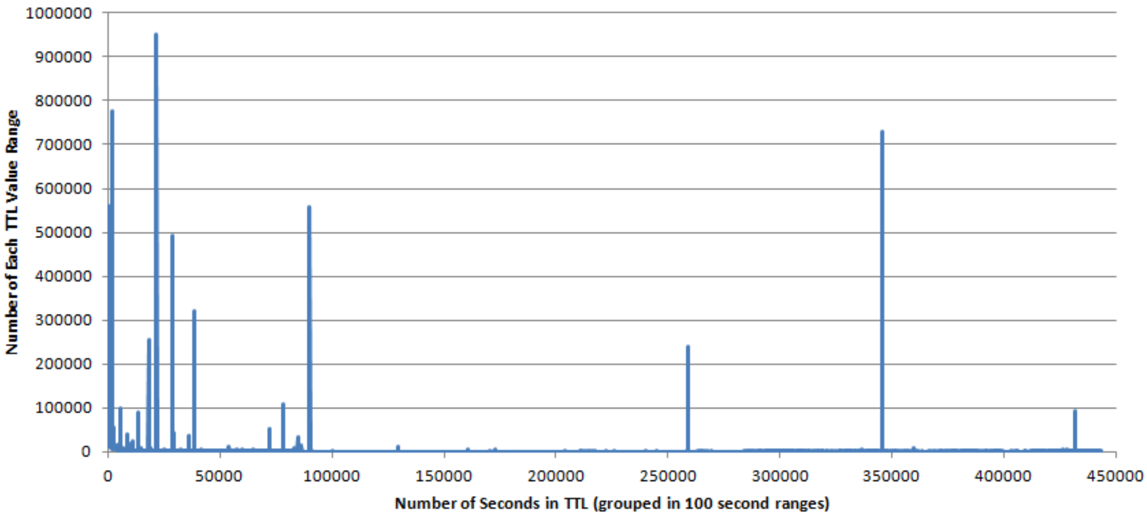
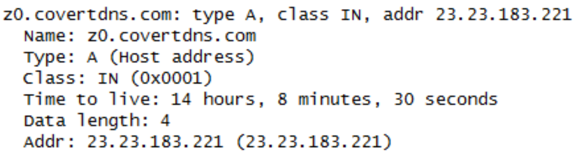


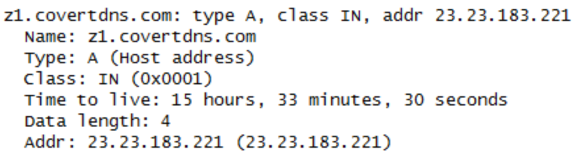
Fig. 1. Plotting the number of occurrences of TTL values from CAIDA datasets

the values will fall within the desired range. Packet captures of this scheme were taken and displayed in Figures 2 and 3 which together form the word ”come”.



```
z0.covertdns.com: type A, class IN, addr 23.23.183.221
  Name: z0.covertdns.com
  Type: A (Host address)
  Class: IN (0x0001)
  Time to live: 14 hours, 8 minutes, 30 seconds
  Data length: 4
  Addr: 23.23.183.221 (23.23.183.221)
```

Fig. 2. Packet capture of covert traffic with the message ”co”.



```
z1.covertdns.com: type A, class IN, addr 23.23.183.221
  Name: z1.covertdns.com
  Type: A (Host address)
  Class: IN (0x0001)
  Time to live: 15 hours, 33 minutes, 30 seconds
  Data length: 4
  Addr: 23.23.183.221 (23.23.183.221)
```

Fig. 3. Packet capture of covert traffic with the message ”me”.

To increase bandwidth, a text only scheme was implemented. A text only message can be sent using a mapping of a=0, b=1, etc with 26=EOF. The EOF character is used at the end of stream to notify the receiver that the message is complete. If the message is an odd length the EOF will appear directly after the message, otherwise it will be sent in the next TTL value. EOF also helps keep a larger TTL as to avoid suspicion as opposed to padding with null values. Using this scheme, a TTL of one day can transmit 3 characters per resource record. Four characters cause the TTL value to move towards a week which is a long time out if a caching server is in the middle. To increase the TTL to fit within the desired range, the value is multiplied by 5 before it is encoded into the packet. This calculation is shown in Equation 2.

$$TTL = (char_1 * 27^2 + char_2 * 27 + char_3) * 5 \qquad (2)$$

This scheme has a one standard deviation range of 0.1-21.05 hours so it falls below the desired range but is acceptable due to the density of low values in the CAIDA dataset.

## IV. Results and Evaluation

Based on a proof of concept server and client, the TTL field of DNS successfully carried a channel between two parties. The experiment was constructed using Amazon AWS to host the DNS server to have the traffic travel across the internet. The client was run on a local host using Wireshark to monitor the traffic between the two locations. To conduct testing through DNS servers, a domain was attached to the server. This ensured that the channel would not be broken during normal communication.

Testing this channel for bandwidth, the client was configured to query 100 new hostnames. Using Wireshark, the duration of all 100 queries were measured and used to find the number of queries per minute. 100 queries took 2.6 seconds which calculates to a bandwidth of 2305 packets per minute.

Using the alphabetic encoding scheme with one resource record, the capacity of the channel is 3 characters per packet. Averaging 2305 DNS interactions a minute, this gives a 6915 character/minute bandwidth. Using the byte encoding alphabet, 2 bytes can be transmitted per packet. This yields 4610 bytes/minute bandwidth.

Both of the proposed alphabets fall within the desired range gained from the statistical analysis of CAIDA data. If the covert traffic is viewed on its own it may look suspicious but with overt traffic it will be sufficiently masked.

Investigating rules from Snort [10], a popular IDS, rules do not exist for specifically limiting DNS traffic based on TTL. There exist a set of rules that do block a DNS packet with TTL as a parameter of the condition, it looks for a TTL less than or equal to one minute. However, it also has a condition that the reply is coming from a non-authoritative server. This rule is to alert to the presence of a DNS spoofing attack and so the DNS covert channel would not trigger that alarm. To test, a snort IDS was configured on the network where it could see the covert channel traffic. The IDS did not detect the presence of the covert channel.

To detect the presence of this channel, a stateful system would need to be employed that would compare the TTL of similar resource records. If the values appeared to fluctuate in a non-uniform manner, there could be suspicion that covert communication is taking place. Such a system would generate a number of false negatives that it may be deemed unusable.

Bromberger alerts to the possibility of covert channels using DNS and gives an outline on how to protect against them [11]. A few of the characteristics that could indicate a DNS channel are lookups that are composed of hexadecimal strings, lookups with long 3rd and higher level names, multiple queries to non-obvious or foreign domains, responses with loopback addresses or other non-routable network addresses, queries to DDNS providers, and requests that are not followed by a requested connection to the requested addresses. As long as the parties communicating are within the same country and limit the number of interactions, the only characteristic my covert channel would fail is the client not requesting a connection to the requested addresses although this can be solved by having the client generate a connection to the addresses it received.

Once the channel is discovered, the channel can be linked to the sender. The receiver is safer from being detected in that the communication stream would need to be tracked through multiple DNS servers to get back to the receiver. Anonymity can be attempted by each party using the follow techniques. For the sender, a rogue DNS can be configured inside a network for sending information. Using spoofed IP and MAC addresses that keep changing, the server can avoid being located. However, the network could be configured as to only allow DNS replies from certain internal machines. The user could also obtain a public addressable DNS server under fraudulent credentials to avoid being tracked if the channel is found.

The receiver can also attempt to remain anonymous by spoofing their address on a request and passively listening for a reply. The caveat here is that the receiver might not see the return traffic if they are on a switched network that will send the reply directly to the address of the spoofed request. If the receiver is on a hubbed network and is able to see the traffic, it could only be increasing the chance of the channel being detected. If the spoofed address does not exist, it would be suspicious that communication is taking place with a non-existent entity. If the address does exist, a host-based IDS might trigger an alert that it received DNS traffic it did not request.

The channel is robust when the reply is from an authoritative server as opposed to a caching server. DNS traffic is necessary for network communication, even if only allowing DNS queries to an internal server. During normal traffic, DNS TTL does not get altered except during caching. When caching occurs, the TTL value gets decremented which will break the channel. During normal request/reply interaction, the TTL doesnt change.

## V. Future Work

DNS uses caching along the path to reduce the amount of queries needed and speed up resolutions. When an answer is received from a caching server, the TTL value will be altered from the value given by the authoritative server. This change can break the channel since the time until it is requeried is unpredictable. Creating a channel that would be robust over caching would allow multiple clients to access the covert channel after a single client resolved the domain name. Although this would not be robust for the full life of the entry, it would allow some flexibility to who could access the channel.

This channel can also be used from an intermediate station [1], so instead of the channel endpoint generating traffic, the response can be encoded into existing traffic. This would require a router in the data flow stream to intercept and alter packets. This could use a legitimate DNS server and limit the linkability of the communicating parties.

## VI. Conclusion

The TTL field of DNS was created to give a lifespan to a domain name, address pair. However, this value can be altered by the user to create a covert channel. This works well because the remainder of the packet is legitimate traffic and TTL does not have an expected value. This creates a robust channel as DNS traffic is largely unobstructed. At the current time, the channel is not robust enough to survive caching.

## Acknowledgements

Support for the IPv4 Routed /24 Topology Dataset is provided by the National Science Foundation, the US Department of Homeland Security, the WIDE Project, Cisco Systems, and CAIDA Members.

## References

[1] S. Zander, G. Armitage, and P. Branch, "A Survey of Covert Channels and Countermeasures in Computer Network Protocols," *Communications Surveys Tutorials, IEEE*, vol. 9, no. 3, pp. 44 –57, quarter 2007.

[2] D. Johnson, B. Yuan, and P. Lutz, "Behavior-Based Covert Channel in Cyberspace," *Intelligent Systems and Knowledge Engineering*, 2009.

[3] R. C. Newman, "Covert Computer and Network Communications," in *Proceedings of the 4th Annual Conference on Information Security Curriculum Development*, ser. InfoSecCD '07. New York, NY, USA: ACM, 2007, pp. 12:1–12:8.

[4] L. Nussbaum, P. Neyron, and O. Richard, "On Robust Covert Channels Inside DNS," in *Emerging Challenges for Security, Privacy and Trust*, ser. IFIP Advances in Information and Communication Technology, D. Gritzalis and J. Lopez, Eds. Springer Boston, 2009, vol. 297, pp. 51–62.

[5] S. Omar, I. Ahmedy, and M. Ngadi, "Indirect DNS Covert Channel Based on Name Reference for Minima Length Distribution," in *Information Technology and Multimedia (ICIM), 2011 International Conference on*, Nov. 2011, pp. 1 –6.

[6] S. Zander, G. Armitage, and P. Branch, "Covert Channels in the IP Time To Live Field," in *Australian Telecommunication Networks and Applications Conference*, Dec. 2006.

[7] S. Zander, G. Armitage, and Branch, "An Empirical Evaluation of IP Time To Live Covert Channels," in *Networks, 2007. ICON 2007. 15th IEEE International Conference on*, Nov. 2007, pp. 42 –47.

[8] H. Qu, P. Su, and D. Feng, "A Typical Noisy Covert Channel in the IP Protocol," in *Security Technology, 2004. 38th Annual 2004 International Carnahan Conference on*, Oct. 2004, pp. 189 – 192.

[9] Y. Hyun, B. Huffaker, E. Aben, and M. Luckie, "The CAIDA IPv4 Routed /24 Topology Dataset - 3/26/2012-3/27/2012,4/7/2012-4/8/2012," http://www.caida.org/data/active/ipv4_routed_24_topology_dataset.xml.

[10] Snort IDS, http://www.snort.org/.

[11] S. Bromberger, "DNS as a Covert Channel Within Protected Networks," in *National Electric Sectory Cybersecurity Organization*, 2011.

# Database Security Requirement on Cloud Computing

**Hana Do, Hoon Jeong, and Euiin Choi**
Dept. of Computer Engineering, Hannam University, Daejeon, Korea

**Abstract** - *Existing database technology has cost or performance issues, scalability issues due to number of concurrent users and data scale in mobile service, internet, and cloud environment. Hence, system has to provide separate data storage management system but also user scalability for service to fit each environment. Of course it has to consider data management function to fit performance with function to require a service application, too. In this paper, suggest requirements to fit the security of databases (Cloud environment) based on protection profile issued by existing security secretariat, and SPP guideline of ISO / IEC 19791. Also, to allow easy access for some people that CC, security assessment, and management does not have expertise.*

**Keywords:** Database security, Cloud service, Authorization, Authentication, Data protection

## 1 Introduction

Existing database technology has cost or performance issues, scalability issues due to number of concurrent users and data scale in mobile service, internet, and cloud environment. Also, these environment provides overfull many function than required by the service, lacks flexibility to be optimization to fit application. And existing relational database systems has been a kind of overfull function such as transaction processing and join operation this cause of poor performance. Hence, system has to provide separate data storage management system but also user scalability for service to fit each environment. Of course it has to consider data management function to fit performance with function to require a service application, too.

At this time, the occurring problem is how protect database used to the service. Korea and the EU have got dual access method concerning the protection of databases. First is how to protect the database by copyright. Second is protection independently the database from copyright protection.

In this paper, suggest requirements to fit the security of databases (Cloud environment) based on protection profile issued by existing security secretariat, and SPP guideline of ISO / IEC 19791. Also, to allow easy access for some people that CC, security assessment, and management does not have expertise.

## 2 Related Work

Database security can be divided two kinds. First is database security through encryption. Second is data protection via strict cipher key management. Database Encryption cannot be decoding by data encryption moreover secure cipher key management even though data is spill, only a few licenser and record for cipher key is allow to access[1].

Data encryption is just a beginning. If there are no authority management and authentication about a strong cipher key, it be leaked easily by insiders, then encrypted data will be hacked. Hence, it needs to consider important three requirements for database encryption. There are three important elements. One, how do database performed encryption, another, what will be to the flow of data within the corporation IT infrastructure, the other, how do database encryption interlocked to security rules within company. When database encryption is in progress, but it is likely that source database size will be change. Some encryption modules create a fixed size even after encryption, and it can be made with the same size as before encryption.

Database encryption should be protected intelligent property against large-scale data breaches rather than the typical encryption and ensured reliable security measures. Next is proposition for security requirements and component from the encryption architecture on database security.

## 3 Database Requirements Analysis

Cipher system architecture for general database security is composed as seven-component[2].

- Cryptographic engine : Carry out encryption

- Key value : Secure key storage

- Key manifest : Details information tracks of the key including aliases, family, status, engine

- Key manager : key management in the key store with list of key

- Protected data : Protected data through encryption

- Cryptographic consumer : Management and handling for data required encryption and decoding

• Cryptographic provider : Connection between the cryptographic engine and cipher consumer

## 3.1 Security Guidelines on Cloud Computing

We analyzed adopted NIST SP 800-53A and SR_MCS as security guidelines on cloud computing[3, 4].

### 3.1.1 NIST SP 800-53A

The purpose for the NIST SP 800-53A is a guideline for evaluation to effectiveness of info protection control within the Federal Information System is as follows:

• Evaluation about information protection control available for comparison, repeatable, and consistent

• Effective cost-evaluation about Information protection control effectiveness

• Recognition and understanding about the risks caused from management of information system

• Providing information more reliable and complete for judging of information protection authorization and FISMA criteria

### 3.1.2 SR_MCS

First, function requirements of SR_MCS are as follows:

• Use the SPP/SST paradigm (Concepts and terminology) of ISO/IEC 19791, and it has to compatible with other security evaluation standards

• Establish secure objective and security measure in cloud system through analysis results of risk and vulnerability, value of assets which realizes assets associated with mobile cloud system. And selection for security function required from ISO/IEC 19791(in accordance with concept of the SPP of ISO/IEC 19791)

• Used to content of PKB data and traditional PP to maintain compatibility with traditional PP with PKB of the existing CC-Tool Box

• Considering assumption in existing PKB, threats, and policy sentence with sentence of traditional PP, then building for database by pre-define to generally an object-oriented assumption(or real environment), risk, and policy sentence

• It needs a writing way to common, formal reusable risk sentence with security policies sentence

• To make available when to develop to security requirements specification sheet for security products not only security requirements of similar operating system-level besides mobile cloud system

• Providing tools to support to SR_MCS Process

• Providing groupware functions for many developers to develop to the security function requirement specification sheet

Second, the performance requirements are as follows:

• Select the optimal security functions(security measures) by considering risk level of organization combined to mobile cloud system resources value, risks, and vulnerabilities

• Tool need not to operate in real time

• Must support levels of various user

• Portability of tool should be higher because should be used in each organization

Third, the security requirements are as follows:

• SR_MCS is need to security function as access right system control about type of users tool, each user's role-based data and functions

• When the SR_MCS is developed, a info about vulnerability, asset value of the organization is confidential data, so it required "confidentiality" function

• SR_MCS document should be maintained "integrity" and maintained "availability". And "accountability" function of the participant during the developing to SR_MCS should be offered, too

The figure below shows the derivation process according to SR_MCS process and authoring tool of security function requirements specification. In particular, the mobile cloud systems analysis, real environment, risk, and security policy is said "mobile cloud system security environment". In order to analyze correct security environment, "risk analysis" must be conducted beforehand. Risk analysis is the matter outside the scope of this paper.

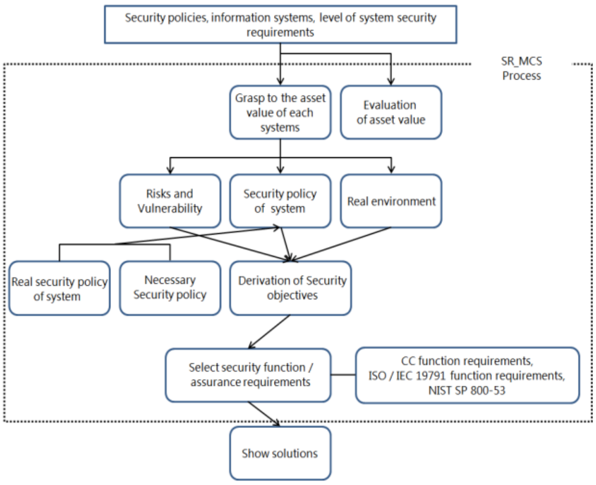
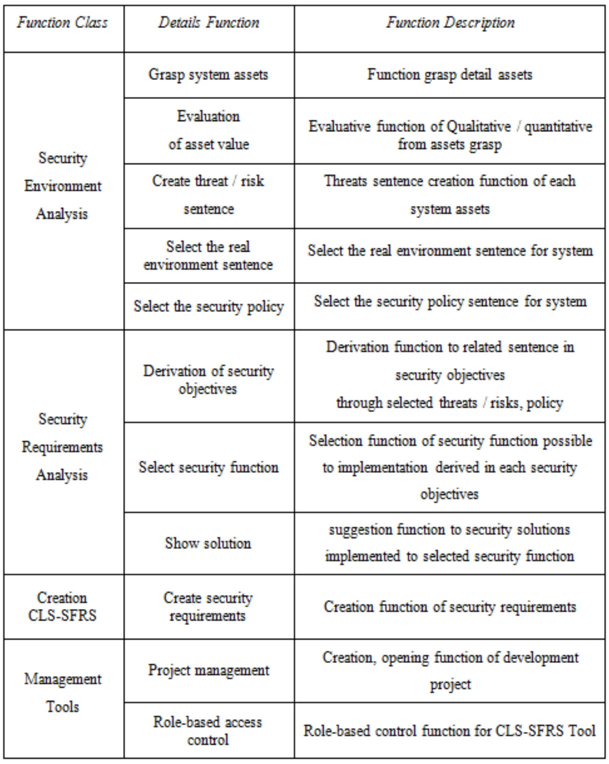

Security policies, information systems, level of system security requirements
SR_MCS Process
Grasp to the asset value of each systems
Evaluation of asset value
Risks and Vulnerability
Security policy of system
Real environment
Real security policy of system
Necessary Security policy
Derivation of Security objectives
Select security function / assurance requirements
CC function requirements, ISO / IEC 19791 function requirements, NIST SP 800-53
Show solutions

Fig.1 SR_MCS Process

Table1 Major function of SR_MCS Tool



| *Function Class* | *Details Function* | *Function Description* |
|---|---|---|
| Security Environment Analysis | Grasp system assets | Function grasp detail assets |
| | Evaluation of asset value | Evaluative function of Qualitative / quantitative from assets grasp |
| | Create threat / risk sentence | Threats sentence creation function of each system assets |
| | Select the real environment sentence | Select the real environment sentence for system |
| | Select the security policy | Select the security policy sentence for system |
| Security Requirements Analysis | Derivation of security objectives | Derivation function to related sentence in security objectives through selected threats / risks, policy |
| | Select security function | Selection function of security function possible to implementation derived in each security objectives |
| | Show solution | suggestion function to security solutions implemented to selected security function |
| Creation CLS-SFRS | Create security requirements | Creation function of security requirements |
| Management Tools | Project management | Creation, opening function of development project |
| | Role-based access control | Role-based control function for CLS-SFRS Tool |

## 3.2 Security Requirements on Cloud Computing

Cloud computing environment has feature to store key info of company in the space managed by cloud service providers. Depending on the characteristic of the cloud computing environment, we are need new security requirements. It is subject to different setup, management environment, service installation, apply to laws and regulations, and action environment other than info system used by currently companies[5, 6, 7].

### 3.2.1 Authentication and Identity Management

Identity management function plays important role on the cloud computing. It is essential security service for data access with use to the cloud services as a set of user attributes used for authentication and authorization services. Solutions have been proposed to provide SSO and identity management functions in a more typical environment. One is that federation-based identity management used to SAML(Security Assertion Markup Language) as based on a trust relationship between companies. The other is user-centered id management method possible to control the id flow between info system by users and to select running system of authentication service in process of identity management and authentication[8].

### 3.2.2 Authorization and Log Management

A security policy in cloud computing environments is define by the company's security administrator or individual users using cloud services and has a structure that apply by the cloud service provider[9, 10]. At this time, the security policy must have a common structure possible for authorization attributes on the cloud services. It should be supported to RBAC(Role-Based Access Control), ABAC(Attribute-Based Access Control), besides DAC(Discretionary Access Control) as the traditional authorization policy.

### 3.2.3 Privacy Protection

When to provide personal info storage and management service from cloud service providers or to discarded, to use personal info, and provided to other systems according to privacy policies set by companies or individuals, It should receive consent of the owner of personal information, must be able to provide notification function with function to fulfill obligations.

# 4 Conclusion

Cloud computing will bring big changes on info management and utilization by companies and normal users. Current, like cloud hard disk service widely used by normal users, it looks a relatively simple structure that users independent use services offered by each cloud service providers. In case of companies, it will be use seamlessly and cost-effective services of several cloud service providers for new business creation.

Another important issue to be considered with the expansion of cloud services is security and privacy protection about the stored information, in place beyond the scope managed of the user. Security requirements which to consider in cloud computing environment is similar to common information system. It is authorization, identity management, authentication, log management, encryption, security and privacy policy creation and enforcement, compliance with applicable laws and regulations, etc.

But, services used by users or businesses on cloud computing environment can become constructed as form provided by interworking between various cloud service providers. Therefore, it is important the standard decision for interworking of the security services between cloud service providers. Services of a public institution besides services provided by common business, for providing safely as cloud form, is necessary to establish authentication standards about software of the cloud that cloud service providers to use. Also, adding of items related to security and privacy protection should be considered in service level agreement content focused on current service quality mainly.

A study on cloud computing, services, and security is the beginning. In overseas, currently studies about technical and managerial countermeasures to respond to cloud security requirements are underway. Our nation is also necessary to establish relevant standards and regulations with cloud computing security-related technology development based on the government, industry, academia, and research cooperation.

## 5 Acknowledgments

This work was supported by the Security Engineering Research Center, granted by the Korea Ministry of Knowledge Economy.

## 6 References

[1] Lee Jeong-a, "Current Status mobile cloud services, domestic and foreign policies", KT Institute of Economics and Business Administration DigiEco Focus, 2010 (http://www.digieco.co.kr)

[2] Kim Young-sun, Go Gap-seung, Sin Jae-in, Lee Gang-soo, "Information Security operating system Level of security functional requirements specification support tool development ", Journal of Security Engineering, Vol.7, No. 1, 2010.2

[3] Bang Young-hwan, Jung Sung-jae, Hwang Seon-myung, "Mobile Cloud System Security Requirements Specification Development Support Tools", Information and Communications Magazine,Vol.28, No.10, 2011

[4] ISO/IEC DTR 15443-3, "Information technology-Security techniques-A framework for IT security assurance", July. 2005

[5] Wayne Jansen, Timothy Grance, "Guidelines on Security and Privacy in Public Cloud Computing", National Institute of Standards and Technology, Draft Special Publication, 2011.1

[6] Neal Leavitt, "Is Cloud Computing Really Ready for Prime Time", IEEE Computer, 2009. 1

[7] Hassan Takabi, James B.D. Joshi, Gail-Joon Ahn, "Security and Privacy Challenges in Cloud Computing Environments", IEEE Security and Privacy, 2010 11/12

[8] OpenID, www.openid.net

[9] Amazon, "Amazon Elastic Compute Cloud (Amazon EC2)", aws.amazon.com/s2

[10] Amazon, "Amazon Simple Storage Service (Amazon S3)", aws.amazon.com/s

# Security Authentication using Improved RBAC Model

**Hana Do, Hoon Jeong, and Euiin Choi**
Dept. of Computer Engineering, Hannam University, Daejeon, Korea

**Abstract** - *Recently, With the IT technique growth, there is getting formed to convert to database environment that means it can access information everywhere and every-time using various devices. But, in this computing environment will be connected to wireless network and various devices. Random access approaches of information resource make trouble to system. So, access authority management is very important issue both information resource and adapt to system through founding security policy to need a system. In this paper supposed to RABC model that based on security about which user's information which provide efficiently access control to user through active classification, inference and judgment about user who access to system and resource.*

**Keywords:** RBAC Model, security, Access Control, Authority management

## 1 Introduction

Recently, With the IT technique growth, there is getting formed to convert to database environment that means it can access information everywhere and everytime[1, 2]. Since the advent of the ubiquitous environment, the user can connect computing environment every time using various devices and the computer can decide to provide useful services to users according to profiles. But in this computing environment will be connected to wireless network and various devices. According to random access approaches of information resource make trouble to system. So access authority management is very important issue both information resource and adapt to system through founding security policy to need a system. But existing access control model is available to approach information resource or computing system by using simply user ID and Password. So this model has a problem that is not concerned about user's context information as user's profile[3, 4].

Computing environment's access control model that has different existing security access control model in which is authorized by user's simple information(e.g,. ID and Password) and it has to add user's profile, location information and time information during realtime. And then it provides service about user's access request by environment. In case that authorized user who is certificated to use service, non authorized user access to data who even if has authentication about access authorization of requesting information. So Access control model has to integrity and automation about resource by control according to user environments and profiles.

In this paper supposed to dynamical RABC model that based on profile about which user's information and environment. Systems provide efficiently access control to user through classification, inference. Therefore we suggested Security Authentication using RBAC Model which stored location and time, frequency information of often used service.

## 2 Related works

### 2.1 RBAC

RBAC is access control model that is more popular in commercial area as alternative model of MAC(Mandatory Access Control) or DAC(Discretionary Access Control). The best feature of the RBAC is not directly allowed to user who is available for performance of operation about information that is assigned by role that is point obtaining access authority of user through assigned role to user. As management for access authority as relation of role and entity role, a number of user and entity, can be managing authority and authorization efficiently in distributed computing with occasionally changing component. Also, assigned low role by between roles of hierarchy architecture provide to authority inheritance that is available of upper role. Using authority inheritance can more efficiently perform to authorization about role that is consisted of hierarchy architecture. This method has advantage of not only simplifying authority management, but also offering flexibility to implementation of security policy[5, 6, 7].

### 2.2 GRBAC(Generalized-RBAC)

GRBAC(Generalized RBAC) model use subject role, object role, environment role in access control decision. And that added context information to existing role based access control. Through modeling in role of subject, object and environment entity offer simplicity and flexibility of access control policy. Security manager describe access authority policy through five attribute that is subject role, object role, environment role, operation, sign.

<<SRole, ORole, ERole, op>, sign>

<<doctor, history case, weekends, read>, ->

Example above the expression show not reading history case on weekend of assigned user as role of doctor. Also, in order to solve inexplicit authorization through role hierarchy architecture use authority inheritance concept of role hierarchy architecture. Authority inheritance is divided into three type of standard, stink, lenient. GRBAC model is handling request of user's access through context information as defining environment role after describing access control policy and adapting transmission rule. But GRBAC is not presented resolve a problem about collision between authority as causing access authority transmission. And GRBAC is difficult to management that occur a plenty of hierarchy as defining user's condition to environment role[6, 8].

# 3 Authentication using User Profile

We have to infer what service is fit to user well for serving various services to user based context-aware information which arose in ubiquitous environment. Generally, we stored profile for using user’s inclination and information. Also, because services that often used have high probability which continuously using, if the services stored in profile, we could reduce time of service using. Therefore, previous technique which information and time of often using services stored in profile was suggested. But there are need to information of user’s location and time for providing more correct services. For example, we assume that a service was used 10 times on a day, and if time of service using is 3 P.M, we should infer that the service almost would use to afternoon. And time and frequency of information is important in ubiquitous. But location information is very important also. Even if services was same, frequency of service was different each other. Therefore we suggest technique that providing the service which demanded by user to store location information with time and frequency in profile and that put the service in location on time to using. Figure 1 shows how we structured Authentication using User Profile

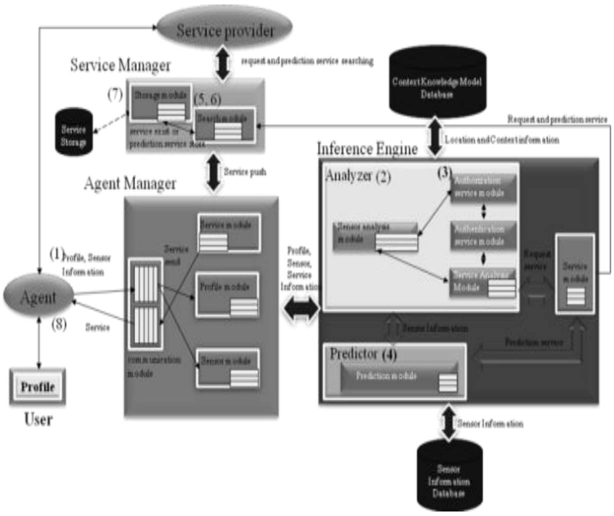


**Fig. 1** Authentication system using User Profile

Fig-1 was consists of 3 modules, such as Service manager, Agent manager, Inference engine, and each modules consist of sub-modules. Service manager is responsible for processing the services which user was requested or predicted by authentication system. And then the predicted services stores on service storage. If users request the predicted service to authentication system, we will directly search it on service storage without searching works. So it is more quickly find service which user requested, and is able to provide it to user. Agent manager is responsible for receiving information from authentication system, and then send to inference engine. Also, it is send services which was find from service provider to service handler on agent. Analyzer within inference engine is responsible for analyzing context with profile and sensor information to provide suitable service for user, processing access control of users. And predictor estimates services which user is going to use service on other place. In inference engine, authorization service module performs that is both in charge of management and treatment in context information of subject and confirming identification about subject that accessible of context-aware access control system. Also, Authorization service module provides service of assignment as dynamically about role of user through analysis of access policy and acquiring added information that is access location, access time, spatial area about context information of subject which is access of resource. And Authorization service module performs for role of access control through comparison and analysis of security policy with both user role of activated user and activated context role in present. Authentication service module performs for monitoring function of user's access control. Authentication service module acquires context information by surround sensor or device besides access information of approached subject. And then, through comparison and analysis of context information about surround environment of accessed user is in charge of pre-processing about authority level of user who wants access. And, through authorization service is in charge of function that provide to data about authority of user to access.

# 4 User Profile

## 4.1 Definition of User Profile

User profile specifies information of interest for an end user. So the profile was structured user information part and service information part.

Structure of user profile was follow:

- User Information: User name, User ID, Personal inclination, hobby

- Service Information: Service Name, Service Provider, Service context, Service frequency value

We assumed that the services will use this place and time next time. If service was demanded in specific location and time. So we used the information of time, location, frequency to provide services to user more correctly and suggested technique which using recently access time, access time, frequency of access, location value, weekend value. And the values stored in profile.

• Recently access time(t): This value stored time when service used recently, and use for finding service which not used for a long time.

• Access time(a): This value have to 24 from 0, and if service was used on 1 P.M, it's value has 13.

• Frequency of access (f): This value stored frequency of service how many users used the service.

• Location value(l): This value have unique number of place where service was used. For example, if user used a service in house and office, location value of A service which used in house is 1, other is 10.

• Weekend value(e): This value have to 7 from 1, if service used on Monday, weekend value is 1. Generally, people's life pattern was repeated per week. So we use the value for analyzing service frequency of user per week.

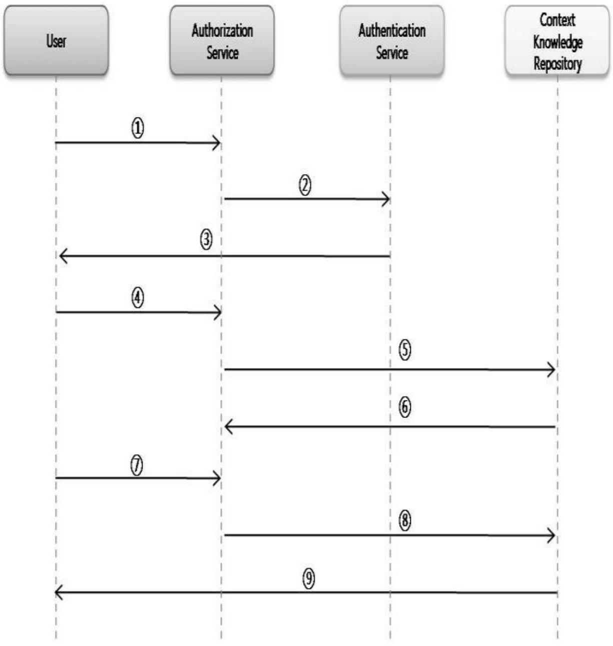


**Fig. 2** Process architecture

### 4.2 Access Control Processing Workflow in Authentication

• User make an approach to authorization service for authority of authentication to access in resource. User utilize for application in order to access of resource.

• Authorization service is call up service for authorizing of authority about user in present. Authentication service is collecting context information of user's surroundings about approach of resource in present.

• Users approach of resource and context-aware service that ask for context information.

• Acquired information by context information of user's surroundings transfer to authorization service module and authorization service module transmit information about receiving of acquired information to authentication service module.

• Acquired authorization service module by context information of user's surroundings try to access of resource that is approach to context knowledge repository for performing access control and role assignment of user.

• It's request data of access policy and information about role-assignment of user from context knowledge repository. Authorization service is granting access authorization by access policy and role of user who want to approach of resource in present.

• User request's to service through acquisition of access authority about assigned role.

• Authorization service module make request to service and authorization service module make an approach to suitable resource in level of access authority through level of authority and role by assigned resource of requiring to user.

• Context knowledge repository can be approached to suitable resource about level of access authority by assigned of authority, security policy and context of user in present.

## 5 Conclusion

Meaning of database computing environment where is available to use for computer conveniently and naturally in common life which is without constraint of location or time. Thus, in distributed computing environment such as ubiquitous environment, user is efficiently available to use and to share of resource between user and other user. Also, we need to access control model to control that is available to access of user that is possible to access in case of sharing resource. And, for using of efficient resource that need to access control model which is able to control of approach to user without authority. Therefore, in this paper is proposed to model that have advantage of which active authorization is more possible then existing access control model as adding a function of authorization about collaborative resource control about other subject in different with RBAC and GRBAC. Proposed model will be making system of active access control that is based on suitable context information in computing environment. We assign to role of access authority about information resource and user to assign of suitable role. And then we provide to service that can be available to information resource through valid access authority of user who is suitable. For using information resource, we will be implementing active access control based on context aware that is estimation of validity about acquired access control through checking satisfaction of security policy about context role in present(although user have a assigned role). And for adapting service along to context transition, we will provide to service which must provide to user in specified context with security policy through aware of automatically about transition of context role.

# 6 References

[1] K. Lyytinen and Y. Yoo, Issues and challenges in ubiquitous computing, In Communications of the acm, Vol. 45. pp. 62-96. 2003

[2] Schilit, BN., Adams, N., Want, R., Context- aware computing applications, Proc. IEEE Workshop on Mobile Computing Systems and Applications, pp. 85-90, 1994

[3] Olivier Potonniée, A decentralized privacy-enabling TV personalization framework, In 2nd European Conference on Interactive Television: Enhancing the Experience (euroITV2004), 2004.

[4] G. Klyne, F. Reynolds, C. Woodrow, H. Ohto, J. Hjelm, M. H. Butler, and L. Tran. Composite Capability/Preference Profiles (CC/PP): Structure and vocabularies 1.0. W3C Recommendation , W3C, 2004

[5] D. F. Ferraiolo, J. A. Cugini and D. R. Kuhn, Role-Based Access Control(RBAC) : Features and Motivations 11th Annual Computer Security Application Conference, November, 1995.

[6] R. S. Sandhu and E. J. Coyne, Role-Based Access Control Models, IEEE Computer, 20(2),pp.38-47, February, 1996.

[7] R. S. Sandhu, D. Ferraiolo and R. Kuhn, The NIST Model for Role-Based Access Control:Towards a Unified Model Approach, 5th ACM Workshop on RBAC, August, 2000.

[8] G. Neumann and M. Strembeck, An Approach to Engineer and Enforce Context Constraints in an RBAC Environment 8th, ACM Symposium on Access Control Models and Technologies(SACMAT2003), pp65-79, June, 2003.

# A Proposed Method for Examining Wireless Device Vulnerability to Brute Force Attacks via WPS External Registrar PIN Authentication Design Vulnerability

**S. Aked**[1], **C. Bolan**[1,2] **and M. Brand**[1,2]
[1]School of Computer and Security Science, Edith Cowan University, Perth Western Australia
[2]secau – Security Research Centre, Perth Western Australia

***Abstract -*** *Wi-Fi Protected Setup (WPS) is a certification scheme introduced in 2007 to ensure that wireless SOHO (Small Office, Home Office) and home networks could be connected to in a trusted, yet user friendly manner. Recently, WPS was shown to have a design and implementation flaw which makes the feature highly susceptible to attack. Although open-source tools have been written and released, no formal testing methodology has been developed. This research presents a proposed method for the testing of this vulnerability in a measured and systematic way.*

**Keywords:** Wireless LAN, Computer Security, Data Security

## 1 Introduction

Access to Local Area Networks (LAN) have traditionally been restricted to wired connections via coaxial cable, CAT 3, 4, 5 or 6 cables, and Unshielded Twisted Pair (UTP). However, in 1997 the Institute of Electrical and Electronics Engineers (IEEE) helped establish the 802.11 set of standards by which communications could be facilitated wirelessly [1]. Alongside the evolution of these standards came an increase in the adoption of this technology in both both consumer and enterprise grade wireless communication devices. Today, mobility has become a significant component of the high consumers demand for electronic devices such as mobile phones, notebook computers and tablets. Worldwide shipment of Wi-Fi integrated circuits increased 28% between 2008 and 2009, with wireless integrated circuits in mobile handsets alone increasing by 50% in 2009 [2]. Instat predict that more than a billion Wi-Fi chipsets will be shipped in 2012 alone, with Wi-Fi chipsets for mobile phones and notebook computers to exceed one billion dollars in 2015 [3].

Alongside this surge in usage has come an increased awareness of security and its associated issues. In an effort to ensure the security of the technology when used in consumer devices the Wi-Fi Alliance drafted the Wi-Fi Protected Setup (WPS) specification and certification in 2007 [4]. Three methods of using WPS were created:

*1. Push Button Configuration Method* - A physical or virtual button is pushed on both the wireless client that wants to join the network, and the wireless router or access point that will be the gateway into the network.

*2. Personal Identification Number (Internal Registrar)* - The PIN of the wireless client that wants to join the network is entered into a web interface of the wireless gateway. The PIN can be written on the wireless device, or may be generated in software.

*3. Personal Identification Number (External Registrar)* - The PIN of the wireless gateway that allows access to the wireless network is entered into an interface of the wireless client.

The PIN (External Registrar) method of authentication was found to be vulnerable to a brute force attack in late 2011 [5]. This vulnerability allows for an attacker to gain unauthorized access to a wireless network within a matter of minutes to days, no matter how strong the Wi-Fi Protected Access (WPA) passphrase is.

## 2 Background

### 2.1 Significance

The aim of this research was is produce a rigorous and comprehensive methodology and procedure that will allow for a wireless device to be tested for its susceptibility to the WPS external registrar PIN authentication design vulnerability. Currently there is no formal testing methodology that may be applied to a wireless device that will give a comprehensive and detailed view of its susceptibility to the vulnerability.

Once in place the method will allow for the systemized testing and evaluation of wireless consumer devices. To date only sporadic data from unnamed sources has been available. Whilst such data of an unknown nature claims to prove vulnerability for a given device, the methodology used is not disclosed or document. Therefore, the method proposed in this work may be used to verify such results ensuring both consistency and reliability in the gathered data.

The information obtained by this research may be useful to owners of wireless devices as a credible and reliable guide to the vulnerability of their devices. It may also be used to expose manufacturers that have yet to patch the vulnerability in their products. This is of particular concern due to the rise of wireless related attacks becoming a feature in modern criminal enterprise [6].

## 2.2 WPS & Vulnerability

Wi-Fi Protected Setup (WPS) is an optional certification from the Wi-Fi Alliance, a non-profit organization that promotes the adoption of 802.11 wireless devices. It has almost 500 members and has certified well over 9,000 products. The standard was introduced in 2007, and currently has over 2,000 certified devices [7]. The standard purports to allow for the setup of wireless devices to be easier for the average consumer, providing for wireless access without the need for a complex passphrase exchange.

Although WPS was an optional certification, the more recent Wi-Fi Direct certification ( has a mandatory requirement that WPS be included in any device that is to be certified [8]. Wi-Fi Direct is designed to allow devices to talk directly to each other, to replace situations where cables are traditionally used. This requirement means that any device that bears the Wi-Fi Direct logo will have WPS capabilities, and will likely have WPS enabled by default.

However, in 2011 a detailed a flaw in the design and implementation of WPS was discovered [5]. The flaw allows for the brute force of the WPS PINs used in Wi-Fi Alliance certified devices. The approach is based on flaws within authentication when using a PIN via an external registrar, and the timing of EAP-NACK messages that reduce the searchable key space of the attack from 108 to 104+104. This keys pace is further reduced as the 8th digit of the PIN is a checksum of the previous seven numbers. Thus, the effective key space is actually only 104+103.

Whilst initially claimed that the WPS vulnerability appears to be widespread a limited number of devices were included in testing [5]. Thus, whilst it is suspected that a significant amount of devices would include this vulnerability, it is difficult to ascertain from current literature the true scope of the problem. As many devices may allow for the disabling of the feature it has yet to be conclusively determined if this approach represents a true solution to the issue. Therefore the method proposed in this work would allow for a true quantification of the issue and the subsequent questions that arise.

## 2.3 Reported Mitigations

As mitigation, some claim that WPS is a secure channel by which to authenticate wireless devices with active brute force protection [9]. Although briefly mentioned, it is stated that the registrar will warn a user, and will not automatically reuse the PIN if a PIN authentication or communication error occur. Whilst it appears that some manufacturers have implemented a delay when an incorrect PIN is used, the length of this timeout is manufacturer and perhaps device or firmware specific.

Microsoft's implementation of WPS in their operating systems released after Windows XP is Windows Connect Now-NET [10]. The feature allows for the same in-band PIN authentication scheme that has been found to be vulnerable to a brute force attack. Microsoft's specification is very detailed and shows the steps that are taken by both the Enrolee and Registrar to authenticate via a PIN. Microsoft note that the "AP Setup Locked" attribute may be set at the access point, and that "The access point should enter this state if it believes a brute force attack is underway against the access point's PIN" [10]. It is further stated, "…the use of the access point's PIN for adding external registrars is disabled in this state" [10]. However, the strength of the implementation and the extent to which supposedly compliant manufacturers implement this timeout, and the duration of the timeout have yet to be investigated. Again a standardized approach would be integral to any research on this subject.

As with most known vulnerabilities, a United States Computer Emergency Readiness Team (US-CERT) Vulnerability Note was created when information of the vulnerability was disclosed [11]. Such alerts are accompanied by a recommendation to disable WPS as a workaround, however as mentioned previously, this may not guarantee the cessation of the attack vector. A second online vulnerability database entry was created with the reference CVE-2011-5053 [12][13]. No workarounds or recommendations are provided.

Recently, an effort to crowd source the detection of the vulnerability across devices and firmware versions has arisen online [14]. The list is fairly comprehensive, with 133 entries covering most router and wireless access point vendors. However, whilst the information is presented in a coherent and uniform way, the accuracy of the data cannot be verified. It must be noted that the information does seem to support the theory that the WPS PIN vulnerability is widespread.

Since the discovery of the WPS PIN vulnerability a number of open source tools have surfaced that allow for testing and exploitation such as Reaver and WPSCrack [15][16]. Thus in conjunction with these tools a standardized approach to testing the vulnerability would allow both individual and systematic audit of all devices giving clear quantification of the problem as well as certifiable testing of mitigation approaches. It is therefore proposed that once established, the methodology described in this research will be utilized to audit and report on the security of popular Wi-Fi devices.

# 3 Proposed Method

As wireless devices that are to be audited may either be delivered to the customers with any version of publicly (and privately) available firmware, it is important that as many versions as possible are tested. It is not enough to assume that if the vulnerability that is to be tested is patched in one version, that all subsequent versions will also not be vulnerable.

Flashing the device to its factory default is an important step, as it is in this state that the initial customer will receive it. It also negates the chances that, if the device was not purchased new, the previous owner changed settings that would affect the results of an audit. Testing devices with both WPS enabled and disabled will ensure that the device manufacturer has not made an error, and that disabling the WPS feature in the configuration truly does disable the feature. This is important as it is logical for a consumer to assume they are not vulnerable if the vulnerable service is not seemingly enabled.

The wash tool was designed to identify wireless devices that have WPS enabled. Proving the effectiveness of this tool in identifying devices that have WPS enabled may help reduce time spent running Reaver against devices that do not have WPS enabled [15]. Reaver has been in development for over a year, and was publicly released in December 2011. The tool is designed to audit wireless devices for the WPS brute force vulnerability. Reaver may either fail to probe a device (either due to the device not having WPS enabled or having other protection mechanisms enabled), succeed but be rate limited (due to the device implementing brute force protection mechanisms), or succeed with little to no impedance.

The proposed method is illustrated below. The method proposes a systematic approach to the testing of any Wi-Fi device allowing for consistency and repeatability. It is envisaged that the implementation of this method will produce a significant volume of reputable data on the WPS vulnerability issue. To this, a study is now underway to verify and utilize the approach against a body of commercial devices.

# 4 Conclusion

The WPS external registrar PIN authentication design vulnerability is a dangerous security hole for home and SOHO users of wireless devices. The public has been lead to believe that as long as their WPA/WPA2 passphrase is complicated enough, then their networks are safe from unauthorized access. Clearly this is no longer the case, but the scale of the vulnerability has yet to be fully examined.

The development of a reliable WPS external registrar PIN authentication design vulnerability testing methodology will allow for a standardized way to test for weak implementations of WPS by device manufacturers. It will allow for current and future devices to be tested, with reliable results generated from an audit.

The results found from applying the developed auditing methodology to wireless devices will not only allow for the detailed examination of data, but will allow members of the public to easily and reliably ensure the security of their own devices.

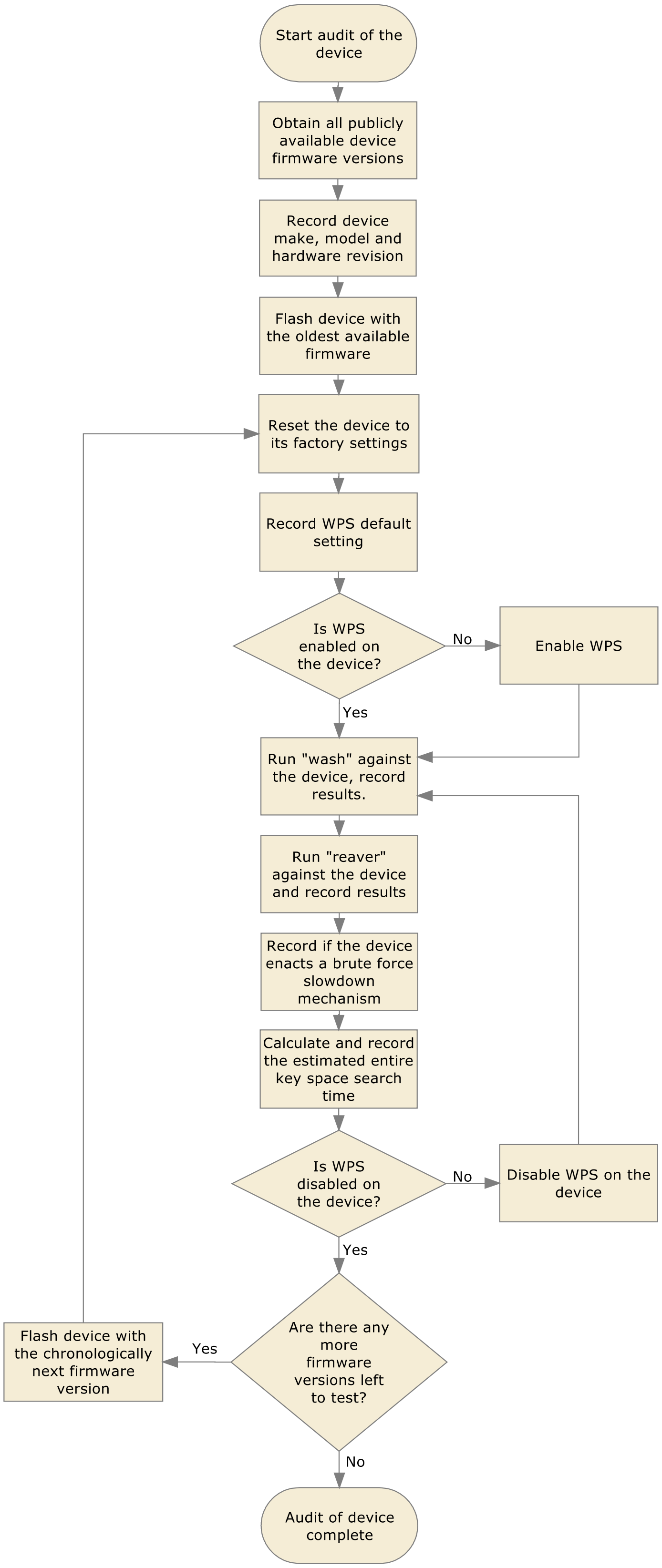

Start audit of the device
Obtain all publicly available device firmware versions
Record device make, model and hardware revision
Flash device with the oldest available firmware
Reset the device to its factory settings
Record WPS default setting
Is WPS enabled on the device?
No
Enable WPS
Yes
Run "wash" against the device, record results.
Run "reaver" against the device and record results
Record if the device enacts a brute force slowdown mechanism
Calculate and record the estimated entire key space search time
Is WPS disabled on the device?
No
Disable WPS on the device
Yes
Are there any more firmware versions left to test?
Yes
Flash device with the chronologically next firmware version
No
Audit of device complete

Figure 1 – The Proposed Testing Methodology

# 5 References

[1] The Economist. (2004). *A brief history of Wi-Fi*. [Online]. Viewed 2012 April 01. Available: http://www.economist.com/node/2724397

[2] Electronics News. (2010, January). *Wi-Fi IC shipments up 28 per cent*. [Online]. Viewed 2012 April 01. Available: http://www.electronicsnews.com.au/news/wi-fi-ic-shipments-up-28-per-cent

[3] J. Happich. (2010, September). *WiFi chipets to pass a billion units per year by 2012*. [Online]. Viewed 2012 April 01. Available: http://www.microwave-eetimes.com/en/wifi-chipsets-to-pass-a-billion-units-per-year-by-2012.html?cmp_id=7&news_id=222901091

[4] Wi-Fi Alliance. (2010, December). *Wi-Fi CERTIFIED Wi-Fi Protected Setup*. [Online]. Viewed 2012 April 01. Available: https://www.wi-fi.org/register.php?file=wp_20101216_Wi-Fi_Protected_Setup.pdf

[5] S. Viehböck. (2011, December). *Brute forcing Wi-Fi Protected Setup*. [Online]. Viewed 2012 March 25. Available: https://sviehb.files.wordpress.com/2011/12/viehboeck_wps.pdf

[6] myPolice QPS News. (2012, March). *War Driving Project to help prevent identity theft*. [Online]. Viewed 2012 April 01. Available: http://qpsmedia.govspace.gov.au/2012/03/22/war-driving-project-to-help-prevent-identity-theft/

[7] Wi-Fi Alliance. (2011, April). *Wi-Fi Alliance Member Symposium*. [Online]. Viewed 2012 March 25. Available: http://www.wi-fi.org/files/20110421_China_Symposia_full_merge.pdf

[8] Wi-Fi Alliance. (2010, October). *Wi-Fi CERTIFIED Wi-Fi Direct*. [Online]. Viewed 2012 March 25. Available: http://www.cnetworksolution.com/uploads/wp_Wi-Fi_Direct_20101025_Industry.pdf

[9] N. Turab & F. Moldoveanu. (2009). A Comparison Between Wireless LAN Security Protocols. *Universitatea Politehnica Bucuresti Scientific Bulletin*. [Online]. 71 (1), pp 61-80. Available: http://www.scientificbulletin.upb.ro/rev_docs/arhiva/full7970.pdf

[10] Microsoft. (2006, December). *Windows Connect Now–NET*. [Online]. Viewed 2012 March 25. Available: http://download.microsoft.com/download/a/f/7/af7777e5-7dcd-4800-8a0a-b18336565f5b/WCNNetspec.doc

[11] J. Allar. (2011, December). *Vulnerability Note VU#723755*. [Online]. Viewed 2012 March 25. Available: http://www.kb.cert.org/vuls/id/723755

[12] Common Vulnerability and Exposures. (2012, January). *CVE-2011-5053*. [Online]. Viewed 2012 March 25. Available: https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2011-5053

[13] National Vulnerability Database. (2012, January). *Vulnerability Summary for CVE-2011-5053*. [Online]. Viewed 2012 March 25. Available: http://web.nvd.nist.gov/view/vuln/detail?vulnId=CVE-2011-5053

[14] Jagermo. (2012). *WPS Flaw Vulnerable Devices*. [Online]. Viewed 2012 March 25. Available: https://docs.google.com/spreadsheet/lv?key=0Ags-JmeLMFP2dFp2dkhJZGIxTTFkdFpEUDNSSHZEN3c

[15] C. Heffner & P. Eacmen. (2012). *reaver-wps*. [Online]. Viewed 2012 March 25. Available: https://code.google.com/p/reaver-wps/

[16] S. Viehböck. (2011). *WPScrack*. [Online]. Viewed 2012 March 25. Available: http://dl.dropbox.com/u/22108808/wpscrack.zip

[17] C. Heffner. (2012, January). *README.WASH*. [Online]. Viewed 2012 March 25. Available: https://code.google.com/p/reaver-wps/source/browse/trunk/docs/README.WASH

# Utilizing the RFID LOCK Command Against Multiple Targets

**C. Bolan**
School of Computer and Security Science, Edith Cowan University, Perth, Western Australia
secau – Security Research Centre , Perth, Western Australia

**Abstract -** *An unlocked Electronic Product Code (EPC) tag allows for issuance of most commands without the need for any authorization. This means that a system with unlocked tags would allow any attacker to modify tag data at will, whilst also opening the door to a range of other misuse. One possible avenue of active misuse against unlocked tags would be to issue LockID commands and 'permanently' lock some or all of a system's RFID tags. As this attack is simply an issuance of a valid command it fits firmly in the category of an active misuse and could also be considered a limited form of DoS as future valid commands would be ignored and limit or cripple the functionality of a system dependent on operation. This paper details an experiment using the LockID command to lock multiple tags within range.*

**Keywords:** Radiofrequency Identification, RFID Tags, Information Security

## 1 Introduction

Radio Frequency Identification (RFID) relies on transponders which are incorporated into an object for the purpose of identification or tracking [1]. The transponder (or tag) may be used to store information and will respond to signals sent by a transceiver (RFID reader) [2]. Increasingly such technology is being incorporated into supply chain management systems throughout the world and is expected to eventually replace traditional bar-coding systems [3].

"*The Electronic Product Code is an identification scheme for universally identifying physical objects via Radio Frequency Identification tags and other means*" [4]. The electronic product code (EPC) standards were created by EPCglobal as an open, community based approach to promote the use of RFID technology in supply chain management., while not explicitly focused on security, the standards purport to promote a secure environment for RFID use and protect both individual and organizational privacy.

Whilst EPC tags were primarily designed for write once / read many time applications they are able to be used in a variety of means across their four states of operation (un-programmed, programmed, locked and killed). These states dictate the behaviour of the RFID Tag when a given command is issued. The focus of this research was to investigate the use of the lock state and its related LockID command and builds upon previous work into directed LockID attacks.

## 2 The LockID Command

According to the EPC standards [5], the LockID command precludes further modification of values contained on an RFID Tag. The command based upon a more specific version of the ProgramID command whereby the [PTR] value points to the most significant bit of the password location and the [Value] must be equal to 0xA5 (hex value A5).

Given this command, the locking of an RFID tag may be achieved through the following steps:

1. Program the KILL code and leave the lock code at 00h;
2. Verify the EPC code by issuing a ScrollallID or VerifyID command;
3. Lock the tag by programming A5h to the Lock location;
4. Check that the tag is locked by issuing a VerifyID command. N

Note: If the tag is locked, the reader will receive no response to this command.

Accordingly, once the tag has been locked it will no longer respond to any programming commands, including the verify command. This suggests that, as the tag does not respond to the programming command, the lock code cannot be removed making it permanently locked. Thus it has been suggested that the only way to modify the tag at all is to utilize the kill command with the programmed password, which will render the tag inactive 'forever' [6]. Subsequent research has demonstrated that resurrection after a tag has been killed is possible – which has the duel effect of resetting the lock but at a significant time cost for any significant tag volume [7].

## 3 The Attack

To date a range of attacks have been developed against systems utilizing this standard, but the LockID based attack differs as it requires no password cracking or additional equipment. Rather, the purpose behind this attack is to utilize the existing controls of the standard to circumvent the systems normal functionality.

The single lock attack is based on the principle of an attacker selecting a single tag and locking that tag. At its base level this attack is no different to a legitimate user locking a single tag in any valid application. To test the validity of this attack a standard tag / reader setup was created in the Faraday cage as illustrated in figure 1.

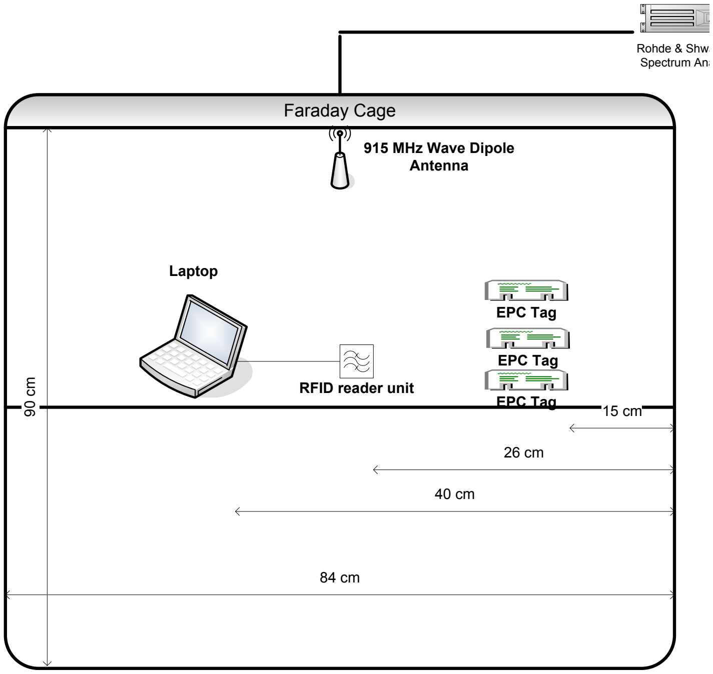


Figure 1 – Experimental Setup

The experimental setup included the use of three EPC RFID tags at a single time; this setup meant that a single tag from the selection may be targeted and locked and the other two may be tested to see if they remain unaltered, showing that a targeted attack against a single tag is viable. As there were three positions that could be occupied by the tags, it was decided that the position of the tag to be locked would be rotated amongst the three positions with each group.

Previous research on this targeted attack showed that the attack was highly effective, In essence the researcher was able to target a specific tag and lock the tag at will. This is demonstrated in figure 2. This new extension of the research was intended to demonstrate that such an attack would be highly scalable. Whilst a large variety of tag numbers were successfully attempted the discussion of results will be limited to three tag setups for the sake of visual clarity.

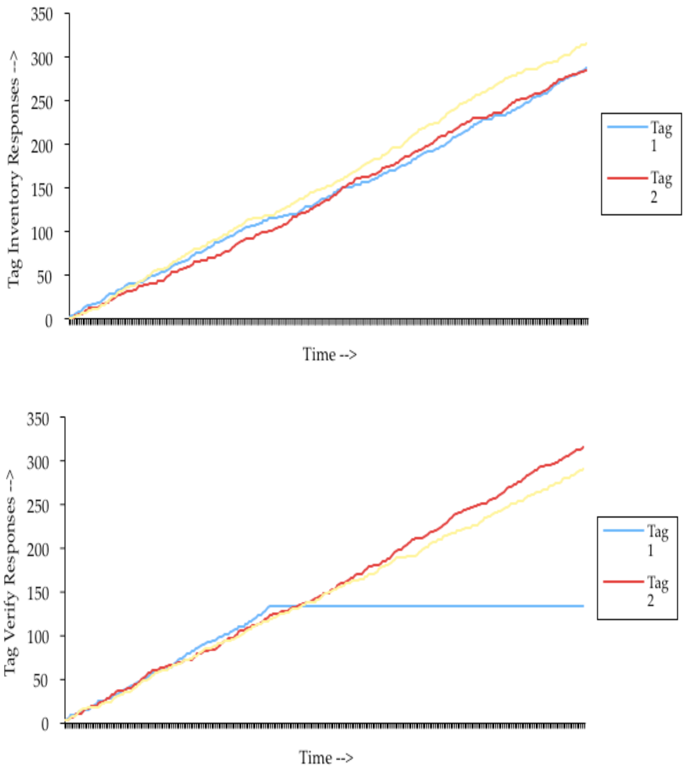


Figure 2 – Single Lock Results

Through experimentation it was found that their were two feasible methods of issuing the attack. The first would be to sequentially issue lock commands for every possible tag id within the tag space. To determine the validity of this attack a few simple calculations were conducted as detailed below:

Tag Identifier Space = 96 bits = $2^{96}$ =
79228162514264337593543950336

*Assuming 100 Lock operations a second:*

Complete Lock Attack Time = $2^{96}$ / 100 =
220078229206289826648734 hours
= 9169926216928742777031 days
= 25123085525832171992 years

Clearly such an attack method would be infeasible, even if a reader could somehow be increased tenfold to allow 1000 lock operations per second, the attack would still take far beyond a single lifetime. With this established, the second more feasible approach was considered whereby the reader would first conduct an inventory of the tags in the area of the attacking machine and then use the list of detected tags to issue focused lock commands.

In figure 3 below, three tags are setup in the cage with an attacking transceiver set to commence operation at a specified interval. From the figure it is clear to see that the attacks occur almost in parallel. Such results were paralleled with every size of victim sample trialed in the study.

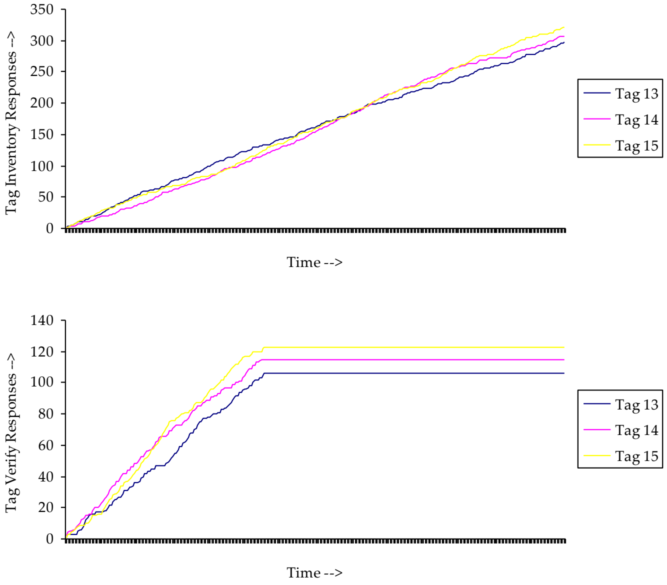


Figure 3 – LockID against multiple tags

Whilst either approach would constitute an active attack (i.e. one that requires direct interaction), it would be possible to integrate the collection method with an eavesdropping attack. Whilst such an attack may take longer to directly target all transponders that were contacted by legitimate users. Such attack blending could limit the detectability of the attack and reduce the likelihood of countermeasures being successfully created and deployed.

# 4 Conclusion

The paper presented a small but significant extension to the previously documented single lock attack. The simplicity of this approach is that like its directed counterpart the attack requires nothing beyond the standard equipment. The single LockID research demonstrated how the command may be targeted to an individual tag without altering the standard functionality of the RFID reader. Similarly the multiple vector attack was shown to work without the modification of the attacking transceiver.

Through the evidence of the multiple attacks efficacy and the two feasible methods of target gathering it is clear that such a method may be employed to attack complaint RFID systems. An example of this attack would be a Supermarket whereby the attacker could lock all tags in the store preventing prices (stored on the tag) from being altered. In this scenario, every affected transponder would need to be killed and resurrected to return to normal operation. Even using fairly conservative figures there would likely be a significant cost in time and lost revenue.

In the standards used for this experiment, the victim would likely find it difficult if not impossible to defend or detect the attack in time to make a difference. Whilst the EPC standard is rapidly evolving many existing setups may be the target of similar attacks though their currently seems to be a lack of evidence for such attacks taking place within the wider community. This may either be due to a lack of motive in the attackers or the rarity of such setups.

# 5 References

[1] Y. Zhang and P. Kitsos, Security in RFID and Sensor Networks. Boca Raton: Auerbach Publications, 2009.

[2] D. Hunt, A. Puglia, and M. Puglia, RFID: A Guide to Radio Frequency Identification. Hoboken, New Jersey: John-Wiley & Sons, 2007.

[3] A. Juels, ""Yoking-proofs" for RFID tags," in International Workshop on Pervasive Computing and Communication Security - PerSec 2004, R. Sandu and T. Roshan, Eds., ed Orlando, Florida, USA: IEEE Computer Society, 2004, pp. 138-143.

[4] EPCglobal, "EPC Generation One Tag Data Standards," EPCglobal 1.1 Rev 1.27, 2005.

[5] EPCglobal, "EPC Radio-Frequency Identity Protocols Class-1 Generation-2 UHF RFID Protocol for Communications at 860 MHz - 960MHz," EPCglobal 1.0.9, 2005.

[6] M. Rieback, "RFID Security and Privacy," PhD, Vrije Universiteit, Amsterdam, 2008.

[7] C. Bolan, "The Lazerus Effect: Ressurecting Killed RFID Tags," in 4th Australian Information Security and Management Conference, Perth, Western Australia, 2006.

# Trusted Discovery & Policy Negotiation for Smart Phones

**J. Porekar**[1] **and M. Vardjan**[1]
[1]Research Department, SETCCE, Ljubljana, Slovenia

**Abstract** - *In this paper we present a framework for boosting trust in discovery of business people, services and pervasive objects for emerging mobile based dynamic business environments. Our approach is based on (i) using authentic discovery URIs and verification methods that can be achieved using digital signatures, followed by (ii) negotiation of micro-policies describing data practices that services promise to deliver. Negotiations are realized through (iii) user-agent assisted evaluation of suitability of micro-policies.*

**Keywords:** Secure discovery; policy negotiation; QR codes.

## 1 Introduction

The business landscape is changing and today business people often use their smart phones in both personal and business contexts. Often the devices are used to discover and initiate interaction with other business people, network connected physical objects, location specific services and social business communities (see [1]). These micro-interactions (see [2]) are often performed in an ad-hoc manner: either using NFC tags, QR codes or other similar technologies that use URIs to redirect user to particular service. The outcome of such micro-interactions is often that certain apps get installed to the smart phone of the user and later in the process these apps are consequently granted rights to certain part of the user's subsystem. This may pose a threat to both the business user and the organization he is working for as it may lead to personal and corporate information leaks.

## 2 Trust Related Requirements

**Authentic discovery URIs**: URIs stored in discovery mechanisms (such as NFC tags or QR codes) need to be authentic. Additionally it needs to be clear which party has produced the discovery URI and which party is behind redirecting the user to particular service.

**Transparency of service provider**: When the user is redirected to the location based service, physical object, business community or a business user acting on behalf of another company (from here on all these will be related to as service) it needs to be clear which entity is behind the service. This information needs to be authentic.

**Authentic transparent policies and agreements**: the party providing a service needs to transparently present the practices to the user. The information on how the service will handle user's data is negotiated and at the end of the negotiation process both parties should be holding the same signed agreement.

**Usability**: Although the process of matching service policies to user preferences may be complex it is necessary for software agent on smart phone to automate it to minimize required interaction with the user.

## 3 Boosting Trust in Service Initiation

### 3.1 Authentic Discovery of physical objects services

As a solution for the requirements described we propose a framework that allows for i) authentic discovery and for ii.) trusted authentic service initiation. The authentic discovery mechanisms relate to digitally signed redirection URIs, which are deployed through discovery mechanisms, such as NFC tags, QR codes and similar. Conceptually the solution is simple: URI, its digital signature and information about party that signed the URI (such as certificate containing public key) is embedded into the payload space. However, due to the limited storage space available on the tags and QR codes, the actual implementation is not trivial to achieve (see Fig. 1 for details). Other similar approaches that relate to authenticity of QR codes and NFC technology are found in [3]

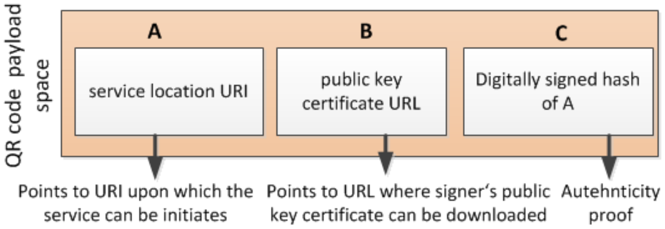

QR code payload space
A
B
C
service location URI
public key certificate URL
Digitally signed hash of A
Points to URI upon which the service can be initiates
Points to URL where signer's public key certificate can be downloaded
Autehnticity proof

Figure 1: Secure QR code structure

### 3.2 Policy Negotiation or Policy Selection

Formal policies are used to legally define relationship between the parties and to specify conditions under which a service or a community membership is provided by one party (the provider) and consumed by the other party (the requester). This is especially important in dynamic environments where the provider's policy may have internal rules that are incompatible with requester's company's policies ([4], [5]).

To minimize number of network transfers, a 3-step policy selection process is developed as shown in Fig. 2. Each step is secured using digital signatures. In first step, the requester fetches provider's XML document with multiple alternatives for the policy. Because the policy options are signed by a verifiable identity, the requester can trust the offer is real and the provider will not only collect the requests, possibly associate them with requester's data like his identity, and then not even provide the advertised service or membership. X.509 [6] certificates and XML-DSig [7] are used in the prototype.

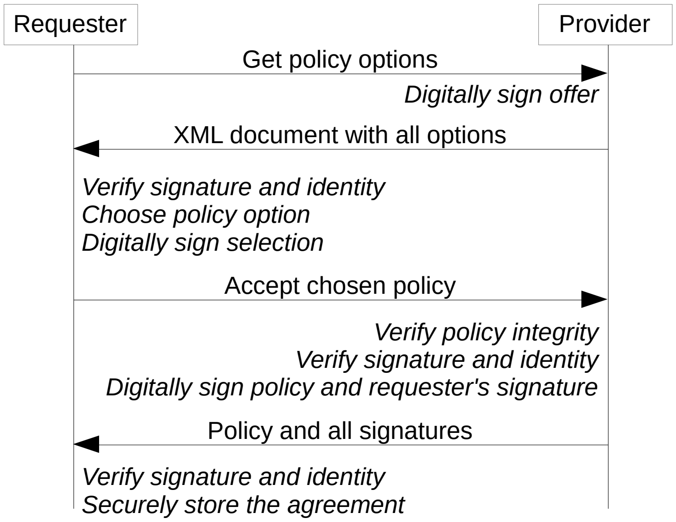
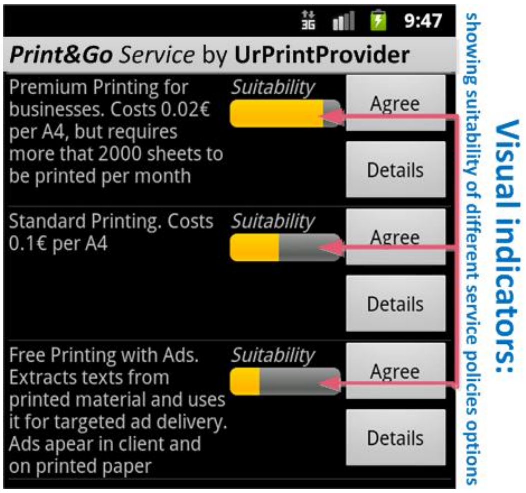

Requester
Provider
Get policy options
Digitally sign offer
XML document with all options
Verify signature and identity
Choose policy option
Digitally sign selection
Accept chosen policy
Verify policy integrity
Verify signature and identity
Digitally sign policy and requester's signature
Policy and all signatures
Verify signature and identity
Securely store the agreement

Figure 2: Simplified negotiation sequence diagram

In step two, the requester a.) locally verifies digital identity and cryptographic signature, b.) chooses a suitable policy and c.) signs it. If Transport Layer Security (TLS) is used, explicit verification of provider identity is not necessary. The requester then sends the provider his signed choice. XML-DSig [7] was chosen to add requester's signature into XML-based policy. In last step, the provider locally verifies policy consistency and requester's signature. The provider signs the requester's signature of the policy and appends it to the final XML-based policy which is then sent to requester. This last signature proves that the chosen policy and requester's acceptance of the policy were successfully received by the provider. When requester holds this proof, the provider can not dispute the policy validity on grounds that the requester did not sign it or did not send it back. Finally, the requester verifies all signatures and verifies the policy content has not changed from the previous step. If at any step a fraud is suspected by either party, the process terminates. The negotiation process is described in more details in [8].



Print&Go Service by UrPrintProvider

Premium Printing for businesses. Costs 0.02€ per A4, but requires more that 2000 sheets to be printed per month

Standard Printing. Costs 0.1€ per A4

Free Printing with Ads. Extracts texts from printed material and uses it for targeted ad delivery. Ads apear in client and on printed paper

Figure 3: An example of User Agent in action: negotiating a policy with NFC discovered "Print&Go" Service

### 3.3 User Agent

The mobile phone user agent is intended to be a semi-automatic helper tool. It evaluates policies and matches them to preferences of the user (see [9]). The user agent provides assisted service initiation / negotiation dialog that contains visual indicators determining suitability of different options under which the service can be provided. (see Fig. 3)

## 4 Conclusions

A secure discovery and trustworthy policy negotiation framework and its user agent were presented. The negotiation process and its implementation are simplified to be used on smart phones and the user selects the policy in a single step. This approach still addresses the needs of most providers in the real world. By using digital certificates and signatures, it assures policy consistency during the negotiation process.

## 5 References

[1] S. Gallacher, E. Papadopoulou, N. K. Taylor, I. Roussaki, N. Kalatzis, N. Liampotis, F. R. Blackmun, M. H. Williams, and D. Zhang, "Personalisation in a system combining pervasiveness and social networking," in 2011 Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN). IEEE, Jul. 2011, pp. 1–6. DOI.1109/ICCCN.2011.6005900

[2] Ben Dodson, Hristo Bojinov and Monica S. Lam, "Touch and Run with Near Field Communication (NFC)", http://mobisocial.stanford.edu/papers/nfc.pdf, accessed 2012-04-19.

[3] Warasart M., Kuacharoen P, Paper-based Document Authentication using Digital Signature and QR Code, 2012 4th International Conference on Computer Engineering and Technology (ICCET 2012) IPCSIT vol.40 (2012) © (2012) IACSIT Press, Singapore

[4] J. Porekar, K. Dolinar, A. Jerman-Blažič, and T. Klobučar, Pervasive Systems: Enhancing Trust Negotiation with Privacy Support. Boston, MA: Springer US, 2007, ch. 2, pp. 23–38; DOI 10.1007/978-0-387-71058-7_2

[5] K. E. Seamons, M. Winslett, T. Yu. Limiting the Disclosure of Access Control Policies during Automated Trust Negotiation, Proc. symposium on network and distributed systems security, NDSS, 2001.

[6] X.509 standard recommendation, http://www.itu.int/rec/T-REC-X.509/en, accessed 2012-04-19.

[7] XML-DSig, XML Signature Syntax and Processing, 2nd Edition http://www.w3.org/TR/xmldsig-core/, accessed 2012-04-19.

[8] M. Vardjan, M. Pavleski, J. Porekar, "Securing Policy Negotiation for Socio-Pervasive Business Microinteractions" in 2012 Proceedings of The Sixth International Conference on Emerging Security Information, Systems and Technologies, SECURWARE Aug, 2012, Rome, Italy

[9] P. Bonatti, S. De Capitani di Vimercati, and P. Samarati, "An Algebra for Composing Access Control Policies," in ACM Transactions on Information and System Security (TISSEC), vol. 5, no. 1, pp. 1-35, February, 2002.