

**SESSION**  
**SECURITY EDUCATION**

**Chair(s)**

**Dr. Gregory Vert**

**Texas A and M Univ. - USA**

**Dr. Syed Rizvi**

**Pennsylvania State Univ. - Altoona - USA**



# Cybersecurity Awareness in Organizations: *a case study of University of Venda*

Isong Bassey<sup>1</sup>, Murendeni Randiman<sup>2</sup>, Kudakwashe Madzima<sup>3</sup>

*Department of Computer Science & Inform. Systems, University of Venda  
Thohoyandou, South Africa*

{<sup>1</sup> bassey.isong, <sup>3</sup> kudakwashe.madzima}@univen.ac.za

<sup>2</sup> mrandima2010@gmail.com

**Abstract**— The swift growth of global interconnectivity, especially the Internet has been valuable and has created positive impacts in today's e-world. This has been witnessed in all spheres of lives. In particular, organizations today are using Internet alongside computers, software, social networks, phones and emails to share data and access information to which University of Venda (UNIVEN) is not an exception. The Internet today provides a common platform through which anyone can virtually take part in globalization. While these innumerable benefits are essential, this well-known interconnectivity poses a myriad of significant security risks and challenges. This has made cybersecurity one of the most critical concerns of the information age in several organizations today. Though, organizations have invested much in security measures to protect their information, employee and computers, the situation is seen skyrocketing worldwide, instead of declining. This shows that being secure is not only a function of advanced security technologies or tools but rather people's knowledge about security within an organization. In this case, we believe investigating the state of cyber security awareness at UNIVEN would assist us to uncover most of its security vulnerabilities in an effort to strengthen and improve its cyber security strategies. To achieve this, we collected data from students and staff using questionnaires. The participants in the survey were randomly selected. Data collected was analysed and presented quantitatively. The results obtained were promising in terms of cybersecurity awareness.

**Keywords** – Awareness, Cybersecurity, Cyberspace, Threats, Attacks

## I. INTRODUCTION

The swift growth in computer connectivity has transformed the way information is shared, the way people, government and nations communicate and conduct business all over the world [1][2]. Today, organizations are using computers, internet, software, social networks, phones, emails, and so on to share data and access information. For instance, data [3] from Internet World Stats in early 2011 shows the number of Internet users has amplified by 445% over the past 10 years for a global penetration level of 29%. In addition, more countries around the world are also working round the clock to get their unconnected citizens online. Accordingly, a recent survey by [4] indicate that telecommunication is one of the fastest growing sectors of the economy in the South African perspective. This is driven by the explosive growth in mobile

telephony and broadband connectivity, with a network that is 99.9% digital and having the most developed telecoms network in Africa. However, while the benefits have been huge, this prevalent interconnectivity also poses substantial risks to people, governments, nations and business's computer-dependent operations [1] [2]. In this case, the more online we are, the more vulnerable we become to cyber threats [2]. Such a massive increase in telecommunication and Internet access increases vulnerabilities and attacks such as phishing, scams, data theft, spyware, viruses and many attacks found in cyberspace. Therefore, strategies have to be put in place in order to raise awareness of cybersecurity in the public sector and organizations as a means to guard against cyber threats.

Today, cybersecurity is one of the most critical concerns of the information age [2]. It is defined as measures to protect information technology (IT), the information it contains, processes, and transmits, and associated physical and virtual elements [2]. All these together comprise what is called cyberspace [2]. Cybersecurity forms the foundation of a connected world and is geared towards ensuring confidentiality, integrity, and availability [1][5]. Thus, great efforts are required to ensure an acceptable level of cybersecurity for either an organization or nation. In the organizational perspective, much has been invested on cybersecurity policies and measures ranging from firewalls, antivirus, intrusion detection systems and other security technologies. These security tools, policies and techniques have proven their applicability in real-world scenarios, effectively mitigating threats as they appear, but organizations still suffer severe cyber threats and the problems are getting worse day by day [6]. The nature of these problems to a larger extent suggests that being secure is not only a function of having all the latest and advanced security technologies or software in place but rather having people with the right knowledge about security and its importance. That is, people who understand the risks of using the Internet, the importance of securing their personal information or information system and the consequences if not done properly. This goes with training people to be cyber security conscious. As we know, cyberspace has no architecture and it only exists virtually making it intangible in nature [2]. In this case, cyber criminals and cyber terrorists can launch attacks on their victims

(individual, organizations and nations) from anywhere as long as they are connected to the Internet and cyberspace [7]. This shows that people in an organization have a key role to play in an effective cybersecurity strategy because they are the ones who are using the internet, networks and other digital devices for the functioning of their organizations.

Securing an organization against cyber-attacks is the highest priorities for which each organization must ensure. To achieve this objective, it is imperative that people found within the organization must vigorously defend themselves and the organization's assets against a variety of internal and external threats by way of having the basic knowledge of cyber security issues as well as measures to avert cyber threats. This requires putting a cybersecurity policy in place and measuring the level of awareness on a regular basis. In the context of this paper, we focus on level of cybersecurity awareness within the University of Venda (UNIVEN) community. In this case, the research questions we want to answer are:

- 1) *How responsive are the members of the University of Venda community to current cyber security threats?*
- 2) *How does the different faculty in the university affect the way members react to cyber threats?*

The goal is to measure their level of cybersecurity awareness and propose approaches that are proven to reduce the existing cyber threats and recommend the best and appropriate measures to enforce security of information systems. We believe by so doing, we will be able to understand better the current security trends and ways of improving them. In this case, we performed a questionnaire survey on members of the university community and the data collected was analyzed and quantitatively presented. The results showed a promising level of cybersecurity awareness. In addition, we proposed measures to improve the current cybersecurity situation in UNIVEN based on the results obtained.

The rest of the paper is structured as follows: Section II is the importance of cybersecurity in UNIVEN, Section III is the methodology used in the study, and IV is the analysis and results. Accordingly, Section V is the discussions while VI is paper conclusion.

## II. WHY CYBERSECURITY AWARENESS IN UNIVEN?

Today's world is increasingly dependent on light-speed communications and ubiquitous networks, and is likely to become ever more reliant on cyberspace as the information society emerges [2]. The exponential use of cyber space throughout the world coupled with globalization has increased the complexity of cyber threats. This is as a result of the exponential growth of the internet [1]. The Internet offers a common platform through which anyone can virtually take part in globalization with ease. However, as stated by [8], "...though the internet may open our minds to new possibilities, it also exposes us to the pitfalls and dangers of cyber threats" such as the vulnerability of computer systems including Internet websites, against unauthorized access or attack, or the policy measures taken to protect them [1][2].

Therefore, it is essential to establish capabilities to thwart these cyber threats and attacks. This should be the most important priority of every organization today. However, to effectively manage against cyber threats in an organization one must first fully understand them and be involved throughout [9][10]. This means assessing people with regards to their awareness and reaction on cyber threats and the organization must be knowledgeable about the current threats in existence and the ways in which to help protect against these threats. We considered this literature very useful in answering the research questions as stated above. At UNIVEN, as an institution of higher education in South Africa, many aspects or activities rely on the Internet and computers, including communication (email and cell phone). This means that, UNIVEN relies on global interconnectivity and as such, it is not isolated from cyber threats and attacks. Despite several security measures in place both logical and physical, it is essential that members of UNIVEN community have to play an important role in the protection of their personal information and that of the university. However, the success of these measures is dependent on their cyber security consciousness. In this case, this paper assesses the current cybersecurity awareness of the UNIVEN community members. Their reaction in this regard will play an important part in determining whether improvement on the current cybersecurity situation is needed or not. This will further assist in deciding which security measure is appropriate to improve their awareness.

## III. RESEARCH DESIGN AND METHODOLOGY

The methodology used in performing our study is discussed as follows:

### A. Participants

Participants of this study are the students and staff members (both academic and non-academic) of the university. The research was conducted using a sample of 160 participants drawn at random from the entire UNIVEN population. Members were subdivided into four groups A, B, C and D. This was done to assist us to answer research question 2. The groups were thus, structured as follows:

- Group A consisted of non-academic staff and lecturers from IT department (10 members)
- Group B consisted of non-academic staff and lecturers from non-IT department (50 members)
- Group C consisted of students who are studying towards IT related degree (40 members)
- Group D consisted of students who are studying non-IT degrees, (60 members)

### B. Data Collection Method

An evaluation questionnaire was designed and used in collecting data from the participants. In order to respect the confidentiality of participants, a closed-ended questionnaire was used and was considered appropriate for the study, albeit we could get more valuable information using personal

interviews only. The questionnaire was subject to strict evaluation by an independent expert to ascertain the suitability and reliability of the questions. The validity was checked by testing the questionnaire with eight participants, two from each group. The validated questionnaires were personally distributed by to the UNIVEN community members to complete at random. The questions were based on their *online activities, Internet safety and security perception, Internet privacy issues, leaving computer unlocked, cyber security consciousness, issues of anti-virus usage, Issues of virus infection, knowledge of fake and legitimate website, Trust on Social Networks* and a host of others.

#### IV. ANALYSIS AND RESULTS

This section presents quantitative analysis of the results obtained in this study as structured in the questionnaire. However, due to space constraint, only few results analysis will be presented as follows:

##### A. Participants' Online Activities

This subsection performed the elicitation of the type of activities members of UNIVEN community are engaged in while on the Internet. As presented in Figure 1, analysis indicates that members of UNIVEN are involved in several activities while online. Accordingly, a majority of the participants used social networking sites, used Internet to do work related things, educational research, checking email, and others.

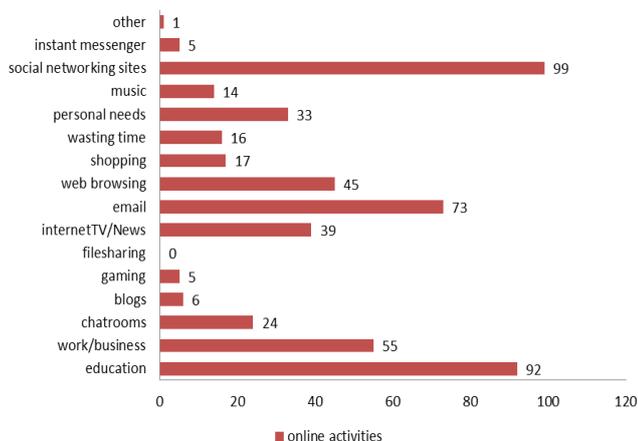


Fig. 1: Participants' online activities

##### B. Internet Safety and Security Perception

In this section, participants were asked to indicate their perception towards Internet safety and security. Their perception could help determine whether they are actually aware of cyber security or not. In this case, analysis revealed that the participants from different groups seem to have different perceptions (See Figure 2). The result indicates that group A and B participants responded "neither not safe or safe" (70%) whereas group C and D mostly responded "somewhat not safe" (35%). Accordingly, it was found that

only participants in B and D indicated the Internet is not safe at all (15%). This finding shows that members of UNIVEN community are somewhat cyber security aware, but perceived safety and security of the Internet differently.

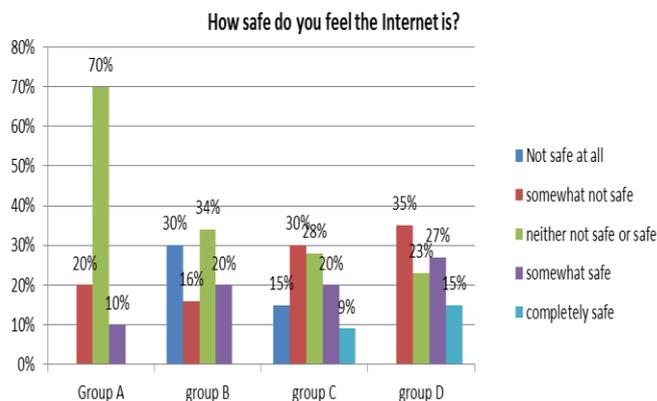


Fig. 2: Participants feelings about safety on the Internet

##### C. Internet Privacy

Here, participants were asked to indicate their concern about Internet privacy. The essence was to know whether they are aware or not of certain cyber threats such as social engineering, cyber criminals and others which can have access to their personal information. The results indicate that a majority of the participants from all the groups seems to be very concerned about the Internet privacy (100%). However, only a few participants from B, C and D were somewhat concerned about Internet privacy (36%). (See Figure 3)

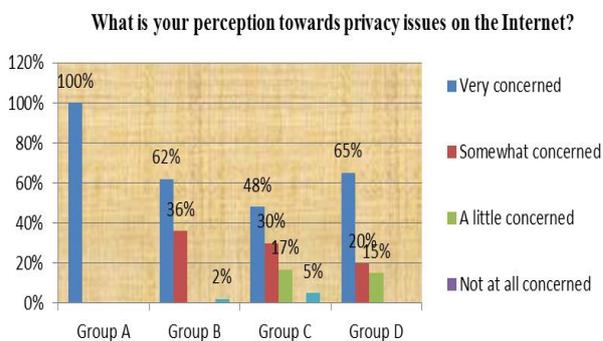


Fig.3: Perception of internet Privacy

##### D. Leaving Computer Unattended

In this section, participants were asked about their feelings when leaving their computers unattended. In this case, by assessing their feelings, we can understand their behaviour in terms of cyber threats awareness. The results analysis shows that many of the participants indicated that they don't feel free when leaving their computer unlocked (100%, "NO"). However, only few participants from C and D indicated that they don't have a problem leaving computer unattended (20%, "YES"). (See Figure 4) These results show that many

of the respondents are aware of the dangers of leaving computer unattended.

Do you feel free when leaving computer unattended?

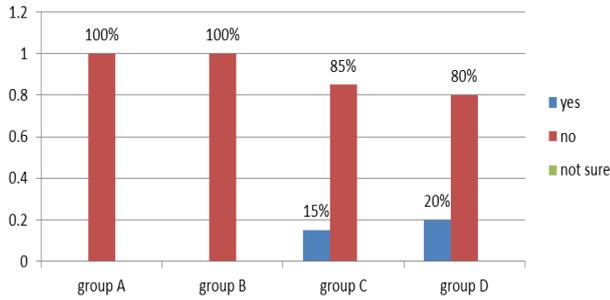


Fig. 4: Participants feeling when leaving computer unattended

E. Issues of Anti-virus Usage

In this section, we explored participant’s views about using computers or systems without an anti-virus installed. The basis was to know whether the participants view anti-virus software as important or not. The results show that several participants are aware of the importance of having an anti-virus on their computers, about 100% in A. Nevertheless, only a few participants indicated not sure whether there is a problem of using a computer without an anti-virus installed on it (14%) while few participants in C and D indicated there is no problem when using a computer without an anti-virus (11%) (See Figure 5). The results showed that most people in the institution are aware while only few are still not aware of the importance of anti-virus.

Do you foresee any problems using computer without an anti-virus?

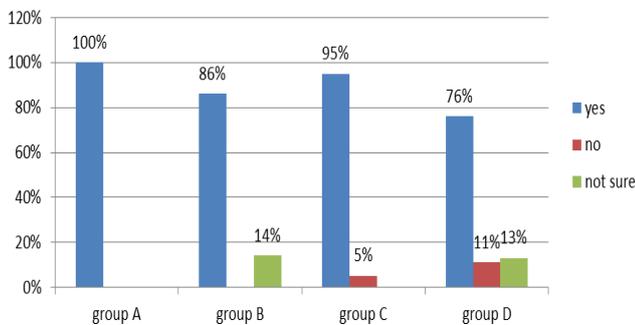


Fig. 5: problem of using computer without an anti-virus

F. Virus Infections Issues

We also went further and assessed participants’ awareness of virus infection in order to know if they have knowledge of being infected by viruses when browsing or downloading certain files from the Internet. This is captured in Figure 6a and 6b respectively. In Figure 6a, we present the results of participants’ level of awareness about getting computer virus from browsing certain websites. The analysis shows that many of the participants in all the groups are aware of virus

infection when browsing certain websites, about 100% in group A. However, only a few participants in groups B, C and D, about 34% indicated they are not aware at all of being infected by virus when browsing.

Do you foresee any problems using computer without an anti-virus?

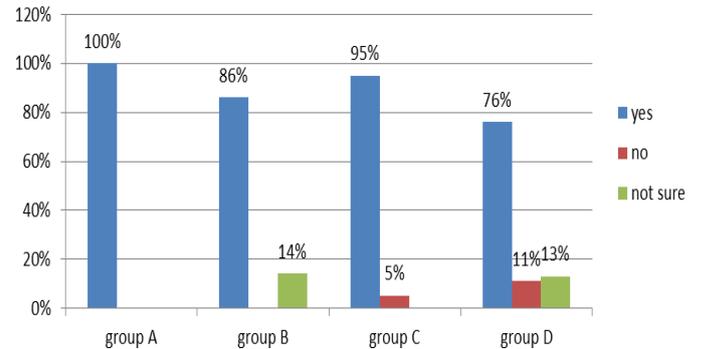


Fig. 6a: level of awareness about getting infected when browsing websites

In the same vein, Figure 6b shows the results of participants’ virus awareness when downloading email attachments. Analysis shows that participants in group A actually check for viruses when downloading files or email attachments (100%) while groups B, C and D participants have different views. In this case, some of them do check for virus while some are not even sure whether to check or not (54%). Accordingly, many of them don’t check at all (47%).

Do you think it is possible to get virus from browsing websites?

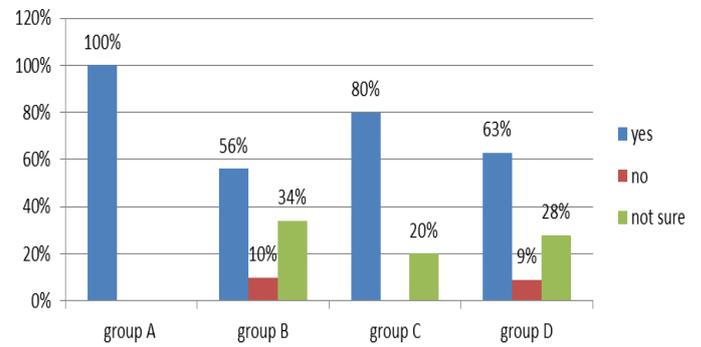


Fig. 6b: Participants’ virus awareness when downloading files or email attachments

G. Fake and Legitimate Websites Identification

In this section, participants were asked to indicate if they can identify which website is fake or legitimate. The basis was to know if they are aware of these websites in order not to fall prey to cyber criminals. From the analysis of the results, we found that about 100% of the participants in group A, 42% in B, 55% in C and 46% in D knew which websites are fake or legitimate. Accordingly, only 58% in B and others indicated they are not sure of how to identify such websites. (see Figure 7).

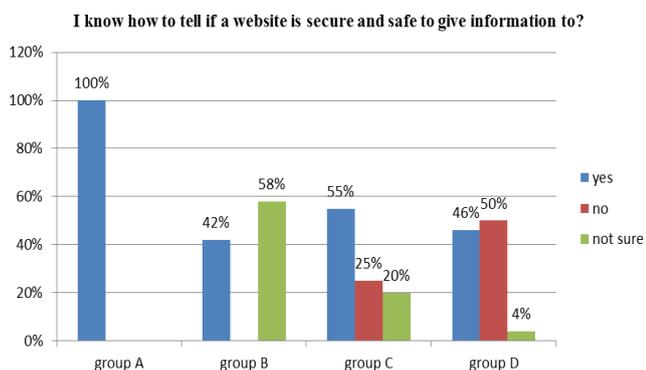


Fig. 7: Knowledge of fake and legitimate website

H. Trust on Social Network

This section deals with participants’ trust level when chatting on social networks. The aim was to know about their views of trust on the social media. Analysis shows that, irrespective of whom they chat to online, many of them (100%) don’t trust them in terms of disclosing their personal information. However, some participants in group D, about 64%, do trust unknown people (see Figure 8). These results show that majority of UNIVEN members are actually security conscious when online when it comes to trust, though few do trust the people they chat with on social media.

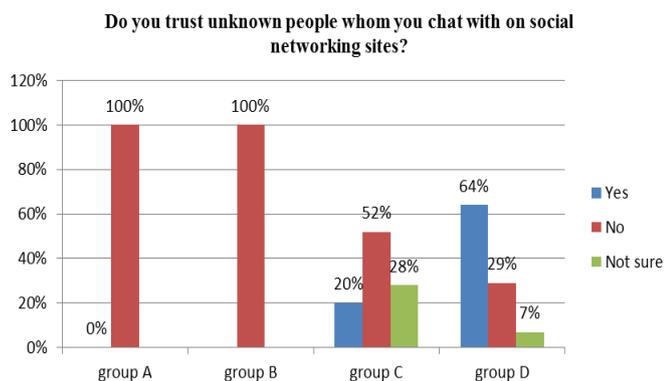


Fig. 8: Participants trust towards unknown people online

I. Unexpected Files and Emails

In this section, we assessed participants’ behaviour when receiving unexpected files and emails. The aim was to find out how participants react when receiving such information they never expected. Analysis shows that the majority of participants delete such files immediately without opening them (70%) while a few participants in groups B, C and D do open the file to see what it is (35%) (see Fig. 9a). Further analysis indicates that, when asked about their reactions towards unexpected emails claiming they won lottery or other financial benefits, majority of the participants in A, (90%), B, C and D actually delete such mails immediately without

opening them (see Figure 9b). However, 8% and 6% in C and D respectively indicated that they do open the mail and send their personal information to the mail sender. Though, some of the participants indicated that they need orientation on the issues of cyber security.

When you receive a file which you are not expecting, (for instance, that you have won \$500,000 USD from a lottery you did not play for, and you should send your personal information), what do you do?

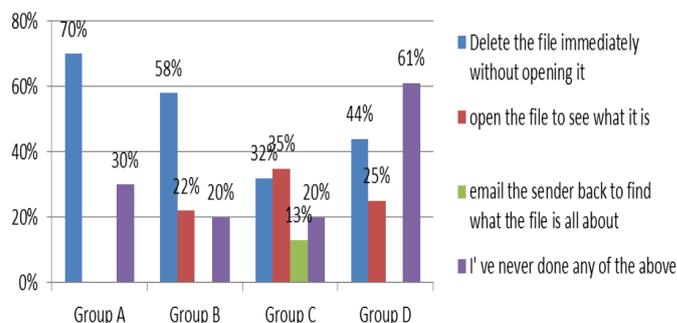


Fig. 9a: Participants’ behaviour when receiving files

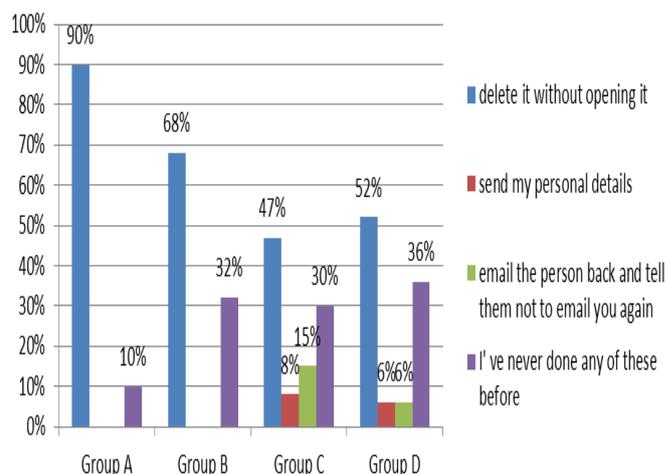


Fig. 9b: Reaction towards unexpected email

V. IMPACT OF DISCIPLINE ON CYBER THREAT’S AWARENESS

In this section, we tried to explore how different groups responded to the current cyber threats based on their responses from the administered questionnaire. Basically, considering group A, we expect that participants in this group will respond in a positive way due to their background knowledge in IT than any other group. In this same vein, group C is formed by students from IT departments and, like group A, there is also indication that their responses are more likely to be positive due to the IT background. Lastly, groups B and D are formed by non-academic and lectures from non-IT department and students respectively with non-IT background.

For this case, we selected nine of the most important and technical questions in the questionnaire. The questions are captured in Table 1. The basis was to enable us to understand what factors affect their responses and the relationship between one group and the other. The descriptive statistics of these questions are captured in Table 2.

**Table 1:** Summary of Technical Questions

Technical Questions	
1.	Anti-virus usage
2.	Virus Infection from browsing
3.	Virus Infection from attachment and downloads
4.	Fake or legitimate websites identification
5.	Receiving unexpected file
6.	Receiving unexpected SMS
7.	Receiving unexpected email
8.	Trust on social networking
9.	Patches usage

**Table 2:** Descriptive Statistics of Technical Questions

Group	Min	Max	Mean	Std. Error	Std. Dev.
GA	30	100	87.7778	7.95435	23.86304
GB	24	100	56.7778	8.36623	25.09869
GC	30	95	56.2222	7.11371	21.34114
GD	29	76	48.5556	4.79036	14.37107

As shown in Table 2, the mean, standard deviation and the standard error of the mean are shown. They represent the positive responses from the different groups considered in this study. Group A has the highest average of 87.8 with and standard error of 7.95. This is followed by groups B and C. In the same vein, group D has the lowest standard deviation showing how consistent were their responses on the part of unawareness of cyber threats. Group B has the largest standard deviation of 25.1 followed by A and C.

Table 3: Correlations

Group	Min	Max	Mean	Std. Dev.
GA	1			
GB	r=.016 p=.968	1		
GC	r=.664 p=.054	r=.146 p=.707	1	
GD	r=.383 p=.309	r=.067 p=.863	r=.795* p=.010	1

\*. Correlation is significant at the 0.05 level (2-tailed).

In order to find the relationship between the way different groups or faculties react to current cyber security threats, we computed a Pearson correlation coefficient using IBM SPSS with a significant value of  $p = 0.05$ . This is captured in Table 3. Based on the results, firstly, there was a positive correlation

between the participants in groups A and C,  $r = 0.664$ ,  $n = 9$ ,  $p = 0.050$ , where  $r$  is the correlation coefficient and  $n$  is the number of questions considered. In addition, a scatter plot that summarizes the results is shown in Figure 10. In general, there was a positive statistical significant correlation between responses of participants of groups A and C (GA and GC). This shows that the way IT personnel and lecturers responded were correlated with the way students with such backgrounds answered.

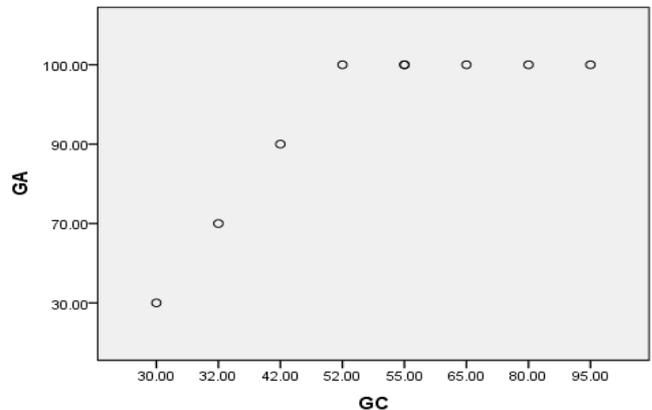


Fig. 10: Plot between group A vs C

Accordingly, there was also a positive correlation between the participants of groups C and D,  $r = 0.795$ ,  $n = 9$ ,  $p = 0.010$ . A scatterplot of these results is shown in Figure 11. The overall conclusion is that, there was a strong, positive statistical significant correlation between responses of participants of group C and D (GC and GD). This shows that the way IT students responded were correlated with the way students with non-IT backgrounds answered. The reasons for this could be due to the exponential growth of the Internet coupled with mobile applications that places no restriction on usage irrespective of the background. In this case, both IT and non-IT students are aware of the current cyber threats in order to protect themselves as well as their devices from being attacked.

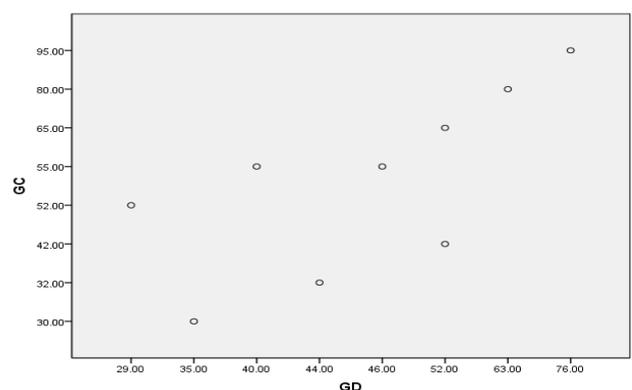


Fig. 11: Correlation between group C vs D

In the same vein, there was a weak correlation between the participants of groups A and B,  $r = 0.016$ ,  $n = 9$ ,  $p = 0.968$ ,

and A and D,  $r = 0.383$ ,  $n = 9$ ,  $p = 0.309$ . A scatterplot of these results is shown in Figure 12 and 13 respectively. With these results, there was no statistical significant correlation between responses of participants of these groups.

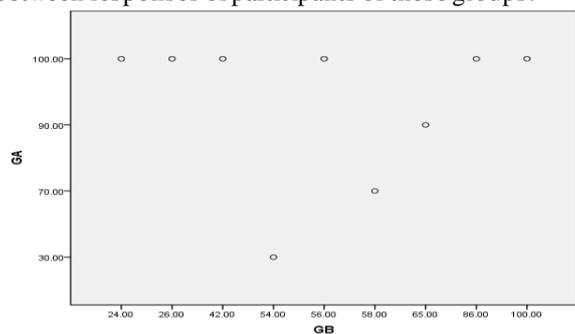


Fig. 12: correlation between group A vs B

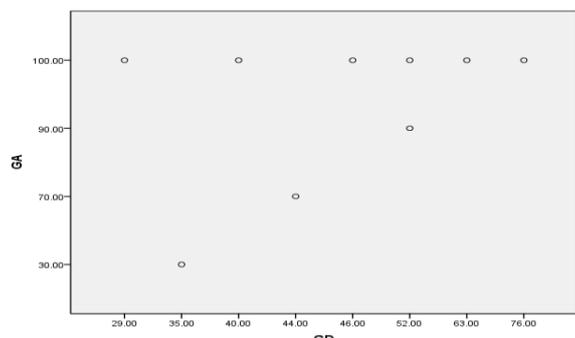


Fig. 13: correlation between group A vs D

### VI. DISCUSSIONS

Based on the results of the analysis we conducted in this research, it is interesting to know that the majority of the members of the UNIVEN community have access to Internet which is mostly done using UNIVEN network on campus. We also found that the majority of those who have access to the Internet were mostly engaged in online activities such as communicating on social networking sites, receiving and sending emails, doing education related activities and so on. When it comes to concern about the Internet safety and privacy which is at the core of this research, we found it amazing that most of them are very concerned about the security of the internet. However, on issues such as the level of knowledge about virus infection either through browsing and downloading attachments from email and websites, as well as how they respond to receiving unexpected file and emails we found out that responses were different between groups. We believed the differences could be as results of their different backgrounds with respect to IT.

In addition, there was a correlation between how group A and C responded as well as how C and D responded. Group C and D were basically students from the IT and non-IT related departments respectively. Thus, the correlation could be as a result of the unrestricted proliferation of mobile applications and the zeal to protect them irrespective of the background. We identified one of the factors that contributed to the

difference in results to be the previous lessons on cyber threats awareness received by participants. Based on this, though analysis not presented in this paper, analysis indicates that A and B had received lessons on how to use and staying safe on the Internet when compared to group C and D participants as well as the usefulness of the lessons. With these findings, though their responses were satisfactory, lots still have to be done. The goal is to ensure that every member of UNIVEN community be fully aware of the currents cyber threats in order to protect themselves as well as UNIVEN information.

#### A. Validity Threats

In this research, there are several issues that may affect the results of the study. Firstly is the attitude of the participants during the administration of the questionnaire. Some of the participants were seen as being too busy and the way the filled the questionnaire showed they were not interested which could affects their responses. Also, in terms of the groups, not everyone in each group participated in the study and it could be that those who did not participated may answer differently if they were present. Therefore, care must be taken to generalize these results to each group. During the course of the analysis of the data collected, there is the tendency that some data were missing, double computed or misrepresented for another group. However, we took every vital measure to ensure that we included the right participants and the analysis was carefully performed. If there is anything not considered in this study, we are confident such information will have no significant impact on the results presented.

### VII. CONCLUSION

Cybersecurity is a new global security concern and the awareness of cyber threats and attacks in organizations should be of high priority. Based on the results in this study, we found that majority of members of UNIVEN community are aware of current cyber threats and are acting accordingly to protect themselves as well as the university information from being victims. However, their responses were background dependent. In a more general way, we found that non-academic staff from the IT department and lecturers from IT departments and their students were the ones found to be more aware of current cyber security threats. For the students, there still exists those who are careless or not well-informed about cyber threats issues. Non-academic staff from non-IT departments and students from non-IT students were also found to respond positively, though their non-IT background limited the way they responded. We also found that there was a relationship between responses from the entire students irrespective of the background. We believed this could be linked to the growth of the Internet and mobile applications which requires them to have such knowledge in order to protect themselves and assets.

As can be seen in today's world, technology is growing very fast and in the same way explosive attackers are

becoming sophisticated and targeting more and more users. Thus, we have to devise new protection methodologies that would focus on training users and alongside put in place security technologies to protect the network at UNIVEN. In this case, we also propose six-ways to make UNIVEN community members aware of attacks directed to them. These include:

- 1) Using Mail Merge in MS Outlook
- 2) Placing security instruction documents in the computer Lab such as “Do not allow the use of USB drives”, “Do not have Windows systems set to automatically boot when a USB drive is inserted” and so on.
- 3) Putting a link on the university website of some trusted websites
- 4) Addressing students during orientation program at university
- 5) Block all access to untrusted websites
- 6) Having the capabilities of cyber analysis and warning of Monitoring, Analysis, Warning and Response [1] in place by the IT department in order to be constantly be aware of the new threats and alerting the university community of the cyber threats.

We believe these methods will help system users to become aware about the important role they play in the information security program, the skills needed in order to survive in the cyberspace and the basic lessons of security vulnerabilities.

We therefore recommend that these be implemented at UNIVEN as it could make a world of differences.

#### REFERENCES

- [1] Nordwood, K.T. and Catwell, S.P. *Cybersecurity, Cyberanalysis, and Warning*, Nova Science Publishers, Inc. 2009, ISBN: 978-1-61728-218-8
- [2] Andreasson, K. *Cybersecurity: Public Sector Threats and Responses* CRC Press, Taylor & Francis Group, 2012, ISBN: 13: 978-1-4398-4664-3
- [3] Internet World Stats. 2010. <http://www.internetworldstats.com/stats.htm> (18/11/2013)
- [4] Anon, South Africa's telecommunications. Available at: <http://www.southafrica.info/business/economy/infrastructure/telecoms.htm#UpeOt3QaLIU> (13/08/2013)
- [5] Liu, Pei-Wen, Jia-Chyi Wu, and Pei-Ching Liu. TWNCERT Social Engineering Drill: The Best Practice to Protect against Social Engineering Attacks in E-mail Form. 2008, <http://www.first.org/conference/2008/contest>.
- [6] Anon, Why do clients need to address information security awareness? Available at: [http://www.noticebored.com/html/why\\_awareness.html](http://www.noticebored.com/html/why_awareness.html) [Accessed August 26, 2013].
- [7] Mitra, A. *Digital Security: Cyber Terror and Cyber Security*, 2010. [http://books.google.co.za/books?hl=en&lr=&id=kzR5nRJRBF4C&oi=fnd&pg=PP1&dq=related:KsLDyTWgjxkJ:scholar.google.com/&ots=-19zqMe8vV&sig=Dshnfp3mzGfMglafNdJ9ch\\_EU](http://books.google.co.za/books?hl=en&lr=&id=kzR5nRJRBF4C&oi=fnd&pg=PP1&dq=related:KsLDyTWgjxkJ:scholar.google.com/&ots=-19zqMe8vV&sig=Dshnfp3mzGfMglafNdJ9ch_EU) [Accessed August 15, 2013].
- [8] Andress, J., 2011. *The Basics of Information Security: Understanding the Fundamentals of InfoSec in Theory and Practice*,
- [9] McCumber, J., 1991. *Information systems security: A comprehensive model ... of the 14th National Computer Security ...*. Available at: <http://trygstad.rice.it.edu:8000/GovernmentDocuments/NSTISS/NSTISSI4011Annex.rtf> [Accessed November 19, 2013].
- [10] SANS, SANS: Information Security Resources. Available at: [http://www.sans.org/information\\_security.php](http://www.sans.org/information_security.php) [Accessed October 7, 2013].

# The UWF Cyber Battle Lab: A Hands-On Computer Lab for Teaching and Research in Cyber Security

Chris Terry, Angelo Castellano, Jonathan Harrod, John Luke, and Thomas Reichherzer  
Department of Computer Science, University of West Florida, Pensacola, FL, USA

**Abstract** - *With a dramatic increase in cyber threats over the last decade, government and industry alike have recognized the pressing need to combat the ever growing cyber attacks on networks and systems. Educational institutions play an important role in researching technology that improve resiliency of systems as well as growing a workforce that understands cyber security challenges and can study and combat cyber attacks. The Computer Science Department at the University of West Florida (UWF) has built a Cyber Battle Laboratory to support undergraduate and graduate education, faculty research and public/private partnerships. Faculty and students can freely experiment with methods of attacks, detection and prevention in a controlled and isolated environment without affecting the campus network or the Internet. The lab is equipped with state-of-the-art technology to assist faculty in reconfiguring the environment for instructional and research purposes. It has been successfully used for classroom instruction and outreach activities at UWF.*

**Keywords:** cyber security; computer networks; virtualization, educational technology, laboratories

## 1 Introduction

The increasing attacks on systems and networks over the last decade disrupt our daily life and threaten the operations of public and private sector organizations. Constantly in the news are stories of attacks to businesses, universities, and government systems. The Government Accountability Office reported that the number of cyber threats increased by 680% from 2006 to 2011 with hackers attacking the integrity of systems and networks to gain access to private data and disrupt services for personal and political gains [1]. According to a recent report by Norton, the damage to consumers world-wide due to cyber attacks is estimated to be \$113 billion in 2013 [2]. The escalation of hacker attacks on our systems continues to be a major concern to businesses and governments that have invested in recent years significant amount of resources to harden system security and train their workforce to deal with the barrage of cyber attacks we experience. Educational institutions have stepped up to the growing need for IT professionals with cyber security background by offering specialized degree programs and certifications to students and professionals seeking to advance

their career. In the U.S. alone the number of IT programs that specialize in cyber security has increased substantially over the past 5 years addressing needs in the public and private sector [3].

To address the workforce needs in the Northwest Florida region in cyber security, the Computer Science Department at the University of West Florida (UWF) has developed undergraduate and graduate specializations that combine traditional core areas in computer science with topics in cyber security including system and network security, digital forensics, cyber warfare and gaming. To support the new programs and faculty research, a new laboratory has been constructed that offers state-of-the-art network and computer systems to build large-scale computer networks and computing environments for experimentation.

The remainder of this paper is organized as follows. The next section describes our motivation for building the UWF Cyber Battle Laboratory (CBL) as well as ethical concerns that arise when students are exposed to methods of cyber attacks and malware. The paper then describes the infrastructure of the CBL detailing the emulation of an independent computer network and systems for experimentation before discussing hacker and support software tools for demonstrating cyber attacks to students. The paper concludes with a discussion of challenges, related work in cyber security education and future outlook for the cyber battle lab.

## 2 Motivation and Ethical Concerns

As part of a long-term teaching and research effort in cyber security, the Computer Science Department has established a new computing lab that creates its own network infrastructure disconnected from the campus network and the Internet. The lab is designed to provide students an interactive learning experience allowing them to solve practical, real-world problems to complement theoretical concepts discussed in classroom and textbooks. It supports lab exercises, capstone projects and thesis research, competition for students in cyber security as well as outreach programs offered to a broader audience in the regional community to spark interest in cyber security and foster new partnerships with business and industry. The cyber lab provides special software tools, pre-configured virtual computing environments and network services, as well as tutorials for instructors to demonstrate

cyber threats through viruses, worms, botnets, and a variety of man-in-the-middle and denial-of-service attacks.

In an effort to ensure that malware remains within the confines of the cyber lab, the lab's network infrastructure is completely isolated from the campus network. In addition, all workstations in the lab have been configured to limit access to data from local machines or the network by disabling all USB ports and removing CD drives. Any file transfer from the external world to the lab must be done through a special podium desktop computer that restricts access to files on computers in the lab to designated IT personnel. When students are given access to the lab they are instructed on the proper use of the laboratory and the potential danger in either bringing outside equipment into the lab or removing data from the lab. The United States Military Academy takes a similar approach with their Information Analysis and Research Laboratory, describing it as the IWAR Range, and instructing students to treat it the same way as they would treat any live fire weapons range [4]. Because security topics taught in the laboratory include attacks and penetration testing, one concern is ensuring that courses also cover the proper ethical uses for the tools the students are being exposed to. Just as it would be unthinkable for a cadet to remove a loaded weapon from a firing range to practice outside the range, the same mentality must be fostered in students when thinking about the tools used in the laboratory for attacks.

For exploring the ethics of computer security, a good example is the story of Randal Schwartz, an engineer working at Intel, who performed unauthorized penetration testing against their network and was eventually prosecuted [5]. While it is important that students are properly briefed on the importance of not using the tools they are exposed to outside the lab environment, a broader ethical discussion may be worth saving for the end of the course. Locasto and Sinclair have noted that an ethical discussion at the beginning of the course would “lack detail” and “focused more on emotional argument rather than informed debate” [6]. Students of the UWF CBL receive special instructions at the beginning of each course that gives them access to the lab. However, these instructions are put in a broader legal and ethical context at the beginning of the course to raise concerns for harm that the methods may cause to others and themselves should they be applied outside the lab.

### 3 The Cyber Battle Lab Infrastructure

The next section describes the configuration of computer workstations and servers in the lab and the physical networks that support cyber security experiments. The current lab has been constructed as a pilot lab to A) gain experience with hardware resources and virtualization software for simulating networks and B) predict resource needs to scale-up the lab's emulated wired and wireless computer networks for more realistic experimentation. It serves as a blue-print for building an expanded version of the lab in the summer of 2014.

#### 3.1 Workstation & Server Environment

The pilot lab provides seats for 24 students that may be assigned to 12 computer workstations in teams of two for lab exercises. The workstations are equipped with dual monitor to provide better visualization of different virtual environments and two network interface cards each to connect the workstations to an experimental and a management network. The workstations run Windows 7 Enterprise as their host operating system with VMware Workstation 9 for virtualization of different machines. A workstation connects with the first network interface card (NIC) to the management network for general lab management services such as authentication and file sharing. VMware Workstation running within the host operating system is configured to use the second NIC installed in the system to connect to the experimental network. Figure 1 illustrates the configuration of the guest and host operating system and the separation between management and experimental network. The next section describes how the virtual environment and the host machine connects to two physical networks.

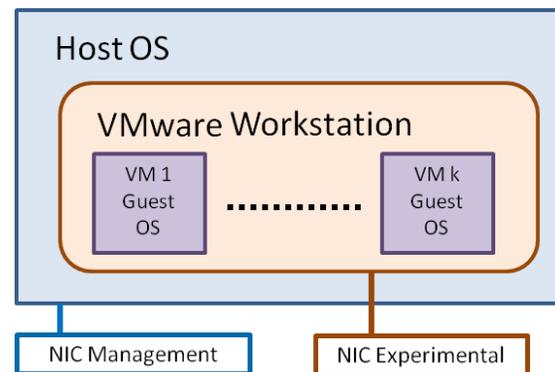


Figure 1. Workstation Network Connections.

In addition to the creation of two independent physical networks, the lab uses VMware Workstation to separate the experimental computing environment from the host environment. VMware Workstation runs virtual machines that are part of the various cyber experiments in the lab, while the host machine joins a general computer lab that supports centralized user authentication, network file systems, file sharing, and more. Each virtual machine executed within VMware Workstation implements a specific computing environment for launching or monitoring cyber attacks or defending against attacks emanating from a remote environment. Virtual machines may also be used to study strengths and weaknesses of operating systems or specific software application such as PDF viewers with known security weaknesses. Virtual machines may also be used to subject students to certain environments where malware has affected their systems to study how users respond to malware on their computer system or simply give users an opportunity to experience malware on a computer system without being harmed.

The lab includes two servers, one R210 II PowerEdge server equipped with an Intel Xeon E3 Quad Core processor and 8 GB of RAM purchased from Dell and one custom built 24 core AMD Opteron system with 64 GB of RAM, deployed with VMware ESXi. The R210 II PowerEdge server provides virtual servers for the management network to facilitate basic lab management and student file storage through a network file system. The custom server provides a virtualized lab environment on the experimental network that can be easily reconfigured to support a wide variety of research and educational activities with a minimal upfront investment in equipment. The VMware servers allow the multiple operating systems and network configurations to be simulated without physically reconfiguring the lab environment.

Using software from the open source Quagga project [7] the custom server runs several virtual routers to emulate large, interconnected wide-area networks. Each router executes within a single Linux server to emulate a network node within a wide-area network. Details of the wide-area network emulation and the physical network configuration follow below.

### 3.2 Wide-Area Network Simulation & Physical Network Configuration

In an effort to setup a more realistic setting for experimentation, the lab was designed to provide an accurate snapshot of a portion of the Internet infrastructure and a diverse set of emulated networks and hosts. We created a wide-area network that replicates parts of the Florida Lambda Rail education network, Internet Service Providers (ISPs), corporate networks (e.g. Google), and three Internet peering points including Equinix Chicago, Telix Atlanta, and NOTA Miami. Links between network nodes are mapped to Virtual Local Area Networks (VLANs) in the physical network discussed below. The virtual routers deployed in the emulated network implement common routing protocols such as the Border Gateway Protocol (BGP) and the Open Shortest Path First (OSPF) for routing data packets throughout the entire emulated network. Using published information from the Internet, the lab uses the same IP addresses used by the ISPs, corporate networks and Internet Backbone Peering Points to create the illusion for students and researchers to experiment with the actual Internet. Figure 2 gives a high-level description of the emulated network implemented by the lab.

In addition to the virtualized network, the lab implements network services such as Domain Name Services (DNS) for name resolution in corporate networks, universities, and ISPs as well as Dynamic Host Configuration Protocol (DHCP) to assign machines joining different networks automatically the corresponding IP addresses of the joined networks. Workstations in the lab may join specific ISPs, corporate or university networks as needed for implementing different attack scenarios or to collect network traffic data for analysis.

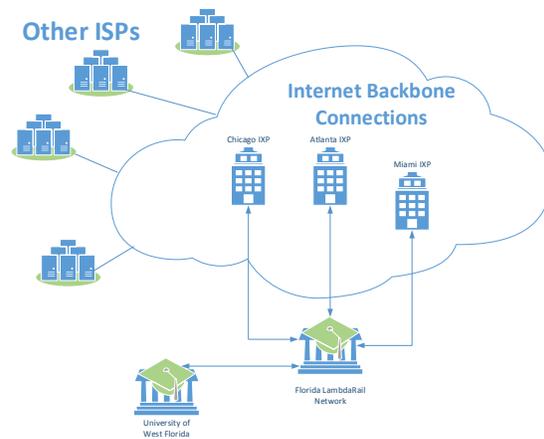


Figure 2: Simulated Wide-Area Network

The lab is designed to run a multiple well known services, including the DNS and TLD root servers, well known Web sites such as *google.com*, and corporate networks running Linux and Windows servers. This provides a familiar environment for students and researchers to use that accurately mirrors the real world Internet infrastructure. This infrastructure is designed to work in conjunction with additional virtual routers and hosts that can be run for individual experiments and classroom activities.

The physical network consists of a total of 3 switches, one layer-3 and two layer-2 switches. Each workstation connects to both layer 2 switches with their corresponding NICs to join the management and experimental network implemented by the switches. The R210 II PowerEdge server and the custom-built server connect to the layer-3 switch, which connects to the experimental network and management network switch. Traffic within the management network and experimental network is kept separate through VLANs implemented by all three switches. The custom-built server runs virtual routers on virtual machines that also use VLAN's to implement links between routers. **Error! Reference source not found.** Figure 3 shows the physical network configuration in the current CBL.

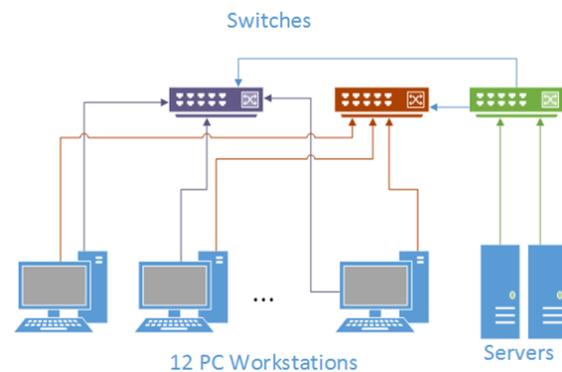


Figure 3: Physical Network Layout

VLAN tagging is used to allow the experimental network interface on the lab workstations to be able to run virtual machines on multiple networks simultaneously giving students and researchers the ability to join different networks with their virtual machines as if they were users on those networks or hackers that have successfully penetrated those networks.

The creation of a management and an experimental network serves two purposes. First, it ensure that the lab provides basic network services such as centralized account management via Active Directory and file sharing services needed for distributing virtual machines to workstation. Second, the separation ensures that malware is not affecting the host computers or the management network needed to run day-to-day lab operation and provide access to the simulated environments. Keeping the experimental and management network separate also gives instructors and researchers the power to refresh the lab after an experiment is completed or reconfigure the lab for an alternate experiment by simply starting and stopping virtual machines running virtual services in the network as needed. The following section discusses the tools and pre-configured virtual environments and tutorials that are available for teaching and research purposes.

## 4 Tools and Resources for Cyber Research & Education

CBL provides a number of software tools, customized environments and tutorial resources allowing users to demonstrate attacks, collect data from attacks or even implement their own attacks. The software tools implement widely known man-in-the-middle attacks that exploit weaknesses in network protocols and services. The tools are written in C using the raw socket interface available in a Linux environment to bypass existing protocols and create customized data packets for an attack.

In its current configuration, CBL implements the following attack methods:

- TCP SYN Flood
- DNS Query Flood,
- ARP Reply Flood,
- DNS and ARP Cache Poisoning.

These attacks have been discussed widely in the literature [8-12]. For space reasons, the paper will focus on a single attack method implemented for the CBL that exploits an ARP cache poisoning strategy combined with a DNS cache poisoning to re-route traffic to a fake Google server that runs on the hacker machine. Figure 4 shows the steps a hacker may take to redirect traffic to his own machine.

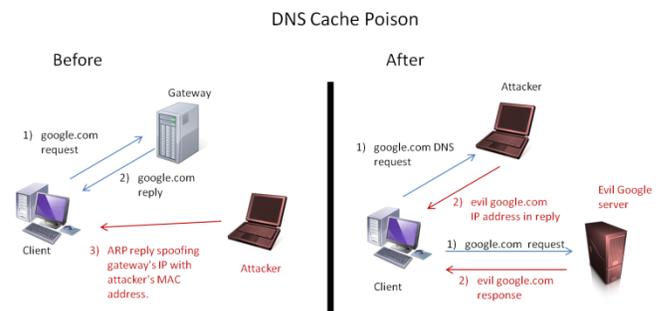


Figure 4: ARP Cache Poisoning attack to redirect traffic to a fake Google server.

The attacks are implemented on a virtual machine from where they can be deployed into any network. Figure 5 shows what happens on a lab workstation before and after the attack is executed. Note that the original Google server is a server that is replicated in the simulated computer network (the server only serves the home page for the Web search) on a Google corporate network and DNS queries are properly resolved by root and top-level DNS servers available in the network.

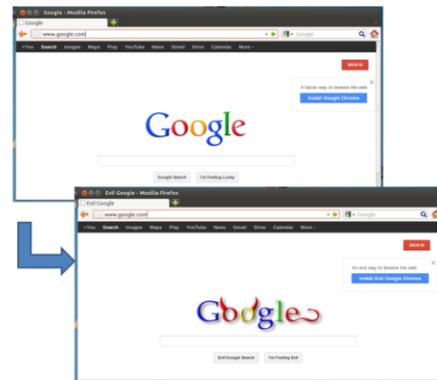


Figure 5: Google Web page as seen by a client before and after the attack.

The majority of software tools created for the CBL have been fully integrated into a Web application that implements a simple dashboard for launching attacks and performing various monitoring services. This dashboard is designed for instructors to control experiments in the lab for educational purposes. Besides attacks the dashboard also allows instructors to control startup of certain applications on the lab workstations such as Wireshark [13] or a Web browser to demonstrate remotely to users the successful execution of the attack or the content of certain data packets. The Web application is deployed on a virtual machine joining the network that is being studied for network attacks. Figure 6 shows a screenshot of the Attack Panel of the dashboard with an ARP Cache Poisoning attack executing.

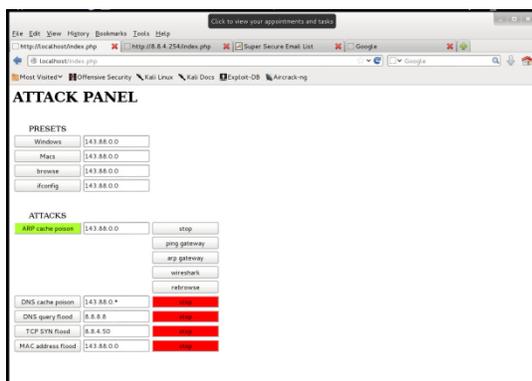


Figure 6: Attack Panel of the dashboard for launching attacks and remote controlling lab workstations.

To enable the dashboard to remote control applications on lab workstations, it uses a custom-built software installed on the virtual machine running the dashboard to send commands to the lab workstation via a UDP socket. Each workstation has the corresponding receiver software installed that listens for incoming commands from the dashboard to either open a Web browser, launch Wireshark, or run other visualization tools helping students understand changes in the environment that are illustrating the attack being studied.

All attack methods, fixes, and monitoring services are fully documented in a Wiki that instructors may access from the CBL podium desktop computer, to assist them with demonstrating cyber experiments in the lab, step-by-step.

## 5 Discussions

The CBL has been used for the first time in a cyber security class in the fall of 2013 and is currently being used in outreach activities. Students of the class last fall performed various lab activities to investigate network architectures and services to which the instructor exposed them. Students were also able to launch attacks and develop and implement defense methods. Finally, students were able to participate in a Red-team vs. Blue-team competition as a culminating project for the class. The hands-on experiments in the CBL environment was well received by students and the instructor of the class. Feedback solicited from participants of the lab showed that students had a good grasp of the security concepts such as man-in-the-middle or denial-of-service attacks because they were able to relate them to experiments they conducted in the class.

Building the software tools that implemented the various attack methods for CBL presented a number of development challenges for the programmers. When an attacker sends a spoofed DNS reply to a victim's DNS query, the spoofed DNS reply must arrive to the victim before its DNS server delivers a DNS reply back to the victim. Additionally, the DNS reply must contain a correctly matching DNS transaction ID and source port field or otherwise the victim's machine will declare it as invalid and discard the reply. Guessing the transaction ID or the source port information is not feasible as

there are too many possibilities. However, by successfully completing an ARP reply flood to the victim's machine, all traffic will be re-routed to the hacker's computer and so any DNS query can be captured by the hacker to extract the transaction ID and source port information for generating a spoofed reply. And by pinging computers in a network, the attacker can extract a victim's valid MAC address to be used in subsequent ARP reply flooding. In multiple experiments run in the lab, our software tools implemented for the CBL have a 100% success rate in executing the cyber attacks named above. They create the basis for developing future attack, defense, and monitoring methods for the lab.

The current CBL supports 12 workstations that can be used by students working in teams of two. It simulates a small but realistic network environment by virtualizing routers and network services. Since the lab uses VMware workstation to virtualize network nodes and servers including DNS, file, Web, directory, mail servers and more, the entire environment can be easily recorded and archived via VMware snapshots. This allows for the network to easily be preserved and reset after various cyber attack experiments have been conducted giving users the needed flexibility to allow for changes to occur as caused by malware without suffering from permanent damages to the network and its services. The next step in the evolution of the CBL is to expand its hardware resources significantly so that the lab can replicate additional peering points, ISPs, corporate networks as they exist on the real Internet making the space more realistic for experimentation. Funding has been secured to include a total of 40-60 servers similar in their configuration to the custom-built server that would allow a total of up to 1000 virtual machines to be executed simultaneously implementing various network resources on an emulated large-scale network.

Most importantly, by creating a realistic environment of real-world services, the lab provides a suitable space for students to learn about cyber security threats. It has the capability to safely host demonstrations of attacks targeted at consumers and end-users, such as phishing attacks, which can be demonstrated against recognizable services that people use, such as banking Web sites. Students majoring in Information Technology can learn about such attack methods and how to defend systems and networks, while students from other disciplines can experience the effects of such cyber attacks and learn about best practices to avoid becoming a victim.

To present date, we are not aware of similar efforts to ours that create autonomous, fully-functional networks with support tools for automatic experimentation and management. The majority of projects discussed in the literature describe a small-scale environment with few machines deployed to examine specific network topologies and services [14-17]. Our lab offers a flexible environment through the virtualization of network routers and services to create wide-area networks linking corporate networks, ISPs, and educational institutions as they exist today on the Internet together in a single network. This environment allows for hands-on cyber experimentation giving students and researchers the opportunity to study cyber attacks and

methods of monitoring and prevention. With the additional hardware resources to be deployed this summer and the implementation of more sophisticated attack methods, the UWF CBL will be creating an environment that allows for realistic cyber experimentation and the development of new technology to monitor threats and better defend against them.

## 6 Conclusions

We present in this paper the design and implementation of a cyber security lab that plays an active role in cyber security education and research at UWF. The lab provides a safe environment for faculty and students to experiment with cyber threats and learn how to detect and defend against possible attacks on systems and networks. The cyber lab has been used for the first time last fall semester by students of a cyber security class to explore widely used methods for studying weaknesses of systems and experiment with a number of cyber attacks playing the role of both an attacker and defender.

The initial use of the lab for classroom activities has shown the effectiveness of using special lab environments for cyber security education and research. The lab allows instructors to demonstrate live attacks on systems and networks to their students and discuss methods to defend against and trace the attacks giving students much needed practical experience. It allows faculty to customize networks and setup experimental environments for collecting data and testing new methods for monitoring and defending attacks. The lab built in the summer of 2013 is now being significantly expanded to create larger networks and allow for more realistic experimental settings to drive research and engage students in cyber security at UWF.

## 7 References

- [1] United States Government Accountability Office. (2012). *Cybersecurity: Threats Impacting the Nation*, GAO-12-666T. [Online]. Available: <http://www.gao.gov/products/GAO-12-666T>.
- [2] M. Merritt, and K. Haley. (2013). "2013 Norton Report", Symantec Corp. [Online]. Available: [http://www.symantec.com/about/news/resources/press\\_kits/de tail.jsp?pkid=norton-report-2013](http://www.symantec.com/about/news/resources/press_kits/de tail.jsp?pkid=norton-report-2013).
- [3] W. A. Conklin, R. E. Cline, and T. Roosa, "Re-engineering Cybersecurity Education in the US: An Analysis of the Critical Factors," in *Proceedings of the 2014 47th Hawaii International Conference on System Sciences*, pp. 2006 – 2014, 2014.
- [4] J. Schafer, D. J. Ragsdale, J. R. Surdu and C. A. Carver, "The IWAR range: a laboratory for undergraduate information assurance education," *Journal of Computing Sciences in Colleges*, vol. 16, no. 4, pp. 223-232, 2001.
- [5] T. Wulf, "Teaching ethics in undergraduate network security courses: the cautionary tale of Randal Schwartz," *Journal of Computing Sciences in Colleges*, vol. 19, no. 1, pp. 90-93, 2003.
- [6] M. E. Locasto and S. Sinclair, "An Experience Report on Undergraduate Cyber-Security Education and Outreach," in *The Second Annual Conference on Education in Information Security (ACEIS 2009)*, Ames, 2009.
- [7] K. Ishiguor, "Quagga Software routing Suite." [Online]. Available at: <http://www.quagga.net>.
- [8] S. McClure, J. Scambray, and G. Kurtz, *Hacking exposed: Network security secrets and solutions* (6<sup>th</sup> Ed.), New York, NY: McGraw-Hill Co., 2009.
- [9] A. Harper, S. Harris, J. Ness, C. Eagle, G. Lenkey, and T. Williams, *Gray hat hacking: The ethical hacker's handbook* (3<sup>rd</sup> Ed.), New York, NY: McGraw-Hill Co. 2011.
- [10] D. Kennedy, J. O'Gorman, D. Kearns, and M. Aharoni, *Metasploit: The penetration tester's guide*. San Francisco, CA: No Starch Press, 2011.
- [11] C. Schuba, J. Krsul, D. Kearns, and M. Kuhn, "Analysis of a Denial of Service Attack on TCP" in *Proceedings of the 1997 IEEE Symposium on Security and Privacy*, Lafayette, IN: Purdue University, 2011.
- [12] M. Simpson, K. Backman, and J. Corley (2011). *Hands-on Ethical Hacking and Network Defense*. Boston, MA: Course Technology.
- [13] Wireshark: The World's Most Popular Network Protocol Analyzer. [Online]. Available: <http://www.wireshark.org>.
- [14] M. Micco and H. Rossman, "Building a cyberwar lab: Lessons learned teaching cybersecurity principles to undergraduates," in *Proc. of 33rd SIGCSE Tech. Symp. Computer Science Education*, Northern Kentucky Convention Center, February, pp. 18-22, 2002.
- [15] P. Mateti, "A virtual environment for IA education," in *Proc. of the 2003 IEEE Workshop on Information Assurance U.S. Military Academy*, West Point, NY, pp. 17-23, 2003.
- [16] R. T. Abler, D. Contis, J. B. Grizzard, and H. L. Own, "Georgia Tech Information Security Center Hands-On Network Security Laboratory," in *IEEE Transaction on Education*, Vol. 49, No. 1, February, 2006.
- [17] S. Standard *et al.*, "Network reconnaissance, attack, and defense laboratories for an introductory cyber-security course," in *ACM Inroads*, Vol. 4, Issue 3, pp. 52-64, September, 2013.

# Bite-sized Learning of Technical Aspects of Privacy

S. Peltsverger and G. Zheng

Information Technology Department, School of Computing and Software Engineering,  
Southern Polytechnic State University, Marietta, GA 30060 USA

**Abstract** -- Research has shown that bite-sized learning cater better to students with short attention spans and help instructors and students to stay focused on the course objectives. To address the difficulty of teaching technical implementation of privacy, the authors designed an assessment for one of the privacy learning modules. The assessment contains the analysis of the subject and implementation of the results using Google's tool Oppia. The paper reports the experiences of using the tool by both students and instructors and how the use of Oppia increased student understanding of the topic and their ability to think critically and creatively.

## 1. Introduction

Many of the current courses address privacy as a legal and a policy issue and do not cover technical details, as technical implementation of privacy is more difficult to teach and learn without supporting environments. The authors had identified this deficiency and proposed a coherent and consistent curriculum framework on teaching privacy and used this framework as a guide to design and develop privacy learning modules with technical details. The purpose of these technical learning modules is to demonstrate what happens "behind the scene," and how technology can be used to protect privacy. The first four labs have been developed with technical competencies [1] that not only demonstrate how to protect customers' privacy and write privacy policies, but also how to develop technical/automatic procedures for their enforcement.

In order to better teach students those technical details, authors have experimented with a bite-site learning concept. Many proponents of bite-sized learning [2], including successfully running since 2003 BBC Bitesize free online study support program, emphasize flexibility in time, better support for non-traditional students and reduced cost of development and provisions. The bite-size learning concept was applied to the assessment of the first learning module "Tracking Keyboard and Mouse Action on the Web". In this paper, authors discuss the development and the use of this approach in the module.

## 2. Bite-sized Learning and Oppia

In the assessment the authors have utilized a new tool called Oppia developed by Google [3]. The tool was released earlier this year and allows creation of interactive learning objects called explorations. It is possible to create your own instance of Oppia or build and host your creation at <http://oppia.org>.

According to Google, "Oppia models a mentor who poses questions for the learner to answer. Based on the learner's responses, the mentor decides what question to ask next, what feedback to give, whether to delve deeper, or whether to proceed to something new. You can think of this as a smart feedback system that tries to "teach a person to fish," instead of simply revealing the correct answer or marking." [4].

This tool and a novel approach have opened new ways for student-centered learning, especially, in evaluation of student learning [5]. Interactive learning, in contrast to the instructor-led lectures, provides learners many options and control over aspects of the learning process, pace, and sequence. The interactive features of a learning environment enable the learning to be individualized and are believed to enhance learning outcomes and learning satisfaction [6].

Google studied click-stream analysis [7] of courses they offer and found that more students completed interactive activities than watched videos or read lesson text. Google used Wilcoxon signed-rank test to confirm that there were more students who completed activities than those who just viewed lessons. The test also found that adding interactive activities to a learning module keeps students engaged in the course.

Interactive hands-on exercises are crucial in subjects like Computer Science and Information Technology that require practical skills. They help students to acquire problem solving skills. Educators know that one of the best ways to learn is to explain it to someone else. The two main ideas are: if you cannot explain it to others, you have to learn more about the topic; and what you understand you can communicate to others. Class discussions partially based on

this principle [8]. Students express ideas using their own words and find a way to share what they have learned with their peers.

### 3. BUILDING EXPLORATIONS

The module of “Tracking Keyboard and Mouse Action on the Web” demonstrates how user keyboard and mouse actions can be tracked using client scripts and commonly available tools. Upon completion of this module, students should be able to

1. examine client scripts that can track keyboard actions;
2. examine client scripts that can track mouse actions;
3. evaluate potential privacy risks arising from user tracking.

In the tasks of this module, instead of designing discussion topics, authors developed an assignment that let students develop an exploration to present one of the topics in the module to other students. Building an exploration take learning to another level. Students not only have to understand the topic and know correct answers, but they also have to find incorrect answers for case studies and questions. Those incorrect answers must be realistic and foresee how the presented content can be misunderstood. The feedback later from students showed that the process of designing incorrect alternatives contributed the most to their learning. The incorrect answers should be designed to destruct learners and would be selected by students who did not completely master all learning outcomes. According to students, when they tried to teach what they learned, they found new views on the topic. Writing an Oppia exploration is more involved than participation in a discussion. Oppia must respond dynamically to any input according to the rules in each state of an exploration. Oppia supports branching and looping that allow creation of a custom path for each learner. Figure 1 shows the result of choosing to start from section 1, with one incorrect and one correct answer to a fill-in-the-blank question.

Oppia developers recommend to develop a flowchart (Figure 2) on the paper before designing an exploration. This step is important because it separates the content and the logic of the assignment from its implementation with the Oppia. An exploration can have multiple branches allowing learners to complete their customized learning experience in their own unique way. The exploration can be designed to give learners a choice of topic access order (Figure 3). The fact that

majority of students implemented some ways of changing the order of topics, shows that students prefer having this flexibility.

In this exploration we will consider two section

1. Understanding keyboard and mouse tracking
2. Preventing companies from tracking you

You can take both the section together or after completing one section you can come back later and finish the other section.

Select one of the following to continue

Section 1

Welcome to section 1

\_\_\_\_\_ fires when the mouse pointer moves over an element

onmouseout

Wrong answer, please try again!

onmousemove

Right answer, lets move on to the next question!

Fig 1. Correct answer to a multiple choice question and loop for a fill-in-the-blank question.

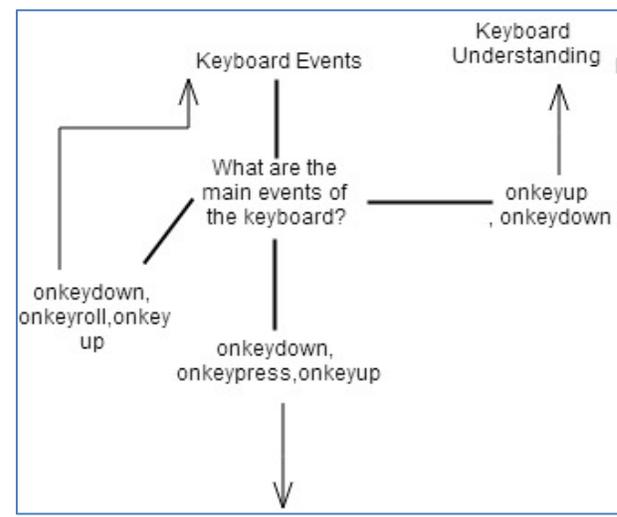


Fig 2. A fragment of a flowchart.

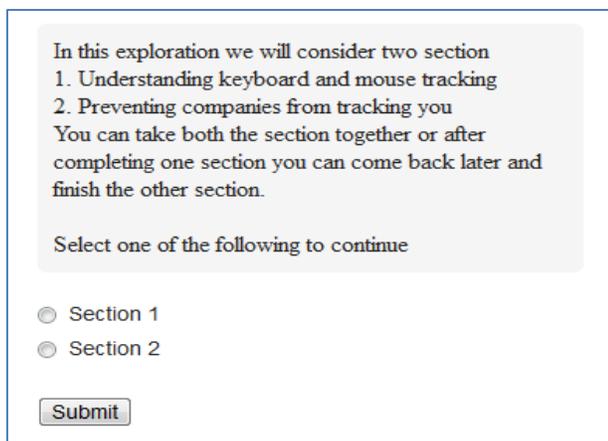


Fig 3. Flexible path of learning.

An exploration is designed as a diagram of interconnected states (Figure 4). Every state in Oppia contains an interactive widget including text input fields, multiple-choice inputs, and clickable Google Maps (to show the location of tracking servers). Each state has one default rule that identifies next state and might contain an optional feedback. The author of the exploration can define additional rules. Explorations can be exported as YAML format files, modified and uploaded to the same or another Oppia installation (Figure 5). YAML, a recursive acronym for Ain't Markup Language, is a human friendly data serialization standard for many programming languages [9].

The authors installed Oppia on a local server for student experimentation in this module. The Oppia uses Django on Google App Engine and requires Java and Python. The application is still in the beta version and can use better documentation, but the authors were successful in installing, deploying and configuring the software, and uploading student explorations.

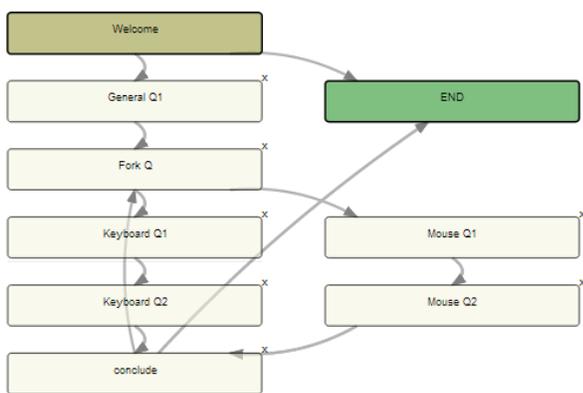


Fig 4 The exploration diagram showing states.

```

default_skin: conversation_v1
init_state_name: Welcome!
param_changes: []
param_specs: {}
schema_version: 2
states:
  Estimate 100:
    content:
      - type: text
        value: What is 10 times 10?
    param_changes: []
    widget:
      customization_args: {}
      handlers:
        - name: submit
      rule_specs:
        - definition:
            inputs:
              x: 100.0
            name: Equals
            rule_type: atomic
            subject: answer
            dest: Numeric input
    [...]
    
```

Fig 5 Example of YAML file

The developed exploration can be embedded in a webpage or even Course Management System (e.g. Desire2Learn). It is very similar to embedding YouTube videos (Figure 6).

```

<oppia oppia-id="hc-jGHKNupXI"
exploration-version="1"
src="http://webapp.spsu.edu:8181">
</oppia>
    
```

Fig 6 Embedded code.

### 4. Findings

At the end of the module, students completed a survey that collected feedback in the following four areas:

- Learning material organization, instruction and logic flow.
- Achievement of learning outcomes.
- Satisfaction with learning process.
- Effectiveness of assessments that cover technical details.

The survey consists of 18 questions or statements. The first group consists of thirteen statements that measure students' perceptions and learning effectiveness. Students evaluate these statements using a 5-point Likert scale. The next group

consists of five open ended questions designed to collect qualitative comments and feedback. The last three open ended questions specifically ask student experience with the Exploration and the tool Oppia. These three questions are:

1. Do you think you learned more about privacy and keyboard/mouse tracking while you worked on the exploration assignment?
2. If your exploration is included as one of the learning resources in the privacy module, how it will improve student learning?
3. Do you have any feedback on the tool Oppia?

A total of nineteen students have completed the module and the survey. Students' qualitative feedback provide valuable insights that will be used to improve the use of Oppia and the Exploration assignment in current and future modules on technical aspects of privacy. We have selected some representative student comments arranged in a number of themes to illustrate student attitude toward this exercise.

Theme #1: having a clearer understanding of the technical details behind the scene.

Students have repeatedly mentioned about the use of the tool help most in understanding the sequence of events that happened in normal mouse clicks and keyboard typing. As one student noted:

*"Working on this assignment helped me have a clearer picture of the sequence of events involved in tracking and privacy thus improving my understanding of the topic."*

Theme #2: students learned more if they have to work on answers.

*"I learned more about privacy and keyboard/mouse tracking as I had to write questions and its answers. I had to visualize the flow of content in order to decide on the flowchart."*

Theme #3: the flow chart seems to be a good tool to help learning.

*"I think a look at the flowchart will help in getting the overview on this module. Sometimes you need to read the entire module before you understand how they are related or to get a bigger picture. ... I feel the flowchart is good way to start any module. Students can get an*

*overview and then each question will help them get a brief understanding."*

Theme #4: the Exploration features interactive learning.

*"This will allow students to utilize their knowledge gained from lessons in an interactive manner. This provides students with an interesting and different way of learning course material which I think might benefit their learning experience."*

Theme #5: students like the tool Oppia.

Some students reported that the creation of the first successful state was a steep learning curve. However, the process became easy after that. The tool gave them a different experience.

*"Absolutely loved working on the project. It was a little confusing at first. ... The thing I love the most is, you can see the results right away as you edit the explorations. What a cool tool!!! Would love to do more such exercises."*

## 5. Conclusion

Bite size learning method using Oppia Exploration received positive evaluation in our module of Tracking Keyboard and Mouse Action on the Web. The feedback provided valuable insight and presented many opportunities for creating an optimal assessment exercises for teaching technical implementation of privacy. The authors plan to apply the bite-sized concept to other modules and collect more feedback. The best explorations created by students will be added to the learning module to help future students. Additionally, the authors hope to implement an element of competition between students for the best exploration on a given topic.

## 6. APPENDIX

Module evaluation survey questions/statements (students were asked to evaluate the following statements using a 5-point Likert scale from mostly disagree to mostly agree):

1. This module has clear objectives and learning outcomes.
2. The readings in this module are adequate to achieve learning outcomes.
3. The tasks are helpful in applying knowledge.
4. These laboratory exercises are crucial in understanding technical aspects of privacy.

5. The assessment indeed tests what I have learned through the module.
6. Seeing these examples motivated me to learn more about technical aspects privacy.
7. I learned in other classes how web apps could undermine privacy.
8. I fully understood before starting this module how JavaScript could be used for mouse tracking.
9. I fully understood before starting this module how JavaScript could be used for keyboard tracking.
10. Seeing examples from this module gave me better grasp of how keyboard tracking works.
11. Seeing examples from this module gave me better grasp of how mouse tracking works.
12. Analysis of logs generated by my interaction with a browser gave me better understanding of how user tracking is done.
13. I will be more aware of the tracking posed through a website and act accordingly.

## 7. REFERENCES

- [1] S. Peltzverger, G. Zheng, "A Virtual Environment for Teaching Technical Aspects of Privacy" ACM Press 2013, pp. 49–52.
- [2] M. Athitakis, "Bite-Size Learning for a Digital World ," Associations Now, November/December 2013, Available: <http://www.asaecenter.org/Resources/ANowDetail.cfm?ItemNumber=441641>
- [3] "Getting Started - Oppia - Downloading and Running Oppia - Tool for creating interactive educational content - Google Project Hosting." [Online]. Available: <http://code.google.com/p/oppia/wiki/GettingStarted>. [Accessed: 18-Mar-2014].
- [4] "Google's Oppia Offers Feedback Loop For Curriculum." [Online]. Available: <http://www.educationnews.org/technology/googles-oppia-offers-feedback-loop-for-curriculum/>. [Accessed: 28-Mar-2014].
- [5] M. Weimer, *Learner-centered teaching: Five key changes to practice*. John Wiley & Sons, 2013
- [6] S. D. Brookfield, S. Preskill, *Discussion as a way of teaching*, Jossey-Bass, 2005
- [7] J. Wilkovski, A. Deutsch, D. Russell, Student Skill and Goal Achievement in the Mapping with Google MOOC, *Proceedings of L@S ACM* 2014.
- [8] D. Jonassen, M. J. Spector, M. Driscoll, M. D. Merrill, and J. van Merriënboer, *Handbook of Research on Educational Communications and Technology: A Project of the Association for Educational Communications and Technology*. Routledge, 2007.
- [9] "The Official YAML Web Site." [Online]. Available: <http://www.yaml.org/>. [Accessed: 19-Mar-2014].

# Privacy Incongruity: An analysis of a survey of mobile end-users

Mark Rowan and Josh Dehlinger

Department of Computer and Information Sciences  
Towson University, Towson, MD, USA

**Abstract** – *Many students are not fully aware of the personal information collection, transmission and storage capabilities of the mobile applications in which they interact with on a daily basis. This poses a serious security and privacy concern that needs to be addressed in the classroom. This paper provides a snapshot about the actual behaviors of students as mobile device end-users and their concerns about the privacy of their personal information. The results show that students are providing personal information regardless of their stated privacy concerns. This may indicate that current privacy policies are not effective. There is a need to better understand the perception of mobile device end-users and improving usable privacy and security. Greater emphasis on the value of personal information and decision making with sharing personal information should become a part of the general education computer science curriculum to better prepare students for their mobile lifestyles.*

**Keywords:** Usable privacy, mobile applications, end-user education, human computer interaction

## 1 INTRODUCTION

Mobile Internet capable devices and software applications allow people to communicate, socialize and share personal information anytime and nearly anywhere. Many people agree that privacy is an important concept, but the challenge is to clearly define exactly what should be private and then implement policies to protect private information. The U.S. Department of Homeland Security's Stop.Think.Connect. campaign website warns that cybercriminals do not always discriminate and offers some advice to minimize their chances for an incident, to include: "Use privacy settings and limit the amount of personal information you post online" [1]. Victims of identity theft have lost money due to fraud, credit histories ruined, denied opportunities and even been temporarily arrested for crimes they did not commit [2]. The U.S. Federal Trade Commission recommends that everyone reads privacy policies to understand how the site maintains accuracy, access, security, and control of personal information it collects and whether it provides information to third parties [3].

A survey looking into the perceptions of privacy by college students in social networking indicated that students routinely do not read privacy policy statements [4]. A collaborative survey between Pew Research Internet Project and Berkman Center for Internet & Society at Harvard found that 70% of teens seek guidance from family and friends about online privacy, but only 9% have asked a teacher [5]. Romney and Romney point out that

most students normally rely on their teachers to prepare them for the environment they will work and function [6]. Teachers need to become a more reliable source for students to approach with their online security and privacy concerns and questions.

A systematic literature review on using mobile computing as a learning intervention did not find courses focused on privacy issues [7]. Recent research has shown how privacy education can be effectively integrated as a course or as modules at the undergraduate level [8], [9], as well as proposals for the graduate level [10]. Privacy as a topic may be able to follow the successful integration of computational thinking into the general education curriculum, similar to the approach by [11]. A recent literature review analysis of mobile application software engineering identified a lack of publications dealing with privacy related issues [12]. [13] opines that learning privacy from a technical perspective is needed, and the knowledge is expected by students' future employers and society.

There is a need to better understand the perception of mobile device end-users and improving usable privacy and security. Greater emphasis on cybersecurity awareness, the value of personal information and decision making with sharing personal information should become a part of the general education computer science curriculum to better prepare students for their mobile lifestyles.

This work presents the results of a paper survey of self-reported Internet related concerns and behaviors of 584 undergraduate students. The results of this research will contribute to an improved understanding of real-world software engineering challenges with end-user privacy concerns, and may also indicate a need to increase privacy awareness as a general education topic.

The contribution of this paper is to present a current snapshot of mobile end-user privacy concerns and their actual behaviors, and to suggest improvements for preparing students for making privacy decisions in their mobile lifestyles as end-users or mobile application developers.

The rest of this paper is organized as follows: survey methodology, data cleaning, data analysis, reported concerns and behaviors can be found in Section 2, Section 3 provides a discussion on the findings with respect to the three research questions, Section 4 presents limitations and Section 5 offers a conclusion and suggests future work.

## 2 METHODOLOGY

Problem formulation was the first step in this research to better understanding why students do not consider teachers a primary source of information for online privacy concerns. A survey was selected as a research method because it quickly captured how individuals are interacting with mobile technology, what problems they are facing and what actions they are taking [14]. The paper survey was composed of seventeen closed questions with dichotomous and nominal polytomous response scales. The subjects in the survey were students in general education computer science classes at a doctorate-granting, public U.S. university during the Spring 2013 semester. An attempt was made to avoid biasing responses (e.g., priming for privacy) by stating the anonymous survey was about Internet end-user behavior and attitudes. Students were compensated for their time with candy. The results were stored in a digital spreadsheet utilizing data validation for data entry and pivot tables for data analysis. In compliance with the Institutional Review Board the surveys were voluntary.

The survey sought to answer the following questions:

Q1: How concerned are student mobile end-users about the privacy of their personal information?

Q2: What are some of the behaviors of mobile end-users that may divulge personal information to third parties?

Q3: What is the relationship between privacy concerns and reading privacy policies?

The answers to these research questions may lead to an improved understanding of mobile end-users or become a catalyst for improving usable privacy and security.

### 2.1 Data Cleaning

A response rate could not be determined because an unknown number of blank surveys were returned and class attendance was not considered. 617 surveys were originally received and 584 remained after data cleaning for the following reasons:

- Seven surveys were removed because students reported not using the Internet at least occasionally to send or receive email.
- Twenty-one surveys were removed because they reported not using mobile computing devices.
- Five surveys were removed because they were unclear or incomplete.

After data cleaning, the data set was manually entered into an electronic spreadsheet using data validation. Each anonymous survey was given a unique identifier for traceability purposes.

### 2.2 Data Analysis

The survey demographics consisted of a gender breakdown that was composed of 301 (51.54%) females, 283 (48.46%) males. The ages of reported males and females (cf. Table 1) ranged from 18 to 53.

Table 1. Age Breakdown by Gender

	Female		Male		Total	
<b>18-23</b>	278	92.36%	245	86.57%	523	89.55%
<b>24-29</b>	20	6.64%	27	9.54%	47	8.05%
<b>30-35</b>	2	0.66%	7	2.47%	9	1.54%
<b>36-41</b>	1	0.33%	2	0.71%	3	0.51%
<b>42-47</b>	0	0.00%	1	0.35%	1	0.17%
<b>48-53</b>	0	0.00%	1	0.35%	1	0.17%
<b>Total</b>	<b>301</b>	<b>100.00%</b>	<b>283</b>	<b>100.00%</b>	<b>584</b>	<b>100.00%</b>

### 2.3 Reported Concerns

When asked, "How concerned would you be if your service provider sold your phone or email contacts data to another party?", female students reported more concern than their male counterparts (cf. Table 2).

Table 2. Concern about contacts sold to a third party

	Female		Male	
<b>Very Concerned</b>	194	64.45%	157	55.48%
<b>Concerned</b>	96	31.89%	84	29.68%
<b>Total</b>	<b>290</b>	<b>96.35%</b>	<b>241</b>	<b>85.16%</b>

As listed in Table 3, there was a slightly larger gap between males and females for the question, "How concerned would you be if your service provider used your photos or other data in marketing campaigns?" (cf. Table 3).

Table 3. Concern about personal data to marketing campaigns

	Female		Male	
<b>Very Concerned</b>	216	71.76%	164	57.95%
<b>Concerned</b>	69	22.92%	88	31.10%
<b>Total</b>	<b>285</b>	<b>94.68%</b>	<b>252</b>	<b>89.05%</b>

Female survey respondents expressed the most concern when asked, "Do you think the exact location of a personal mobile computing device (i.e., Android smartphone, iPhone, Windows 8, iPad, etc.) should be treated as private information?" (cf. Table 4).

Table 4. Believe that location should be private

	Female		Male	
<b>Yes</b>	280	93.02%	246	86.93%
<b>No</b>	21	6.98%	37	13.07%
<b>Total</b>	<b>301</b>	<b>100.00%</b>	<b>283</b>	<b>100.00%</b>

Female survey respondents further emphasized their concerns more than their male counterparts when asked, "How concerned

would you be if your service provider shared the exact location of your personal mobile computing device (i.e., Android smartphone, iPhone, Windows 8, iPad, etc.) with third parties?" (cf. Table 5).

**Table 5.** Concern if location data was shared with third parties

	Female		Male	
<b>Very Concerned</b>	197	65.45%	144	50.88%
<b>Concerned</b>	85	28.24%	91	32.16%
<b>Total</b>	282	93.69%	235	83.04%

## 2.4 Reported Behaviors

A majority of the students, 578 (98.97%), reported posting pictures online or sending them via email. Six female students reported not posting pictures online or sending them via email and it is possibly due to modesty or religious convictions.

Most students, 353 (60.45%), reported having used an online backup service to protect their data. Anecdotal evidence indicated that some students were simply referring to a free backup service offered by their phone service provider.

The results for the question, "Have you read a privacy policy for a software application, website or social networking site?" (Table 6) are similar to research results by previous studies. Grossklags and Good's research found that most users do not read End User License Agreements when indicating consent to the software installation process [14].

**Table 6.** Read Privacy Policies

	Female		Male	
<b>Always, completely review</b>	2	0.66%	5	1.77%
<b>Always, partially review</b>	29	9.63%	19	6.71%
<b>Sometimes, completely review</b>	12	3.99%	19	6.71%
<b>Sometimes, partially review</b>	166	55.15%	149	52.65%
<b>Never</b>	92	30.56%	91	32.16%
<b>Total</b>	301	100.00%	283	100.00%

Female respondents (84.05%) outnumbered male respondents (79.15%) in stating that they have shared personal information (i.e., name, address, and/or email address) with a company for their customer loyalty or rewards programs (cf. Table 7).

**Table 7.** Share personal information for rewards

	Female		Male	
<b>Yes</b>	253	84.05%	224	79.15%
<b>No</b>	48	15.95%	59	20.85%
<b>Total</b>	301	100.00%	283	100.00%

A majority of the students, 477 (81.68%), of the survey respondents reported using location awareness technology on a

smartphone. 318 (66.67%) of these respondents have turned off their phones or disabled location awareness technology so that their location was not known (cf. Table 8).

**Table 8.** Location Awareness use

	Used location awareness	Made location unknown
<b>Female</b>	242	158
<b>Male</b>	235	160
<b>Total</b>	477	318

Male respondents (49.12%) outnumbered female respondents (42.52%) in stating that they have uninstalled a mobile application because they found out it was collecting personal information that they did not want to share.

## 3 DISCUSSION

Overall, there is a disconnect between student concerns about the usage of their personal information and their behaviors as mobile device end-users. This disconnect could be more actively addressed in general education computer science courses with case studies, discussions and role-playing in three ways: Students as End-Users, Students as Software Developers, and Students as Business Owners or Chief Information Officers. Students should be encouraged to discuss real world cases and their experiences of sharing personal information as end-users. Students in the role of software developers should plan for privacy as a non-functional requirement and consider the full life cycle of collected data. Students as business owners or Chief Information Officers should consider the legal, security and ethical implications of collecting personal information, creating enforceable privacy policies, as well as terms of service.

In response to Q1: How concerned are student mobile end-users about the privacy of their personal information? The survey results reflected that the students consistently reported being very concerned or concerned about the privacy of their personal information. This was best observed when 517 (88.53%) students reported being very concerned or concerned about their personal information being shared with third parties but the overwhelming majority, 577 (97.57%) do not always completely read privacy policies. This could be reflective of the costs of reading privacy policies as found by [16], end-user comprehension challenges as discussed by [17], the ubiquity of End User License Agreements have trained even privacy-concerned users to click on "accept", which thwarts the very intention of informed consent as observed by [18], or some end-users may feel like they do not have a real choice due to social pressures to participate.

In response to Q2: What are some of the behaviors of mobile end-users that may divulge personal information to third parties? Students reported using email services, posting photos online, using online backup services, enabling location awareness technology and providing personal information in exchange for loyalty/rewards memberships. A major challenge faced by all

end-users is the complexity involved in making decisions to share information without knowing about the positive and negative consequences of disclosure [19]. The survey results reflect that a disconnect exists between student stated concerns and their behaviors. The survey results support previous studies that indicate a difference between stated individual privacy preferences and their actual behavior [20], [21]. This disconnect may be better addressed by encouraging students to better understand their individual mobile devices and mobile cybersecurity threats in relation to certain behaviors that may make them more vulnerable. The Federal Communications Commission offers information and a customizable tool for creating security tips based on specific mobile operating systems which could be beneficial for improving cybersecurity awareness [22].

In response to Q3: What is the relationship between privacy concerns and reading privacy policies? Students should be reminded that they risk their online security if they have not taken proper precautions to protect themselves and their personal information. The survey results showed that 94 (31.23%) female students reported to have never read a privacy policy, but 86 (91.48%) of these female students went on to report being very concerned or concerned if their service provider shared their exact location with third parties. Similarly, 97 (34.28%) male students reported to have never read a privacy policy, but 77 (79.38%) of these male students went on to report being very concerned or concerned if their service provider shared their exact location with third parties. This seems to indicate that privacy policies may not be an effective tool for notifying end-users about their commercial data practices or a general lack of cybersecurity awareness. Also, it would be instructive to differentiate the concept of privacy from confidentiality. As [8] points out, privacy is related to a person and their sense of control of access that others have; whereas confidentiality relates to data about an individual and the security goal that refers to limiting information access and disclosure of data to authorized users. This disregard of privacy policies could be more actively addressed in general education computer science courses with case studies, discussions and role-playing. Students could discuss the value of collecting personal information for business purposes. Also, students should consider the legal, security and ethical implications of collecting personal information, creating enforceable privacy policies and terms of service. Greater emphasis on the value of personal information and decision making with sharing personal information should become a part of the general education computer science curriculum to better prepare students for their mobile lifestyles.

## 4 LIMITATIONS

Citizenship was not considered in the survey and international students with different cultural, political or privacy expectations may have had a marginal impact on the results. Only general questions were asked about privacy concerns and behaviors. More specific questions related to particular social networking applications, mobile gaming applications, etc., may lead to different results due to student decisions for short term benefits

versus long-term costs. Strict random sampling was not applied, but these survey results should be considered valid and acceptable, in describing the behaviors and concerns of this sample [14]. This survey did not capture if any cybersecurity events (e.g., identity theft, media reports about commercial data breaches, etc.) may have influenced student concerns. Follow-up surveys or interviews would offer a more precise understanding of some students' decision making about their behaviors and attitudes.

## 5 CONCLUSION

Overall, some students did report taking proactive steps in protecting their personal information (i.e., disabling location awareness, always partially reading privacy policies and uninstalling applications). The privacy incongruity between their stated concerns and their actual behaviors should still be a concern for future employers, and should be more actively addressed by educators. Security managers should also be aware of this incongruity because these students represent the future work force, which may need additional training in understanding employment policies (e.g., bring your own device, operational security, etc.) or a need for improved cybersecurity awareness. Female respondents reported more concerns about their privacy but did not necessarily report more conservative behavior when sharing information or reading privacy policies. Future work could examine if social pressures impact female students to feel obligated to share more personal information. Entire courses on end-user privacy and security may not be necessary, but course injections on usable privacy and security could be enhanced throughout the general computer science education classes. Increased emphasis could be placed on encouraging students to increase their cybersecurity awareness by proactively taking responsibility for their personal information, as well as understanding legally binding privacy policies or terms of service agreements. Future work could investigate privacy by design concepts and the emphasis on usable privacy and security education in software engineering, networking and database courses. Specifically, there should be an investigation into the training of new software developers in creating effective privacy policies that could be evaluated in a usability study by mobile end-users.

## 6 REFERENCES

- [1] "Stop.Think.Connect. Cyber Tips," (U.S. Department of Homeland Security), [online] <http://www.dhs.gov/stopthinkconnect-cyber-tips> [Accessed: 17 April 2014].
- [2] "Office of Justice Programs: Identity Theft," (U.S. Department of Justice). [online] <http://ojp.gov/programs/identitytheft.htm> [Accessed: 17 April 2014].
- [3] "How to keep your personal information secure," (Federal Trade Commission), [online] <http://www.consumer.ftc.gov/articles/0272-how-keep-your-personal-information-secure> [Accessed: 17 April 2014].

- [4] J. Lawler and J. Molluzzo, "A survey of first-year college student perceptions on privacy in social networking," *Journal of Computing Sciences in Colleges*, vol. 26, issue 3, pp. 36-41, January, 2011.
- [5] A. Lenhart, M. Madden, S. Cortesi, U. Gasser, and A. Smith, "Where teens seek online privacy advice," (PewResearch Internet Project), [online] 15 August 2013, <http://www.pewinternet.org/2013/08/15/where-teens-seek-online-privacy-advice.html> [Accessed: 10 April 2014].
- [6] V. Romney and G. Romney, "Neglect of information privacy instruction – A case of educational malpractice?," in *Proceedings of the 5<sup>th</sup> Conference on Information Technology Education*. Salt Lake City, UT, 2004, pp. 79-82.
- [7] M. Rowan and J. Dehlinger, "A systematic literature review on using mobile computing as a learning intervention," in *Proceedings of the 18<sup>th</sup> ACM Conference on Innovation and Technology in Computer Science Education*, Canterbury, UK, 2013, p. 339.
- [8] J.Vaidya, B. Shafiq, D. Lorenzi, and N. Badar, "Incorporating privacy into the undergraduate curriculum," in *Proceedings of the Information Security Curriculum Development Conference*. Kennesaw, GA, 2013, pp. 1-7.
- [9] S. Ovaska and K. Raiha, "Teaching privacy with ubicomp scenarios in HCI classes," in *Proceedings of the 21<sup>st</sup> Annual Conference of the Australian Computer-Human Interaction Special Interest Group*. Melbourne, AUS, 2009, pp. 105-112.
- [10] L. Cranor and N. Sadeh, "A shortage of privacy engineers," *IEEE Security & Privacy*, vol. 11, no. 2, pp. 77-79, March-April, 2013.
- [11] C. Dierbach, H. Hocheiser, S. Collins, G. Jerome, C. Ariza, T. Kelleher, W. Kleinsasser, J. Dehlinger, and S. Kaza, "A model for piloting pathways for computational thinking in a general education curriculum," in *Proceedings of the 42<sup>nd</sup> ACM Technical Symposium on Computer Science Education*. Dallas, TX, 2011, pp. 257-262.
- [12] M. Rowan and J. Dehlinger, "Research Trends and Open Issues in Mobile Application Software Engineering," in *Proceedings of the 2013 International Conference on Software Engineering Research and Practice*, Las Vegas, NV, 2013, pp. 38-44.
- [13] S. Peltsverger and G. Zheng, "A virtual environment for teaching technical aspects of privacy," in *Proceedings of the Information Security Curriculum Development Conference*. Kennesaw, GA, 2013, pp. 49-53.
- [14] J. Lazar, J. Feng, and H. Hocheiser, "Surveys," *Research Methods in Human-Computer Interaction*, West Sussex, UK, Wiley & Sons, 2010, pp. 100-107.
- [15] J. Grossklags and N. Good, "Empirical studies on software notices to inform policy makers and usability Designers," in *Proceedings of the 1<sup>st</sup> International Conference on Usable Security*, Scarborough, Trinidad and Tobago, 2007, pp. 341-355.
- [16] A. McDonald and L. Cranor, "The Cost of Reading Privacy Policies," *I/S: A Journal of Law and Policy for the Information Society*, vol. 4, no. 3, 2008, pp. 540-565.
- [17] G. Meiselwitz, "Readability assessment of policies and procedures of social networking sites," in *Proceedings of the 5<sup>th</sup> International Conference on Online Communities and Social Computing*, Las Vegas, NV, 2013, pp. 67-75.
- [18] R. Bohme and S. Kopsell, "Trained to accept?: a field experiment on consent dialogs," in *Conference on Human Factors in Computing Systems*, Atlanta, GA, 2010, pp. 2403-2406.
- [19] A. Acquisti and J. Grossklags, "What can behavioral economics teach us about privacy?," *Digital Privacy: Theory, Technologies, and Practices*, Boca Raton, FL, Auerbach Publications, 2008, pp. 363-380.
- [20] A. Acquisti and J. Grossklags, "Privacy and rationality in individual decision making," in *IEEE Security & Privacy*, pp. 24-30, January, 2005.
- [21] B. Berendt, O. Gunther, and S. Spiekermann, "Privacy in e-commerce: stated preferences vs. actual behavior," *Communications of the ACM*, vol. 48, no. 4, 2005, pp. 101-106.
- [22] "Create your own smartphone security checklist," (Federal Communications Commission), [online] <http://www.fcc.gov/smartphonesecurity> [Accessed: 15 April 2014].

## **SESSION**

# **SPECIAL TRACK ON WIRELESS NETWORKS SECURITY + MODELING OF INFORMATION SECURITY**

### **Chair(s)**

**Dr. Hanen Idoudi**

**Manouba Univ. - Tunisia**

**Dr. Samiha Ayed**

**Telecom Bretagne - France**



# Security Considerations in WSN-Based Smart Grids

Hanan Idoudi<sup>1</sup>, Mustafa Saed<sup>2</sup>

<sup>1</sup>National School of Computer Science, University of Manouba, Tunisia  
hanan.idoudi@ensi.rnu.tn

<sup>2</sup>Electrical and Computer Engineering, University of Detroit Mercy, USA  
saedma@udmercy.edu

**Abstract**—Wireless Sensor Networks (WSNs), which are composed of battery powered devices, are attracting a tremendous attention owing to their wide range of applications. Recently, their use in the smart grid to respond to several communication needs was stressed. A Smart Grid is an innovative paradigm to enhance the power grid system with communication capabilities in order to perform several tasks of monitoring and surveillance. Despite their advantages, there are several challenges facing WSN applications in the Smart Grid. Security is one of the most critical challenges. In this paper, the application of WSNs in smart grid is reviewed, and the security issues accompanying their use are discussed.

**Keywords**—Smart Grid, Wireless Sensor Networks, Security Threats

## I. INTRODUCTION

The Smart Grid is intended to provide a next-generation electrical network supported with configurability, assessment and self-monitoring features. This new paradigm relies on information and communication technology to perform these tasks.

Wireless technologies and Wireless Sensor Networks (WSNs) for instance, were proposed as an interesting and efficient support for communications in Smart Grid.

Wireless Sensor Networks consist of small resource constrained devices that can organize themselves into a multi hop wireless network [1]. WSNs are used in a wide range of potential applications including military, medical systems, and robotic exploration. This in turn explains why researchers in the field were motivated by these types of networks in their research work.

Recently, WSNs have been adopted in Smart Grids [3, 4] because they feature rapid deployment, low cost and flexibility, and aggregated intelligence via parallel processing. While existing remote sensing, monitoring, and fault diagnostic solutions are too expensive, WSNs provide cost-effective sensing and communication solution in a remote and online setting.

WSNs are applied by utility companies and suppliers for

substation automation management in addition to wireless automatic meter reading (WAMR) system. They are intended to measure and monitor the energy usage or the power lines quality and in some cases, provide a real time feedback on systems misbehavior to allow timely fault detection and avoidance tasks. Advantages of WSN applications in Smart Grid are numerous since they can offer timely and efficient monitoring of the power grid and can reduce operational costs by eliminating the need for human readers and provide an automatic pricing system for customers [5].

On the other hand, WSNs brought new challenges and issues when applied to the Power Grid. Serious security threats induced by these types of communication technologies must be dealt with efficiently to prevent the malfunctioning of the vital power grid.

The rest of this paper is organized as follows: the overview of the Smart Grid paradigm is presented in section 2. Section 3 discusses the WSNs characteristics and their applications in Smart Grid, and section 4 reviews the WSN-based Smart Grid security issues.

## II. SMART GRID OVERVIEW

The smart grid is a new architecture for electric power grid infrastructure using sophisticated transmission and distribution communication networks to deliver electricity [9], [10]. The goal of smart grid is to improve the efficiency, reliability and safety of the electric systems through the new architecture of communication technologies, automated control, and dynamic optimization of electric system operations, maintenance, and planning. The main characteristics of the smart grid include information technology, automation, and interaction to provide the two-way data communication. Fig. 1 shows typical smart grid infrastructure (the combination of power system and the information

system of smart grid). It illustrates network connections of the end to end smart grid system. These connections pinpoint the communication and data management from the customer to collector to utility control center and to transmission and distribution substations where the electronic controllers are located.

The electronic controllers manage the generation and flow of electrical power. The residence block in the figure demonstrate the home area network and the components are used to communicate to smart grid network, such as smart thermostat, smart water heater, smart appliances and smart meter. There are two ways of the wireless connection in the home area network devices to a smart controller/meter through a network; Zigbee and Wi-Fi.

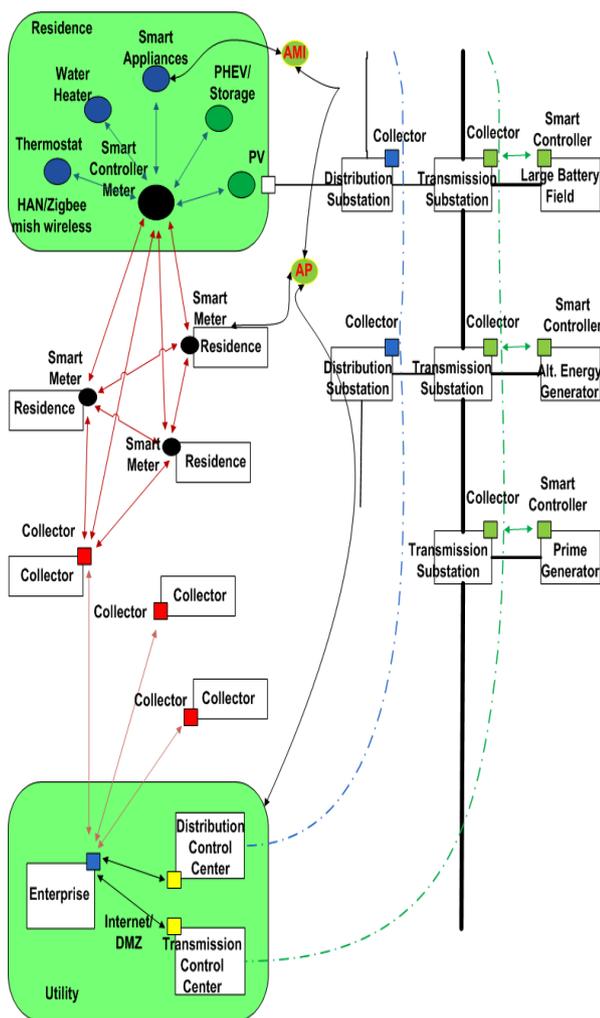


Fig. 1. A block diagram for typical smart grid

Collector nodes communicate with the utility through common communication mechanisms to provide readings and troubleshooting via Internet. In the utility block the communication will be accomplished through intranet by using Zigbee wireless technology including a Demilitarized Zone (DMZ), which is designed to prevent the flow of unauthorized messages. The distribution and transmission control centers have legacy communication paths and additional smart grid communication paths [11].

### III. WIRELESS SENSOR NETWORKS AND SMART GRID

Wireless sensor networks (WSNs) are composed of small and resource-constrained devices communicating wirelessly and organized according to a multi-hop wireless network to route information to a specific control unit called sink. A node in the WSN has one or more sensors, embedded processor, moderate amount of memory and transmitter/receiver circuitry. Because of their wide range of potential applications including military, medical systems, precision agriculture and robotic exploration, wireless sensor networks have become a promising technology for Smart Grids.

In this section, the characteristics of WSN when applied in Smart Grids are reviewed, and then an overview of their applications in a Smart Grid context is covered.

#### A. WSN-based Smart Grid benefits and challenges

To perform communications in the Smart Grid, several technologies can be used [3]. However, given the adverse environment in which the Smart Grids are deployed, using a wireless infrastructure is the most suitable solution for this type of networks. Moreover, owing to their autonomous behavior, large-scale pervasive and inexpensive nature, wireless sensor networks (WSNs) are good candidate in such context.

Advantages of WSN applications in Smart Grid are abundant. These applications can offer timely and effective monitoring of the power grid, lower operational costs by abolishing the need for human readers, and offer an automatic pricing system for customers. WSNs are characterized by a straightforward and prompt deployment, low cost and flexibility and accumulated intelligence via parallel processing. They can deliver an interesting alternative to current remote sensing, monitoring and fault diagnostic solutions which are too expensive. On the other hand, WSNs provide cost-effective sensing and communication solution in a remote and online style.

Since sensor nodes are usually battery powered, conserving their energy and prolonging the network life time are considered primary goals for most of their applications when designing protocols for those networks. However, with the Smart Grids, this goal is no longer an issue since sensors can benefit directly from the power grid to be continuously powered. Hence, when establishing communications through WSNs within the context of Smart Grids other concerns must be taken into account. Deploying WSNs in a Smart Grid context introduces a number of challenges including:

- The consideration of the noisy environment produced by the electrical elements within the Smart Grid sensor nodes at the time of transmitting their data on the network.
- Taking into account the heterogeneity of both the sensor nodes and the traffic sent by the nodes

according to the features for which the sensor node was designed.

- The need to ensure interoperability between WSNs and other existing wireless technologies to improve the reliability of data transmission to the control center.
- The security consideration which is already critical in Smart Grid and more critical with WSN.

#### B. WSNs applications in Smart Grid

WSNs are anticipated to be used essentially for Wireless automatic meter reading (WAMR) and Remote System Monitoring and Equipment Fault Diagnostics. Liu [5] classified the applications of WSNs into three categories: power generation, power delivery, and power utilization.

Since sensors are low cost and easy to deploy devices, they can be used in power generation units to measure several parameters such as steam, temperature, and air/fuel flow rates. This information is fed into the data acquisition system in the power plant for monitoring purposes to control generators' operation and prevent any malfunctioning through timely reporting of any faulty component or misbehavior.

WSNs can avoid or greatly alleviate power-grid and facility breakdowns when deployed along the power delivery systems for monitoring purposes. They can report on outage, abnormal activities and parameters thresholds, and allow timely maintenance. WSNs are envisioned to reduce electric utility operational costs by eliminating human readers. They can offer an online pricing system based on online energy consumption monitoring of customers. Besides, several home and building automation applications can benefit from WSNs including recommendation or regulatory systems for controlling power consumption of building and facilities [2],[6],[7].

### IV. SECURITY CONSIDERATIONS IN WSN-BASED SMART GRID

Security of the WSN communications is one of the most important issues to deal with. WSNs suffer from many security threats due to their inherent characteristics. As is the case with all kinds of wireless networks, WSNs are more vulnerable to security threats originating from the open communication environment than wired networks. Unlike other wireless technologies, such as Wi-Fi, applying advanced and complex security mechanisms is not relevant due to their physical limitations. This complicates protection measures and makes WSNs more vulnerable to external attacks. Thus, WSN-based Smart Grid suffers from all the security threats facing classical WSN communications in addition to new security vulnerabilities. This complication will result in a large set of vulnerabilities to overcome.

#### A. Security threats in WSN

Many types of attacks on WSN exist. In a selective

forwarding attack, malicious nodes may refuse to forward certain messages and destroy them, ensuring that they are not propagated any further. A simple form of this attack is when a malicious node behaves like a black hole and refuses to forward every packet it receives. By this, neighboring nodes will conclude that the communication has failed and decide to seek another route.

In sinkhole attack, attacker advertises incorrect information, such as high quality route to a sink. An attacker can actually provide this kind of route connecting all nodes to real sink and then selectively drop packets. Because of the communication pattern (all traffic is directed to sink), WSNs are highly susceptible to this kind of attack since all the surrounding nodes of the adversary will start forwarding packets destined for a sink through the adversary, and also propagate the attractiveness of the route to their neighbors.

Sybil attack consists of a single node that pretends to be present at different parts of the network. The malicious node illegitimately presents multiple identities to other nodes in the network. The Sybil attack can significantly decrease the effectiveness of fault tolerant schemes, such as distributed storage, disparity and multipath routing, and topology maintenance.

Wormholes may convince two nodes to be neighbors when in fact they are far away from each other. Wormholes may convince distant nodes that they are close to sink. This may lead to sinkhole if a node on the other end advertises high-quality route to that sink. Well placed wormhole can completely disorder routing since attackers may influence network topology by delivering routing information to the nodes before it would really reach them by multi hop routing. Wormholes may be used in conjunction with Sybil attack.

Nodes in WSNs learn about their neighboring nodes through HELLO packets, which are required by many WSN routing protocols. In a Hello flood attack, attackers can broadcast HELLO message to nodes and then advertise high-quality route to sink. Some routing protocols use link layer acknowledgments. Attackers may spoof acknowledgements to convince other nodes that a weak link is strong or that a dead node is alive. Consequently, a weak link may be selected for routing causing packets sent through that link to be lost or corrupted.

#### B. Security threats in WSN-based Smart Grid

The smart grid inherited all of the vulnerabilities of the traditional internet as a result of connecting the power grid to the network. These vulnerabilities demand high security measures to protect the smart grid. Through information technology, the smart grid allows customers to manage their energy services and access smart grid convenience features. However, this can cause damage to the smart grid system and

increase the possibility of cyberattacks and cascade failures propagating from one system to another [12]. The current electronic devices of the power grid do not support cryptography and data security. When the power grid was built, engineers and designers did not consider the security implementation in the electronic design. This was because there was no external communication to these electronic devices. Also, the communication between the customers and the power grid is a one-way communication, from the power grid to the customers. Therefore, the electrical power system does not have extra capacity to perform any security function. New security concerns arise when switching to the smart grid as a result of merging industrial control system (ICSs) and information technology (IT) [13]. Industrial control systems have been built without any consideration to security concerns. However, Information Technology considers security as a high priority. This requires changes to the ICS to upgrade the hardware and software of the power grid. Extra care must be taken to ensure these components will not increase vulnerabilities. Both ICS and IT personnel need to communicate with each other. There are considerable differences between the two technologies. The IT staff uses the patching server to update the system or add another level of access to the system and then forward the required update to all the hosts. This will cause the system to restart. Consequently, this will force the power grid to eventually shut down and cause blackout.

Another issue has to do with the power grid using a dedicated serial line that has low speed and very limited access. IT uses the TCP/IP protocol for communication in the Ethernet and Wi-Fi connection. This is characterized by high-speed communication and multiple accesses at the ftp server, telnet client, and web server. Consequently, many benefits to remote access and troubleshooting will be provided. Unfortunately, this will also create a new vulnerability for cyber-attacks.

The smart grid technology uses many new devices, such as Advance Meter Infrastructure (AMI), Smart Meter, and Demand Response, which allow customers to access their bills anytime in order to monitor the power consumption of home appliances and decide which one to shut down, turn on, or even schedule the appliances operation time. This will create security and privacy concerns. For example, if nobody is at home, a sophisticated cyber-attacks may result in tampering with electrical appliances. Advance Meter Infrastructure (AMI) device will be installed in residences to provide two-way communication between the customer and the power grid control center

#### 1) Advanced Metering Infrastructure Security

Advanced Metering Infrastructure (AMI) is a part of the smart grid architecture and has been used to intelligently monitor and control power distribution and consumption [14]. AMI acquires and delivers fine-grained electricity measurements using two way communications through smart meters or other energy management devices to support real data reading in real-time services and dynamic load control in the end to end communication between end-users and energy providers.

The AMI plays a major role in controlling communications with the consumers. In AMI architecture there are two ways of the connection of the smart meters to collectors using either indirect connection or direct connection. In indirect connection, some smart meters are connected to other smart meters, which are in turn connected to the collector. In direct connection all smart meters are directly connected to the collector. In both connections, the collectors are connected to substations. Securing the AMI is vital to convince residents and business owners that smart grids are reliable and trustworthy. This implementation requires an AMI security profile to define and address all security concerns and find the solutions for these concerns. Setting up an AMI in each house and having it acting as an access point will demand monitoring, managing, and maintaining such devices [15]. The smart grid should deploy cryptography and key management to build a robust security implementation. Some of the smart grid devices permit physical access to the customer, such as smart meter, and advanced meter infrastructure. This may possibly lead to potential cryptanalysis. Furthermore, cryptography implementation requires complex hardware and software design support. All customer information and the power system must be encrypted to protect privacy and maintain integrity, and confidentiality [16]. Cryptography and key management shall be an appropriate solution to protect the smart grid from cyber-attacks, and tackle the physical access concerns of the smart grid security.

Fig. 2 depicts the direct and indirect smart meter-to-collector communication topology. The collector (C) is the center point between the substation and the smart meters (SMs) in direct connection, but in indirect connection the smart meters are connected in series to the collector, which is in turn connected to the substation.

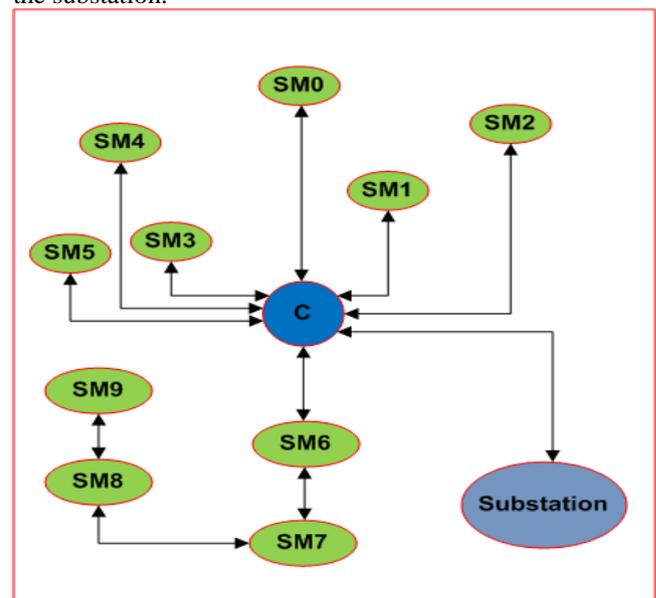


Fig. 2 AMI Architecture- Smart Meters Connections Topology

## 2) Utility Security

The software and the firmware will be managed by the utility, and data management between the customers and the substation will be carried out in the utility as well. This will increase the security concerns of hacking the system and vulnerabilities. Some signals travel with various vulnerabilities to many end-user devices through media networks, such as controlling and monitoring. In addition, concerns have been raised regarding the resistance of the smart grid and how it repairs itself without resulting in equipment or infrastructure damage, or blackout. The software and firmware in the system are responsible for protecting the system from unauthorized access by people, who are trying to intrude the system and tamper with databases and other information. For any application that needs to be applied in the smart grid, the secure software development life cycle should be taken into consideration. This will help to avoid the lack of oversight in this area and mitigate possible vulnerabilities to achieve security such as authentication vulnerability, authorization vulnerability, cryptographic vulnerability, input and output validation, password management vulnerability, link vulnerability and protocol errors.

## 3) Substation, Distribution, Transmission and Generation Security

Combining the information technology and electricity grid into smart grid network provides more efficient and reliable grid operations, and thus leads to a great many benefits, ranging from energy savings to a high degree of home automation. In spite of all benefits of the smart grid, the new technology it introduced exposes the system to many new threats. The traditional power grid automation system has been physically isolated from the corporate network. However, smart grid networks shall allow power grid automation to connect to public networks. This kind of connection will increase the vulnerability of hacking the power grid automation.

The current architecture of the power grid automation does not support the security needed to deter cyber-attacks. Wei et al. [17] proposed a new framework by introducing an additional layer of security to protect the power grid automation system from hackers and unauthorized people. They presented the needed security and safety when the power grid automation system is connected to public networks or the cloud, and divided the security layer into three major parts:

- i. Security agents: The agents provide protection to the edge of the system to secure the network from the cyber-attacks. The security agents in the Control Center are more intelligent and complex than the security agents in the Intelligent Electronic Devices (IEDs).
- ii. Managed Security Switch: This connects the Substations in the Control Center.
- iii. Security Manager: The manager is located in the automation network and connected to switches using current IT security implementation.

## V. DISCUSSION

Much work is yet to be done to enhance the security of the WSN-based smart grid network. In particular, the following should be taken care of:

- 1) Deploying the public key infrastructure (PKI) in the smart grid [18]. Using sensors brings limitations when using cryptographic solutions since lightweight cryptographic mechanisms should be designed to cope with physical and logical limitations of sensors.
- 2) Securing the trusted device profile and devising the smart grid certificate lifetime.
- 3) Ensuring the privacy concerns regarding customer information are resolved [19].
- 4) Following a robust security style for the smart grid as a future priority to achieve appropriate authentication for any device communication via the smart grid.
- 5) Addressing all the newly discovered vulnerabilities of the WSN-based smart grid by monitoring and tracking the communication and the data flow through the smart grid.

## V. CONCLUSION

Smart grid is the next generation of power line networks. Supplying them with sensors is a key solution to support efficient and timely communication in Smart grid. Despite their advantages, WSN bring new security challenges that add to the existing smart grid security concerns. In WSN-based smart grid, both intrinsic smart grid security threats and WSN vulnerabilities should be addressed.

## REFERENCES

- [1] Akyildiz, I., Su, W., Sankarasubramaniam, Y., and Cayirci, E. Wireless sensor networks: A survey. *Computer Networks* 38 (4), 2002, 393-422.
- [2] M. Erol-Kantarci, H.T. Mouftah. Wireless sensor networks for cost efficient residential energy management in the Smart Grid. In: *IEEE Transactions on Smart Grid*, 2011, June, No. 2, p. 314-325.
- [3] V.C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella; C. Ceca-ti, G.P. Hancke, "Smart Grid technologies: communication technologies and standards," In: *IEEE Transactions on Industrial Informatics*, 2011, November, No. 4, p.529-539.
- [4] V.C. Gungor, Bin Lu, G.P. Hancke, "Opportunities and challenges of wireless sensor networks in Smart Grid," In: *IEEE Transactions on Industrial Electronics*, 2010, October, No. 10, p. 3557-3564.
- [5] Yide Liu. *Wireless Sensor Network Applications in Smart Grid: Recent Trends and Challenges*. International Journal of Distributed Sensor Networks, Volume 2012.
- [6] G. Kalogridis, C. Efthymiou, S.Z. Denic, T. A. Lewis, and R. Cepeda. Privacy for smart meters: towards undetectable appliance load signatures. *IEEE International Conference on Smart Grid Communications*, pp. 232-237, USA, 2010.
- [7] Joel Hoglund, Dejan Ilicy, Stamatis Karnouskosy, Robert Sauterz, and Per Goncalves da Silvay. Using a 6LoWPAN Smart Meter Mesh Network for Event-Driven

Monitoring of Power Quality. IEEE International Conference on Smart Grid Communications, 2012.

[8] Michael Zillgith, Simon Fey, Pascal Benoit, Stefan Feuerhahn, Robert Kohrs. Wireless IP Networks in Smart Grid Applications. IEEE International Symposium on Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems, September 2012, Germany.

[9] V. C. Gungor, et al., "Opportunities and Challenges of Wireless Sensor Networks in Smart Grid - A Case Study of Link Quality Assessments in Power Distribution Systems," Industrial Electronics, IEEE Transactions on, vol. PP, pp. 1-1,2010.

[10] H. Khurana, et al., "Smart-Grid Security Issues," Security & Privacy, IEEE, vol. 8, pp. 81-85,2010.

[11] F. Cleveland, "Enhancing the Reliability and Security of the Information Infrastructure Used to Manage the Power System," presented at the IEEE PES General Meeting, Tampa, FL, Jun. 24-28, 2007.

[12] L. Wang, C. Li, H. Cheung, C. Yang, and R. Cheung, "PRAC: A Novel Security Access Model for Power Distribution System Computer Networks," presented at the IEEE PES General Meeting, Tampa, FL, Jun. 24-28, 2007..

[13] S. Clements, and H. Kirkham, "Cyber-Security Considerations for the Smart Grid," presented at the IEEE PES

General Meeting, Minneapolis, MN Jul. 25-29, 2010.

[14] H. Sui, H. Wang, M.-S. Lu, and W. Jen Lee, "An AMI system for the deregulated electricity markets," IEEE Transactions on Industry Applications, vol. 45, no. 6, pp. 2104-2108, 2009.

[15] Guidelines for Smart Grid Cyber Security, NISTIR 7624. 1-2010.

[16] G. N. Sorebo, and M. C. Echols, Smart Grid Security: An End-to-End View of Security in the New Electrical Grid. Boca Raton, FL: CRC PRESS, 2012, pp. 1-79.

[17] D. Wei, Y. Lu, M. Jafari, P. Skare, and K. Rohde, "An Integrated Security System of Protecting Smart Grid against Cyber Attacks," presented at the 2010 Conf. Innovative Smart Grid Technologies (ISGT), Gaithersburg, MD.

[18] M. Zhao, S. Smith, and D. Nicol, "Evaluating the Performance Impact of PKI on BGP Security," presented at the 2005 4th Annual PKI Research and Development Workshop, Gaithersburg, MD.

[19] A. Barengi and G. Pelosi, "Security and Privacy in Smart Grid Infrastructures," presented at the 2011 22nd Int. Workshop Database and Expert Systems Applications, Toulouse.

# Security of Online Social Networks

Rihab Ben Aicha<sup>1</sup> and Hanen Idoudi<sup>2</sup>

National School of Computer Science, University of Manouba, Tunisia

<sup>1</sup>ben.aicha.rihab@gmail.com, <sup>2</sup>hanen.idoudi@ensi.rnu.tn

**Abstract**—Online Social Networks are among the most used Internet services. Nonetheless, their increasing popularity is facing a tremendous increase in security breaches that threaten these applications. Being a huge data warehouse for personal and very private information, ensuring privacy is the most important challenge for OSN.

In this paper, we present some security aspects and some privacy issues for Online Social Networks. We review some important vulnerabilities and threats. Then, we show security analysis results of some popular OSN.

**Keywords**—OSN; web application security; privacy; Web Vulnerability scanners.

## I. INTRODUCTION

Over the last few years, a surprising number of web applications that are vulnerable to hackers was increasing permanently and proportionally to the number of the Internet users. This could be a result of minimal attention given to security risks while developing and deploying web applications. Online Social Networks (OSN) in particular were one of the hackers' favourite targets. In fact, OSNs were dedicated to help people stay in touch by reconnecting with friends from way-back-when, establish new relationships with others and exchange knowledge with them as well as share posts publicly [1]. To join an OSN you will have to create an account and provide a massive amount of personal information that will be stored in their servers and never be removed which could threaten your private life. People don't give enough importance to those threats [2]. They believe that they don't have anything to hide or that is worth hacking or, they have enough trust in the web sites they use. So they don't worry about security but they forget that nowadays we are connected to each other more than ever. By ignoring security you are not only putting yourself at risk but others as well.

Despite the diversification of OSNs purposes, they are offering a synchronization that means bringing together all your activities and your contacts from different OSNs in one central location. As the OSNs aim at collecting a lot of personal information, the risk of leaking this data gets higher.

Therefore, this can be considered as a threat for privacy especially when it comes to entities interested in collecting personal data either for legal or illegal purposes. They can

easily get all they want without making huge efforts.

The success of these web sites makes them gaining a large popularity. They are the first destination of the Internet users especially teenagers whose lack of awareness and absence of parental control makes them the most vulnerable to privacy breaches. Many of the famous OSNs such as Facebook, Twitter, LinkedIn and MySpace got involved in privacy breach issues. They are facing accusations of intercepting communications for their profit at the expense of users or non-users. Facebook, for example was sued many times in 2011 which makes the firm pay about \$20 million to compensate its users for using their data without explicit permission.

In this paper, we will give an overview of the threats facing OSNs, then, we will analyze some OSNs' vulnerabilities and threats that can be used against the users' privacy.

The remainder of this paper is organized as follows. Section 2 will present some threats and vulnerabilities that threaten OSN's security. Next, we will introduce some of the well known web vulnerability scanners with a comparison between their main functionalities. Then, in section 4, we will give results of vulnerabilities analyses of some OSNs with an interpretation of the outcome and a comparison between those web sites. Finally, we will conclude in section 5 and summarize our findings.

## II. OSN SECURITY THREATS AND VULNERABILITIES

OSN security is about protecting data and sensitive information from those with malicious intentions. To protect this information we will have to know about the different vulnerabilities and attacks that could threaten us as OSN users. In this section, we will go through the basics of information systems security as well as the vulnerabilities and attacks that threaten a web application in general and Online Social Networks in particular.

### A. Fundamentals of information security

Information systems security relies on the following core principles:

- **Confidentiality** is about restricting the access to the users' sensitive information to those who have permissions.
- **Integrity** is about ensuring that information remains

consistent and never been modified by unauthorized people.

- **Availability** defines the fact that the information is accessible in the right time and place by authorized people.
- **Authenticity** is guaranteeing that information is from the source that is claimed to be.
- **Non-repudiation** is a service that generally relies on a digital certificate used to ensure that the sender and the recipient of a message are both purported to be, so that they cannot deny it later.

These five points represent the requirements that have to be guaranteed for information security. The absence of one of them can make the system vulnerable to attacks and threats.

*B. Web applications Attacks and Vulnerabilities*

Vulnerability is a web application weakness that could be exploited by attackers to compromise the system security. For OSN, users should be aware about different risks that could threaten their privacies.

Open Web Application Security Project community (OWASP) is a non-profit organization that aims to improve Internet software's security. The following table presents the top 10 threats' ranking for both current and previous versions [3].

TABLE 1: OWASP RELEASE NOTES

OWASP TOP 10 -2010 (Previous Version)	OWASP TOP 10 -2013 (Current Version)
A1-Injection	A1-Injection
A3-Broken Authentication and Session Management	A2-Broken Authentication and Session management
A2-Cross Site Scripting (XSS)	A3-Cross-Site Scripting (XSS)
A4-Insecure Direct Object Reference	A4-Insecure Direct Object References
A6-Security Misconfiguration	A5-Security Misconfiguration
A7-Insecure Cryptographic Storage - Merged with A9 -->	A6-Sensitive Data Exposure
A8-Failure to Restrict URL Access - Broadened into -->	A7-Missing Function Level Access Control
A5-Cross Site Request Forgery (CSRF)	A8-Cross-Site Request Forgery (CSRF)
<buried in A6: Security Misconfiguration>	A9-Using Components with Known Vulnerabilities
A10-Unvalidated Redirects and Forwards	A10-Unvalidated Redirects and Forwards
A9-Insufficient Transport Layer Protection	Merged with 2010-A7 into 2013-A6

Differences between these two versions are caused by the evolution of the technology's weaknesses and the number of technology users. Increase of the complexity of

systems and the tremendous growth of number of CyberCriminals, attackers and security geeks are among reasons for this difference. Vulnerabilities could cause different damage with different effects and in different levels once they are exploited by CyberCriminals to launch different attacks.

Here-after are some of the notable vulnerabilities that should be taken into consideration.

- **Information Leakage** is a weakness that affects applications. It means that sensitive information is being accessed and used for a third-party benefit.
- **Spams** are unwanted messages sent to users by some internet users who use electronic messaging systems. OSN spammers use social networks to send advertising messages to other users by creating fake profiles [4].
- **Phishing attack** is a sort of a scam on the Internet. It is about stealing users' valuable information by deceiving them and making them believe that they are receiving Emails from legitimate web sites [4].
- **Sybil attack:** is a security threat based on forging identities. This type of attack relies on the fact that a user pretends to be multiple nodes in the system in order to use these fake identities in malicious activities [4].
- **Malicious software** generally known as Malware is a software designed to sneak into networks. It exploits the OSN structure to spy on the user's connections and gather their sensitive information.
- **XSS** is an abbreviation of Cross-site Scripting. It is one of the common attacks used against web applications. This type of vulnerability allows attackers injecting malicious scripts into the application web pages. Those scripts will execute some unwanted tasks on the client-side which causes information theft.
- **SQL injection** is one of the hackers favourite attacks used against web applications. It is based on injecting SQL queries into the application in order to execute them on the database and steal valuable information.
- **Clickjacking** or Hijacking clicks on the Internet is a malicious technique used by an attacker to make the user click something (link, button, page...) different from what they intended to click and get tricked.
- **De-anonymization** is an attack based on many techniques such as network topology. Attackers use the De-anonymization to uncover the real identity of users who want to stay anonymous in social networks in order to protect their privacies using pseudonyms instead of using their real names.
- **Identity clone attacks:** Attackers duplicate user's presence in the same OSN in order to convince his friends and make with them a trusting relationship which facilitates the collection of information.
- **Botnet attack**, short for robot and network, is about taking control over several computers using malicious software. Attackers will manage these computers remotely and run some attacks like the denial-of service (DOS). The risk of being infected by such an attack is higher over OSNs which offer an open platform to their users in order to deploy applications on it [5].

### III. COMMON OSNs ATTACKS AND VULNERABILITIES

OSNs are more vulnerable than any other web application because of the great amount of personal information they get. In this sub-section, we will discuss some specific OSN vulnerabilities and we will review the most known attacks against the most popular OSNs today, namely, Facebook, Twitter and LinkedIn.

Actually, children and teenagers are among the large number of victims in the OSNs. The information they share is usually misused and potentially abused, which jeopardize their privacies. This is due to unawareness of technological weakness and their naivety so they tend to trust easily social network relationships. For instance, they are highly threatened by the Internet paedophiles or online predators [6].

A surprising recent Facebook statistics were published about this issue by the SociallyActive [7] in 2013 and shows that 55% of teens have shared personal information, photos and even physical descriptions with strangers. 24% of teens get embarrassed due to a leak of some private information on their behalf and over one in four teens were victims of stalking on Facebook.

Facebook, as one of the most popular OSNs, broke the privacy policies many times and was sued [8]. Let's specify briefly some of its users' claims. The firm used users' names and photos in advertisement without any permission in addition to giving personal information like age, phone number and addresses to application developers. A Facebook user can delete his account but actually, Facebook never remove its content from their servers. Furthermore, the firm released a new facial recognition system for photos that can tag a user automatically and without his confirmation. Facebook friends could check each other locations without prior permission. Also, a user can be added to groups without asking for consent. A user can manage posts visibility to friends, friends of friends or public; but his posts would be shared with anyone who followed him. In 30, December 2013, Facebook was sued by two users for scanning [9] their private messages and sharing the information with advertisers and marketers for a profit.

Facebook still breaches users' privacy but it is not the only OSN that was sued for alleged privacy issues. Twitter and LinkedIn also faced lawsuits for the same reasons. Last year, LinkedIn was sued for 'Hacking' Users' Emails to spam their friends. The plaintiffs said that by providing their emails to LinkedIn the site will bombard [10] their friends with up to three email invitations on their behalf.

Hackers exploit OSNs' vulnerabilities to collect information about their victims in order to use them in several attacks against users. That's why security is very important on social networks more than any other web application.

A few other common attacks on some of the popular OSNs are mentioned as follows.

#### A. Facebook

##### Likejacking attacks

The main idea is that attackers create interesting posts using social engineering tactics [11]. This technique is based on the use of intriguing posts that rely on rumours, celebrity news and even disasters. By clicking the link, some malicious scripts would automatically re-post the image or video on their contacts' walls and even in some groups that they joined. This attack could also make users like a Facebook page on their behalf.

##### Rogue applications

Facebook allows anyone to develop an application and submit it on its open platform to make it accessible to other users. Cybercriminals use this opportunity to collect sensitive information about people such as their email addresses and their Facebook IDs in order to use them later in spamming and Phishing attacks.

##### Chat Attacks

Hackers use the chat feature in some Phishing attacks and even to launch DOS attacks, although they are not friends of the target.

#### B. Twitter

##### Spammed Tweets

The common way of attacking users is using tweets. A user that you are following could post a link and by curiosity you click on it and you get spammed.

##### Malware downloads

Attackers could also use Twitter to affect your machine. They could trick you and make you click on a post that starts a download of a Malware.

##### Twitter bots

In addition to attacks mentioned before, Cybercriminals use twitter bots which means controlling Botnet zombies that infects users' machines.

#### C. LinkedIn

##### Invitation to connect

Invitation to connect is the most common attack in LinkedIn. It is a Phishing email asking the recipient to click on a button to decide about an invitation to connect to a user. This message is not from LinkedIn but from an attacker.

##### "Mail delivery failed"

This is one of the techniques used by Cyber criminals and that you have to be careful about. The user gets a message with this subject and clicks on the link to see which of his messages bounced back but he sends off a Phishing attack.

##### "Dear Customer"

Is another type of message received by LinkedIn users. This message includes a link used for Phishing. By clicking on the link, you will be redirected to a page that looks genuine but it is fake and it will extract the target information.

##### "Comunicazione importante"

It is an Italian expression meaning "important communication". This message contains links to scam web sites and sometimes links to download Malware applications.

IV. WEB SECURITY SCANNERS

In the previous sections, we discussed some possible attacks and vulnerabilities that could present a real threat against OSN users' privacies. In this section, we are going to describe some tools that may help analyze security vulnerabilities of web applications. Web security scanners or Web Vulnerability Scanners (WVS) provide information about the existing vulnerability of a given web site in order to help fix them and prevent attackers from misusing it. We will also present a brief description of some of the famous WVSs [12] and we will provide a short comparison between these tools.

A. Acunetix WVS

Acunetix is among the leaders in web applications scanning. It is a commercial tool but offering a trial version with limited capabilities. The tool operates as follows. It starts by collecting useful information about the web application. It detects the type of the server, the technology and even its scripts and links so that it can define the structure of the application. Based on the structure established previously and the scan profile selected by the user initially, the tool will start a list of tests that will crawl the application and provide a detailed report of the detected vulnerabilities with indications about the infected files [13].

B. Wapiti

Wapiti is an open source software running under Windows, Unix/Linux and Macintosh. This tool is designed to perform black-box tests which means testing the application independently of its structure and code. It starts collecting and identifying vulnerable scripts by injecting them with data [14].

C. Skipfish

Skipfish is an Open source tool designed to run under Linux, FreeBSD, MacOSX and Windows (Cygwin). It is a dictionary based scanner. It is a fast and easy to use tool. Actually, Skipfish is more helpful while scanning server-side vulnerabilities and wide range of flaws. It generates reports by the end of each scan. These reports provide useful information to developers so they can decide facing these weaknesses [15].

D. BurpSuite

Burp Suite is an Open Source tool (set of tools) designed to support Windows, Linux and MacOS X. The main goal of this tool is to inspect data exchanged between the browser and the web server. To start you will have to set the scanner and configure your browser. Once launched, Burp Suite goes on all the functionalities of an application and starts examining the content using its various ranges of tools. [16].

E. OWASP ZAP

ZAP (Zed Attack Proxy) is an open source tool running under Windows, Unix/Linux and Macintosh. This Tool is very powerful scanner offering a lot of functionalities and it is completely free. ZAP starts by crawling the application and launching an active scan which is done by attacking the application in order to detect any vulnerability. The scanner presents various features like the possibility of being configured as a proxy and launching passive scans which are based on inspecting responses sent from the server [17].

F. Summary of WVSs differences

The short description of the tools mentioned previously gave us a view on their performance when scanning web applications. Table 2 summarizes and gives a brief comparison between those scanners facing the top 10 vulnerabilities published by the OWASP community for the last year.

TABLE 2: WVSS COMPARISON

OWASP TOP 10 -2010 (Previous Version)	OWASP TOP 10 -2013 (Current Version)
A1-Injection	A1-Injection
A3-Broken Authentication and Session Management	A2-Broken Authentication and Session management
A2-Cross Site Scripting (XSS)	A3-Cross-Site Scripting (XSS)
A4-Insecure Direct Object Reference	A4-Insecure Direct Object References
A6-Security Misconfiguration	A5-Security Misconfiguration
A7-Insecure Cryptographic Storage - Merged with A9 -->	A6-Sensitive Data Exposure
A8-Failure to Restrict URL Access - Broadened into -->	A7-Missing Function Level Access Control
A5-Cross Site Request Forgery (CSRF)	A8-Cross-Site Request Forgery (CSRF)
<buried in A6: Security Misconfiguration>	A9-Using Components with Known Vulnerabilities
A10-Unvalidated Redirects and Forwards	A10-Unvalidated Redirects and Forwards
A9-Insufficient Transport Layer Protection	Merged with 2010-A7 into 2013-A6

In fact, the five tools present a good performance against the different vulnerabilities mentioned in the previous table. This outcome shows that these scanners had significant changes, what means that each of them is working on improving its functionalities compared to the evolution of the new vulnerabilities ranking.

V. VULNERABILITIES ANALYSIS OF SOME OSN

In this section, we present and describe a vulnerability analysis based on the Acunetix WVS. Our choice is explained by the fact that Acunetix has a good score and showed a great improvement once compared to other WVSs

as it was published in the Security Tools Benchmarking [18]. In addition to the tool score in this benchmark, Acunetix took first place in the WindowSecurity.com Readers' choice Awards [19] for this year. These were among the reasons that encouraged us to provide the vulnerabilities analysis using it.

Thus, we launched some scans using the trial version of Acunetix as it is summarized in the following table. Each of those scans took one hour and gave us the result described below.

TABLE 3: VULNERABILITIES ANALYSIS.

Vulnerability level	Facebook	Linkedin	Twitter	MySpace
High	0	2	1	96
Medium	22	213	71	10
Low	251	37	6	17
Informational	538	27	21	34

Acunetix classifies threats into four categories. The first level is the informational alert which means that the web application scanned is at a risk of information disclosure. The second is the low risk alert or the level one and it shows up when your application data traffic is not encrypted or when your application reveals its directories paths. The third is the Medium risk alert or level two. This alert is the one where the vulnerability is caused by the server. The final level and the most dangerous threat is the High risk alert. It reports a potential vulnerability that could cause information leakage and could lead to your web application hack.

The assessment presented in this table shows that Facebook is the most secure OSN between the four mentioned previously and MySpace is the most vulnerable OSN with 96% higher attacks. LinkedIn and Twitter are vulnerable too and presented diverse level of weaknesses. All these results should be detailed to know more about the risks that the most used social networks are running.

1). MySpace

MySpace is one of the popular social networks with a lot of new features and functionalities otherwise it is among the most vulnerable web sites. The result of the scan with this limited capabilities trial version of Acunetix presented in the above table and in the following screenshot shows that about a 96 High risk alerts was detected in addition to 10 medium risk alerts, 17 low risk alerts and 34 informational alerts.

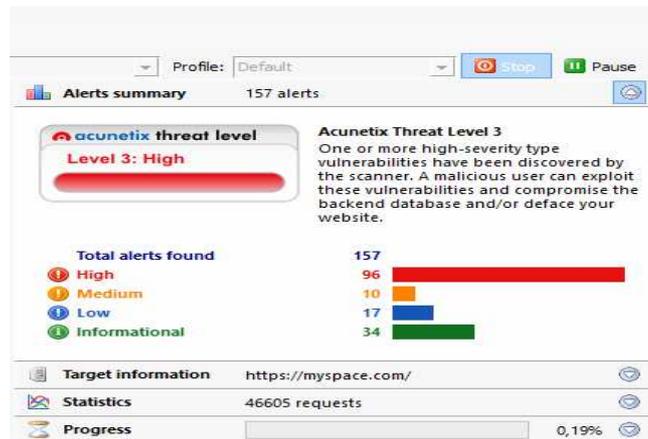


Figure 3 : MySpace Scan Result.

The informational level presents mainly broken links alerts, which means that some pages are not linked and lead to errors. The first level shows a possibility of a Clickjacking attack and a possible sensitive information disclosure. For the second level, Cross-site request forgery vulnerability was the unique threat in this category and it reflects the possibility of a wicked exploit of the application. Finally, the high level risk alert presents both breach attacks, which lead to information leakage and XSS vulnerabilities, as explained previously [20].

2). Twitter

Twitter is a popular OSN with millions of users and fans around the world. it offers new features such as sharing, reading and writing tweets which need to be well protected to mitigate the risk of attacking the users' privacies. Twitter is in permanent progress to ensure the safety of its users but like any other OSN it presents a few vulnerabilities and security weaknesses. The result of the scan presented in the previous table shows broken links as informational alerts, in addition to a possibility of information leakage. in The low level category the risk is the presence of unprotected session cookies. For the medium risk, the application shows up a possibility of malicious exploit caused by the CSRF vulnerability. Finally, the high level alert is a Breach Attack which may cause information leakage in this application [21].

3). LinkedIn

LinkedIn as an OSN for professionals and companies has its own security strategies. A lot of data are stored in its servers that should be out of reach from attackers. The scan on this OSN as presented in the above table gives us information about the web application security status. As informational level, broken links shows up in LinkedIn too. In the low level threat, a possibility of unprotected session cookies in addition to a possible sensitive information disclosure which could lead here to more advance attacks. The medium level reveals a possibility of a misuse of the application. And the final level, High risk alerts present a server-side vulnerability, which is the Denial of server as well as a configuration file disclosure which here helps attackers

gathering sensitive information in order to use them in more dangerous attacks [22].

#### 4). Facebook

Facebook with such number of users and a huge worldwide popularity is in a progress to improve its security strategies. The OSN showed the best result in the previous scan with no high risk alert. However, it is always vulnerable, and the possibility of exploiting these weaknesses is getting higher since it hosts a great amount of personal information. The scan showed more than 500 informational alerts about broken links in the application. In addition to that, a low risk alert of information leakage shows up here with the Session token in URL threat coupled with file upload risk which allows an attacker uploading malicious files and compromise users' information. As a Medium threat, Facebook presents CSRF vulnerability and a possible sensitive information disclosure by hijacking the OSN pages [23].

## VI. CONCLUSION

Nowadays, Online Social networks are becoming a fundamental way to communicate and share with others. Based on exchanging personal and private information, this makes them one of the most vulnerable Internet services and the most targeted by hackers. Protecting users' privacy while using OSN is of at most important and yet a challenging issue.

In this paper, we discussed OSN security aspects and some privacy issues. We reviewed some of the most common threats and attacks of popular OSN. We used then a Web Vulnerability Scanners (WVS) to compare and analyze 4 OSN performances regarding security vulnerabilities. Our work showed that despite their popularity and wide use, these OSN still present serious privacy and security threats. There are a lot to do to enforce the security of OSN and to raise users awareness about the risk they face if they misuse these tools.

## REFERENCES

- [1] Chi Zhang, Jinyuan Sun, Xiaoyan Zhu, and Yuguang Fang, "Privacy and Security for Online Social Networks: Challenges and Opportunities", IEEE Network, Vol.24, No.4, pp.13-18, July-August 2010.
- [2] Carlos Laorden, Borja Sanz, Gonzalo Alvarez, Pablo G. Brngas, "A Threat Model Approach to Threats and Vulnerabilities in On-line Social Networks".
- [3] www.owasp.org/index.php/Top\_10\_2013-Release\_Notes accessed in Jan 2014.
- [4] Hongyu Gao, Jun Hu, Tuo Huang, Jingnan Wang, Yan Chen, "Security Issues in Online Social Networks," IEEE Internet Computing, vol. 15, no. 4, pp. 56-63, July-Aug. 2011, doi:10.1109/MIC.2011.50.
- [5] M.A. Devmane and N.K Rana, Security Issues of Online Social Networks, P.V.P.P College of Engineering, Theem College of Engineering, Mumbai.// i didn't find a reference for this.
- [6] M A Devmane and N K Rana. Article: Privacy Issues in Online Social Networks. International Journal of Computer Applications 41(13):5-8, March 2012. Published by Foundation of Computer Science, New York, USA.
- [7] Fire, M.; Goldschmidt, R. & Elovici, Y. (2013), 'Online Social Networks: Threats and Solutions Survey',CoRR abs/1303.3764.
- [8] R. Gross and A. Acquisti, "Information Revelation and Privacy in Online Social Networks," Proc. ACM Workshop Privacy in the Electronic Soc. (WPES 05), ACM Press, 2005, pp. 71-80.
- [9] www.HYPERLINK "http://sociallyactive.com/facebook-and-kids-a-parents-guide-to-facebook-privacy-and-security/"sociallyactive.com/facebook-and-kids-a-parents-guide-to-facebook-privacy-and-security accessed in Feb 2014.
- [10] www.bbc.com/news/technology-25584286 accessed in Feb 2014.
- [11] http://www.huffingtonpost.com/2013/09/20/linkedin-sued-hacking-emails-spam\_n\_3963195.html accessed in Feb 2014.
- [12] E. Fong and V. Okun, "Web application scanners: definitions and functions," in Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS '07), Waikoloa, Hawaii, USA, January 2007.
- [13] www.acunetix.com/ accessed in March 2014.
- [14] www.wapiti.sourceforge.net accessed in March 2014.
- [15] www.HYPERLINK "https://code.google.com/p/skipfish/"code.google.com/p/skipfish/ accessed in March 2014.
- [16] www.HYPERLINK "http://portswigger.net/burp/"portswigger.net/burp/ accessed in March 2014.
- [17] www.HYPERLINK "https://code.google.com/p/zaproxy/"code.google.com/p/zaproxy/ accessed in March 2014.
- [18] www.windowsecurity.com/news/WindowSecurity-Readers-Choice-Award-Web-Application-Security-Acunetix-Web-Vulnerability-Scanner-Jan14.html accessed in March 2014.
- [19] www.HYPERLINK "http://sectooladdict.blogspot.com/2014\_02\_01\_archive.html"sectooladdict.blogspot.com/2014\_02\_01\_archive.html accessed in March 2014.
- [20] www.myspace.com accessed in March 2014.
- [21] www.twitter.com accessed in March 2014.
- [22] www.linkedin.com accessed in March 2014.
- [23] www.facebook.com accessed in March 2014.

# Security Concepts and Issues in Intra-Inter Vehicle Communication Network

Mustafa Saed, Scott Bone, John Robb  
 Hyundai-Kia America Technical Center, Inc.  
 Automotive Company  
 Superior Township, MI 48198, USA  
 {msaed, sbone, jrobb}@hatci.com

**Abstract**—It is demanding to provide secure communication among vehicles in Vehicle to Vehicle (V2V) and Vehicle to Infrastructure networks (V2I). Vehicles need to authenticate each other and verify the integrity of the shared safety information which is critical. Adversaries can masquerade as real subscribers in V2V/V2I networks and broadcast bogus messages before to destroy the system with such as sending inaccurate safety information to other vehicles. The intent of this paper is to survey the attempts that have been made to tackle vehicle security, and present the security approaches necessary to enforce tough security measures that fully protect the vehicle security architecture.

**Keywords**—V2V, V2I, CAN Bus, Security Network, Security Threats

## I. INTRODUCTION

In the past decade, automotive companies that develop telematics systems have been faced with critical security and privacy issues related to everyday applications that allow interfacing between vehicles and humans. Interaction between people and vehicles will lead engineers to creative thinking of ideas and quite possibly a paradigm shift in determining the methods to provide a sufficiently secure system that will cover all the accessible gaps to unauthorized users. Security concerns were less prevalent in the past due to technological gaps, but with advancements in technology (e.g. ability to develop ad-hoc interfaces, easily accessible hardware and software, etc...) every day computer hackers and cryptologists have created a tremendous amount of concern. Current vehicle architectures are at risk for wireless security break-ins, but future vehicle architectures and systems will increase the risk and this risk needs to be mitigated. These risks will be enabled with the use of embedded phones and wireless protocols containing private information (e.g. financial records, pin numbers, credit card information, birth dates, etc...) [1]. Protecting the customer's private information and the vehicle systems from "hackers" and the infectious viruses these software programs produce will have a direct effect on trustworthiness and the quality from the view of the consumer as well as the vehicle's safety dynamics, such as the multiple Electronic Control Units (ECU's) and its associated systems relying on accurate and uncorrupted information.

The most probable scenario for hacking to occur is to take full advantage of the telematics /wireless feature embedded in the vehicle and performing the function of the electrical

system brain; therefore, allowing this module to become the "open input" to the world. There are currently 2 identifiable probable solutions in closing the technological gap:

- Inter- Vehicle Communication: Secure the communication/protocol between the vehicle and the infrastructure through the wireless network.
- Intra- Vehicle Communication: Secure the communication/protocol between the telematics unit and the ECUs connected through the CAN Bus.

The first solution Inter-Vehicle communication incorporates the use of cryptography and data security with the packet data session over (TCP/IP) and the voice service. The planned proposal(s) include piggy backing off the developed concepts that have been somewhat successful in the security arena. There are a number of researches that are attempting to secure the V2V Networks, but these are still not strong enough to provide effective security and safety. There have been many attempts regarding this technique, such as:

Roshan Duraisamy et al [2] proposed a new hardware implementation, which uses Elliptic Curve cryptography and Digital Signature Algorithms (ECDSA), by enabling two parties "a remote agent and network embedded system" to initiate a 128-bit symmetric key, and make all transmitted data encrypted via the Advance Encryption Scheme (AES). Chenxi Zhang et al [3] introduced a new technique Identity-based Batch-Verification (IBV), which uses a private key for pseudo identities; therefore, the certificates are not required. Each received signature will be verified within 300 ms intervals, but this depends on the Dynamic Short Range Communication (DSRC) protocols. Yi Qian et al [4] proposed how much Medium Access Control (MAC) the layer protocol can achieve through both Quality of Service (QoS) and security requirement for vehicular networks safety application, and designing of efficient MAC protocol to achieve the safety related vehicular networks.

Based on the techniques introduced above, there is a critical concern with securing the vehicle to infrastructure communication. This is due to the fact that all communication between the vehicle and roadside units are implemented with wireless technology; thus, allowing for the probability of numerous attacks or viruses being injected into the unprotected system if security is not enforced. Avoiding these problems and creating a secure,

effective, and yet useful Inter-Vehicle communication method will ensure system reliability [5].

The second solution, Intra-Vehicle Communication requires the protection of data transmission between the vehicle ECU's through the Controller Area Network (CAN) Bus which is an open and unsecured automotive protocol. To this point, automotive companies haven't had concerns with securing this type of communication due to the low risk of infiltrating the CAN Bus remotely. In fact, the only way of accessing the CAN Bus is by connecting a diagnostic tool physically to the vehicle through an On-Board Diagnostic (OBD) connector, so that authorized technicians can perform troubleshooting analyses. However, with the ability to easily develop hardware interfaces and software application layers, automotive companies are implementing to access the CAN Bus remotely through the telematics ECU by using Wi-Fi, BT and cellular network. With this possibility existing, security risks are now increased to the point of allowing "unauthorized systems and network access, Auditability and compliance, Customer data breaches, Internal and external sabotage, and the Theft of intellectual property and confidential business information" [6]. This paper will present background into the specifics of CAN applications, provide the work that has been done so far in the field of vehicle security, and then the security recommendations in the vehicle security field. The rest of the paper is organized as follows: Section 2 focuses on automotive multiplexing methods, classifications and protocols. Possible vulnerabilities in vehicle communication are presented in section 3. Section 4 discusses vehicle communication security issues. The future of vehicle security is introduced in section 5. Finally section 6 concludes the paper.

## II. AUTOMOTIVE MULTIPLEXING METHODS, CLASSIFICATIONS AND PROTOCOLS

### A. Multiplexing Methods and Classifications

Multiplexing in automotive technology has become the greatest achievement in the struggle to make vehicles more efficient by reducing the weight in the power distribution system (i.e. wire harnesses) and keeping modules embedded for security. Multiplexing in terms of technological specifics utilizes a single or dual wire (i.e. bus) connecting multiple ECU's and their corresponding messages and signals through two primary methods; time division or frequency division process. The time division strategy inserts a sample of each channel onto the data stream and the channels are selected for a short period of time. This uses the most accurate form of time sharing amongst various channels and is the method most prevalent in the automotive industry [7]. Frequency division, however, uses a different approach which shares the process amongst various channels where information can be designated by a carrier frequency via each channel to modulate the sinusoidal signals [7].

For the purpose of accurately determining the protocol for developing multiplexing strategies between the various vehicle sub-systems, the Society of Automotive Engineers (SAE) divided the automotive communication sector into three classes. These classes are described as:

- Class A can support 100 nodes and is categorized to handle data speeds (i.e. baud rate) up of 1 kilobit per second (kb/s). However, the lag time, which is the time delta between a transmission request and transmission initiation, is 50 ms. Class A baud rate is used in the following systems: tail light, turn signals, driver convenience features, and entertainment systems [7].
- Class B can support 50 nodes as it is categorized as an information system with data speeds up to 100 kb/s [7].
- Class C is mainly used in real time events that require urgent speed with high accuracy values. Its data rate is in up to 1 Mb/s. Class C baud rate is used mainly in powertrain systems. Class C does not accommodate new systems such as Intelligent Vehicle Highway System (IVHS), collision avoidance system, Global Position System (GPS), and many other related systems [7].

The various types of communication signals are transmitted and received by many types of network nodes; known as protocols. These protocols are created by a set of rules for coding, address structure, transmission sequence, error detection, and handling. When associated with automotive networking, protocols cover a majority of functions assigned to the various layers of the Open System Interconnection (OSI) model. When involved in a noisy surrounding, a multiplexing protocol would be optimized to meet the technical and functional specifications of the system.

### B. Protocols

Inter-Controller Area Network (ICAN) is a network protocol designed primarily for the vehicle networking environment. A CAN controller acts as mediator to alleviate the node processor from over-working itself from the high speed of message transfer. In CAN, disputes between messages are determined on a bit-by-bit basis in a non-destructive arbitration, which result in the highest priority message gaining access to the bus. The CAN protocol supports 2,032 different messages of up to 8 bytes of data. Unlike many serial communication protocols, CAN message data contains no information related to the destination address. The message contains an identifier which indicates the type of information contained within. This feature allows for convenient addition or deletion of the intelligent nodes in an automotive system. Also, each node decides whether to read or ignore a CAN message. A message may be broadcasted to multiple nodes by using the CAN protocol [7].

### III. VULNERABILITIES IN VEHICLE COMMUNICATION

Most products are designed to stop good people from unintentionally doing bad things. This has led to situations in which product security is sometimes an afterthought resulting in frequent redesigns. Making security decisions as early as possible in the design phase makes it easier to avoid costly redesigns that are difficult to both manage and implement.

Threat Modeling, an essential design practice used during all stages of product development, is a practice that can be used to ensure that all security threats have been realized, documented, and mitigated. This practice can also help device makers ensure all stakeholders have considered security as part of the overall product design and part of the developmental process. Since products are often built by several parties, successful use of Threat Modeling requires that all involved parties adopt this practice [8].

Threat Modeling usage scenarios define the scope of the design. All possible usage scenarios including any scenarios perceived as out-of-scope should be listed and marked accordingly. Scenarios should cover all features used by the system, not just the scenarios used by a car. The following are examples of usage scenarios:

- A car used by a home user
- A car used as a taxi
- A car used as a rental device
- Car Wi-Fi connected to a home access point
- Car Wi-Fi connected to a home access point and roams onto public hotspots
- A car at a dealership used as a demonstration vehicle
- User installs third party electronics device on a CAN bus

A definitive list of Usage Scenarios allows all development process stakeholders to know how the device can and will be used, and this also helps teams identify scenarios not previously considered. The following examples are threat categories found with mitigation strategies defined:

- Threat categories: Tampering, information disclosure, and denial-of-service, elevation of privilege: Address book entries are sourced from untrusted external sources and stored in a user's address database. External sources include; USB devices, Memory cards, Bluetooth technology, Wi-Fi, HMI editing, Internet (navigation traffic data), and eCards.
- Threat categories: Information Disclosure: Device crash dumps and device logs are memory and file system dumps whose primary purpose is to aid system debugging, which may contain Personally Identifiable Information (PII) such as phone numbers and SMS text messages. Such files may be obtained from customer vehicles in the field to debug difficult to replicate or high severity issues. If PII is included in those files, it can be viewed by parties outside of the private individual the crash dumps were taken from.

Crash dumps, including PII could be supplied to 3rd party application developers by the Vehicle Manufacturer for review, possibly unintentionally disclosing PII [9].

### IV. SECURITY ISSUES AND THREATS IN VEHICLE COMMUNICATION

While the need for advanced telematics systems continues to drive consumer interest, automotive manufacturers are equally pressured to provide workable systems that can guarantee no unauthorized entry from hackers and other cryptology experts. These types of concerns were never a problem in the past due to gaps in electrical knowledge and technology, but with the sudden advancements in IT, and the ability to develop ad-hoc interfaces with easily accessible hardware and software, developers are suddenly overwhelmed with concern fueled by the increased knowledge and capabilities of the average hacker. These concerns are focused on some of today's vehicles, but even more so in regards to the development of future vehicles.

#### A. Boot Loader

At the time of power-on of any embedded system a Boot loader will be invoked directly from memory. The security issues for the boot loader update process can be addressed as follows:

- An update process that can be used to downgrade as well as upgrade any components (i.e., install v1.2 code over already installed v1.3 component) could be used by a user to install a less secure component, making the device easier to exploit
- The update process can be used to update the Boot loader itself.
- The update requires settings to be changed or recalculated (updating certificates, or updating stored security hashes).
- If the update fails part way through the update process (due to loss of power, failed write to memory or Denial-of-Service) the install process requires a means to back out any changes.
- If any part of the update fails, the update will leave the device vulnerable or malfunctioning. An example of this situation is an update that contains security certificates that need to be stored in a hardware security store (a special area of memory, or separate memory entirely that is not accessible via the data/address bus, such that applications are not able to read/write to it) and files that are signed with the new certificate.
- The update process keeps a log of features updated and configuration changes. The update process modifies user data.

### B. Privacy

Personally Identifiable Information (PII) is any information such as one's name and phone number that can be used to distinguish or trace an individual's identity. Information that can be linked to an individual such as location, favorite shops, and music is also classified as PII. For devices utilizing PII, the recommended security approach is to first consider all information private and be able to be associated with a person or individual object and to then justify why any data is considered not private. PII must always be secured, but note that secured does not necessarily mean disclosed. PII may be disclosed, but only with the individual's knowledge and consent [10].

### C. Operating system OS

Most common embedded operating systems are those that have been thoroughly developed and designed by large corporations such as Microsoft® (Microsoft Windows Embedded, Windows Embedded Automotive) and QNX (QNX CAR Application Platform). The security issues for the operating system can be addressed as follows:

- Developers add their own custom encryption or wrap an existing encryption type in their own code leading to unforeseen weaknesses in encryption and security problems.
- Developers add their own versions of standard functions (strcmp(), strcpy(), memcpy(), etc.). Custom implementations may include quirks that can lead to unforeseen performance, stability, and security issues. Developers misuse high risk functions (for example: sscanf(), strcat()). Incorrect use will not be detected via build warnings [11].

### D. Application

Application types can include native applications, Java Virtual Machine (JVM) and Adobe Flash Player (FP) applications. These application types differ, in-part, based type of device access. Typically, native applications will be written because they need a higher level of access to the device or because of performance reasons. The security issues for the application program interface (API) can be addressed as follows:

- API functions call directly into hardware devices
- API functions allow access to system configuration files
- API functions trigger other system applications to be executed. This can lead to privilege escalation threats and vulnerabilities
- API functions have access to system events. If events are influenced by the application rather than the API, unforeseen instabilities and threats arise within the application as well as in system services [11].

### E. Communication

Any device that connects to external sources whether trusted or untrusted sources will have inherent security threats that need to be mitigated. Many of these threats are not device specific but are effectively communications specific. Therefore, whenever reviewing security aspects of communications systems, extreme care should be taken to not make any assumptions about anything. Many communications systems and protocols were designed and developed before secure design techniques existed (these practices were developed because of the lack of security in design and development). Also, adding security at a later date often does not enhance the security of a system. Accordingly, it is easy to understand why communications systems pose the highest security threats to a system. The security issues for the communication can be addressed as follows [11], [12], and [13]:

- Connection to unauthenticated user – protocols such as ARP and DHCP do not authenticate the server to which they are connecting. Connection is usually assigned by the quickest response. On networks, a malicious device, if able to respond quicker than the intended server, can cause the target to connect to it rather than the intended server, this type of threat can lead to information disclosure, Denial-of-Service, spoofing and tampering.
- Name resolution services are easily confused – typical systems use domain name servers (DNS) to identify target machines by name rather than an IP address, these services can easily be exploited and have had multiple vulnerabilities in them.
- Network Bridging – When multiple methods to connect to the Internet are possible (Wi-Fi, cellular, Bluetooth technology, USB modem), it is possible that several connections may be active at the same time. This is an extremely risky practice do to name resolution issues and other vulnerabilities may occur as a result of the simultaneous connections.
- Weak configuration authentication: many of the Internet configuration settings are accessible to the whole system. An application may be required to setup the configuration of another application or service, but the system may require a different configuration. Thus, an application needs to be held responsible for ensuring it has the correct configuration at all times. Often, all applications are given access to facilities to release and/ or renew the device IP and other network configuration items, thus, allowing them to disconnect other services that require specific connections.
- Local host easily exploitable – many applications in the past used the local host IP for inter-process communications (IPC) since unauthenticated user applications may be able to gain access to or block access to these services. They may also be able to gain elevation of privilege or cause a Denial-of-Service to system features.

- Poorly implemented networking code – Code, if badly written to perform mutual authentication over SSL, can leave a system vulnerable to man-in-the-middle attacks.
- Blocking sockets – Poorly written code utilizing blocking sockets can cause local Denial-of-Services threats and vulnerabilities.
- Raw sockets – Raw sockets allow applications to see all network traffic on a device. Some of this information may include security or private data, which can be used to exploit other vulnerabilities. Badly implemented raw sockets could also open up other applications to see additional data causing stability issues and security vulnerabilities.
- Shortened URLs – Extreme care should be taken when developing, reviewing and using these URLs since they are frequently used to direct the user to install malware or to download a virus. Shortened URLs are most commonly used on social networking sites.
- ASCII/Unicode Threats – Many Internet protocols (example http) were developed based on these character sets. Many threats and vulnerabilities exist through exploitation of interpretation of character strings that contain ASCII control codes and non-displayable characters.
- Network Firewall rule errors – Poorly implemented firewall rules or applications being able to modify a network firewall can render the firewall useless and expose the system to further exploits. A common rule error is caused when the rule priority or order of execution is modified to insert an allow-all rule somewhere in the rule sequence.
- Data cost/charge – Many Wi-Fi and cellular data systems have an associated cost of usage, either monthly or per mega-byte. Any external data usage should be controlled by the system such that the user does not incur unexpected charges due to extreme data usage [13].

#### F. USB

It is (currently) a wired technology that allows many different device types to be connected to the system. Typical USB devices include: cameras and web cameras, flash memory cards, Wi-Fi adapters, Bluetooth technology interface, cellular internet adapters, hubs, phones, media, players, personal computers, mice and keyboards, GPS devices, serial port devices and multifunction adapters. The security issues for using the USB can be addressed as follows [10]:

- Device insert/ejection – USB devices can be inserted or ejected at any time. When any service is utilizing the feature, care should be taken to ensure no disruption of service. A malicious user can repeatedly insert and then remove a USB device at a rapid rate causing the system to go into an unstable state and a Denial-of-Services. Performing this action while the vehicle is in motion could cause driver distraction and an accident; therefore, device detection may be

restricted to when a vehicle is stationary. So hacking the system will cause safety issues.

- USB Flash Memory devices – Adding flash memory devices to the system add areas to the file system. Extreme care should be taken since any contents should be considered untrusted with no guarantee of reliability can be assumed. With continued usage device sector read and write failures can occur, when any system application reads or writes to a device, it should protect against this type of failure.
- USB Multifunction devices – A current trend in USB devices is to provide, for example, a USB cellular modem with build in flash memory device (and a micro SD card slot), the flash memory device usually contains auto-runnable code to install a driver for the cellular modem. Since the modem would typically be manufactured for use on a desktop system, it is highly unlikely that the available drivers would operate on the vehicle system. Also, the USB modem memory card could have been modified by a malicious user and replaced with malware or a virus.
- Multiple identical USB devices – It may be possible to connect two USB Flash memory devices to the system, but developers should be aware of security issues arising from device insertion order. It is also not frequently accounted for that two USB communication devices are connected at the same time (e.g., two cellular modems).

#### G. Wireless

The security issues for the Wireless can be addressed as follows [10]:

- Wireless connection information is stored in weakly protected databases that can be accessible to untrusted applications. If an untrusted application is given full access to the database where all previous wireless networks are stored. Existing entries can be deleted, Existing entries can be tampered with and new entries may be added. This may lead to a Denial-of-Service or granting access to an untrusted network.
- Wireless connection passwords and keys are stored in clear text. If wireless keys are stored in clear text and untrusted applications have the ability to read the data, further exploitation is possible.
- To allow greater areas to be covered by Wi-Fi, it is a common practice for a network to have multiple access points (AP's). Each AP will contain an identical AP name, but have a different MAC address. When the Wi-Fi device moves from an area, where the signal is currently connected to is stronger than the other Wi-Fi AP, to a location where the current connection is weaker, then it will attempt to connect to the stronger signal. Security considerations need to be enforced when this occurs. A number of questions need to be asked: is the new Internet path as secure as the previous one? Have any security protocols changed as a result of the change in AP? Can the new AP be trusted?

Often the transfer from AP to AP occurs internally to the Wi-Fi subsystem and is transparent to a user. However, consideration of this needs to be made during threat modeling and system development. Wireless roaming also exists between different network types. Some examples include, Wi-Fi to Cellular Internet USB modem, Wi-Fi to Bluetooth PAN Internet, Bluetooth PAN Internet to Cellular Internet USB modem, and Wi-Fi home network to a Coffee Shop Public Wi-Fi network. Each wireless network type has individual security implications that need to be considered.

## V. THE FUTURE OF VEHICLE SECURITY

The future of vehicle security seems promising. The following security improvements should be taken care of [14], [15], [16], [17], [18] and [10]:

- Enhancing the vehicle security approach by adopting the Internet protocol version 6 (IPv6) in the vehicle communication protocol, synchrophasor security/NASPInet, anonymization, behavioral economics/privacy, and cross-domain security involving it.
- Using the public key infrastructure (PKI) in the vehicle security, and addressing all the related security requirements of the operation and devices of the vehicle communication.
- Securing the trusted device profile and implementing and developing the vehicle security certificate lifetime.
- Resolving the privacy concerns regarding customer information in the vehicle.
- Preventing the transfer of some critical data, such as the business location or cross border data transmission.
- Implementing a robust security approach for the vehicle communication as a future priority to achieve proper authentication in any device communication via the vehicle.
- Addressing all the newly created vulnerabilities of the vehicle communication by monitoring and tracking the communication and the data flow through the vehicle.

## VI. CONCLUSIONS

This paper presented the work that has been done so far in the field of vehicle security. It also introduced the future approaches, techniques, and methods needed to improve and enhance this security. All of the security features that the vehicle needs to cover were addressed. The paper provided a broad view of how we can make the vehicle a very secure system to take full advantage of all its features. Our future work will focus on extending security requirements to all the vehicle communication, including in vehicle communication, vehicle to vehicle communication, vehicle to infrastructure communication,

and third party access. The main focus will be on how to implement powerful cryptographic protocols to achieve outstanding security.

## REFERENCES

- [1] S. Lee, G. Pan, J. Park, M. Gerla and S. Lu, "Secure incentives for commercial and dissemination in vehicular networks," in Proc. the 13 Annual International Conf. Mobile Computing and Networking, 2007, pp. 150-159.
- [2] Roshan Duraisamy, Zoran Salcic, Maurizio Adriano, and Miguel Morales-Sandoval, "Supporting Symmetric 128-bit AES in Networked Embedded Systems: An Elliptic Curve Key Establishment Protocol-on-Chip," University of Auckland, University of Rome, National Institute for Astrophysics, Optics and Electronics, 2006.
- [3] Chenxi Zhang, Rongxing Lu, Xiaodong Lin, Pin-Han Ho, and Xuemin (Sherman) Shen, "An Efficient Identity-based Batch Verification Scheme for Vehicular Sensor Networks," University of Waterloo, 2008.
- [4] Yi Qian, Kejie Lu, and Nader Moayeri introduced paper, "Performance Evaluation of Secure MAC protocol For Vehicular Networks," National Institute of Standards and Technology, University of Puerto Rico, 2008.
- [5] U.S Department of Transportation, National Highway Traffic Safety Administration, "Vehicle Safety Communications Project; Task 3 Final Report; Identify Intelligent Vehicle Safety Applications Enabled by DSRC," Notional Technical information service (22161), Virginia, March 2005.
- [6] William Stallings, "Cryptography and Network Security Principle and practices," New Jersey, NJ: Pearson Prentice Hall, 2010.
- [7] Paret, D. and Riesco, R., "Multiplexed Networks for Embedded Systems: CAN, LIN, FlexRay, Safe-by-Wire," SAE International, June 20, 2007.
- [8] Koscher, K., Czeskis, A., Roesner, F., Patel, S. et al., "Experimental Security Analysis of a Modern Automobile," IEEE Symposium on Security and Privacy, 2010, pp. 447-462.
- [9] Checkoway, S., McCoy, D., Kantor, B., Anderson, D. et al., "Comprehensive experimental analyses of automotive attack surfaces," Proc. of USENIX Security, 2011.
- [10] Forouzan, B., "Cryptography and Network Security," New York, NY: Mc Graw Hill, 2008.
- [11] The Telegraph, "Thieves placed bugs and hacked onboard computers of luxury cars," 02 July 2012. Available: <http://www.telegraph.co.uk/news/uknews/crime/9369783/Thievesplaced-bugs-and-hacked-onboard-computers-of-luxury-cars.html>
- [12] Wright, A., "Hacking cars," Communications of the ACM, Nov. 2011.
- [13] Bouard, A., Schanda, J., Herrscher, D., and Eckert, E., "Automotive proxy-based security architecture for CE device integration," Proc. of Mobileware, 2012.
- [14] Zagar, D. and Grgic, K., "IPv6 security threats and possible solutions," WAC, July, 2006, pp. 1-7.
- [15] Zhao, M., Smith, S., and Nicol, D., "Evaluating the Performance Impact of PKI on BGP Security," PKI Research and Development Workshop, Gaithersburg, 2005.
- [16] Wolf M. and Gendrullis, T., "Design, Implementation, and Evaluation of a Vehicular Hardware Security Module," ICISC, 2011.

- [17] Hersteller Initiative Software, "SHE Secure Hardware Extension V1.1," 2009. Available: <http://www.automotive-his.de>
- [18] Hoppe, T., Kiltz, S., and Dittmann, J., "Security Threats to Automotive CAN Networks Practical

Examples and Selected Short-Term Countermeasures," Computer Safety, Reliability, and Security, 2008.

# Negotiation of sensitive resources using different strategies for policy's protection

Diala Abi Haidar

MIS Department, Dar Al Hekma University, Jeddah, Saudi Arabia

**Abstract**—*In recent security architectures, it is possible that the security policy is not evaluated in a centralized way but requires negotiation between the subject who is requesting the access and an access controller. This negotiation is generally based on exchanging credentials between the negotiating parties so that the access controller can decide to accept or deny the access. Such a negotiation presumes that policies or part of policies are exchanged between negotiating entities. In some situations, not only the requested resource but also its corresponding access control policy may be sensitive. This requires that such security policies cannot be revealed before some obfuscation is applied on them. In this paper, we present our approach for the negotiation of sensitive resources, mainly policies, by using different strategies including the obfuscation and revealing strategies. As such, a sensitive security policy is divulged following a specific revealing strategy and after an obfuscation is done. Such approach ensures that no sensitive information is exchanged before its corresponding requirements are fulfilled.*

**Keywords:** Access Control, Trust Negotiation, OrBAC, Policies

## 1. Introduction

Traditionally, access control is enforced by centralized stand-alone architectures. In this case, the access controller “knows” all information necessary to evaluate the access control policy. As a consequence, when a subject sends a query to the access controller, this access controller does not need to interact with this subject to decide if this query must be accepted or rejected.

However, in more recent architectures, such a centralized evaluation of the access control policy is no longer appropriate. When a subject sends a query to the access controller, this controller needs to interact with the subject through a negotiation protocol. The objective of this protocol is to exchange additional information necessary to evaluate the policy. This information generally corresponds to credentials the subject has to provide to prove that he or she satisfies the requirements to execute the query.

Notice that the negotiation protocol can actually behave in a symmetric way in the sense that the access controller may also exchange credentials to provide the subject with guarantees that this subject can interact securely with the controller.

To automate the negotiation, each entity should define its access control policy that protect its sensitive resources including its credentials. A credential is a signed document that assert a binding between between an entity and some of this entity's attributes [6]. A negotiation policy should specify which credentials the other entity should present that respond to the access control policy protecting the requested resource. In the negotiation process, credentials will be exchanged until each entity satisfies the access policy of the requested resource. In this case the negotiation is successful otherwise it will fail.

During the negotiation, policies or part of policies are exchanged between negotiating entities in order to request given credentials. Such policies may themselves be sensitive; that is, each entity may not want to disclose its policies or requirements to the other entity. An obfuscation needs to be applied in this case to ensure that the revealed policy does not divulgate any protected information. Such obfuscation can be done by hiding the policy through the addition of some useless parts (*i.e.* this technique is what we define to be *Obfuscation strategy*) or by negotiating parts of the policy gradually (*i.e.* defined as the *Revealing strategy*). These strategies will be discussed in the following sections.

We first start by defining a classification for the resources in section 2. In section 3 we differentiate between negotiation policies and access control policies to be able to clearly define what will be used as conditions during the negotiation process. The negotiation policies will be further specified in section 4. This specification will lead to a formalization of the policies as conjunction and disjunctions of elementary conditions over some attributes. After such a specification is done, the negotiation strategies including the *Obfuscation* and the *Revealing* strategies will be fully explained in section 5. The introduced concepts will be illustrated through an example in section 6. Section 7 discusses basic criteria for choosing between the different revealing strategies. Finally, section 8 is dedicated to a related work study and section 9 concludes this paper.

## 2. Resources

We define as protected resource all sensitive information, *e.g.* services, policies, credentials used to certify attributes, *etc.* All the protected resources are managed by access control policies. We classify protected resources in three classes [4]:

- **Class 1 -“Resource with direct access”:** All protected resources that belong to this class are managed by policies that do not trigger a negotiation process. If the attributes that are needed to verify the resources’ access control policies are given in the request, the evaluation may be possible. If they are not given, they will not be collected through a negotiation process.
- **Class 2 -“Resource with direct negotiated access”:** If the request for access to resources from this class does not include all the needed attributes, the missing attributes may be *directly requested* from the requestor. *Directly requested* means that there is no strategy to hide the resource’s negotiation policy.
- **Class 3 -“Resource with indirect negotiated access”:** We consider that the resources that are classified in this class are managed by a policy that should be kept secret. Thus, missing attributes may be indirectly requested from the requestor. That is, a strategy should be applied to obfuscate part (or all) of the negotiation policy.<sup>1</sup>

An example of *class 1* resources is the *never-accessible* resources. We define these resources as those that can not be accessed under any circumstances. The resources that are *always-accessible* also belong to the *class 1*. The corresponding access control policies allow access to them without restrictions.

Note that *class 2* and *class 3* resources do not necessarily imply a negotiation between the requestor and the service provider, if all the needed attributes are given by the requestor directly. Furthermore, starting or not a negotiation do not reveal to the requestor the classification of the resource. Any classified resource may be seen as a *class 1* resource whenever all the needed attributes are given by the requestor, since no negotiation is needed. Similarly, a requestor cannot easily distinguish between *class 2* and *class 3* resources.

Finally, deducing the classification of the resource does not call into question the sensitivity of this resource. We should note that our classification is not based on the sensitivity of the resources. That is, a *class 1* resource is not necessarily less sensitive than a *class 2* (or *class 3*) resource and vice versa. As such, a given *never-accessible* resource might be a very sensitive resource while classified as *class 1* resource.

### 3. Access control and negotiation policies

We must clearly differentiate between access control policies and negotiation policies. Access control policies are policies defined internally to any organization and managing the access to its internal resources. However, negotiation policies are policies that are exchanged during the negotiation process. The negotiation policy may be the same as

<sup>1</sup>Negotiation policy is obtained from internal access control policies and used during the negotiation process (see section 5).

the access control policy, in case this latter is revealed as it is.

A negotiation policy is generated from the internally defined access control policy. This generation must derive all the required credentials that need to be requested from the remote entity in order to be able to evaluate locally the access control policy. For instance, if an organization is basing its access control on a Role-Based model (eg. RBAC [3]), the negotiation policies should specify the conditions over some required attributes (revealed in credentials) that need to be satisfied to be able to map a given user in its corresponding role. We will assume that a derivation process<sup>2</sup> is applied to the internal access control policies in order to generate the negotiation policies.

A negotiation policy can be seen as a Boolean condition<sup>3</sup>. Giving such a Boolean condition, a negotiation strategy should be applied to define which part(s) of this condition should be disclosed and in which order.

A negotiation strategy (further defined in section 5) is the approach chosen to negotiate some credentials during the negotiation process. It defines the step by step sequence of messages as well as the content of messages exchanged between the negotiating entities.



Fig. 1: The strategy application function

Before going further in our reasoning, we need to define the following:

**Definition 3.1: Potentially negotiated condition:** We consider as potentially negotiated condition, the condition that is obtained from the application of a derivation process over an access control policy.

**Definition 3.2: Negotiated condition:** We define as negotiated condition, the remaining unsatisfied condition of a potentially negotiated condition after evaluating it against the local information.

**Definition 3.3: Strategic condition:** We define as strategic condition, the parts of the negotiated condition after the application of a given negotiation strategy (output of figure 1). It corresponds to the condition that is actually negotiated within the negotiation process.

<sup>2</sup>To express our negotiation policies a derivation process must have been done internally to deduce the attributes that need to be negotiated. Such derivation process is outside the scope of this paper.

<sup>3</sup>The terms negotiation policy or negotiation condition are both used in this paper to refer to the negotiation policies exchanged during the negotiation process.

We need to note that some conditions may be evaluated locally because they might be related to some contextual attributes (*i.e.* time of the access request), or already received information within the initial access request, or they are received in a previous negotiation process with the same entity. In such cases, a *potentially negotiation condition* might be evaluated without the needs to undertake any negotiation. Thus, no negotiated condition exists in this case.

## 4. Specification of negotiation policies

### 4.1 Negotiated attributes

All the access control policies in our study are based on the OrBAC model [9] and expressed as:

$$SR(Decision, R, A, V, Ctx)$$

where SR is a given security rule stating that the decision *Decision* is applied whenever a role *R*, is requesting to perform the activity *A* on the view *V* in the context *Ctx*.

In OrBAC, policies can be expressed using *organizational entities* (Role, Activity and View) as well as *concrete entities* (Subject, Action and Object). Using concrete entities is useful whenever one need to grant exceptional authorizations to specific users. For instance, exceptional policies are used if we want to give a *Subject* that has the role *R* some specific permissions regardless of the role *R* enabled within the organization.

The objective of the negotiation is to exchange credentials in order to decide if a query must be accepted or not. When the access control policy corresponds to a set of *access control rules*, this consists in determining if the query matches one of these rules.

That is, supposing that an access control rule

$$SR_1(Permission, Teacher, A, V, WorkingHours)$$

states that the role *Teacher* is permitted to have the activity *A* over the view *V* in the context *WorkingHours*. Now, consider that a subject *John* sent a request to read a record *Record1* as follow:

$$Request(John, Read, Record1).$$

In order to be able to evaluate such a request against the access control rule  $SR_1$ , we must be able to determine if:

- *John* is a *Teacher*,
- the *Read* action is an activity *A*,
- the *Record1* belongs to the view *V* and
- the context *WorkingHours* is active.

Thus, some conditions over some attributes related to the subject *John*, the action *Read*, the object *Record1* as well as the current time need to be satisfied so that the assignments to the corresponding role, activity and view is done. The negotiation policies must be expressed using these conditions so that missing attributes are collected through the negotiation process.

Each organization manages its local resources. As such, the mapping of an accessed resource *Record1* into its corresponding view *V* is only based on local attributes of the *Record1*. Thus, we assume that the attributes used in the definition of the view are not negotiated. Consequently, these objects' attributes do not appear in the obtained negotiation policies. Furthermore, actions that are allowed to be done over a given resource are also defined by the organization and grouped into activities. Any not defined action will not be considered as accepted and consequently the request is denied. Thus, there will be no action's attributes in the obtained negotiation policies.

We presume that some contextual information (such as the local time of the request, the IP address of the received request, etc) can be evaluated locally. Thus, such information is not considered in the negotiation process. However, other conditions such as emergency context, delegation authorization, and so on can be negotiated. Such context related attributes should be included in the negotiation policies. As one can notice, such information may be considered as attributes assigned to the user requesting the access. That is, we will consider that only subject's attributes used in the definition of the role will be negotiated.

### 4.2 Specification using Boolean conditions

As previously stated, any negotiation rule is defined as a condition over the attributes related to the subject of the request. That is, a negotiation's security rule can be formulated as follows:

$$NC^i = C_s^i.$$

where  $NC^i$  is a *potentially negotiated condition* (input of figure 1) representing the negotiation security rule derived from an access control policy  $SR_i$ .  $C_s^i$  is the conjunctions and/or disjunctions of conditions over subjects' attributes relative to a security rule  $SR_i$ .

A resource's access control policy may contain more than one access control rule. From each of these rules, a given *potentially negotiated condition* is obtained. If we consider all the access control rules relative to one given resource (*R*), we will obtain the *potentially negotiated condition* relative to *R* as follows:

$$(1) \quad NC(R) = \bigvee_{i=1}^n NC^i = \bigvee_{i=1}^n C_s^i$$

where  $NC(R)$  is a Boolean condition that expresses the conditions over subjects' attributes that should be satisfied in order to access *R*. It is obtained by applying a disjunction between all the conditions  $NC^i$  for all the access control rules within *R*'s access control policy.

**Definition 4.1: Elementary Condition (EC):** We define as elementary condition, a Boolean condition over one unique attribute.

Given the above definition, the  $C_s^i$  may be written as:

$$C_s^i = \bigvee_{i=1}^n \bigwedge_{j=1}^m EC_s^i$$

where m, n are positive numbers,  $EC_s^i$  is an elementary condition over one unique attribute of the subject (eg. age, profession, delegation, ...).

Finally, we can reformulate the  $NC(R)$  as follows:

$$(2) \quad NC(R) = \bigvee \bigwedge EC$$

Notice that this  $NC(R)$  is the *potentially negotiated condition* as per definition 3.1. After the verification of the locally available attributes, the remaining non-satisfied conditions of  $NC(R)$  will be considered for negotiation as *negotiated condition*. The *strategic condition* that will be effectively negotiated is obtained after the application of one of the negotiation strategies explained in section 5.

## 5. Negotiation Strategies

### 5.1 Resource classification based negotiation

In section 2, we have classified our resources into three classes. The *class 1* resources do not require any negotiation process when accessed. When a request to access a given *class 2* or *class 3* resource arrives to the system, a negotiation process might be conducted if the required attributes to evaluate such a request are not available. For *class 2* resources, the negotiation policy can be revealed as it is without applying any strategy. That is, the *negotiated condition* is similar to the *strategic condition* (refer to figure 1). *Class 3* resources' negotiation policies contain sensitive information. Thus, a strategy should be applied to hide parts of these policies. Such a negotiation strategy is based on two different strategies:

- *Obfuscation strategy*: The applied strategy to hide information, i.e. it dictates the content of the messages.
- *Revealing strategy*: The way the information is revealed, i.e. it dictates the sequence of messages.

These strategies need to be applied on the *negotiated condition* relative to a given resource R (i.e.  $NC(R)$ ) in order to deduce the conditions that needs to be negotiated.

### 5.2 Obfuscation strategy

An obfuscation strategy is a way to hide a given sensitive information while revealing it. Consider for instance that you need to convey a secret message to a given person, you can use all the cryptography techniques to hide this message or you will only need to reveal this message in clear while hiding it within a lot of other information seen as bulk or noise information. This is typically seen in telecommunication, where a given clear signal can be hidden

within a signal containing a lot of noise<sup>4</sup>. The concept is similar to the principles of hiding a signal that is detailed in [1] as *substitution method of invisibility*.

This approach, that we call *Noise Introduction*, is the obfuscation strategy that we used to hide some sensitive information within a negotiation policy. For each condition or part of condition that may reveal sensitive information, a *family* is defined. What we understand by family is some related or similar conditions. For instance, suppose that a university is going to post a job where only PhD holders are eligible for interviews. If the university consider that this information (i.e. the condition of being PhD holder) need not to be revealed, it may ask any applicant to choose between the following regarding his/her educational level:

- Bachelor degree holder
- Master degree holder
- Doctorate degree holder
- None of the above

All these items in such a list are related and considered as a *degree family*. The applicant may not deduce from this list that there are different rules to be applied for each case; where only one rule is a permission rule (the one relative to Doctorate degree holders) and three other prohibition rules. Other examples of families may be *religion family*, *nationality family*, etc. Although the above example is not relative to a negotiation situation but one can apply such approach on a *negotiated condition* by simply asking for additional conditions over attributes taken from the sensitive attribute's family. In this case, families of attributes should be defined with a limited number of members. Whenever one condition over one attribute should be sent to the negotiating entity, the same condition over all the family members is also sent.

### 5.3 Revealing strategy

The way the information is revealed is another way to hide such information. It is evident that for a *class 2* resource where the negotiation policy is not sensitive, the whole *negotiated condition*  $NC(R)$  can be revealed. In this case, there is no need to apply any obfuscation neither specific revealing strategies. In other cases of *class 3* resources, some information may (or may not) need to be first hidden, i.e. additional noise is added, then a specific revealing strategy should be applied.

Given the Boolean condition

$$NC(R) = \bigvee \bigwedge EC$$

We can have four revealing strategies:

- Strategy 1: The negotiator chooses one of the conjunctions  $\bigwedge EC$  in the condition and sends it to the requestor as a whole.

<sup>4</sup>The concept can be taken similar slightly to what is also known to be as *Steganography*.

- Strategy 2: The negotiator chooses one of the conjunctions  $\wedge EC$  in the condition and send each of its elementary condition EC one by one.
- Strategy 3: The negotiator chooses one disjunction of elementary conditions  $\vee EC$  to be sent as a whole.
- Strategy 4: The negotiator chooses one disjunction of elementary conditions  $\vee EC$  and send each of its EC one by one.

We need to note that in the case of *Strategy 1*, the negotiator requires a set of credentials to be revealed by the negotiating entity in order to satisfy the negotiated condition. These credentials, after probable subsequent negotiation, should be revealed as a whole set or none is revealed. That is, at the level of the requesting entity, after one requested credential's policy is checked; if such credential is unlocked (*i.e.* could be revealed) it is put in a queue until all the rest of the credentials requested in the received policy (*i.e.* condition) are unlocked. This approach will preserve the completeness property of the strategy [13].

To better illustrate these different strategies let us consider the following NC(R):

$$(I) \quad NC(R) = [(a \vee b) \wedge c \wedge (d \vee e)] \vee (f \wedge g)$$

where a, b, c, d, e, f and g are elementary conditions.

This condition can be rewritten in its Disjunctive Normal Form (DNF) *i.e.* as disjunctions of conjunctions:

$$(II) \quad NC(R) = (a \wedge c \wedge d) \vee (a \wedge c \wedge e) \vee (b \wedge c \wedge d) \\ \vee (b \wedge c \wedge e) \vee (f \wedge g)$$

as well as its Conjunctive Normal Form (CNF) *i.e.* as conjunctions of disjunctions:

$$(III) \quad NC(R) = (a \vee b \vee f) \wedge (c \vee f) \wedge (d \vee e \vee f) \\ \wedge (a \vee b \vee g) \wedge (c \vee g) \wedge (d \vee e \vee g)$$

If such a negotiation condition manages a *class 3* resource for which we have decided to apply given revealing strategy to hide the sensitive information, thus the negotiated conditions will be as follows:

- if strategy 1 is chosen, the negotiator chooses one of the below conjunctions from the condition (II). If the negotiation does not succeed another conjunction from the list will be chosen until no more are available. In this case, the negotiation process fails.
  - 1)  $a \wedge c \wedge d$
  - 2)  $a \wedge c \wedge e$
  - 3)  $b \wedge c \wedge d$
  - 4)  $b \wedge c \wedge e$
  - 5)  $f \wedge g$
- if strategy 2 is chosen, the negotiator chooses one of the conjunctions listed above and reveals each of its elementary condition one by one. That is, if the

conjunction  $b \wedge c \wedge d$  is chosen, *b* is first negotiated then *c* then *d*. If one elementary condition within the chosen conjunction is not satisfied, thus the whole conjunction is discarded and another one is chosen. In the same way as strategy 1, if none of the conditions is satisfied, the whole negotiation process fails.

- if strategy 3 is chosen, the negotiator chooses one disjunction from the condition rewritten as in (III). Thus, the negotiator can choose to start by one of the below conditions. The chosen condition will be revealed as a whole and expected to be satisfied by the submitted credentials from the negotiating party. If this condition is not satisfied, the whole negotiation process fails.

- 1)  $a \vee b \vee f$
- 2)  $c \vee f$
- 3)  $d \vee e \vee f$
- 4)  $a \vee b \vee g$
- 5)  $c \vee g$
- 6)  $d \vee e \vee g$

- if strategy 4 is chosen, the negotiator chooses one disjunction of elementary conditions from the above list and send each of its elementary conditions one by one. Thus, if  $a \vee b \vee f$  is chosen, it is not revealed as a whole but instead the negotiator first reveals *a* and if not satisfied then *b* is negotiated. *f* is negotiated in case *b* is not satisfied. Furthermore, if the whole chosen condition is not satisfied, the negotiation process fails.

## 6. Illustrating example

Let us consider the following example where Alice possesses credentials<sup>5</sup>  $C_1, C_2, C_3, C_4$  and  $C_5$ . Bob has the credentials a, b, c, d, e and f. Both Alice and Bob have defined their access control policies as well as the class associated with each of these credentials. Consequently the following negotiation policies are derived (*i.e.* *potentially negotiated condition*). We need to mention that for simplicity reasons and to better illustrate this example, we have noted a condition over some attributes that are exchanged in a given credential by the names of the corresponding credential. As such, within the policy of *a*,  $C_1$  and  $C_2$  refer to conditions over some attributes that need to be satisfied. Such attributes are in fact certified in credentials  $C_1$  and  $C_1$  respectively.

$C_1 : (a \vee b) \wedge c$	(class2)	$a : C_1 \vee C_2$	(class2)
$C_2 : d \vee (e \wedge a)$	(class2)	$b : C_3$	(class2)
$C_3 : f$	(class2)	$c : C_5 \wedge C_3$	(class2)
$C_4 : a \vee b \vee (c \wedge d)$	(class3)	$d : C_2 \vee C_4$	(class3)
$C_5 : True$	(class1)	$e : C_1 \wedge C_5$	(class3)
		$f : True$	(class1)

<sup>5</sup>In this example we consider the credential equivalent to an elementary condition EC. That is, receiving a credential will satisfy the corresponding condition.

Let us now illustrate the application of the negotiation strategy at each side. Before doing so, we must first notice that  $C_5$  as well as  $f$  are *class 1* resources for which no negotiation is required. For simplicity we consider these two credentials as always-available credentials. From the other side,  $C_1, C_2, C_3, a, b$  and  $c$  are *class 2* resources for which the whole policy, after possible local evaluation (*i.e. negotiated condition*), is revealed as it is. If we consider the case of credential  $c$ , we must note that in order for this credential to be revealed, Bob should receive the credentials  $C_5$  and  $C_3$ . That is, Alice will receive the policy  $C_5 \wedge C_3$ , if  $C_5$  is already unlocked (*i.e. can be revealed*) it will never be revealed until  $C_3$  is unlocked.  $C_5$  will be kept in the queue until  $C_3$  is negotiated. At that time, all the credentials in the queue are revealed at once. If one credential cannot be revealed, thus the policy cannot be satisfied, none of the credentials in the queue is revealed. Credentials  $C_4, d$  and  $e$  are *class 3* resources. The negotiation policies for these resources should not be revealed before an obfuscation strategy is applied.

The *Request* step corresponds to the initially requested resource; it is the launch point of the negotiation protocol. The *Reveal* step corresponds to an exchange of negotiation conditions to be satisfied. Such conditions could be part of or a whole negotiation policy depending on the requested resource's (*i.e. credential*) classification. Finally, the *Send* step corresponds to the fact of unlocking resources (*i.e. a credential or the originally requested resource*) and sending them to the requesting entity. Finally, according to our classification, *class 2* resources should always be first negotiated before *class 3* resources. Refer to the third step in the negotiation protocol of figure 2 where we started by negotiating  $a$  before  $e$  since this latter is a *class 3* resource.

### 7. Strategy discussion

In section 5, we have discussed some strategies to negotiate *class 3* resources. These resources are managed by sensitive negotiation policies that needs to be hidden. Such a hiding can be done through the obfuscation strategy or one of the four different revealing strategies. The question that may arise is when to use one given strategy and not another. We will try to answer this question in this section. Furthermore, we need to mention that so far we were focusing on how to hide sensitive information within a given negotiation policy. We have not yet elaborated about the level of trust between the negotiated entities. That is, a given sensitive information may become accessible with a more flexible negotiation if the negotiating entity becomes trusted or a previous negotiation process took place between both entities. Given such a trust concern, we can discuss our revealing strategies.

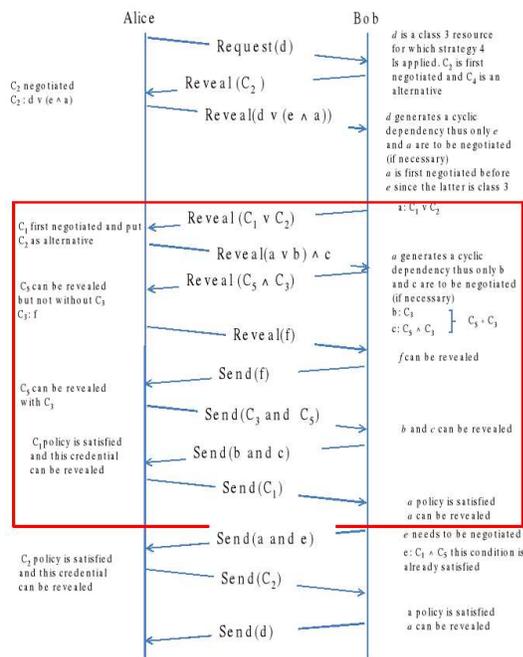


Fig. 2: Example of negotiation

If Alice started by requesting resource  $d$  then figure 2 shows the corresponding negotiation. One can see in the boxed area the part that corresponds to Alice requesting resource  $a$  from Bob. In this process we have mentioned the cyclic dependency detection. This is done whenever a previously requested credential is requested again. We did not elaborate on this issue in this paper since it has been widely studied in the literature [8], [12], [5], [16]. We must note that the negotiation protocol is based on *Request, Reveal*

#### 7.1 Revealing strategy 1

This strategy requires all the conditions that constitute one revealed conjunction to be satisfied otherwise another conjunction is negotiated. Such a constraint requires that the other negotiating entity accepts to reveal a big amount of data at one time although sometimes there is not sufficient trust between both entities yet. There is a big risk that the negotiation does not succeed. Such a strategy might be chosen between an entity and another one with which a previous successful negotiation took place. That is, the process can be accelerated with less risk of failure.

#### 7.2 Revealing strategy 2

Choosing the revealing strategy 2 allows a more smooth and flexible negotiation since the conditions are negotiated one by one. This does not mean necessarily that it takes longer to terminate the negotiation process since if one of the elementary conditions is not satisfied, the whole conjunction is discarded and another one is chosen. Such a strategy may be chosen to negotiate with an entity with which no

previous successful negotiation exists<sup>6</sup> but such entity is somehow trusted ( *e.g.* belonging to a trusted organization, or previously proven trust through cooperation). Conditions will be revealed from the less sensitive to the more sensitive and trust will be gained incrementally.

### 7.3 Revealing strategy 3

This strategy should be adopted whenever the negotiating entities had no previous interaction but a given level of trust exists. Such a trust might be gained by local information gathered about the negotiating entity such as IP address, domain name, *etc.* That is, using this strategy, both entities will be revealing a given set of conditions to be satisfied.

### 7.4 Revealing strategy 4

This strategy should be adopted whenever the negotiating entities had no previous interaction and no information about trust can be generated. In fact in this case, conditions are revealed one by one within a chosen disjunction. The order of choice of the revealed condition always goes from left to right.

## 8. Related work

Among other works done on negotiation of security policies, many were basically focused on the negotiation of trust. TrustBuilder [15], [11] is a system for negotiation of trust in dynamic coalitions. It allows negotiating trust across organizational boundaries, between entities from different security domains. Using TrustBuilder, parties conduct bilateral and iterative exchanges of policies and credentials to negotiate access to system resources including services, credentials and sensitive system policies.

The TrustBuilder approach consists in gradually disclosing credentials in order to establish trust. Only policies that are relevant to the current negotiation may be disclosed by the concerned parties. Since these policies may contain sensitive information, their disclosure can also be managed by some strategies [10]. The authors in [10] have proposed two negotiation strategies; one does disclose some credentials more than required to speed up the negotiation process, however, it does not disclose any access control policy. The other strategy strictly discloses policies as well as credentials that the negotiating entity has gained access to. Although this latter strategy might have some similarities with our proposed work, however, it does not consider the classification of the resource in order to adapt the negotiation process.

Trust- $\chi$  [2] is another framework for trust negotiation specifically conceived for a peer-to-peer environment. Trust- $\chi$  proposes a language for the specification of policies and credentials needed in the negotiation process. Furthermore,

<sup>6</sup>We can imagine that after a trial with strategy 1 that leads to a failure, the next time the entities will negotiate they will adopt the strategy 2.

it provides a variety of strategies for the negotiation. Trust- $\chi$  introduces *trust tickets* that are issued after a negotiation process succeeds. Such trust tickets reduce as much as possible the number of credentials and policies needed in subsequent negotiation processes relative to the same resource thus speeding up these processes. Similarly to TrustBuilder, the Trust- $\chi$  disclosure policies state the conditions under which a resource can be revealed. Furthermore, *prerequisites* (*i.e.* set of alternative policies to be disclosed before the policy they refer to) associated with sensitive policies manage their disclosure. However, even in this case, Trust- $\chi$  does not handle different resources' classification but the idea of trust tickets might be very helpful in our case if we want to switch from one strategy to another or even decide on a given negotiation strategy. That is, the trust level in our case is considered to be a driver for the choice of the negotiation strategy especially for *class 3* resources.

In [12], authors propose PRUNES as a complete strategy for automated trust negotiation. Such a strategy ensure that no credentials are revealed if the negotiation will not succeed and if these credentials are not relevant to the current negotiation. Both entities do not have a knowledge about each entity's policies but each can build incrementally, during the negotiation, a partial negotiation search tree and apply an efficient backtracking strategy. Before starting the credential exchange, the negotiation process should start by an initial phase called *negotiation phase* where both entities engaged in the negotiation exchange some requests and messages to check if an agreement can be reached. In this phase a policy, that is controlling a credential requested during the negotiation, is revealed. This is a problem once we need to protect also the access control policies that are managing a resource. Furthermore, in PRUNES, the first credential that figures in the policy is always negotiated. In our approach, we have proposed a strategy for negotiation ensuring that sensitive policies are not revealed as a whole. If the policy is a disjunction or conjunction of conditions over credentials, different strategies exist for the choice of the conditions to be negotiated exist.

In [8], another highly efficient strategy for automated trust negotiation called DFANS is proposed. As [12] and [14], this strategy is a complete strategy. It is based on the use of Ordered Binary Decision Diagrams (OBDDs) to represent the boolean conditions, *i.e.* negotiation policies. For each of the policies protecting one negotiated resource and/or credential, an OBDD is build. Many differences exist between DFANS and PRUNES; first of all PRUNES include a first phase of negotiation where no credentials are disclosed and a given disclosure sequence leading to a successful negotiation is searched. In DFANS credentials are disclosed as soon as their policies are satisfied. This is also a difference with our proposed negotiation approach where credentials, even unlocked, might be kept in a queue and not disclosed until all the requested other credentials are

unlocked. Second, PRUNES is based on negotiation trees with backtracking strategy however DFANS does not include backtracking. We should note that the idea of backtracking used in PRUNES is essential in our case when a given negotiated condition part of a disjunction is not successful and another alternative should be negotiated. Third, DFANS deals with solving cyclic dependencies that were leading to unnecessary failure in PRUNES. This is done by using the Obvious Signature Based Envelope (OSBE) [7]. We build our proposed approach on the idea of the existence of a way to detect cyclic dependencies and such a proposed work in [7] might be further considered. However, as in PRUNES and [14], DFANS does not support that both entities use different strategies and does not protect private policies while in our approach we do consider these two concepts. Finally, none of the above listed work has dealt with the introduction of noise as a way of hiding information.

## 9. Conclusion

In this paper we have proposed an approach for hiding sensitive policies during the negotiation of resources. We have based our proposal of negotiation on the classification of the resources between stranger entities; *Class 1* would never be negotiated, *class 2* would be negotiated and their corresponding negotiation policy is revealed and *class 3* would be negotiated after some obfuscation strategies as well as chosen revealing strategies are applied. Future work consists of proposing a formal negotiation protocol. The derivation process to generate *potentially negotiated conditions* as well as backtracking should be further formalized. Finally, the cyclic dependency should be considered in the future proposed formalized protocol.

## References

- [1] Tuomas Aura. Practical invisibility in digital communication. In Ross Anderson, editor, *Information Hiding*, volume 1174 of *Lecture Notes in Computer Science*, pages 265–278. Springer Berlin Heidelberg, 1996.
- [2] E. Bertino, E. Ferrari, and A. C. Squicciarini. Trust-X: A Peer-to-Peer Framework for Trust Establishment. *IEEE Transactions on Knowledge and Data Engineering*, 16(7):827–842, 2004.
- [3] D. F. Ferraiolo, R. Sandhu, S. Gavrila, D. R. Kuhn, and R. Chandramouli. Proposed NIST Standard for Role-Based Access Control. *ACM Transactions on Information and Systems Security (TISSEC)*, 4(3), 2001.
- [4] D. Abi Haidar, N. Cuppens, F. Cuppens, and H. Debar. Resource Classification Based Negotiation in Web Services. *Third International Symposium on Information Assurance and Security (IAS)*, pages 313–318, August 2007.
- [5] Hai Jin, Zhensong Liao, Deqing Zou, and Weizhong Qiang. A new approach to hide policy for automated trust negotiation. In Hiroshi Yoshiura, Kouichi Sakurai, Kai Rannenberg, Yuko Murayama, and Shinichi Kawamura, editors, *Advances in Information and Computer Security*, volume 4266 of *Lecture Notes in Computer Science*, pages 168–178. Springer Berlin Heidelberg, 2006.
- [6] Adam J Lee. Credential-based access control. In Henk C A van Tilborg and Sushil Jajodia, editors, *Encyclopedia of cryptography and security*, pages 271–272. Springer, New York, 2011.
- [7] N. Li, W. Du, and D. Boneh. Obvious signature-based envelope. In *Proceedings of the 22nd ACM symposium on principles of distributed computing*, PODC '03, 2003.
- [8] H. Lu and B. Liu. Dfans: A highly efficient strategy for automated trust negotiation. volume 28, pages 557 – 565, 2009.
- [9] A. Miège. *Definition of a formal framework for specifying security policies. The Or-BAC model and extensions*. PhD thesis, ENST, June 2005.
- [10] K. Seamons, M. Winslett, and T. Yu. Limiting the Disclosure of Access Control Policies During Automated Trust Negotiation. In *Network and Distributed System Security Symposium*, San Diego, CA, April 2001.
- [11] K.E. Seamons, T. Chan, E. Child, M. Halcrow, A. Hess, J. Holt, J. Jacobson, R. Jarvis, A. Patty, B. Smith, T. Sundelin, and L. Yu. TrustBuilder: negotiating trust in dynamic coalitions. *Proceedings DARPA Information Survivability Conference and Exposition*, 2:49–51, April 2003.
- [12] X. Ma T. Yu and M. Winslett. Prunes: an efficient and complete strategy for automated trust negotiation over the internet. In *Proceedings of the 7th ACM conference on Computer and communications security*, CCS '00, pages 210–219, New York, NY, USA, 2000. ACM.
- [13] W. H. Winsborough, K. E. Seamons, and V. E. Jones. Negotiating Disclosure of Sensitive Credentials. In *Security and Cryptography for Networks*, 1999.
- [14] H. Yan and M. Zhu. A complete and efficient strategy based on petri net in automated trust negotiation. In *Proceedings of the 2nd international conference on Scalable information systems*, InfoScale '07. ICST, 2007.
- [15] T. Yu, M. Winslett, and K. E. Seamons. Supporting structured credentials and sensitive policies through interoperable strategies for automated trust negotiation. *ACM Transactions on Information and System Security (TISSEC)*, 2003.
- [16] Charles C. Zhang and Marianne Winslett. Distributed authorization by multiparty trust negotiation. In Sushil Jajodia and Javier LÃşpez, editors, *ESORICS*, volume 5283 of *Lecture Notes in Computer Science*, pages 282–299. Springer, 2008.

# MPEG-21 Based Approach to Secure Digital Contents Using DC Metadata

Samiha Ayed<sup>1</sup>, Muhammad Sabir Idrees<sup>1</sup>, Nora Cuppens-Boulahia<sup>1</sup> and Frédéric Cuppens<sup>1</sup>

<sup>1</sup> Telecom Bretagne, Rennes, France

**Abstract**— *With the proliferation of the use of digital resources, many metadata were created in order to describe as detailed as possible these assets within their different contexts. The Dublin Core set of elements can be considered as the most widely used metadata standard of digital resources. This standard is expressive enough to deal with resource characteristics. However, it is less expressive to manage and take into account security access control to these resources. In this paper, the contribution is twofold. First, we show that the Dublin Core standard can be extended to provide a mapping process to other existing metadata specification languages in order to have a generic representation of electronic resources. Second, we suggest managing access control to these resources based on the MPEG-21 norm. For this purpose, we show how the Dublin Core elements can be used to provide inputs.*

**Keywords:** MPEG-21, Dublin Core, Security policy, Metadata

## 1. Introduction

New technologies make significant inroads into daily lives of consumers. Today, digital resources (music, video, ebooks) can be carried anywhere with a diversity of players becoming smaller and smaller with very high performance. Managing digital resources seems an easy task. However, due to the many types of attacks these resources are exposed to, the security of digital contents represents a challenge and it is always an open issue. The representation of digital resources is based mainly on two elements: (1) The description of the content to give an idea about the structure of the digital resource and (2) The description of security policy that we should associate with the resource. These descriptions are inter-related.

On the one hand, to handle the document life cycle, a set of data has been identified and derived from existing models that have proven their efficiency. All the required data, so called "metadata" are used to precisely control how a given document is created, modified, disseminated while preserving the initial security policies. Metadata are keys to ensuring that resources will survive and continue to be accessible in future. On the other hand, the security policy defines different rights depending on

the document content and on the user profile. This policy mainly describes the access and usage control for the content. The current enforcement of security within digital contents is based on the DRM (Digital Rights Management) mechanism. Before the enforcement phase, the specification phase must formally describe policies to be applied.

Actually, many metadata exist in order to describe the digital contents. Each of them is related to different aspects of the resource. To include security aspects, each metadata has its own specificity. However, the whole solutions are not sufficient to express the constraints related to the use of a specific content. In this paper, we show how to define generic metadata that takes into account security aspects. Thus, the contribution of this paper is twofold: First, we propose to converge different existing metadata to the Dublin Core standard to have a generic model to represent the resources and second, in order to take into account the security aspects to be applied on the resource, we make the link with MPEG-21, the most common standard for digital contents. Such an approach avoids having a huge diversity of metadata to express contents and different ways to introduce security control in these contents. The approach aims to be generic and compliant with standards to secure digital contents.

## 2. State of The Art

Digital contents can be described based on many metadata types. There are mainly three different trends for these descriptions. The first type is used to describe e-books for e-commerce. ONIX metadata is an example of these metadata. ONIX (Online Information Exchange) is an international standard [1] used to describe, in a structured and xml based way [6], the set of information and properties related to a specific document. The second trend of metadata is resource oriented. IMS template is an example of this kind of metadata description language. The description of the resource based on this template is more based on details related to the e-learning environment such as the fields in which the resource should be used, to whom the resource is created, how teachers and students can use it, etc. The last metadata type is user oriented. It describes and focuses more on the description of the user characteristics. VCard [3] is such a metadata description language. This template is the first format

used to exchange personal information between different entities. In order to take into account security aspects, these descriptive metadata are based on a very simple approach. Rights associated with a resource are described as any other information describing the digital resource. For example, ONIX in its version 3.0 has defined some specific tags defining in the same way different properties of the resource and indicating by codes the different types of usage permitted on the resource. Similarly, IMS and Vcard are using tags to indicate some security aspects related to the resource that they are describing. Introducing tags related to security aspects within metadata is not an expressive approach to specify security requirements that can be associated with a specific content or resource. Moreover, this approach does not show how to enforce these security aspects. In real use, digital contents are secured with DRM (Dynamic Rights Management) mechanisms [11], [12] which are defined and coded by default without any relation with the resource description based on the metadata. The specification of these rights should be formally defined using a standard of rights expression language to express these security requirements through the license concept and based on the initial description of the resource using metadata. Resource metadata serves only to give, through the template description, the set of rights that are defined on this specific electronic resource. However, many metadata defining rights expression languages exist separately. We denote especially ODRL [5] of OPA-DRM standard and MPEGREL [17] of MPEG-21 standard [7]. In this work, we propose to use MPEG-REL as the rights expression language to secure the descriptive metadata defining the digital resources. The formalization of the security policy derived from this definition will be introduced as an input of the DRM implementation in order to enforce the content rights and security. Verification and execution of this license should comply to the resource and user requirements. This is the main contribution of this paper which is presented in next sections.

### 3. Proposed Approach

In this paper, we propose an approach to define generic metadata which are based on the Dublin Core standard. The elements composing different existing metadata can be expressed using this standard. The approach proposed is mainly based on two steps:

- 1) Transform the description of the digital content, based on ONIX or IMS, to a unified and consistent description based on Dublin core metadata
- 2) Define the license to be associated with the content including the set of rights granted to the user.

#### 3.1 Security-relevant Metadata Information

To express the security policy in this paper we are based on the MPEG-21 norm since it presents the most widely used REL for digital contents. The choice of MPEG-REL is specifically based on the most important difference existing between MPEG-REL and ODRL. For ODRL, conditions and constraints are included in rights. When conditions change we have to define a new action with these new conditions even if we are dealing with the same action or right. However, conditions in MPEG-REL are expressed as an element of the permission like the right. So conditions and constraints have to be dissociated from different rights. The choice of such a language is also motivated by the use of XrML (eXtensible Rights Markup Language) developed by ContentGuard [14] and which has received the support of many technological actors like Adobe, Microsoft, HP labs, Xerox.

The strength of Dublin Core is the definition of different elements required to specify a security policy. Within the metadata description, only some relevant tags may be useful to define the security policy associated with the content. A tag is a term describing part of the digital content. In this section, a strong focus is currently put on tags associated with key elements used to define rights. As presented in section 2, the control of the content is based on the use of a license which describes the rights of a specific user on a specific resource. This license is defined using three basic elements: (1) The resource, (2) The user and (3) The set of rights. Within the Dublin Core template we focus on the description of these elements. The set of these elements is considered immutable information. Thus, each template provides these three basic information to be used to link the document and the set of rights. Rights are actually defined as permissions on the document using ccREL (The Creative Commons Rights Expression Language) which defines licenses based on RDF.

#### 3.2 Mapping Process

During this process, the different descriptive metadata are described using Dublin Core standard. Initially, the standard defined the original 13 core elements which later increased to 15: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, and Rights. The Dublin Core was developed to provide a simple and concise way to describe web-based documents. For this process we defined a mapping algorithm to translate an ONIX or an IMS description to a Dublin Core description. Tables 1 and 2 represents the mapping of different tags generated by the translation. An excerpt of the main algorithm of this transformation is presented here for the translation from the ONIX metadata to the Dublin

Core metadata. The same process can be applied to transform IMS templates. For the Dublin Core tags we classify elements into four classes: (1) Resource definition (2) Resource owner (3) Resource conditions (4) Rights expression.

Listing 1: XSLT transformation ONIX/Dublin Core

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/
  Transform"
3   version="1.0">
4   <xsl:template match="/">
5     <dc:XML><HEAD><TITLE>Dublin Core
      description from ONIX input</TITLE></
      HEAD><BODY> <xsl:apply-templates/></
      BODY></dc:XML>
6   </xsl:template>
7   <xsl:template match="ProductIdentifier"><
      dc:Identifier><xsl:value-of select="
      ProductIdentifier"/></dc:Identifier></
      xsl:template>
8   <xsl:template match="ProductIDType"><dc:Type>
      <xsl:value-of select="ProductIDType"/></
      dc:Type></xsl:template>
9   <xsl:template match="TitleDetail"><dc:Title> <
      xsl:value-of select="TitleDetail"/></
      dc:Title></xsl:template>
10 </xsl:stylesheet>

```

The main idea of the algorithm 1 is to parse the input ONIX XML file and to replace different nodes of this file by corresponding nodes in order to generate the XML output file based on the Dublin Core template. The `<xsl:apply-templates>` element applies a template to the current element or to the current element's child nodes. The attribute "match" of the tag `<xsl:template>` allows the definition (based on the XPath concept [8]) of XML elements to which the transformation is applied. The value of these nodes is evaluated using the `<xsl:value-of>` element which is used to select the value of an XML element and add it to the output. The transformation is presented only for three attributes as example. The whole attributes can be generated in the same way.

The mappings presented here are not a reformatting mechanism. For each initial record, a corresponding record using Dublin Core elements is generated. During this transformation the semantics of the records is kept. Indeed, Dublin Core is known for and has been chosen in this paper because of its genericity. This genericity comes mainly from the two following points:

- 1) Dublin Core elements are optional: if one of the elements does not fit with the meaning of the record or we are not able to find a corresponding field for a specific element then the element is not considered. The transformation is based on an extraction of meaningful, interesting and core fields to be generated based on Dublin Core elements.
- 2) Dublin Core elements are underspecified: the meaning of different elements is not application related. Thus, each element can acquire a specific meaning from the context. For example, IMS records

Element Classification	Dublin Core	ONIX
Resource Definition	Identifier	<ProductIdentifier>
	Type	<ProductIDType>
	Title	<TitleDetail>
	Subject	<Subject>
	Description	<TextContent>
	Format	<ResourceForm>
	Source	<ResourceLink>
Resource Owner	Language	<DefaultLanguageOfText>
	Creator	<SenderIdIdentifier>
	Publisher	<Publisher>
Resource Conditions	Contributor	<Contributor>
	Date	<SentDateTime>
	Coverage	<Territory>
Rights Expression	Relation	<RelatedWork>
	Rights	<EpubUsageConstraint>

Table 1: ONIX/Dublin Core Mapping Table

are describing specific information within the e-learning context. During the definition of Dublin Core elements from IMS description, this meaning can be kept. This is a strong point giving the possibility to do a semantic interpretation based on what we are managing.

Listings 2 gives a simple example of Dublin Core description of a resource. For simplicity, only few fields are considered in the example.

Listing 2: Dublin Core Metadata Example

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <metadata xmlns="http://example.org/myapp/" xmlns:xsi=
  "http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://example.org/myapp/
  http://example.org/myapp/schema.xsd" xmlns:dc="
  http://purl.org/dc/elements/1.1/">
3 <dc:title>Tutorials for programming languages</
  dc:title>
4 <dc:description>Tutorials for programming
  languages gives an initiation for the
  programming basis and approaches.</
  dc:description>
5 <dc:subject>Computer science</dc:subject>
6 <dc:identifier>http://www.tutorials-on-line.
  ProgrammingLanguage.eu</dc:identifier>
7 <dc:format>"text/html"</dc:format>
8 <dc:rights>Permission is granted for students to
  print only one copy of the resource. </
  dc:rights>
9 </metadata>

```

## 4. MPEG-21 Process Using DC Metadata

In this section, we define the securing process of the Dublin Core metadata using the MPEG-21 standard. This process includes many phases that we present below.

### 4.1 Digital Item Declaration

MPEG-21 is not intended to create new formats for content. Thus, it has defined a way to define unambiguously the declaration of different resources. When the Dublin Core metadata is received with the content or the resource, a new "Digital Item" has to be defined to declare this new resource. The definition of this new

Element Classification	Dublin Core	IMS
Resource Definition	Identifier	<Resource.Type>
	Type	<LearningResourceType.Type>
	Title	<General.Type > (element name="title")
	Subject	<General.Type>
	Description	<Resource.Type> (element name="description")
	Format	<Technical.Type>
	Source	<Aggregationlevel.Type>
Resource Owner	Language	<General.Type > (element name="language")
	Creator	<RoleLifeCycle.Type> ("author", "initiator")
	Publisher	<RoleLifeCycle.Type> ("publisher", "editor")
Resource Conditions	Contributor	<RoleLifeCycle.Type> ("content provider", "validator")
	Date	<DateTime.Type>
	Coverage	<General.Type> (element name="coverage")
Rights Expression	Relation	<Relation.Type>
	Rights	<Rights.Type>
	Rights holder	<copyrightAndOtherRestrictions.Type>
	Audience	<IntendedEndUserRole.Type> ("teacher", "author", "learner")

Table 2: IMS/Dublin Core Mapping Table

digital item is based on the information provided within the metadata. This new core concept introduces a new container for content. It is a structured container which defines the resource characteristics (based on the Dublin Core metadata) and also the resource content (based on the received resource or content). To define these digital items, the DIDL (Digital Item Declaration Language) is used. Thus, if we consider the example of the metadata described using Dublin Core presented in section 3.2, the Digital Item to be derived from the metadata description is defined as follows:

Listing 3: DI description in DIDL

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <DIDL xmlns="urn:mpeg:mpeg21:2002:01-DIDL-NS">
3   <Item> <Descriptor> <Statement mimeType="text/
4     plain">Computer Science</Statement></
5     Descriptor>
6   <Descriptor><Statement mimeType="text/plain">
7     Tutorials for programming languages</
8     Statement></Descriptor>
9   <Descriptor><Statement mimeType="text/plain">
10    Tutorials for programming languages gives
11    an initiation for the programming basis
12    and approaches</Statement></Descriptor>
13  <Descriptor><Statement mimeType="text/plain">"
14    text/html"</Statement></Descriptor>
15  <Component><Resource mimeType="Text" ref="
16    http://www.tutorials-on-line.
17    ProgrammingLanguage.eu/"></Component>
18
19  </Item>
20 </DIDL>

```

To define these Digital Items we should apply the following matching. It defines the meaning of different concepts used to construct the Digital Item based on the Dublin Core template.

- **Resource:** It defines an individual identifiable asset. It is described by two parameters: the identifier which provides a value or a reference (URI: Universal Resource Identifier) of the resource and the MIM-TYPE attribute which gives the type (audio, video, text, etc.) of the resource. Based on Dublin Core tags, these two parameters respectively correspond to the Dc:Identifier and the Dc:Type.

- **Statement:** It contains a piece of information defined in any data format and that can be attached to a specific element. The data in a Statement is a piece of information, but not an asset. The same attributes of the resource concept can be used with a statement. Considering again the Dublin Core template, information about the statement concept can be extracted from the tags: Dc:Title, Dc:Rights, Dc:Description. We can notice that these data represent a set of information that we can associate with a resource.
- **Descriptor:** It is used for descriptive data. These descriptive data may be a resource (image, cover page, etc.) or a statement (a description, a title, a subject, etc.). A descriptor is a piece of information related to all or part of a specific resource instance. In general, it contains control or structural information about the resource and not about the content itself. Thus, information described in different statements of Dublin Core metadata can be considered descriptor.
- **Component:** It is the DIDL element that groups Resource elements with Descriptor elements. A component binds a resource to a set of descriptors. Considering Dublin Core template, the component is binding the descriptive data (title, subject, description, etc.) to the resource (the content).
- **Item:** It is a group of components associated with a set of relevant descriptors containing information about the item as a representation of a resource (asset). This is the global structure which includes the different elements.

To recognize the document as a DIDL document, we need to add at the top of this document a mandatory namespace declaration. In our example, we used the original DIDL namespace which requires the URI "urn:mpeg:mpeg 21:2002:01-DIDL-NS" as its namespace. Once the file is generated, we should ensure that the

transformation of elements described by Dublin Core metadata fits to the structure used in the MPEG-21 standard. For that, we should check if the set of DID files are valid with respect to the MPEG-21 DIDL XML schema. For this purpose, we can use XML parsers like Xerces or MSXML.

## 4.2 Digital Item Identification and Description

Once the digital item defining the resource and its content based on the Dublin Core metadata is created, we should make it identifiable and locatable uniquely. This key aspect was defined in MPEG-21 framework to carry unique and persistent identifiers. This identification should link the DI with the related information such as the content and also the descriptive metadata. To achieve this unique identification, the DII specification introduces three different XML elements: the Identifier, the RelatedIdentifier and the Type element:

- **Identifier:** This element is defined based on the statement concept of DIDL. The identifier of a resource in Dublin Core template (Dc:Identifier) is used to fill in the statement about the Identifier element within DID file as a children of the descriptor element.
- **RelatedIdentifier:** This element is also expressed within the DID file by the statement element. It indicates the set of resources with which our asset has a relation. This statement is filled in based on the tag Dc:relation. It can contain the URI of the resources with which the relationship is defined and also the type of such relation.
- **Type:** This element is used to specify the type of the resource. DII Types can define different categories of resources, for example, a music album type or a movie collection type or a specific ebook category. Considering again the Dublin Core metadata, this element is filled in based on the tag Dc:Format.

An example of digital item identification corresponding to the example of listing 2 is provided in listing 4

Listing 4: DII Example

```

1 <didl:Item>
2   <didl:Descriptor>
3     <didl:Statement mimeType="text/xml; charset=
4       UTF-8">
5       <dii:Identifier xmlns:dii="
6         urn:mpeg:mpeg21:2002:01-DII-NS">
          urn:http://www.tutorials-on-line.
          ProgrammingLanguage.eu/</
          dii:Identifier></didl:Statement></
          didl:Descriptor>
5     ...
6 </didl:Item>
```

## 4.3 MPEG - Right Expression Language

So far, we showed how the resource defined by its content and a set of characteristics can be expressed using the MPEG-21 concepts as described in the Dublin Core metadata. This description does not deal with the security aspects. To specify a set of rights, MPEG-21 defines its own rights expression language (MPEG-REL) [17]. Actually, in this paper we express rights in MPEG-REL and associate them with the resource described by a Dublin Core metadata. A detailed description of how to define these licenses using the Dublin Core metadata is given in the next section.

## 4.4 Intellectual Property Management and Protection

The IPMP component is central in the MPEG-21 standard [13]. It is the core component which ensures and executes the link between the different parts already presented. For this purpose, it generates a mapping of elements already defined in Digital Item Description in its own language (IPMP DIDL). These elements are used to check the execution of rights and to check if a specific license refers to the good resource or not.

We now present the global framework which connects the different components together.

## 5. Definition of Global Framework

Figure 1 represents the global picture of the interaction between MPEG-21 components and the DublinCore metadata. The first step is based on the Dublin Core metadata in order to construct inputs to each component of the MPEG-21 standard. This transformation module generates inputs in languages used by each component (DIDL, IPMP DIDL). For links 3, 4, 5 and 6, we indicate the pieces of information to be transferred to each component (it should be based on the language used by that component). The meaning of each arrow is explained below.

(1) The basic elements about the content and the description of the resource are classified using the type of the information required by the MPEG-21 components. We define the fields to be used for the DID, the ones to be used for DII, IPMP and MPEG-REL.

(2) A transformation is defined on these elements and it generates different input files (DID file, DII file, IPMP file). For example, a DID file describes the resource (the content and the description of the content). Algorithm 1 gives an example of the transformation process to be applied in order to generate a DIDL structure for a Dublin Core description. In the same way we can derive inputs for other structures. The algorithm implements the matching explained in section 4.1.

(3), (4), (5) and (6) After transformation, an XML file containing information mentioned on each link is

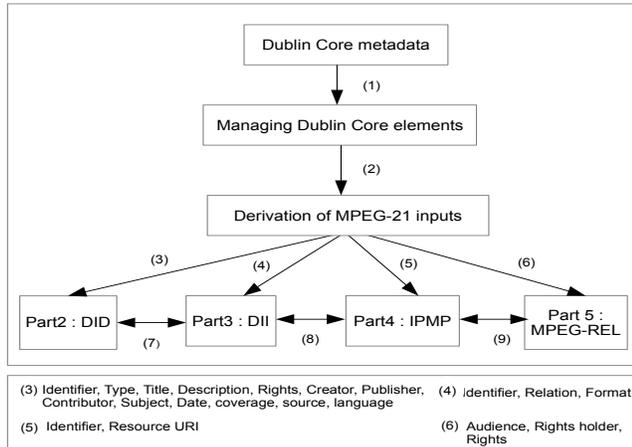


Fig. 1: Different Framework Links

sent to the corresponding component. For example, the identifier, the relation and the format of the resource are generated into DIDL and sent to the DII component. These elements will be used to check conformity of the information received from DID and to uniquely identify the resource.

(7), (8) and (9) The IPMP component links the different components in order to ensure that the identification and the application of rights are done on the right resource. For this purpose, the IPMP component is based on initial inputs extracted from the Dublin Core metadata and makes the link using the identifier of the resource. The same identifier is used in DID and DII to be sure that the definition of the resource during a specific process is the same.

```

Require: Dublin Core Metadata file
Ensure: DID elements
Parse (DC-File.XML)
repeat
    DIDLItem ← GetRootNode()
    for each RootNode.Child do
        if RootNode.Child = Identifier then
            CreateNew (ComponentElement);
            Component.Resource ← ValueOf (Identifier)
        else {RootNode.Child ≠ Identifier}
            CreateNew (StatementElement);
            Statement.Value ← ValueOf (RootNode.Child)
        end if
    end for
until EOF
    
```

Algorithm 1: Generation of DI elements

### 5.1 MPEG-REL License Definition

To introduce security aspects, the Dublin Core standard has only a Rights tag which is not enough expressive. This tag gives information about rights held in and over the resource. Typically, a right element will contain a right management statement for the resource, or a reference to a service providing such infor-

mation. Right information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. Since this element was not enough expressive, the Qualified Dublin Core has been defined in order to refine the metadata Dublin Core. There are 7 new defined elements (Audience, Provenance, RightsHolder, InstructionalMethod, AccuralMethod, AccuralPeriodicity, AccuralPolicy). In this paper we focus on the two elements Audience, RightsHolder and the element Rights presented in Dublin Core. These elements are mandatory to describe the license that we have to associate with the resource. We show the meaning of these elements and we show how their content can be useful to build the MPEGREL license of the resource.

- 1) Rights Holder: a person or organization owning or managing rights over the resource. This tag is used by MPEG-REL to define the issuer part of the license.
- 2) Rights: this tag contains two elements:
  - AccessRights: it describes the set of rights that a user can have on the resource. This field will be used by MPEG-REL to define the set of actions granted by the license.
  - License: it gives the link of the license to be associated with the resource. Actually, Dublin Core is using licenses described in the CcREL language [16]. In our case, this tag will be filled by the link of MPEG-REL that will be defined for the resource.
- 3) Audience: it indicates a class of entities to whom the resource is intended or useful. This element should be used by MPEG-REL to define the set of users to whom we define the license.

Thus, to generate the MPEG-REL license to be associated with the resource we have the following required elements: the resource (DC:identifier), the issuer of the license (Rights Holder), the set of permissions to be defined on the resource (AccessRights) and the condition on the user to whom we will grant access and rights (Audience). The audience field is interpreted as a certificate license. This type of license is defined by MPEG-REL in order to check a specific condition on some user. The license will be valid after verification of the certificate license. For example, if the audience field says that the user of a resource *R* should be a student then, if we grant a license *L1* to a user Bob, we have to check that Bob is a student. This verification is done by the certification license. Certification license is a specific license type defined by MPEG-REL to allow the verification of some properties on a specific user. An example of an MPEG-REL license expressing the right of the Dublin Core metadata example is given in listing 5.

Listing 5: MPEGREL License Example

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <!-- From From: http://mpeg.telecomitalia.com/
   working_documents/mpeg-21/rel/REL_fcd.zip
   2003-05 -->
3 <license xmlns="urn:mpeg:mpeg21:2003:01-REL-R-NS"
   xmlns:sx="urn:mpeg:mpeg21:2003:01-REL-SX-NS"
   xmlns:mx="urn:mpeg:mpeg21:2003:01-REL-MX-NS"
   xmlns:dsig="http://www.w3.org/2000/09/xmldsig#"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-
   instance" xsi:schemaLocation="
   urn:mpeg:mpeg21:2003:01-REL-MX-NS rel-mx.xsd">
4 <r:grant>
5   <r:keyHolder><r:info><dsig:KeyValue><
   dsig:RSAKeyValue>
6     ...
7     </dsig:RSAKeyValue>
8   </dsig:KeyValue></r:info></r:keyHolder><mx:print/>
9   <mx:diReference><mx:identifier>urn:http://www.
   tutorials-on-line.ProgrammingLanguage.eu/</
   mx:identifier></mx:diReference>
10  <sx:ExerciseLimit><sx:count>1</sx:count></
   sx:ExerciseLimit>
11 </r:grant>
12 <issuer>
13   <keyHolder><info><dsig:KeyValue>
14     <dsig:RSAKeyValue><dsig:Modulus>
   X0j9q99yzA==</dsig:Modulus><
   dsig:Exponent>AQABAA==</
   dsig:Exponent></dsig:RSAKeyValue>
15   </dsig:KeyValue></info></keyHolder>
16 </issuer>
17 </license>

```

## 6. Conclusion

In this paper we considered the Dublin Core metadata to be the most generic standard to be used to describe digital resources. We showed that the element set proposed by this template is rich and generic enough to easily define mapping processes. These mappings are defined to transform different metadata into a description based on the Dublin Core structure in order to have a generic representation of digital resources. Moreover, we presented in this work an approach to link the metadata description with the MPEG-21 components. The purpose was to define an integrated model which also cares about the access control policy. These policies are not precisely expressed in the initial description of the Dublin Core metadata. To define the link between the Dublin Core metadata and the MPEG-21 components, we make use of elements defined within the Dublin Core template. This approach gives a more rigorous management of security aspects within metadata. MPEG-21 proposes the most suitable rights expression language to express in a well structured way the licenses to be associated with digital resources. The limitation of these licenses is their atomic definition. Indeed, each license, when generated, should be specified for a specific user, for example Bob. If Alice, another user, needs to access the same resource as Bob and with the same set of privileges then a new license must be generated for Alice. To bring a solution to this limitation, more abstract models to express licenses can be considered. XACML [18] and OrBAC [19] models may

be a solution to this limitation since they are based on the concept of roles. Thus, the same license can be generated for a group of users who needs the same privileges and who may have the same role.

## Acknowledgment

This work has been carried out in the MO3T (Modèle Ouvert 3-Tiers) project.

## References

- [1] <http://www.editeur.org/8/ONIX/>
- [2] IMS Learning Resource Meta-data Information Model - Version 1.2 Final Specification.
- [3] <http://hypercontent.sourceforge.net/docs/manual/develop/vcard.html>
- [4] Cheun Ngen Chong, Sandro Etalle, and Pieter H. Hartel. Comparing Logic-based and XML-based Rights Expression Languages. In Workshop on Metadata for Security, International Federated Conferences (OTM'03), Catania, Italy, November 2003.
- [5] Renato Iannella. Open Digital Rights Management (ODRL). In World Wide Web Consortium Workshop on Digital Rights Management for the Web (W3C DRM'01), Sophia-Antipolis, France, January 2001.
- [6] World Wide Web Consortium, "XML Inclusions (XInclude) Version 1.0," W3C Recommendation, 20 December 2004.
- [7] David Parott. Requirements for a Rights Data Dictionary and Rights Expression Language. Technical report, Reuters, June 2001.
- [8] World Wide Web Consortium, "XML Path Language (XPath) Version 1.0," W3C Recommendation, 16 November 1999.
- [9] ISO/IEC, "ISO/IEC FDIS 21000-2 Information technology - Multimedia framework (MPEG-21) - Part 2: Digital Item Declaration second edition," 2005.
- [10] ISO/IEC, "ISO/IEC 21000-3:2003 Information technology - Multimedia framework (MPEG-21) - Part 3: Digital Item Identification," March 2003.
- [11] A. Arnab and A. Hutchison. Requirement analysis of enterprise DRM systems. In Information Security South Africa, 2005.
- [12] A. Jamkhedkar and G. L. Heileman. DRM as a layered system. In ACM Workshop on Digital Rights Management, pages 11-21, 2004.
- [13] M. Ji, S. M. Shen, W. Zeng, T. Senoh, T. Ueno, T. Aoki, Y. Hiroshi, and T. Kogure. MPEG-4 IPMP extension for interoperable protection of multimedia content. EURASIP Journal on Applied Signal Processing, 2004(14):2201-2213, 2004.
- [14] ContentGuard. XrML Software Development Kit: User's Guide, 2001. [www.contentguard.com](http://www.contentguard.com).
- [15] World Wide Web Consortium (W3C). Extensible Markup Language (XML) 1.0 (Third Edition), 2004. [www.w3.org/TR/REC-xml/](http://www.w3.org/TR/REC-xml/).
- [16] Hal Abelson, Ben Adida, Mike Linksvayer, Nathan Yergler. ccREL: The Creative Commons Rights Expression Language. March 3rd, 2008.
- [17] International Organization for Standardization (ISO). ISO/IEC 21000-5 :2004 Information technology - Multimedia framework (MPEG-21) - Part 5 : Rights Expression Language, 2004. [www.iso.ch/iso/fr/prods-services/popstds/mpeg.html](http://www.iso.ch/iso/fr/prods-services/popstds/mpeg.html).
- [18] eXtensible Access Control Markup Language (XACML) Version 2. Standard, OASIS, February 2005.
- [19] A. Abou El Kalam, R. El Baida, P. Balbiani, S. Benferhat, F. Cuppens, Y. Deswarte, A. Mieke, C. Saurel, and G. Trouessin. Organization Based Access Control (Or-BAC). In IEEE 4th International Workshop on Policies for Distributed Systems and Networks (Policy 2003), Lake Como, Italy, June 2003.

**SESSION**  
**INFORMATION ASSURANCE**

**Chair(s)**

**Dr. Hiroaki Kikuchi**  
**Meiji Univ. - Japan**



# Simple Method to Quantify Audit Findings

Gary Lieberman  
Caldwell University  
Division of Business  
Caldwell, NJ, U.S.A.  
glieberman@caldwell.edu

**Abstract** – Deciding which security assessment findings are important enough to require immediate attention and which are not is challenging at best. In most cases the security assessment results are weighted and prioritized using impact values retrieved from a national database of vulnerabilities, developed by the federal government with little or no consideration given to the business use of the system being assessed or its risk impact on the business itself. The evaluation of the assessment data and the associated remediation decisions are often left to an IT staff that generally has little or no business acumen. This paper presents a method that analyzes and quantifies both the needs and the degree of sustainable business risk against a vulnerability impact scale. A method that allows for the quantitative determination of which elements in a large set of discovered vulnerabilities across numerous systems are important in the context of the company's business risk tolerance and which aren't. This method is designed to allow the handling of large data sets with accuracy and ease.

**Keywords**— *Vulnerability Assessment, Business Risk, Risk Impact, Decision Analysis, Vulnerability Remediation.*

## I. INTRODUCTION

Security assessments are costly and time consuming. They are even more costly when one considers that quite often the focus of the remediation effort is misdirected. Knowing which of those discovered vulnerabilities are important is as critical to a company's ability to reduce risk as knowing how to remediate them. Too often IT departments fail to properly determine the importance of each asset, and in turn, the importance of discovered vulnerabilities relating to those assets [1]. Because the tendency is to assume everything should be fixed rather than concentrating on the most important vulnerabilities, the remediation effort may be spent in the wrong areas [2]. Misdirected remediation efforts can be wasteful, counter-productive and may mean that the most critical vulnerabilities are going unaddressed while the efforts are misdirected. Having a method by which an IT department can quickly and easily determine which systems require immediate attention and then drill down to identify the vulnerabilities within those systems that need attention first is critical to an organization's successful security assessment program.

Risk analysis and security assessments are not something that should be relegated solely to an IT department. The analysis cannot be performed in a technology vacuum either, but must be a collaborative effort by a team of system administrators, security experts and people who understand the business mod-

el that the systems support. The goal is to develop a simple method that bridges the three areas together in a manner that is both functional and understandable to all of the disciplines involved. This method must be able to consider differences in business/technical risk impact for each system on a scale that is based upon national impact ratings, but is still unique to the organization resulting in a qualitative index that applies across all systems within that organization. The method presented in this paper produces an index by which IT and business personnel can identify and judge those systems which require the highest degree of remedial attention. Additionally, those systems with minor index scores can be relegated to the end of the list and addressed when timing is more opportune [3].

The focus of most security assessments is on identifying vulnerabilities, as they should be. Assessment tools are generally agnostic when considering the business risk associated with the discovered vulnerability [4]. *Phase Three* of the COBIT implementation framework [5] suggests that priority be given to areas that are easiest to remediate yet provide the greatest benefit. Furthermore, *Phase Four* recommends that remediation projects be defined that are supported by justifiable business cases. With this in mind the goal of this paper is to develop a method by which the severity of the discovered vulnerability is considered in the same context as that of the business risk assigned to an asset.

The simple qualitative method described in this paper allows for the ranking of systems and their weaknesses according their use and business context. For example, consider three systems with the same two vulnerabilities, one being a weak password policy and the other being an insecure treatment of backup tapes. Each vulnerability may be treated with a different level of reverence depending on several business risk factors. If system #1 is not networked and kept in a secure room with restricted access, the weak password policy may well be deemed less important than the backup tape insecurity. Conversely, if systems #2 and #3 are networked and facing the internet the password policy may be deemed much more important than the backup tape security. Lastly, if systems #1 and #2 are development machines with contrived data and system #3 houses business critical data, another area of consideration is introduced. Even with an assessment containing three systems and two vulnerabilities, the possible combination of risk factors and vulnerabilities to be considered becomes extremely

complex and challenging. Now consider the quandary the IT staff would have if this assessment covered 200 computers with a possible average of 30 vulnerabilities per system. The problem is then deciding which of the 180,000 vulnerabilities should be addressed first. To further complicate matters, consider that most assessments utilize some form of the FIPS Pub 199 security categorization standards classification of High, Medium and Low to rank discovered vulnerabilities [6]. With 180,000 vulnerabilities ranked high, medium and low the natural tendency is to tackle all of the high vulnerabilities first, then the medium ones and so on. But this would then have the IT staff jumping from one system to the next and then back again and not addressing the problem in a controlled and orderly manner. Furthermore, which should be handled first high risk vulnerabilities on a system with no critical data or low risk vulnerabilities on a system with highly sensitive data? Now consider that this decision needs to be made hundreds of times in our example assessment and the complexity of the problem becomes quite clear.

Without a simple qualitative method of evaluating assessment findings the decision as to which vulnerability is most important and should be addressed first is too subjective and does not allow for uniformity in a corporate wide security remediation effort. Especially if one were to consider the context of hundreds of computers with multiple vulnerabilities. Selecting the wrong systems to work on has the potential to deal a fatal blow to the firm's reputation, incur huge losses and even put the firm out of business should a critical vulnerability be exploited while the IT staff is focused elsewhere. There is no escaping the need for a scientific approach to analyze each discovered vulnerability as it applies to a business risk model. However, the more granular the analysis is the higher the level of cohesion with the business model that can be achieved. For instance, with the example mentioned above there needs to be 180,000 individual instances of evaluation for each discovered vulnerability in the context of the business model and the use of the computer on which the vulnerability exists rather than one single analysis of an assessment consisting of 180,000 vulnerabilities.

## II. RELATED RESEARCH

Spanos et al. [7] propose a Weighted Impact Vulnerability Scoring System that enhances the Common Vulnerability Scoring System (CVSS). Their system has significant merit as it considers weights based on vulnerability impact metrics such as Confidentiality Impact, Integrity Impact and Availability Impact. Their method looks at the asset-vulnerability relationship from a vulnerability perspective and does not consider the risk associated with the asset in a business context. However, they conclude their research by stating that their method could be enhanced by the development of an algorithm that considers the corporate value and risk of each information system which is closely aligned with the focus of this paper.

Nath et al. [1] present a novel approach that develops a reconciliation metric system which marries attack graph analysis

and vulnerability scan results. Their perspective is to analyze the network as a whole and identify the vulnerabilities with the highest probability of success and remediate that portion of the network. The goal of their research is aligned with the goal of this paper, which is to reduce the uncertainty surrounding large vulnerability studies and identify which of these vulnerabilities should be addressed first. Their solution seems to be a good fit for attack vectors that are network specific, but not sufficient handling large numbers of system based vulnerabilities.

Lund et al. [8] developed a method that uniquely ties the components of a system to an asset and considers overall risk on an asset by a ranking called an *Asset Level*. This lends itself nicely as a base from which to consider business rules in the assessment of asset risk. The concept of managing risk by asset is sound except the method they present does not allow for the same vulnerability within multiple assets to be evaluated singularly against the business impact on a qualitative scale. However, the evaluation of risk by asset level is a sound concept and would be useful as the basis for determining asset value in the research presented in this paper.

Ruyi et al. [3] propose an improved CVSS (Common Vulnerability Scoring System) which takes into consideration the OS and Server type in calculating a vulnerability score. The weakness in this approach is that two identical servers with identical OS's can have diametrically opposed impacts on the business if exploited by the same vulnerability due to the nature of the system's use. Their approach has merit if we were to use their modified CVSS scores instead of the standard CVSS scores in the calculation of the vulnerability score presented in this paper.

Jiang et al. [4] present a novel approach of vulnerability ranking using a context-aware ranking method aligned with business risk. Their approach is specifically focused on a software design pattern known as Service Oriented Architecture (SOA) and contains a complex algorithm to measure the importance of each service. It considers the exploitability of the vulnerability and the impact of the exploit on the service. A ranking of vulnerabilities is arrived at by the use of scoring functions. Their approach is rather complex, but appropriately so because they are tackling the scoring of vulnerabilities in a complex software design pattern. Their approach contains numerous ways to construct a scoring function unique to SOA vulnerabilities, and therefore, each vulnerability requires its own review. Their approach is commendable, although complex. It serves to illustrate that considering business risk in the scoring of vulnerabilities is indeed a valid pursuit.

Wu and Wang [9] use a model-based automated approach to quantify the overall vulnerability score of a company. They have built a tool that allows the user to model the enterprise vulnerability topology of a company based on business goals and remediate those vulnerabilities that threaten to impact the business goals the most. Their work is very closely aligned

with the work presented in this paper. However, the Wu and Wang EVMAT tool is specifically focused on e-commerce business, but does illustrate the value in considering the business context in the calculation of asset value.

### III. APPLYING DECISION ANALYSIS METHODOLOGY TO THE ASSESSMENT DATA

The goal of the Decision Analysis Methodology is to allow the decision maker the ability to make a productive informed decision using all of the available information and data. Every management decision made requires the decision maker to think in terms of objectives, alternatives and potential risk. Whether the choice involves a few criteria or many, the process remains the same. Given the nature of human fallibility and the inadequacy, or in this case the complexity of the data available, no decision methodology can guarantee the perfect decision every time [10]. By using a systematic framework for evaluating the data from a security assessment the IT staff can increase confidence in the order and ranking of the assets selected for remediation. Further, this method enables the systems with the highest critical risk to the firm to receive the most attention. This, consequently, reduces corporate risk and exposure as expeditiously as possible while assuring that the IT resources are being applied where they are most needed.

Decision Analysis Methodology suggests that any decision process can be reduced to a mathematical formula resulting in a weighted quantitative index of alternatives against which an informed decision can be made. In 1981 Kepner and Tregoe [10] introduced the concept of a *Weighted Decision Analysis Matrix* that facilitates this hypothesis. When used for the purposes of analyzing security assessment output data, the required matrix is primarily an array presenting on the left axis a list of vulnerabilities that will be evaluated regarding a list of systems presented on the opposing axis.

On the left axis, each vulnerability is weighted using the rounded Common Vulnerability Scoring System, version 2.2 score (CVSS) for that vulnerability. The CVSS Impact Scores (I) are available from the National Vulnerability Database (NVD) which is sponsored by the Department of Homeland Security (DHS). "NVD is the U.S. government repository of standards based vulnerability management data. This data enables automation of vulnerability management, security measurement, and compliance" [11]. This system provides base scores which represent the innate characteristics of each vulnerability listed in the Common Vulnerability and Exposure (CVE) database. It is useful here because it represents the perceived impact of a vulnerability as compared to other vulnerabilities in the database without consideration of the business context of the systems being assessed. The CVSS Impact Score scale ranges from one to ten, with ten being the most critical and one being the least.

It should be noted that any vulnerability scoring system will work within the context of this methodology. The Weighted Impact Vulnerability Scoring System (WIVSS) proposed by

Spanos et al. [7] will work equally well. Any scoring system can be used as long as it is consistent across all vulnerabilities. The opposing (top) axis consists of a list of systems. Table 1 shows the initial setup of a decision matrix. The top axis lists the Name or IP address of systems considered in the assessment and their criticality score (C). The left most axis contains vulnerabilities and their associated CVSS Impact Scores (I).

For the purposes of this paper the evaluation of the system risk impact (criticality) was calculated on a scale of one to four with four being the most critical and one being the least. These equate to Critical, High, Medium and Low which are often used in security assessment reports. The larger scale (one to N) can be used to gain a more granular perspective of the business risk impact of each system. An example of evaluating a system criticality score can be seen with a non-networked system that contains no critical information. This system would then have a low criticality score of one. Conversely, a system with that sits on the network facing the Internet and contains sensitive customer data may have a high criticality score of four or more accordingly.

Vulnerability Nessus ID	CVSS Impact Score	20.200.200.20	20.200.200.22	20.200.200.23	20.200.200.24	20.200.200.26	20.200.200.28	20.200.200.29	20.200.200.34
		C	C	C	C	C	C	C	C
10007	I	S	S	S	S	S	S	S	S
10043	I	S	S	S	S	S	S	S	S
10077	I	S	S	S	S	S	S	S	S
10116	I	S	S	S	S	S	S	S	S
10198	I	S	S	S	S	S	S	S	S
10205	I	S	S	S	S	S	S	S	S
10227	I	S	S	S	S	S	S	S	S
10245	I	S	S	S	S	S	S	S	S
10249	I	S	S	S	S	S	S	S	S
10263	I	S	S	S	S	S	S	S	S
10302	I	S	S	S	S	S	S	S	S
10394	I	S	S	S	S	S	S	S	S
10407	I	S	S	S	S	S	S	S	S

Table 1

It is important to note that determining the criticality value for the system is generally done at the time the system is built and put into use. Assessing a system's risk to the business is one of the most important last steps of the system's final rollout checklist and should updated as part of the annual system audit review or sooner when a significant change to the system make-up or use takes place. The criticality index is subjective in nature, but can be made less so through the use of risk assessment formulas such as calculating the Annualized Loss Expectancy (ALE). Determining asset risk requires collaboration between the business owners and the IT department to arrive at index numbers that are realistic. The advantage of this framework is that determining the criticality index is generally performed once at the system's birth, while vulnerability assessments should be yearly at a minimum.

To illustrate this methodology's simplicity and easy to use any preexisting scale by which asset business value is measured

can be used. Methods such as FIPS Pub 199 or the ones referenced in the "Related Research" section fit nicely and work well. Utilization of a monetary scale, such as the ALE, for asset value calculation would also fit nicely with this method. The key is that all assets must be evaluated on the same scale and in the context of business value.

For the purposes of this study a criticality index scale of one to four was chosen for use, with four being the most critical. The data used in this study was from an actual vulnerability assessment output collected from a scan using the Nessus Vulnerability Scanner from Tenable. The vulnerability ID's are those associated with Nessus Vulnerability Plugins. Normally, each vulnerability plugin is assigned a CVSS Impact Score that is made up of a complex formula considering vectors and metrics. A fully detailed description of the formulas for determining the CVSS Impact Score can be found on the NVD website at <http://www.first.org/cvss/cvss-guide.html>. In this study the actual CVSS Impact Scores associated with the vulnerability plugins were not made available to the author, therefore a randomized set of numbers ranging between one and ten were used instead. The impact of this on the results of the study was deemed insignificant, but worth mentioning.

CVSS scores are arbitrary and can be adjusted to better suit the needs of the firm commissioning the vulnerability assessment. Consider Nessus Plugin 20728, a weak SA password that carries a CVSS value of 6.5. Should the commissioning firm feel that strong SA passwords are of the utmost importance; the CVSS value can be adjusted upward and lend greater credence to occurrences of vulnerability plugin 20728. Again, this flexibility adds to usability of this method. As with the Criticality Index, any vulnerability ranking scale can be used in this method. The key being the same scale must be used for all discovered vulnerabilities.

IV. CALCULATING THE VULNERABILITY SCORE

The Vulnerability Impact Score (S) is the product of two independently arrived at numbers, I and C. The score (S) exists at the intersection of the CVSS Impact Score (I) and system Criticality Index (C) for those instances where the vulnerability is

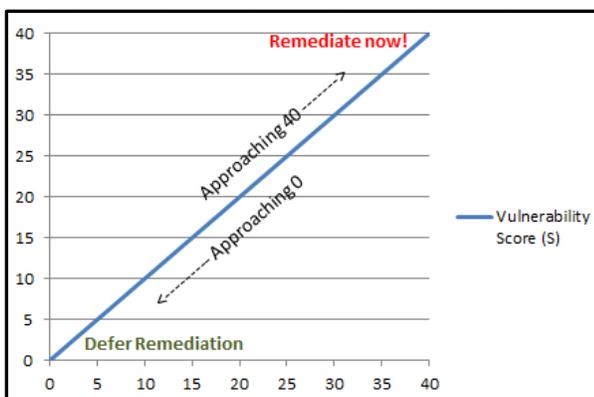


Figure 1

discovered on the system. A value of zero should be entered

for S at the CVSS Impact Score/Criticality Index intersection (I x C) for vulnerabilities that are not present on those systems.

$$I \times C = S$$

The Vulnerability Impact Score (S) represents the impact of a single vulnerability in the context of business risk for a single system. Considering a CVSS Impact Score range of one to ten and a Criticality Index range of one to four, the highest S value would be 40. Therefore, looking at the matrix one could extrapolate that any S value approaching 40 would warrant immediate vulnerability remediation on the associated system. While those S values approaching zero could withstand deferred remediation.

Table 2 demonstrates how a vulnerability with a high CVSS Impact Value (10043) of nine should be considered critical when considering system 20.200.200.20 and much less significant on system 20.200.200.29. System 20.200.200.20 has an S value of 36 which is approaching 40 and system 20.200.200.29 has an S value of 9 which is approaching 0. This is not to say that system 20.200.200.29 should be ignored, just given less consideration than system 20.200.200.20.

Vulnerability Nessus ID	CVSS Impact	System Criticality Index											
		4	3	3	3	2	2	1	4	3	4	4	
10007	2	8	6	6	6	4	4	2	8	6	6	6	6
10043	9	36	0	27	27	18	18	9	36	27	36	27	36
10047	6	24	0	18	18	12	0	0	24	18	24	18	24
10058	2	8	6	6	6	4	0	2	0	6	6	6	6
10085	3	12	9	9	9	6	6	3	12	9	12	9	12
10111	6	24	18	18	18	12	12	6	24	18	24	18	24
10227	9	36	27	27	27	18	18	0	36	27	36	27	36
10245	7	0	21	21	0	0	14	7	28	21	28	21	28
10249	3	0	9	9	9	6	0	0	12	9	12	9	12
10263	8	32	24	24	24	16	16	8	32	0	32	0	32
10302	1	4	3	3	3	2	2	1	4	3	4	3	4
10394	9	36	27	27	27	18	18	9	36	0	36	0	36
10407	7	28	21	21	21	14	14	7	28	21	28	21	28

Table 2

Furthermore, one could gain much greater insight if the S values for each system are summed arriving at a total System Vulnerability Impact Score (S<sub>s</sub>) which is indicative of the total risk impact of all discovered vulnerabilities in the context of business use and the risk for each system as compared to all other systems that are candidates for remediation. This allows one to make a determination as to which systems are wholly deficient and should receive the most attention. This is illustrated in table 3.

V. ANALYSIS OF TEST ASSESSMENT DATA

The data received and used for this study was procured from a global management consulting, technology services and outsourcing company. The data was scrubbed to remove any reference to the company or its customers. It is from an actual security audit performed by one of the company's security

teams. The assessment covered 190 servers, discovered 172 unique vulnerabilities across those 190 servers for a sum total of 27,121 discovered vulnerabilities. The delivered final assessment report was over 1,500 pages long and virtually impossible to discern which of the 190 systems required the most immediate attention or which ones required the least. The report was sorted by machine and listed each discovered vulnerability on an alpha scale of critical, high, medium and low. CVSS impact values were not supplied with the final assessment report so it was unclear as to which high vulnerabilities out ranked each other.

Vulnerability Nessus ID	CVSS Impact Score	CVSS Impact Score									
		20.255.200.202	292.268.70.99	292.268.70.74	20.250.260.200	20.250.260.238	20.250.270.64	20.255.200.224	20.250.272.52	20.250.65.80	
10007	2	8	8	8	8	8	8	8	8	8	
10043	9	0	36	36	0	36	36	36	36	3	
10077	6	24	24	24	24	24	24	0	24	2	
10116	2	8	8	0	0	8	8	8	0		
22532	5	20	20	20	20	20	20	20	20	2	
22536	9	36	0	36	36	36	36	36	36	3	
22538	9	0	36	36	36	36	36	36	36	3	
23643	8	0	32	0	32	32	32	32	32	3	
23645	6	24	24	24	24	24	24	24	24	2	
23646	4	16	16	16	16	16	16	16	16	1	
23647	3	0	12	12	12	12	12	0	12	1	
		3348	3328	3320	3304	3292	3292	3292	3288	328	

Table 3

VI. SIMPLICITY AND EASE OF USE

COBIT 5 requires a single integrated framework that enables a holistic approach to IT governance and management of enterprise IT. By enabling the evaluation of all system assets and all discovered vulnerabilities through a single comprehensive matrix which utilizes already established asset valuation scales and published vulnerability scores the IT staff is enabled to comply with and meet the holistic approach requirements of COBIT 5 [12]. Simplistic, in sense that already defined scales are leveraged for use and easy to use in that a Vulnerability Impact Score indicates which assets and a which vulnerabilities need the highest attention. This method leverages the already ascertained asset business risk and published vulnerability scores and analyzes them through mathematical formula based scientific approach that allows for easy identification of the weakest systems, ones that are in need of the most attention.

VII. CONCLUSION

The IT business decision maker requires information sources that contain complete, extensive and relevant information from which an informed decision can be made [13]. Even though the decision maker attempts to utilize as much of the entire body of available information as possible, it is impossible to thoroughly access all the required information. So IT decisions are constantly made under circumstances of uncertainty and based on sets of incomplete information [14]. The method

described in this paper bridges the gap between a large and completely onerous set of data and the ability to accurately assess that data during the decision process. By reducing the probability of an incorrect decision caused by information overload, the security remediation effort can be more directed and focused on the systems that need the remediation the most and afford the firm the highest degree of reduced risk possible. The last thing any IT manager wants to do is spend his/her valuable resources in the wrong place. The method described here is simple, adaptable and flexible enough to be made to fit almost any assessment data such as the linear assessment data used in this paper or assessment standards such as the NSA's INFOEC Assessment Method (IAM) or the INFOSEC Evaluation Method (IEM).

The major advantage of this method is not only does it allow for massive amounts of data to be analyzed without subjecting the decision maker to data overload and possibly incorrect decisions, but it also pushes the subjective analysis of the discovered vulnerabilities to a much lower level where only a single vulnerability and a single system are being evaluated at a time, regardless of and uninfluenced by other discovered vulnerabilities on other systems. Furthermore, the decisions made are done so based on a sound foundation of business risk rather than the perceived exploitability of the discovered vulnerability as is most often the case.

A further advantage is that the methodology and reasoning behind determining the remediation roadmap is now fully auditable and can clearly represent the logic behind the actions taken. Because the threat of exploitation looms heavy over all IT departments regardless of whether or not vulnerabilities exist, it is paramount to be able to defend to the auditors and regulators why one machine was remediated first and another second and so on.

VIII. ADDITIONAL RESEARCH

Additional focus can be applied to developing a stricter and more quantitative method for assessing business risk in determining the Criticality Score (C). This is by no means a simple task, which is why it is generally performed quite subjectively or better described as *by the seat of one's pants*. However, putting various uncertainties, both aleatory and epistemic, aside it would seem reasonable that risk factors could be boiled down into a formula by which a criticality index could be developed within repeatable process that could be applied to numerous systems. As mentioned in the "Related Research" section, much research is being performed in the area of determining asset risk. The method presented here lends itself to working with almost any method of determined asset risk without having to change the methodology. The key being that these systems can be judged on identical criteria producing a criticality index that is relative to all systems being assessed.

The entire process surrounding the building and populating of the matrix could be baked into an application which automatically gathers the vulnerability and risk assessment data and

displays it graphically on a computer screen. Additional functionality could be developed to allow for *what-if* modeling and automated attack graph creation.

#### REFERENCES

- [1] H. V. Nath, K. Gangadharan, and M. Sethumadhavan, "Reconciliation engine and metric for network vulnerability assessment," presented at the Proceedings of the First International Conference on Security of Internet of Things, Kollam, India, 2012.
  - [2] B. Wu and A. J. A. Wang, "EVMAT: an OVAL and NVD based enterprise vulnerability modeling and assessment tool," presented at the Proceedings of the 49th Annual Southeast Regional Conference, Kennesaw, Georgia, 2011.
  - [3] W. Ruyi, G. Ling, S. Qian, and S. Deheng, "An Improved CVSS-based Vulnerability Scoring Mechanism," in *Multimedia Information Networking and Security (MINES), 2011 Third International Conference on*, 2011, pp. 352-355.
  - [4] J. Jiang, D. Liping, E. Zhai, and Y. Ting, "VRank: A Context-Aware Approach to Vulnerability Scoring and Ranking in SOA," in *Software Security and Reliability (SERE), 2012 IEEE Sixth International Conference on*, 2012, pp. 61-70.
  - [5] ISACA, "COBIT 5: Implementation," ed: ISACA, 2012.
  - [6] F. I. P. S. PUBLICATION, "Standards for Security Categorization of Federal Information and Information Systems," ed, 2004.
  - [7] G. Spanos, A. Sioziou, and L. Angelis, "WIVSS: a new methodology for scoring information systems vulnerabilities," presented at the Proceedings of the 17th Panhellenic Conference on Informatics, Thessaloniki, Greece, 2013.
  - [8] M. S. Lund, F. den Braber, and K. Stolen, "Maintaining results from security assessments," in *Software Maintenance and Reengineering, 2003. Proceedings. Seventh European Conference on*, 2003, pp. 341-350.
  - [9] B. Wu and A. Wang, "A multi-layer tree model for enterprise vulnerability management," presented at the Proceedings of the 2011 conference on Information technology education, West Point, New York, USA, 2011.
  - [10] C. H. Kepner and B. B. Tregoe, "The Uses of Decision Analysis," in *The New Rational Manager*, ed Princeton, NJ: Princeton Research Press, 1981, pp. 103-105.
  - [11] (2008, Januray 23, 2014). *National Vulnerability Database Version 2.2*. Available: <http://nvd.nist.gov>
  - [12] ISACA, "COBIT 5 An ISACA Framwork," ed: ISACA, 2012.
  - [13] B. P. Kumar, J. Selvam, V. S. Meenakshi, K. Kanthi, A. L. Suseela, and V. L. Kumar, "Business decision making, management and information technology," *Ubiquity*, vol. 2007, pp. 1-1, 2007.
- [14] C. Saunders and J. W. Jones, "Temporal Sequences in Information Acquisition for Decision Making: A Focus on Source and Medium," *The Academy of Management Review*, vol. 15, pp. 29-46, 1990.

# Small to Medium Enterprise Cyber Security Awareness: an initial survey of Western Australian Business

Craig Valli, Ian Martinus and Mike Johnstone

c.valli@ecu.edu.au, i.martinus@ecu.edu.au, m.johnstone@ecu.edu.au

Security Research Institute

Edith Cowan University

Perth, Western Australia, Australia

## Abstract

Small to Medium Enterprises (SMEs) represent a large proportion of a nation's business activity. There are studies and reports reporting the threat to business from cyber security issues resulting in computer hacking that achieve system penetration and information compromise. Very few are focussed on SMEs. Even fewer are focussed on directly surveying the actual SMEs themselves and attempts to improve SME outcomes with respect to cyber security.

This paper represents research in progress that outlines an approach being undertaken in Western Australia with SMEs in the northwest metropolitan region of Perth, specifically within the large local government catchments of Joondalup and Wanneroo. The high order goal of the project was to assist with measures to improve their cyber security resilience and resistance to threats. This paper documents outcomes of an initial survey of SMEs and its implications for interventions to improve information security and make the businesses less susceptible to computer hacking incidents.

**Keywords:** cyber security, information security, computer hacking, small to medium enterprise

## 1 Introduction

The largest growth area for targeted cyber attacks in 2012 was SMEs; 31 percent of all attacks targeted them, and they area sustained a 30% increase in attacks overall [1]. These statistics are typical of reports, surveys and studies highlighting the significant threats now being realised on SMEs who use the Internet to conduct business [2]. The effects of cyber attack can destroy businesses financially through loss of bank account details as well as materially through loss of intellectual property or customer data. However, a recent survey by Kaspersky Labs reports that 75% of SMEs believing they are not targets of cybercrime and a further 59% say they have no data of interest for attackers[3].

Due to their need to compete and survive, SMEs are now some of the biggest adopters and users of the Internet and its associated technologies. These technologies include, but are not limited to, social media such as Facebook and Twitter, mobile phones and tablets connected to 3G/4G networks, email, cloud-based applications-often accessing these services on high speed DSL, Cable or Ethernet connections to the Internet [4, 5].

There is little literature available about the ability of SMEs to deploy, use and monitor cyber security countermeasures. Even basic cyber security countermeasures such as virus and malware scanners are increasingly complex and difficult to set up, except for an experienced IT security professional. Many SMEs simply cannot afford these professional services due to fiscal and time constraints [2, 6].

Vendors have not been insensitive to the need to protect operating systems from attack. Many of the popular operating system vendors have provided firewall solutions as system defaults. However, we posit that most SMEs would not know where to check the configurations of said firewall, or how to respond or report any attacks the firewall may have blocked during the course of its operation.

We asserted an approach that actively engages SMEs to help themselves better secure their enterprises can produce significant benefits for all stakeholders. Our plan was to engage with small to medium enterprises in the Wanneroo and Joondalup local government areas in Western Australia to assist with this new economy threat. The aim of the project pilot was to increase SME knowledge about cyber security issues. In addition to knowledge, our aim is to arm them with the tools and techniques to better protect their business by achieving a lower cyber security risk profile through active utilisation. This paper outlines the issues with respect to small to medium enterprise cyber security and analyses initial findings from the survey instrument used with study participants.

## 2 The SMESEC project

The SMESEC project actively involves researchers from the ECU Security Research Institute and the relevant staff from the local government, introducing a program to educate small to medium enterprises about cyber security issues they face and the need to address through positive action. This engagement was achieved through a survey and initial awareness raising workshops, leading to intervention driven by the survey's findings. The survey analysis will allow targeted workshops as the main vehicle for the dissemination of information to SMEs on how best to protect their business.

The four stages of the project are as follows:

### Stage 1 - Initial survey of small to medium enterprises in the catchment areas

The survey established a baseline of user knowledge about cyber security and also the perceived risk that cyber security presented to their businesses. The anonymous survey was completed online or in face-to-face mode.

### Stage 2 Analysis of survey results

An analysis of survey results has been undertaken to find areas of focus to be addressed in the intervention workshop. (The context and content of this paper)

### Stage 3 Conduct of the workshop

This is designed to be a two-hour information session examining key issues identified from the analysis of the survey. The aim of the workshop is to enable businesses to utilise existing tools and freely available, robust software to create a more secure and resilient business.

### Stage 4 Post workshop survey

This will be conducted with the attendees of stage 3, approximately one to two months after the workshop to assess the impact of the intervention.

## 3 Initial Survey details

The survey was designed to collect basic information about:

1. Basic demographics - enterprise type, age of respondents
2. Respondents knowledge of cyber security issues

3. Type and numbers of generic device types used on the Internet
4. Any cyber security protections or countermeasures that may be applied to the devices
5. Update frequency for equipment or operating systems
6. User confidence in using certain applications

This information was collected using 17 exploratory questions; the aim was not to make the questionnaire long or protracted, nor overly technical in emphasis. The survey was deployed as an anonymous online survey and all data were acquired in this fashion. The total number of respondents to the survey so far is 50. The survey was distributed via the respective business associations from the local government regions to approximately 1200 individual email addresses.

## 4 Survey Results

### 4.1 Basic Population Demographics

The industry profile of the respondents was Retail 12%, Services 44%, Manufacturing 6%, Other 38% and no Primary. The age demographic of respondents was 18-35 year old (18%), 35-45 year old (34%), 45-55(26%) and 55 or over (22%).

### 4.2 Technology – Profile and Use Practices

The type of devices used to access the Internet in the surveyed SMEs demonstrated a broad mix and number of device types. There is more than one device type used with the average being 2.92 device types per respondent used for business purposes. It is interesting to note that the desktop computer is still the most single used overall piece of IT device at 84% but only marginally. Smartphone usage is 78%, laptops at 74% and tablets at 54% in the respondents businesses.

Access technology was also identified in the study based on potential business and home use. The results were ADSL Modem 62%, ADSL Wireless 54%, 3G/4G Wireless 50% and 2% who did not know (the labels here are truncated for brevity, explanation between the different types of ADSL was provided in the questionnaire to reduce ambiguity).

Inquiry around cyber security countermeasures deployed by the respondents evinced that Firewall 88%, Virus Scanner 86%, Malware Scanner 43%, Spam Killer 35% and 8% did not know what they were using. Of those respondents who have anti-viral countermeasures 12% indicated they never updated, 20% did not know when they update and 8% said less than once a month since they last updated signatures. This profile represents 40% of SME businesses employing a countermeasure and it been largely ineffective due to poor process. Encouragingly though 37% update several times a week, 12% once a week, 2% 2-3 times a month, 8% once a month.

In response to questions about installing updates of operating systems on PCs, the automatic update functionality was used by 64% of respondents, a further 16% were updating at least weekly. Of the remaining 20%, 14% at least update once a month, with 4% less than once a month and 2% not knowing. Of the Smartphone or tablet owners users 92% had installed updates on their devices, with 8% indicating never having done so.

Phishing emails had been sent to 98% of all respondents, with 98% of respondents also asserting that they knew what one was. The type of phishing received by respondents was identified as financial/banking institution related 86%, free prize 78%, lottery 80% and other topics 40%.

One of the key issues is around security of financial transactions in particular Internet banking. Respondents answered questions about their perceptions of security and safety of using banking on the Internet. Of the respondents 4% never felt secure when using Internet banking, with a further 2% rarely feeling secure and 12% sometimes. Nearly half (48%) felt secure most of the time and the remaining 32% feeling always secure.

## 5 Discussion

### 5.1 Mobile devices proliferate

Mobile devices combined are the dominant platforms outnumbering PCs two-to-one in the surveyed businesses. This dominance presents some interesting issues for cyber security. It is fair to assume mobile devices would see connectivity to multiple networks and also possible types of network channels. There is safety in assuming that even the simplest usage scenarios of business and home use is a connection to two different networks. In the case of smart phones, the use of multiple channels opens up opportunities for potential exploit or compromise. This would include 802.11 wireless, 3G/4G network and the

often forgotten Bluetooth. This protocol is the predominate pairing mechanism when a user is operating their device while in their car.

### 5.2 Multiple Channel, multiple threat

The respondents clearly identified they are using multiple Internet media types for access to business transactions with a large proportion being mobile. This trend is consistent with usage patterns from Australian Bureau of Statistics reports where mixed use is demonstrable as is the ongoing proliferation and penetration of mobile devices [4, 5]. These mobile devices primarily use wireless transmission for communication to networks and other devices e.g. automotive systems. All wireless transmission is susceptible to interception regardless of technology. Basic physics confirms this assumption. The protection for these wireless systems typically relies on protocols and cryptographic countermeasures which are manifestly insecure.

ADSL-based wireless, and wireless used in these mobile devices is typically 802.11 b/g/n and is known to be insecure[7]. Through deduction and implication, 54% of SMEs are vulnerable to exploitation via known documented 802.11 vulnerabilities. Many wireless transmissions are vulnerable to simple but illegal interception of wireless signal, the technology for extraction of cryptographic keys and subsequent decode are effective and well documented. As early as 2001 this has been occurring for the capture of financial credentials by cyber enabled criminals from wireless enabled technology.

The type and value of information disclosed on wireless can cause SMEs to become attractive targets for cyber criminals. We posit that this is a haven for data relating to identity theft. Given that identity theft is the largest and fastest growing crime types in the world this should be considered by SMEs concerned with protection of customers private details. Of the Wanneroo and Joondalup businesses surveyed, they were vocal in their reaction to 'old economy' physical theft, but less aware of the damage of virtual identity theft and the potential for devastating business loss.

The use of personal business devices and subsequently connecting to various outside networks also raises the risk to an SME. Wireless access point (WAP) spoofing techniques are well documented, but awareness is low. Despite an SME providing safe secure networks at the business premise, the level of risk increases dramatically if an employee logs in from a home networks or the "free" wireless at the local cafe, library or city centre 'hotspot'. SMEs can

be seen as a trusted second or third party, and used as "watering holes" to break down the security of other businesses with cyber criminals stalking a business user through the other party[1]. A recent example of customers accessing a Chinese restaurant's web site to order through the online menu and subsequent infection with malicious code demonstrates this point. The 'watering hole' was designed by cyber criminals in order to facilitate an attack on a geographically close oil and gas company[8].

The attendant data leakage possible through a work synchronised/synchronising device being exploited and compromised while in a home/external network scenario was not specifically covered in the survey. This exploitation could occur either through malicious extraction of the data from the device directly using a USB as a result of physical access, or through interception of transmission across insecure or unmonitored network endpoints. There is a colloquial term within cyber criminal networks called "whaling" for the targeting of high worth individuals [9]. It should be noted this high worth often relates to information value not personal financial wealth, making this compromise of most SMEs real and viable.

### 5.3 Basic cyber security countermeasure use and deployment is lacking

It was disappointing to see firewalls as a primary defence not being utilised fully, given that all contemporary operating systems have firewalls as default configurations. Equally, the use of antivirus is also low given the attendant risk these present for conventional PCs. It is again more alarming when taken in the context for mobile devices when 82% of respondents do not install antivirus. Furthermore on the mobile platforms 94% of all respondents have downloaded applications (apps) from the various vendor based platforms. Given that it is known that these apps, even from "reputable" vendor sites, are vulnerable [10] to exploit due to programming flaws or deliberate insertion of malicious code, this raises the risk to SMEs significantly. There appears to be a "lost in translation" event here around transitioning knowledge from PC environments to smartphone and tablet environments. Further research into the penetration of smartphone apps as small business take up increases would be useful to investigate including the loss of business as a result of those dubious downloads.

### 5.4 Patch is starting to match message

The survey responses relating to operating system patching indicated that the messages from various initiatives in industry and government around the need to patch systems regularly are potentially being heard[11, 12]. The results show that the automatic update processes when seamless or made to be "set and forget" are elected as a choice by 64% of end users. The implication *ceterus paribus* here for cyber security is that 80% of operating system related vulnerability or exploit code will prove to be ineffective if patched within a week cycle, and 64% as soon as an update is available via automatic update mechanism. This lessens the effective windows for exploit, threat realisation by cyber criminals and ultimately results in a common good outcome for cyber security.

## 6 Conclusion and Future Work

There is a definite identified need for education and dissemination of cyber security information to SMEs in this project as we move into Phase 3 of the SMESEC project. The initial survey results strongly indicate that there is a pressing need for direct intervention and the wider Wanneroo and Joondalup business group has indicated strong interest in this component through their online and physical enquiry.

This intervention is to be achieved through the facilitation of a targeted practical workshop to enhance understanding and implementation of cyber security countermeasures for SMEs. In addition, post-workshop the provision of supporting process documentation in the form of conventional paper-based and online materials for SMEs is seen as an important support mechanism.

The content of the workshop will be as follows. First, the need to get antivirus countermeasures installed on mobile platforms owned by these SMEs and at the same time reinforcing the messaging about regular patching and updates. This will significantly reduce the risk to the 94% of SMEs who are indicate they are currently using unprotected devices.

Second, considerations about the use of wireless for transmission of critical or sensitive business information across wireless conduits will be explored. This intervention will involve educating the individual business about the significant risk wireless presents when transmitting business data such as email and document attachments. There will be a demonstration of the use of high grade file-based encryption for data storage using for instance

Windows EFS at a file system level. Educating SMEs about the effective use of open source cryptographic solutions such as OpenPGP and Truecrypt to allow for the safe transfer of documents is also incorporated in the workshop agenda.

Finally, a refresher or back to basic assistance on firewall, anti-virus, malware and spam-based applications will close the first stage of the loop, after which point the business community can then become advocates of the process we designed and deployed. The results indicate that SMEs have not transferred skills and knowledge around these technologies to smart phone and tablet in particular and further 'communications' is needed in the marketplace to get the message out there.

As planned in the project, there will be an evaluation of the workshop phase with a post-workshop survey. Furthermore, we are seeking mechanisms to use existing Australian government data on botnet data for instance to demonstrate effectiveness of our interventions. As with any intervention, dissemination to a wider business group beyond the original geographic area is the wider goal.

## 7 References

- [1] Symantec, *Internet Security Threat Report 2013*, 2013, Symantec Corporation.
- [2] J. Hayes and A. Bodhani. (2013) Cyber security - small firms now in the firing line. *Engineering and Technology Magazine*. Vol 8 Issue 6, The Institution of Engineering and Technology: London, UK
- [3] W. Ashford. (2013). *SMEs believes they are immune to cyber attack*. Available: <http://www.computerweekly.com/news/2240216202/SMEs-believes-it-is-immune-to-cyber-attack-study-shows>
- [4] Australian Bureau of Statistics, "8153.0 - Internet Activity, Australia, June 2013," Australian Bureau of Statistics, 2013. Available: <http://www.abs.gov.au/AUSSTATS/abs@.nsf/allprimarymainfeatures/70EF9515319BA B35CA257CB30013246D?opendocument>
- [5] Australian Bureau of Statistics, "8166.0 - Summary of IT Use and Innovation in Australian Business, 2011-12" Australian Bureau of Statistics, 2013.
- [6] Anonymous. (2013). *SMEs must get better at the cyber security basics, ICAEW tells Parliamentary group on IT*. Available: <http://www.icaew.com/en/about-icaew/newsroom/press-releases/2013-press-releases/smes-must-get-better-at-the-cyber-security-basics-icaew-tells-parliamentary-group-on-it>
- [7] Valli, C. and P. Wolski. *802.11b Wireless Networks Insecure at Any Speed*. in *International Conference on Security and Management - SAM'04*. 2004. Las Vegas: CSREA Press.
- [8] N. Perlroth, "Hackers Lurking in Vents and Soda Machines," in *New York Times*, New York Edition, 2014, 8th April, p. A1.
- [9] IBM, "X-Force 2011 Mid-Year Trend and Risk Report," IBM 2011.
- [10] P. Krill. (2012, 31 March). *Google finally scans malware-ridden Android Market*. Available: <http://www.infoworld.com/d/security/google-finally-scans-malware-ridden-android-market-185654>
- [11] Australian Signals Directorate. (2014). *Strategies to Mitigate Targeted Cyber Intrusions*. Available: <http://www.asd.gov.au/infosec/top-mitigations/top35mitigations-2014-table.htm>
- [12] CERT Australia, "The top cyber security tips for small to medium business," CERT Australia, Australia, 2014. Available: <https://www.cert.gov.au/system/files/5/5/CERT-Australia-top-cyber-security-tips-for-small-to-medium-business.pdf>

# Detecting the Vulnerability of Multi-Party Authorization Protocols to Name Matching Attacks

Wenjie Lin (Contact Author)\*, Guoxing Chen\*, Ten H. Lai\*, David Lee†

\*Ohio State University, 2015 Neil Ave, Columbus, OH

†HP Labs, 1501 Page Mill Road, Palo Alto, CA

{linw, chenguo, lai}@cse.ohio-state.edu {david.lee10}@hp.com

**Abstract**—Software as a Service (SaaS) clouds cooperate to provide services, which often provoke multi-party authorization. The multi-party authorization suffers the so-called name matching attacks where involved parties misinterpret the other parties in the authorization, thus leading to undesired or even fatal consequences (e.g., an adversary can shop for free or can log into a victim's Facebook account).

In this paper, we propose a scheme to detect the vulnerability of multi-party authorization protocols that are susceptible to name matching attacks. We implement the detecting scheme and apply it to real world multi-party authorization protocols including Alipay PeerPay, Amazon FPS Marketplace, and PayPal Express Checkout. New name matching attacks are found, and fixes are proposed accordingly.

**Keywords:** multi-party authorization, name matching attacks, protocol analysis.

**Tracks:** Security Applications, Information Assurance

## I. INTRODUCTION

There are a large number of Software as a Service (SaaS) clouds in the market, such as Google Docs (document management), Dropbox (storage and sharing), HP ePrint (printing), Salesforce (customer relation management), Cloud9 (revenue forecast), 1010Data (Big Data analysis), and Alipay (third-party payment). When these clouds cooperate with one another, more advanced services can be provided, and they often require multi-party authorization. For example, a user Alice may authorize HP ePrint to retrieve and print her bank statements stored in her Google Docs—this service would involve three-party authorization. There are services requiring four-party authorization. For example, Alice may order some products from Newegg and ask her husband Bob to pay the bill through Alipay (which is a popular PayPal-like service provider in Asia). Here, Alice and Bob together authorize Newegg to withdraw some money from Bob's Alipay account.

Cloud services requiring five- or more-party authorization are not unthinkable. Envision a new cloud service called SyncData that allows multiple users to synchronize their data stored at various clouds. For example, Alice and Bob may want SyncData to update Alice's dataset in 1010Data according to Bob's sales data in Salesforce. This requires five-party authorization. As another example, suppose it takes several managers' approvals to allow Cloud9 to retrieve a company's sales data in Salesforce, thus triggering an  $n$ -party authorization.

Cisco recently announced plans to build the world's largest global Intercloud—a network of clouds—in the next two

years. As data communications and job migrations among clouds become more efficient and secure, we believe that new services involving multiple clouds and thus requiring multi-party authorizations will be emerging.

In this paper, we investigate multi-party authorization protocols in cloud services, focusing on detecting the vulnerability to the so-called *name matching attacks*, which we believe are main threats to the security of multi-party authorization. Our study is motivated by the following reasons.

First of all, name matching attacks are new threats that typically exist in cloud-based authorization but not in traditional authorization such as RBAC [22], PBAC [20] and ABAC [28]. In cloud applications, a user may use different usernames at various clouds. It is not trivial (if not impossible) to figure out whether the username say aaa at one cloud and the username say xyz at another cloud refer to the same user (especially when privacy is a concern). Thus, when an initiator issues an authorization request with  $n$  parties' names in it, it is important, but nontrivial, for each named party to know who exactly the other parties are — which is the root cause of the name matching attacks.

Secondly, name matching attacks lead to undesired or even fatal consequences to multi-party authorization. In the aforementioned example of HP ePrint, Alice (with username xyz at Google) grants the authorization at Google Docs. If an adversary was able to launch an attack, such that HP ePrint failed to match Alice's name—thought it was bbb (Bob's HP username) who granted the authorization—HP ePrint would print Alice's bank statements on Bob's printer (Figure 1). Alice's personal data would be released. We summarize name matching attacks against three-party authorization in the literature in Table I.

Finally, to the best of our knowledge, although name matching attacks have been vaguely recognized in the literature as violations of “a series bindings” [24] and “association” [26], detecting the vulnerability of multi-party authorization protocols to the name matching attacks has never been singled out as a problem. The existing research focused on three-party scenarios and was in case-by-case fashion. For example, researchers identified attacks [3], [4] against the famous three-party authorization protocol OAuth [12], [13] and its applications to Single Sign On [25], [23], [7], [6]. Various attacks were also identified in PayPal and in Amazon payment services [24]. As far as we know, more-than-three party authorization and its security have not been studied.

This paper has two contributions. First, we propose a

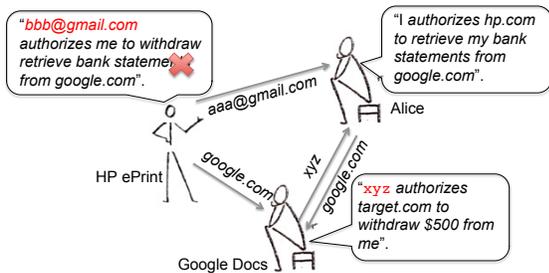


Fig. 1: An example of name matching attack

TABLE I: Name matching attacks in literature

Pub date	Affected protocols	Reference
April, 2009	OAuth 1.0	Session fixation attack [3]
Aug, 2011	OAuth v2-16	Auth code swap attack [4]
May, 2011	NopCommerce integrating Amazon Simple Pay	[24]
May, 2011	Amazon Payment SDK	[24]
May, 2012	Google ID & Smartsheet SSO	[25]
Oct, 2012	One third of studied RPs using OAuth in [23]	Session swapping attack [23]

scheme to detect the vulnerability of multi-party authorization protocols to name matching attacks. Our detecting scheme is inspired by the observation that all current multi-party authorization protocols are composed of five types of three-party primitives. Therefore, by checking the security of each primitive and their composition with Protocol Composition Logic (PCL) [11], [21], we are able to detect the vulnerability. Comparing to other formal method approaches (e.g., [14], [19]), our detecting scheme takes more protocol insights into consideration, and thus can pinpoint the vulnerability in a protocol and suggest a remedy immediately.

Moreover, we implement the detecting scheme and apply it to high-profile three-party and four-party authorization protocols, including Alipay PeerPay, Amazon FPS Marketplace, and PayPal Express Checkout. New vulnerabilities to name matching attacks are found and the remedies are proposed.

The rest of the paper is organized as follows. In Section II, we introduce the system model and the adversary model. We propose the detecting scheme in Section III. The scheme is applied to case study in Section IV.

## II. SYSTEM AND ADVERSARY MODELS

**Entities and names:** There is a universal set  $\mathcal{E}$  of *entities*, which in practice consists of all cloud service providers (e.g., Google, Amazon), end users (e.g., Alice, Bob), enterprise users, traditional service providers (e.g., banks, Visa) and so on. An entity knows another entity by a single *name* such as a user's username or a cloud's URL. (Each entity typically knows only a subset of entities in  $\mathcal{E}$ .) For simplicity, we assume that a user may have at most one account/username at a cloud. (If Alice has multiple accounts/usernames at Google, we will have to treat the "person" Alice as multiple entities, each corresponding to one of her Google username.) Denote the

set of all entities' names as  $\mathcal{N}$ . The acquaintances between entities (i.e., who knows who by what name) are modeled as a function  $\text{Ent}$ , as described below.

**Definition 1: [Function Ent]**  $\text{Ent} : \mathcal{E} \times \mathcal{N} \rightarrow \mathcal{E} \cup \{\perp\}$ . For each entity  $e \in \mathcal{E}$  and name  $m \in \mathcal{N}$ , let  $\text{Ent}(e, m)$  denote the entity that is known to  $e$  by the name  $m$ . If  $e$  doesn't know anybody by the name  $m$ , then  $\text{Ent}(e, m) = \perp$ .

*Example:* If Alice's username at Google is xyz, then  $\text{Ent}(\text{Google}, \text{xyz}) = \text{Alice}$ .

**Multi-party authorization:** In an  $n$ -party authorization, an initiator sends out an authorization request (typically to one of the entities involved in the request), thereby triggering  $n$  entities to exchange messages and reach an authorization decision respectively (e.g., the decisions made by HP ePrint, Google Docs and Alice in the aforementioned example in Figure 1).

The request specifies names of  $n$  entities—authorizers, authorizees, and enforcers—in the authorization. An authorizer (e.g., Alice) is an entity who grants the authorization; An authorizee (e.g., HP ePrint) is an entity who is granted the authorization; An enforcer (e.g., Google Docs) is an entity (often a server) who manages resources and enforces the authorization.

**Reliable and secure channels:** The channel is reliable without message loss (e.g., a TCP channel), and is secure against eavesdroppers and interception (e.g., an SSL channel). However, we do not assume all messages can be authenticated, because it is unrealistic to assume all ordinary users (e.g., Alice and Bob) have a public key known by the rest parties in the authorization.

**Bounded message delay and computation time:** We assume that there are bounds on message delay and computation time. A party aborts the protocol if it does not receive a message or obtain a computation result in time.

**Concurrent self composition:** As there are concurrent multi-party authorizations, an authorization protocol should be secure under concurrent self composition, that is, the protocol remains secure even when it is executed concurrently multiple times [17].

**The adversary model:** We consider the Byzantine adversary model [15] where an adversary can exhibit arbitrarily malicious behaviors. Moreover, an adversary can launch web attacks described in [23]: It can send malicious links via spam or by posting an Ad on a malicious website. If a victim clicks on the malicious link, the (browser of the) victim will send HTTP requests (i.e., GET and POST methods) with the messages crafted by the adversary.

## III. DETECTING THE VULNERABILITIES TO NAME MATCHING ATTACKS

In this section, we introduce a scheme that can detect vulnerabilities to name matching attacks in a class of multi-party authorization protocols, the protocols that are composed by three-party primitives (described next).

### A. Overview of our detecting scheme

To the best of our knowledge, the multi-party authorization protocols we are aware of—which are provided by Google,

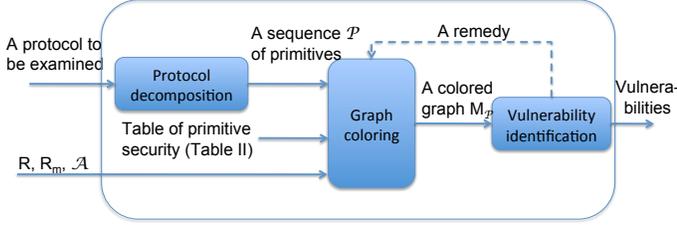


Fig. 2: A scheme to detect vulnerabilities to name matching attacks

Amazon, Alipay, Facebook, and other high-profile vendors—are all composed of three-party primitives (described in Section III-B1). We thus focus on detecting vulnerabilities in this kind of multi-party authorization protocols. The detecting scheme includes three modules (Figure 2): protocol decomposition, graph coloring, and vulnerability identification.

In the protocol decomposition module, a given  $n$ -party authorization protocol is re-written as a sequence of three-party primitives. We describe the primitives and their security in Section III-B1 and Section III-B2. In this step, manual assistance is needed.

In the graph coloring module, the sequence of primitives together with other inputs are fed into Algorithm 1, which thereby outputs a colored graph. If the colored graph has a red or grey edge, a name matching vulnerability is likely to exist.

In the vulnerability identification module, one can identify the vulnerability by investigating how the graph is colored step by step in the graph coloring module. This investigation reveals the root cause of the vulnerability, so that one can construct a name matching attack accordingly. With each vulnerability, one can propose a remedy and verify it by repeating the last two modules. Here manual assistance is needed.

### B. Protocol decomposition

All multi-party authorization protocols that we are aware of can be composed by the following five types of three-party primitives.

**1) Five types of three-party primitives:** A primitive  $(e, c, d)$  is a protocol involving three (not necessarily distinct) entities  $e, c, d$ . The input to the protocol is an entity  $e$  and two names  $m_c$  and  $m_d$  by which  $e$  knows entities  $c$  and  $d$ , respectively. (Note that  $c$  and  $d$  are not part of the input; only their names known by  $e$  are.) At the end of the protocol, entity  $c$  outputs a name  $m'_d$  by which it knows entity  $d$ . That is,  $c$  outputs a name  $m'_d$  such that  $\text{Ent}(e, m_d) = \text{Ent}(c, m'_d)$ . (We could have written the primitive as  $(e, m_c, m_d)$ , but chose to write it as  $(e, c, d)$  for ease of understanding.)

In the following,  $e, c, d$  indicate distinct entities. A session ID is used to distinguish between different sessions or executions of the same primitive.

- 1)  $P_1 = (e, c, c)$ : Suppose  $e$  knows  $c$  by a name  $m_c$ . The protocol is trivial; just let  $e$  send  $c$  a message containing the name  $m_c$  and session ID  $sid$ . After the primitive (protocol),  $c$  outputs the name it calls itself in session  $sid$ :

$$e \rightarrow c : \{m_c, sid\};$$

$c$  outputs its own name in session  $sid$ .

- 2)  $P_2 = (e, c, e)$ : Suppose  $c$  calls  $e$  by a name  $m_e$ . The protocol is straightforward —  $e$  sends  $c$  a message containing the name  $m_e$  and session ID  $sid$ . After the primitive,  $c$  outputs  $m_e$  in session  $sid$ :

$$e \rightarrow c : \{m_e, sid\};$$

$c$  outputs  $m_e$  in session  $sid$ .

- 3)  $P_3 = (e, c, d)$  where  $e, c$  know  $d$  by the same name  $m_d$ . After the primitive,  $c$  outputs  $m_d$  in session  $sid$ :

$$e \rightarrow c : \{m_d, sid\};$$

$c$  outputs  $m_d$  in session  $sid$ .

- 4)  $P_4 = (e, c, d)$  where  $e, d$  know  $c$  by the same name  $m_c$ . The protocol consists of two steps. In the first step,  $e$  sends a message to  $d$  (who is known to  $e$  by the name  $m_d$ ): “Hi  $m_d$ , please ask  $m_c$  to output your name.” In the second step,  $d$  sends a message to  $c$ : “Hi  $m_c$ , my name is  $m'_d$ .” After the primitive,  $c$  outputs  $m'_d$  in session  $sid$  ( $m'_d$  is the name by which  $c$  knows  $d$ ):

$$e \rightarrow d : \{m_c, m_d, sid\};$$

$$d \rightarrow c : \{m'_d, sid\};$$

$c$  outputs  $m'_d$  in session  $sid$ .

- 5)  $P_5 = (e, c, d)$  where  $e, d$  know  $c$  by the same name  $m_c$ , and  $c, d$  knows  $e$  by the same name  $m_e$ . The protocol includes four steps. The first step is similar to the one in  $P_4$ , except that  $e$  generates a secret  $sec$  and sends it to  $d$ . In the second step,  $d$  sends a message to  $c$ : “Hi  $m_c$ , my name is  $m'_d$ . The secret is  $sec$ . Please confirm it with  $m_e$ .” In the third and last steps,  $c$  sends  $sec$  back to  $e$ , who verifies if it has sent the  $sec$  in the first step. After the primitive,  $c$  outputs  $m'_d$  in session  $sid$  ( $m'_d$  is the name by which  $c$  knows  $d$ ):

$$e \rightarrow d : \{m_c, m_d, sec, sid\};$$

$$d \rightarrow c : \{m_e, m'_d, sec, sid\};$$

$$c \rightarrow e : \{sec, sid\};$$

$$e \rightarrow c : \{\text{correct}, sid\};$$

$c$  outputs  $m'_d$  in session  $sid$ .

**2) Security of individual primitives:** We examine if each primitive satisfies the security property that all honest entities output correct names, that is, if  $c$  is honest and outputs a name  $m'_d$ , there must be an entity  $e$  who inputs two names  $m_c, m_d$  such that  $\text{Ent}(e, m_c) = c$  and  $\text{Ent}(e, m_d) = \text{Ent}(c, m'_d)$ .

The security of individual primitives is summarized in Table II. In the table, 0 indicates that a name matching attack can be found and 1 indicates that the security property holds (if certain condition is satisfied). Here  $\mathcal{A}$  is the authentication attributes. For example,  $\mathcal{A}(c, e) = 1$  means that  $c$  can authenticate the messages sent by  $e$ . The proof is based on Protocol Composition Logic (PCL) [11], [21]. Due to space limit, please refer to our full paper [2] for details.

### C. Graph coloring

The GRAPH\_COLORING algorithm (Algorithm 1) takes five inputs: (1) a sequence of primitives  $\mathcal{P}$ ; (2) the set of entities  $R = \{au_1, \dots, au_i, az_1, \dots, az_j, ef_1, \dots, ef_k\}$  in an authorization, where each  $au_i$  is an authorizer, each  $az_j$  is an authorizee, and each  $ef_k$  is an enforcer; (3) the set of entities  $R_m$  ( $R_m \subset R$ ) who may be malicious; (4) the authentication

TABLE II: Security of individual primitives

Primitive	All are honest	$e$ may be malicious	$d$ may be malicious
$P_1 = (e, c, c)$	1 if $\mathcal{A}(c, e) = 1$	0	–
$P_2 = (e, c, e)$	1 if $\mathcal{A}(c, e) = 1$	0	–
$P_3 = (e, c, d)$ , where $e, c$ know $d$ by the same name.	1 if $\mathcal{A}(c, e) = 1$	0	1 if $\mathcal{A}(c, e) = 1$
$P_4 = (e, c, d)$ , where $e, d$ know $c$ by the same name.	1 if $\mathcal{A}(d, e) = \mathcal{A}(c, d) = 1$	0	0
$P_5 = (e, c, d)$ , where $e, d$ know $c$ by the same name, $c, d$ know $e$ by the same name.	1 if $\mathcal{A}(d, e) = \mathcal{A}(c, d) = 1$ and $\mathcal{A}(e, c) = \mathcal{A}(c, e) = 1$	0	1 if $\mathcal{A}(d, e) = \mathcal{A}(c, d) = 1$ and $\mathcal{A}(e, c) = \mathcal{A}(c, e) = 1$

attributes  $\mathcal{A}$ , where  $\mathcal{A}(c, e) = 1$  iff entity  $c$  can authenticate the messages sent by entity  $e$ ; and (5) Table II ( $T$ ) that determines the security of individual primitive.

The GRAPH\_COLORING algorithm outputs a three-color (green, red, gray) graph. The nodes are the entities in the authorization. A green edge  $(c, d)$  where  $c, d \in R$  indicates that  $c$  can correctly figure out the name of  $d$ ; A red edge  $(c, d)$  implies that  $c$  may *incorrectly* match the name of  $d$ —an alarm of name matching attacks; A gray edge  $(c, d)$  means that  $c$  may not be able to find out the name of  $d$ .

Here we briefly illustrate how the algorithm work: GRAPH\_COLORING algorithm checks the sequence of primitives one by one (Line 8–16). If a primitive is simply in one of the five types  $(e, c, d)$ , the color of the edge  $(c, d)$  is determined by the function CHECK\_GREEN (Line 9–13). If  $c$  continues executing the protocol when the name outputted in the primitive is the same with the name  $c$  has known (denoted as  $(e, c, d)$ ), the edge  $(e, d)$  is colored according to Line 14–16.

The correctness of the algorithm is justified by PCL too. Due to space limit, please refer to our full paper [2] for details.

---

**Algorithm 1** GRAPH\_COLORING
**Input:**

- 1: Set of entities:  $R = \{au_1, \dots, au_i, az_1, \dots, az_j, ef_1, \dots, ef_k\}$ ;
- 2: Set of entities that may be malicious  $R_m \subset R$ ;
- 3: Sequence of primitives:  $\mathcal{P} = p_1 \dots p_t$  where each  $p_i = (e, c, d)$  has  $e, c, d \in R$ ;
- 4: Authentication attributes  $\mathcal{A}$
- 5: Security of primitive  $T: \mathcal{P} \times \mathcal{A} \times R_m \rightarrow \{0, 1\}$  (Table II)

**Output:** Colored graph  $M_{\mathcal{P}}$ 

- 6: Color all edges gray in the complete graph  $M_{\mathcal{P}}$
  - 7: Color all edges directed from the initiator to other nodes green
  - 8: **for**  $i = 1$  to  $t$  **do**
  - 9:   **if**  $p_i = (e, c, d)$  **then**
  - 10:     **if** CHECK\_GREEN( $p_i, \mathcal{A}, R_m$ ) **then**
  - 11:       Color the edge  $(c, d)$  green
  - 12:     **else**
  - 13:       Color the edge  $(c, d)$  red
  - 14:     **else**                    $\triangleright p_i = (e, c, d)$  is for verification
  - 15:       **if**  $(e, d)$  is red &&  $(c, d)$  is green &&  $T(p_i, \mathcal{A}, R_m)$  &&  $c \in R - R_m$  **then**
  - 16:       Color the edge  $(e, d)$  green
  - 17: **return**  $M_{\mathcal{P}}$ ;
- 

---

- 1: **function** CHECK\_GREEN( $(e, c, d), \mathcal{A}, R_m$ )
- 2:   **if**  $T((e, c, d), \mathcal{A}, R_m) = 0$  **then**
- 3:     **return** false
- 4:   **if**  $d = e$  &&  $(e, c)$  is green **then return** true
- 5:   **else if**  $d \neq e$  &&  $(e, d), (e, c)$  are green **then return** true
- 6:   **else return** false
- 7: **end function**

---

#### D. Vulnerability identification

A red or grey edge in the colored graph rings an alarm of name matching attacks. To detect the vulnerability and construct possible attacks, one should examine the step-by-step coloring in Algorithm 1.

We summarize common vulnerabilities as follows.

- 1) **Lack of authentication:** For example, in the Single Sign On service provided by Gigya, 13% of studied websites do not authenticate the messages sent by Gigya [23]. Due to the lack of authentication, the edge from Gigya to the authorizer is colored red. Name matching attacks identified in [23] can thus be launched.
- 2) **Misused primitives:** For example, PayPal uses primitive  $P_4$  instead of  $P_5$ . However,  $P_4 = (e, c, d)$  is insecure when entity  $d$  is malicious. This causes a name matching attack (Section IV-C).
- 3) **Missing primitives:** For example, in Alipay protocol a primitive  $P_5$  is missing, which produces two red edges and thus a name matching attack (Section IV-A).

The detecting scheme may have false positives. It is possible a vulnerability is found, but an adversary cannot exploit it due to application constrains. In this case, manual investigation is needed to rule out false alarms.

#### IV. CASE STUDY

We apply the detecting scheme to real world multi-party applications, including Alipay PeerPay, Amazon FPS Marketplace and PayPal Express Checkout. New vulnerabilities and name matching attacks are found. Remedies are proposed accordingly.

Due to space limit, we will introduce the case study on Alipay PeerPay in details and briefly report our findings in Amazon FPS Marketplace and PayPal Express Checkout.

### A. Alipay PeerPay (four parties)

In this section, we apply the detecting scheme to Alipay PeerPay and find a vulnerability susceptible to a name matching attack.

Alipay is a popular PayPal-like payment service provider in Asia, which has 300 millions of registered users. The PeerPay service provided by Alipay enables one to shop online and let someone else to pay the bill.

For example, with PeerPay service Alice can order a \$500 iPad on Yihaodian (an online shopping website owned by Walmart with URL `yihaodian.com`) and let her husband Bob pay for her. In this example, a four-party authorization is needed: Alice and Bob together authorize Yihaodian to withdraw \$500 from Bob's Alipay account. Here Alice and Bob are two authorizers; Yihaodian is the authorizee, and Alipay is the enforcer.

**1) Normal workflow of Alipay PeerPay:** In the above example, Alice first places an order of iPad at `yihaodian.com` and selects the payment method as "alipay". Once Alice clicks on "checkout", she is redirected to `alipay.com` where Alice logs in her account (say with username `aaa@gmail.com`). Upon a successful login, Alipay asks Alice to verify the request: "paying `yihaodian.com` \$500 for an iPad." If the request is correct, Alice specifies her husband Bob to pay the bill by providing his Alipay username say `bbb@gmail.com` to Alipay. Alipay will thereby send an email to `bbb@gmail.com` (Bob) as a notification. Once Bob logs in `alipay.com`, he will review the request "`aaa@gmail.com` is requesting you to pay \$500 to `yihaodian.com`. Do you agree?" If Bob confirms, Alipay will notify Yihaodian of the successful authorization.

**2) The detected name matching attack:** We applied the detecting scheme to the Alipay PeerPay protocol and identified a new name matching attack, with which an adversary can shop online for free. Note that this attack is so easy to launch, that the adversary does not need to have a priori knowledge on computer science. The attack scenario is as follows.

**Step 1: An adversary (say Eve) posts an advertisement:** Eve posts a malicious advertisement on her Weibo page (a Twitter-like social network), claiming that she can order an iPad for buyers at `yihaodian.com` with %5 cash back. Suppose Eve's Yihaodian username is `eee@gmail.com`.

**Step 2: Alice places an order:** If Alice is attracted by the advertisement and places an order for an iPad at Eve's Weibo page, she will be redirected to `alipay.com`, where Alice logs in with her Alipay username say `aaa@gmail.com`.

**Step 3: Alice verifies the request and specifies Bob to pay the bill:** At `alipay.com`, Alice is asked by Alipay to verify the request "paying \$500 to `yihaodian.com` for an iPad"—this is exactly what Alice wants to do. Note that in the request, the money is paid to Yihaodian instead of Eve. Alice then specifies Bob to pay the bill by providing Bob's Alipay username say `bbb@gmail.com`.

**Step 4: Bob verifies the request and grants the authorization:** Alipay sends an email to `bbb@gmail.com`, indicating that a request is waiting for approval. Bob then logs in to his Alipay account where he verifies the request "`aaa@gmail.com` is requesting you to pay \$500 to `yihaodian.com`. Do you

agree?" As this is what Bob wants to do and he trusts Alipay, Bob will grant the authorization.

**Step 5: Eve shops for free:** Once Bob grants the authorization, Alipay allows Yihaodian to withdraw \$500. However Alipay (as well as Alice and Bob) and Yihaodian do not match the names of the first authorizer correctly—Alipay thinks that the first authorizer is `aaa@gmail.com` (Alice), while Yihaodian thinks that the first authorizer is `eee@gmail.com` (Eve). As a result, Alipay will charge the bill to Bob who is specified by Alice, while Yihaodian will ship the iPad to Eve.

**Remark 1:** This attack is different from the traditional phishing attack, because Eve does not masquerade as `yihaodian.com`.

**Remark 2:** The reader may argue that this is not an attack, because Alice and Bob can ask for a refund from Alipay. However, because the refund is charged on Yihaodian, Yihaodian now becomes the victim—it follows the Alipay PeerPay protocol exactly, but suffers the financial loss.

**Remark 2:** The reader may argue that, as Alice places an order at Eve, she deserves the attack. It is not true in this scenario, because Alice does not authorize the adversary Eve to withdraw the money from Alipay. Instead, she authorizes Yihaodian to withdraw the money. However, Yihaodian retrieves the money from her husband Bob, but ships the iPad to Eve.

**3) Techniques to launch the attack:** Eve can easily take four steps to launch the name matching attack without a priori knowledge on computer science. First of all, Eve logs in to her Yihaodian account using her web browser and places an order for an iPad. In the second step, Eve turns on her firewall (e.g., Little Snitch) and blocks any request that is sent to `alipay.com`. In the third step, Eve chooses "alipay" as the payment method at `yihaodian.com` and clicks on "checkout". Because of the firewall, her browser cannot send any request to `alipay.com`. In the final step, Eve records the blocked request (an URL) to `alipay` (Figure 3a) and posts the URL in her advertisement page at Weibo. When a victim (Alice) clicks on "checkout" at Eve's Weibo page, Alice's browser follows the URL and sends the recorded request to `alipay.com`.

**4) Applying the detecting scheme:** Our detecting scheme identifies the above name matching attack as follows.

**Protocol decomposition:** According Alipay official document [1], there are four entities in the authorization: the first authorizer  $au_1$  (e.g., Alice), the second authorizer  $au_2$  (e.g., Bob), the authorizee (e.g., Yihaodian), and the enforcer (Alipay). Here  $au_1$  and  $au_2$  may be malicious. Note that  $A(au_1, az) = 0$  because the first authorizer does not authenticate the messages sent by the authorizee.

The Alipay PeerPay protocol can thus be decomposed to the sequence of primitives:  $p_1 = (az, ef, au_1)$ ,  $p_2 = (az, ef, ef)$ ,  $p_3 = (az, ef, az)$ ,  $p_4 = (ef, au_1, au_1)$ ,  $p_5 = (ef, au_1, ef)$ ,  $p_6 = (ef, au_1, az)$ ,  $p_7 = (au_1, ef, au_2)$ ,  $p_8 = (ef, au_2, au_1)$ ,  $p_9 = (ef, au_2, au_2)$ ,  $p_{10} = (ef, au_2, ef)$ ,  $p_{11} = (ef, au_2, az)$ ,  $p_{12} = (ef, az, ef)$ ,  $p_{13} = (ef, az, az)$ .

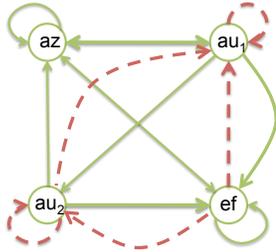
**Graph coloring:** Figure 3b shows the output of Algorithm 1. The red edges indicate that authorizers and the enforcer may match the names of the first authorizer incorrectly. A vulnerability to name matching attacks may exist.

**Vulnerability identification:** According to the process of Algorithm 1, the red edges are caused by  $\mathcal{A}(au_1, az) = 0$  and malicious  $au_1$ . The edges do not turn green when GRAPH\_COLORING algorithm terminates. Therefore, a primitive that can verify the authorizers' names and turn the edges green is missing.

5) **A remedy:** To fix the vulnerability, intuitively, Alipay should verify with Yihaodian that Alice is really the first authorizer, before Alipay shows the request to her. More specifically, the protocol should add the type  $P_5$  primitive ( $ef, az, au_1$ ) right after  $p_3$ , which will turn all red edges to green. We have reported the attack as well as the remedy to Alipay.

Time	Dura...	Total...	Size	Method	Status	Content...	URL
7:51:54.890	1049...	1049...	0	GET	200	applicati...	http://tracker.yihaodian.com/tracker/info.do?1
7:51:54.892	714 ms	714 ms	0	GET	302	applicati...	http://my.1mall.com/gateway/select_gateway.c
7:51:59.112	1055...	1055...	-1	GET	200	text/html	http://netpay.yihaodian.com/online-payment/c
7:52:00.206	412 ms	412 ms	62	GET	503	applicati...	http://netpay.yihaodian.com/favicon.ico
7:52:11.843	0 ms	0 ms	-1	GET	Cancelled	unknown	https://alipay.com/gateway/dop_input_ol...

(a) Eve recording the blocked request to alipay.com



(b) The colored graph

Fig. 3: Alipay PeerPay

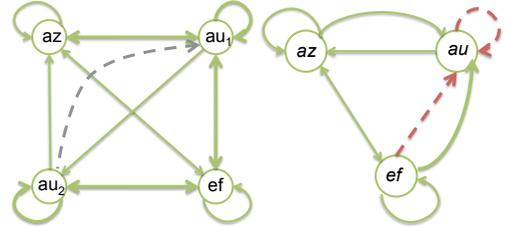
### B. Amazon FPS Marketplace (four parties)

Amazon FPS Marketplace service enables one to setup his own marketplace (say payloadz.com) where buyers and sellers can trade products (e.g., files, eBooks, music) and pay through Amazon. This service requires a four-party authorization: a seller (the first authorizer  $au_1$ ) and a buyer (the second authorizer  $au_2$ ) together authorize Payloadz (authorizee  $az$ ) to transfer a certain amount of money from the buyer's Amazon account to the seller's Amazon account. (Here Amazon is the enforcer  $ef$ .) Both authorizers may be malicious.

*The detected name matching attack (minor):* We found that an adversary can pretend to be a well-reputed seller at *payloadz.com* and sell fair quality products. Suppose Alice is attracted by the adversary's advertisement and places an order for a \$50 file from the adversary who claims he is the well-reputed seller say *bbb* at *payloadz.com*. When Alice is directed to Amazon and logs in there, Amazon will show Alice a request, saying "pay *bbb* via *payloadz.com*" with "total amount: \$50". However, *bbb* at *amazon.com* and the well-reputed seller *bbb* at *payloadz.com* may be two different

entities. As a result, Alice actually pays the money to the adversary and buys the file with poor quality from him.

This name matching attack exploits the vulnerability that the edge from the second authorizer (e.g., Alice) to the first authorizer (e.g., the seller) is grey (Figure 4a), which indicates that Alice cannot know which seller at *payloadz.com* she is paying the money to. This can be easily fixed by the first authorizer verifying the name of the second authorizer at the authorizee.



(a) Amazon FPS Market-(b) PayPal Express Checkout place

Fig. 4: Colored graphs

### C. PayPal Express Checkout (three Parties)

The detecting scheme can also be used in three-party authorization protocols, such as PayPal Express Checkout.

1) *The detected name matching attack:* We are able to identify a new name matching attack, which enables an adversary to shop on *target.com* for free. The attack is similar to the one against Alipay PeerPay.

The adversary first creates an advertisement at Facebook or Twitter, promoting a special offer to get cash back if one orders Target's merchandise through him. Suppose Alice is attracted by the advertisement and places an order for a \$50 Target e-gift card through the adversary. When she chooses to pay by PayPal, Alice thinks that she is authorizing *target.com* to withdraw \$50 dollars from her PayPal account.

Once Alice logs in her PayPal account, she will be asked to verify the request: paying Target (TARGETCORPO) \$50 dollars for a "Seasons Greetings Snowglobal Gift Card"—exactly what Alice wants. Because the page looks exactly the same as if she places the order by herself, and more importantly because Alice trusts PayPal to protect the order, she is very likely to click on "continue". Afterwards, the adversary can construct a valid request to *target.com* and completes the payment.

As a result, PayPal will charge the bill to Alice while Target will ship the gift card to the adversary. This name matching attack exploits a vulnerability in PayPal Express Checkout protocol, indicated by the red edges in Figure 4b.

2) **A remedy:** The vulnerability is caused by a misuse of three-party primitive. To fix it, the  $P_4$  type of primitive should be replaced by the  $P_5$  type of primitive. For more details, please refer to our full paper [2]. The attack and the remedy have been reported to PayPal.

## V. RELATED WORK

Authorization has been studied for decades. The most traditional ones include discretionary access control (DAC) [16], mandatory access control (MAC) [8], role based access control (RBAC) [22], which are widely used in homogeneous environment. In Internet environment (which is no longer homogeneous), authorization is closely related to Single-Sign-On (SSO) authentication. When a user requests services from a website, he is first authenticated by a third-party identity provider, where the user receives tickets (Kerberos [18]), cookies (Microsoft .Net Passport [10]) or encoded URLs (Liberty Alliance Project [5]). The user sends them to the website, who then provides services accordingly. In these authorizations, name matching is typically not an issue.

Security of three-party authorization in cloud environment has drawn great attention of researchers. Researchers found a session fixation attack [3] and an authorization code swap attack [4] against OAuth [13], whose security was further analyzed formally with Communicating Extended Finite State Machines [14], pi-calculus [7], Alloy [19], and Universal composability Security Framework [9]. Since 2011, researchers has inspected OAuth's application on Single Sign On (SSO) and revealed attacks [25], [23], [7]. Besides OAuth, online payment, another three party application, was also studied [24]. Learned from the field study of three-party authorization, researchers built systems to automatically extracting specifications from implementations (AUTHSCAN [6]), to offers authorizes the security protection to vulnerable web API integrations (InteGuard [27]), and to uncovering implicit assumptions of SDK provided by enforcers [26]. To the best of our knowledge, security of four-or-more party authorization has not been systematically studied yet.

## VI. CONCLUSION

In this paper, we proposed a scheme to detect the vulnerability of multi-party authorization protocols to name matching attacks. By applying the scheme, we found new name matching attacks in high-profile three- and four-party authorization protocols such as Alipay PeerPay, Amazon FPS Marketplace, and PayPal Express Checkout.

In our future work, we will investigate name matching attacks when multiple colluded adversaries are present. We are also seeking a secure protocol that can solve the multi-party authorization problem.

## REFERENCES

- [1] Alipay Peerpay. <http://home.alipay.com/bank/paymentPayOther.htm>.
- [2] Full paper. <https://www.dropbox.com/s/hf49zxs3gzknkcy/full.pdf>.
- [3] OAuth Security Advisory: 2009.1. <http://oauth.net/advisories/2009-1/>.
- [4] [OAUTH-WG] Auth Code Swap Attack. <http://www.ietf.org/mail-archive/web/oauth/current/msg07233.html>.
- [5] L. Alliance. Liberty alliance project. *Web page at* <http://www.projectliberty.org>, 2002.
- [6] G. Bai, J. Lei, et al. AUTHSCAN: Automatic Extraction of Web Authentication Protocols from Implementations. *Proceedings of 20th Annual Network & Distributed System Security Symposium*, 2013.
- [7] C. Bansal et al. Discovering Concrete Attacks on Website Authorization by Formal Analysis. In *Computer Security Foundations Symposium (CSF)*, pages 247–262. IEEE, 2012.
- [8] D. E. Bell and L. J. LaPadula. Secure computer systems: Mathematical foundations. Technical report, DTIC Document, 1973.
- [9] S. Chari, C. Jutla, and A. Roy. Universally Composable Security Analysis of OAuth v2.0. Technical report, 0. Cryptology ePrint Archive, Report 2011/526. <http://eprint.iacr.org>, 2011.
- [10] M. Corporations. Microsoft. net passport review guide. Technical report, Technical report, Available at [www.microsoft.com](http://www.microsoft.com), 2003.
- [11] A. Datta et al. Protocol Composition Logic (PCL). *Electronic Notes in Theoretical Computer Science*, 172:311–358, 2007.
- [12] E. Hammer-Lahav. RFC 5849: The OAuth 1.0 protocol. *Internet Engineering Task Force (IETF)*, 2010.
- [13] D. Hardt. Rfc 6749: The oauth 2.0 authorization framework. *Internet Engineering Task Force (IETF)*, 2012.
- [14] Y. Hsu and D. Lee. Authentication and Authorization Protocol Security Property Analysis with Trace Inclusion Transformation and Online Minimization. In *IEEE International Conference on Network Protocols (ICNP)*, pages 164–173. IEEE, 2010.
- [15] L. Lamport et al. The Byzantine Generals Problem. *ACM Transactions on Programming Languages and Systems (TOPLAS)*, 4(3):382–401, 1982.
- [16] B. W. Lampson. Protection. *ACM SIGOPS Operating Systems Review*, 8(1):18–24, 1974.
- [17] Y. Lindell. Lower Bounds for Concurrent Self Composition. In *Theory of Cryptography*, pages 203–222. Springer, 2004.
- [18] J. Lopez et al. Authentication and authorization infrastructures (AAIs): a comparative survey. *Computers & Security*, 23(7):578–590, 2004.
- [19] S. Pai, Y. Sharma, S. Kumar, R. Pai, and S. Singh. Formal verification of oauth 2.0 using alloy framework. In *Proc. of Communication Systems and Network Technologies (CSNT)*, pages 655–659. IEEE, 2011.
- [20] L. Pearlman et al. A Community Authorization Service for Group Collaboration. In *Proc. of Policies for Distributed Systems and Networks*, pages 50–59. IEEE, 2002.
- [21] A. Roy et al. Secrecy Analysis in Protocol Composition Logic. In *Advances in Computer Science-ASIAN 2006. Secure Software and Related Issues*, pages 197–213. Springer, 2007.
- [22] C. E. Sandhu, R.S. et al. Role-Based Access Control Models. *Computer*, 29(2):38–47, 1996.
- [23] S.-T. Sun and K. Beznosov. The Devil is in the (Implementation) Details: an Empirical Analysis of OAuth SSO Systems. In *Proceedings of ACM conference on Computer and Communications Security (CCS)*, pages 378–390. ACM, 2012.
- [24] R. Wang et al. How to Shop for Free Online—Security Analysis of Cashier-as-a-Service based Web Stores. In *Security and Privacy (SP), IEEE Symposium on*, pages 465–480, 2011.
- [25] R. Wang et al. Signing Me onto Your Accounts through Facebook and Google: a Traffic-Guided Security Study of Commercially Deployed Single-Sign-on Web Services. In *Security and Privacy (SP), IEEE Symposium on*, pages 365–379, 2012.
- [26] R. Wang, Y. Zhou, et al. Explicating SDKs: Uncovering Assumptions Underlying Secure Authentication and Authorization. In *Proceedings of the USENIX Security Symposium*. USENIX, 2013.
- [27] L. Xing, Y. Chen, et al. InteGuard: Toward Automatic Protection of Third-Party Web Service Integrations. In *Proceedings of 20th Annual Network & Distributed System Security Symposium*, 2013.
- [28] E. Yuan and J. Tong. Attributed based Access Control (ABAC) for Web Services. In *Web Services, Proceedings. IEEE International Conference on*. IEEE, 2005.

# A Dynamic Approach to Risk Calculation for the RAdAC Model

Roberto Marinho, Carla Merkle Westphall and Gustavo Roecker Schmitt

Informatics and Statistics Department  
Federal University of Santa Catarina  
Florianopolis, Brazil  
{marinho, carlamw}@inf.ufsc.br

**Abstract**— This paper aims to provide a model for dynamic risk assessment for the RAdAC model supported by the use of ontologies to perform the calculation of risk. From the mapping of the different variables involved in the calculation of risk into axioms of an ontology, it is possible to dynamically infer the risk of access to specific data based on the available risk factors and their weights.

**Keywords**—RAdAC; Access Control; Ontology; Risk Evaluation; Cloud Computing.

## I. INTRODUCTION

Normally, traditional access control models are static, i.e., their rules do not change over user access. Thus, traditional models do not have enough flexibility to support existing dynamic environments in pervasive and ubiquitous computation, computational grids and cloud computing [1], [2]. Currently, there are some dynamic access control models in spite of the lack of implementation in the present literature. In the RAdAC (*Risk-Adaptive Access Control*) model, access control is adaptive and based on the calculation of the risk of access to the system that the user performs [3], [4]. This calculation is made in real time and should guarantee the system security.

According to [5], an ontology corresponds to an specification of a conceptualization, describing its concepts and interrelationships. In this context, the main objective of the use of ontologies is to enable knowledge to be shared and reused, as well as allow new information to be collected from the inferences about the information available [6].

The use of ontologies in the development and adaptation of access control models in order to provide flexibility and dynamism in decision making is already exploited in several works [7], [8]. However, the use of ontologies in the context of dynamic risk assessment, anchored in RAdAC access control model consists of a topic not yet explored in the literature.

The present work aims to develop a model of access control based on RadAC, supported by the use of ontology to infer the risk of access to a particular object in varied scenarios and situations. The remainder of the paper is organized in four sections: section 2 presents a brief description of RAdAC access control model, addressing its workflow and decision-making method. Section 3 lists some related work. In Section 4, a model for the dynamic calculation of risk based on

ontology is presented. Finally, section 5 describes the completion of the work and future work.

## II. RAdAC

Risk is the potential damage that can arise in a current or future process and is generally represented by the probability of an unwanted event and its resulting impact [9].

The access control models based on risk conduct a risk analysis in the request to make an access decision. This risk analysis can be qualitative or quantitative. In the qualitative methods different scales of risk, such as high, average and low ranges are used, and usually the valuation is done by an expert opinion. Yet, in quantitative methods there is a way to assign a numeric value that represents the risk of a request for access.

In quantitative methods, the risk of an event is usually represented by the calculation expressed in formula (1).

$$R = P \times I \quad (1)$$

In (1), P is the probability of occurrence of the event and I is the impact of the event occurrence. In situations where there is a history of access and where the impact can be quantified, especially in monetary values, the calculation becomes easier, but there are situations where these items are not easily obtained or it is desired to also consider other features.

In the RAdAC model access control is adaptive and is based on the calculation of the risk of the user access [3], [4]. This calculation is done in real time and must ensure the security of the system. This model originated in the National Security Agency of the United States [3]. The paper proposes the use of access control based on risk for a more effective information sharing in military environments, but does not detail forms of risk evaluation.

The Figure 1 represents the ideal flow of risk assessment in RadAC model and shows the risk assessment is the first step of the process. The security risk represents the damage to the system if the information is released. If the calculated risk is acceptable according to the specified access control policy then the operational need is evaluated. The operational need is the necessity for an entity to have access to certain information to complete a mission and is represented by a value. Thus, access is released if the operational need is greater than the risk of security.

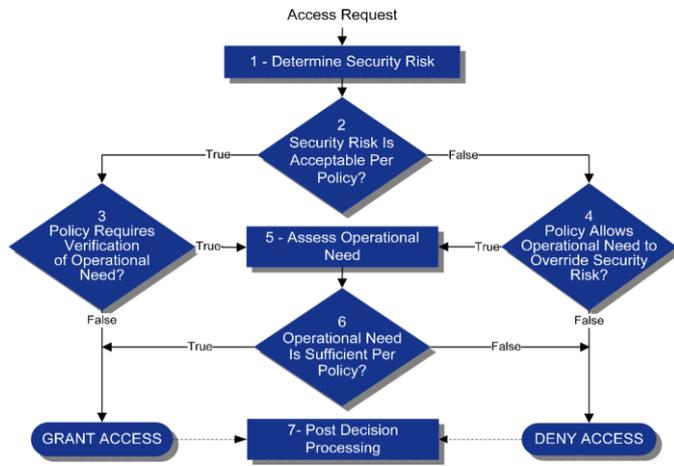


Figure 1- RAdAC Decision Model

### III. RELATED WORK

Dynamic access control systems evaluate each access request in real-time, in a dynamic way, analyzing, beyond pre-defined policies in the system, context information such as the risk of the operation, the user need to perform the operation, the benefit of the operation for the system and the user, among others.

The work [10] describes the current difficulties in federated identity management in the cloud and shows how to use the proposed risk assessment as a method to enable the construction of dynamic identity federations in the cloud. Although quite detailed, the work does not present numerical values to the metrics or how these values should be obtained, presenting only a semantic description of each metric.

Paper [1] presents a dynamic access control model based on risk for cloud computing environments. The work specifies risk policies in the form of XML files (eXtensible Markup Language) using different metrics and risk quantification methods that do not necessarily have to be pre-defined.

In [11] is proposed to consider the calculation of risk components as characteristics of people, IT, objects, environments factors, situational factors and heuristics.

The work [12] presents a quantification method for the RAdAC model based on the opinion of experts. A list of risk factors is compiled and a value assigned to each factor. Subsequently, weights are assigned to each value and the end result is the combination of all factors with their weights. It is one of the works of the literature that presents the more concrete risk calculation form.

[13] describes some practical challenges to implement the RAdAC model, including the real-time calculation of the safety risk for each access decision, the determination of operational risk, the quantification of the level of confidence; the use of heuristics to achieve access decisions and the right to revoke access at any time.

Some papers in the literature consider the properties of confidentiality, integrity and availability for calculating the risk associated with the action that will be taken on appeal [14],

[15]. Other works use fuzzy logic to calculate risk values and learning techniques to perform the access decisions.

[8], in turn, has developed a model in which semantic contexts are employed and represented using ontology to dynamically determine the appropriate allocation of roles in an incident management system.

Considering the studies reviewed in the literature, it was observed that the calculation of risk may consider the context information, the history of previous actions of the subject and also the properties of confidentiality, integrity and availability associated with the actions that will be made on the resources. However, in dynamic contexts, composed of different metrics or variables, perform the calculation of risk is still a challenge. The use of ontologies is an alternative to achieve dynamic contexts due to its adaptability and flexibility.

### IV. A DYNAMIC MODEL FOR CALCULATING RISK

In this paper, the proposed dynamic risk assessment model uses different approaches in the composition of the total security risk.

The following factors compose the dynamic model for calculating the risk:

- Context: characteristics of the subjects requesting access, characteristics of IT components, characteristics of the objects or from the required information, environmental factors, situational factors and heuristics;
- Security Characteristics of the Actions: the characteristics of confidentiality, integrity and availability of the actions on the resource;
- History of the Subject: the subject has previous actions in the system that are stored in the form of a history to reward good use or penalize the misuse of the system, according to the prior behavior of the subject;

The context risks come from the risk factors and their respective weights determined in [12] (Table 1), in which a quantification of the weights of the factors was performed and validated by experts.

For the calculation of the risks involving confidentiality, integrity and availability of the actions on the resources, metrics developed in [15] are used. In this calculation, the risk is calculated based on the impact that a particular approach can lead, divided into: low impact (1-5), moderate impact (6-10) and high impact (11-15).

[15] quantifies the loss of confidentiality, integrity and availability in three levels of impact based on the effect and potential damage that the loss can lead. In the model presented here, the scale of 1 to 15 corresponds to the risk of confidentiality, integrity and availability, which can be adapted to create more impact levels.

Completing the three pillars that compose the Total Security risk, there is the risk based on the history of the subject, comprising a score corresponding to the previous actions executed by the subject.

Actions considered negative increment historical risk of the subject, while positive actions decrement it.

TABLE I.

Risk Factor	Weight
<b>Characteristics of Requester</b>	16.66667
Role	2.777778
Rank	2.777778
Clearance Level	2.777778
Access Level	2.777778
Previous Violations	2.777778
Education Level	2.777778
<b>Characteristics of IT Components</b>	16.66667
Machine Type	2.380952
Application	2.380952
Connection Type	2.380952
Authentication Type	2.380952
Network	2.380952
QoP/Encryption Level	2.380952
Distance from requester to source	2.380952
<b>Heuristics</b>	16.66667
Risk Knowledge	8.333333
Trust Level	8.333333
<b>Situational Factors</b>	16.66667
Specific Mission Role	3.333333
Time Sensitivity of Information	3.333333
Transaction Type	3.333333
Auditable or Non-auditable	3.333333
Audience Size	3.333333
<b>Environmental Factors</b>	16.66667
Current Location	8.333333
Operational Environment Threat Level	8.333333
<b>Characteristics of Information Requested</b>	16.66667
Classification Level	3.333333
Encryption Level	3.333333
Network Classification Level	3.333333
Permission Level	3.333333
Perishable/ Non-Perishable	3.333333

In this context, the risk based on the history of the subject is represented on a scale of 0 to 10. Therefore, after each access attempt, the history of each subject is updated according to the result of the action performed.

The function that calculates the total risk is defined by the formula (2). The weights of each of the factors are represented by p1, p2 and p3, respectively. Possible values for the weights could be 0.5, 0.3 and 0.2, for example, or any other amount considered adequate for the considered system.

$$\text{Total Risk} = p1 * \text{Context Risk} + p2 * \text{Risk Considering Confidentiality, Integrity and Availability} + p3 * \text{Risk Based on the History of the Subject} \quad (2)$$

The formula in (2) emerged after a detailed study of the related work and describes the calculation of risk concisely considering a wide range of the factors involved.

The infrastructure of the developed model is shown in Figure 2. In it, the user requests the application access to a specific data, the application in turn builds an xml file defining the context, which is composed of a set of attributes used to calculate the risk of the user access (total risk). The tag called *context* is the content of that request. In the scenario of Figure 2 are considered the attributes of location of access (accessLocation), user role (UserRole), machine type (MachineType) and the protocol that is being used in access (ApplicationProtocol) to set the context of access to be considered in the calculation of risk.

The risk factors involved in calculating risks of confidentiality, integrity and availability are mapped in a separate xml file named C.I.A. Risk and responsible for providing risk considering confidentiality, integrity and availability. The identities of the subjects and their respective history of access are stored in another file named User History. This file is modified after each access request and provide risk considering the history of the subject.

The ontology used to infer context risk factors from attributes sent by the tag *context* is represented by a file named Context Risk Ontology.

Finally, the risk evaluator server, called Risk Evaluator, retains the information contained in XML files and performs the necessary queries to extract the information contained in the ontology, performing calculations on the risk factors available based on the formula (2).

However, as the availability and validation of all these risk factors in certain environments is complex, and sometimes inapplicable, the model does not require that all factors are provided to perform the calculation of risk. Taking into account only the factors available, the weights of each factor can be redistributed according to the existence of other factors, thus creating a dynamic context.

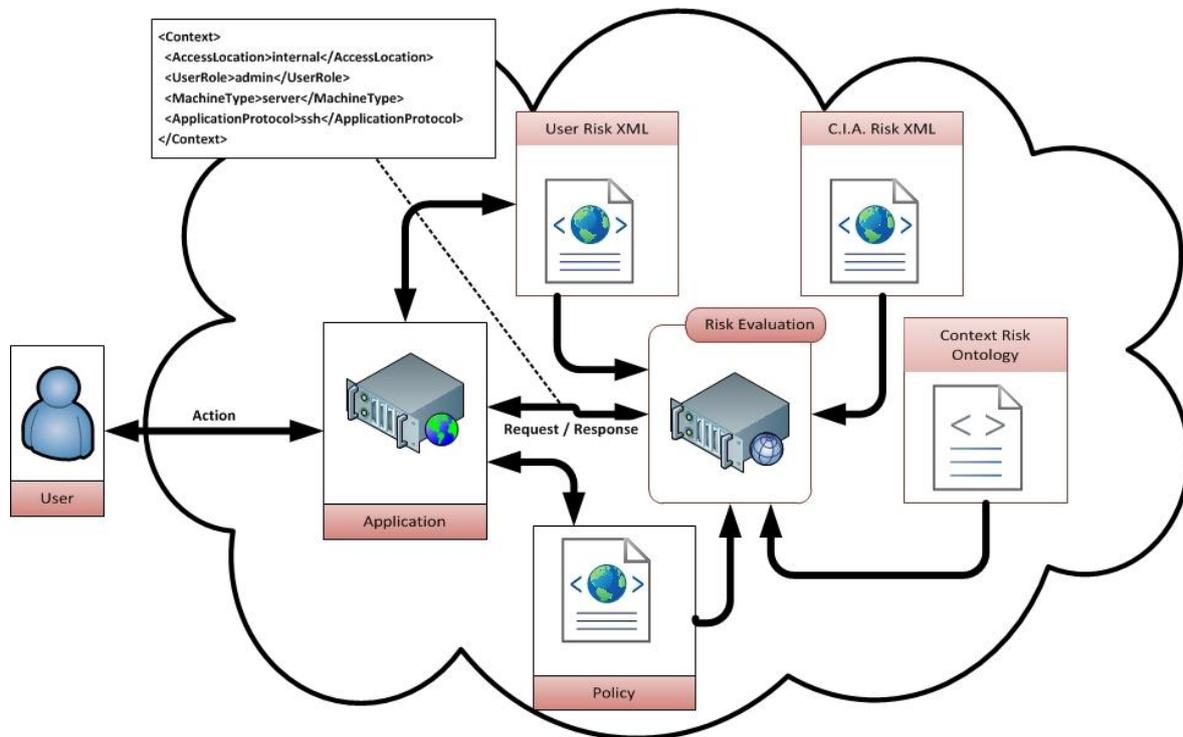


Figure 2 - Model Infrastructure

## V. EXPERIMENTAL RESULTS

A prototype access control system in a cloud environment was created to evaluate access control policies dynamically, using the RAdAC access control model in a cloud computing environment.

An ontology was created with the help of Protégé [16], mapping all risk factors into classes and subclasses, assigning weights defined by [12] for each class that represents a risk context factors.

The main tools used were the XML language for defining data and requests, the Amazon EC2 cloud to host a client application and a server application, the programming language Java EE, the Tomcat application server and web services to communicate between different parts of the application. The APIs *protege-owl* and *Jena2* ontology were used to implement and consult the ontology for the dynamic risk assessment.

The developed application receives the information about the action to be performed by a string. On the server side, considering the context information, the ontology encounters the available risk factors in accordance with the contents of the sent string and calculates the context risk based on Table 1.

Figure 4 shows part of the ontology developed. In it, the risk factors were divided into subclasses of the risk categories involved in the calculation. Instances of subclasses refer to the attributes of the risks sent via XML access request, identified in Figure 2 by the context tag.

Queries on the ontology are performed at runtime using the SPARQL language. The query presented in Figure 3 illustrates

how existing risk are obtained based on the context attributes sent in the request. In the example shown in Figure 2, the *MachineType* attribute with the value *server* is searched on the ontology. The result comprises one or more factors related to *MachineType* attribute that has the value equal to *server*. From the identification of the factors involved, their respective weights are obtained based on Table 1.

```

PREFIX risk:
http://www.semanticweb.org/marinho/ontologies/2014/3/risk-ontology#
[...]
select * where { risk:server rdf:type ?risk_factor
                ?risk_factor rdfs:subClassOf ?risk_group }

```

Figure 3 - Query Example

On a trial basis, queries were carried out on the ontology of Figure 2 after the execution of a reasoner about the same.

In total, the ontology shown in Figure 4 has 6 classes, 27 subclasses and 81 instances, following the model proposed by [12] in the risk assessment context.

```

<owl:Class
rdf:about="http://www.semanticweb.org/marinho/ontologies/2014/3/risk-ontology#Characteristics_of_IT_Components"/>
</owl:Class>
<owl:Class
rdf:about="http://www.semanticweb.org/marinho/ontologies/2014/3/risk-ontology#Machine_Type">
<rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/marinho/ontologies/2014/3/risk-ontology#Characteristics_of_IT_Components"/>
</owl:Class>

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/marinho/ontologies/2014/3/risk-
ontology#server">
<rdf:type
rdf:resource="http://www.semanticweb.org/marinho/ontologies/2014/3/risk-
ontology#Machine_Type"/>
</owl:NamedIndividual>

```

Figure 4- Part of the Context Risk Ontology

According to the number of risk factors for calculation of the Total Measure of Security Risk (in this paper, context risk), the value of the weight of each factor is adjusted in proportion to the values defined in [12].

Due to the small size of the ontology, the execution time of queries is extremely small, being less than 100 ms. This value is directly related to the size of the ontology and the fact that the reasoner has not been executed on-the-fly.

Summing the query time to the ontology and the seek times in the XML files responsible for providing the risk of confidentiality, integrity and availability, and the risk based on the history of the subject, the final execution time remained below one second.

## VI. CONCLUSION

With the creation of new models of access control as RAdAC, the need for new approaches to treat access to dynamic and different environments emerged.

The possibility of establishing metrics for calculating the risk of access through ontologies in the RAdAC model, allows an expansion of the use of this access control model spectrum, since, in different environments, the risk factors are not present in the same form.

The developed implementation in an environment of cloud computing provides dynamic access control in the form of a prototype. Several other tests are necessary to make inferences in real time with a larger number of data.

The scientific contributions of this work that can be cited are: (a) a quantitative method for calculating risk considering the aspects of context, security features (confidentiality, integrity and availability), and history of the subject; (b) development of ontologies to provide control of flexible and dynamic access based RAdAC.

The dynamic and flexible aspect of the proposed model exists because even in environments where the risk factors and their weights are known, the calculation of Total Risk does not strictly depend that all factors are known at the time of access. As a benefit, a value measured informing the risk will always be obtained, leaving to the model developers to develop policies for access control by determining which risk factors are required for the validation of the calculation performed.

In the related work described in Section 3, the [10] does not provide numerical values for the metrics and how these values should be obtained, only a semantic description of each metric. In [1], risk policies in the form of XML files use different risk metrics but the administrator or user must explain the methods of quantification. The work [12] presents the idea of calculation used as the basis of the formula proposed in this

article. In [13], calculating real-time security risk for each access decision is cited as a challenge for the implementation of the model RAdAC. Properties of confidentiality, integrity and availability were also considered in the work [14] and [15]. [8], on the other hand, only uses ontologies to assign roles dynamically. However, none of the related work uses ontologies to make flexible risk calculation in dynamic contexts, composed of different metrics or variables.

Among the future work that can be cited is the implementation of an access control software targeted to the Web environment for testing and performance measurements adapting and refining the model proposed in this paper.

## REFERENCES

- [1] SANTOS, D. R.; WESTPHALL, C.M.; WESTPHALL, C.B. Risk-based Dynamic Access Control for a Highly Scalable Cloud Federation". In: SECURWARE 2013 - The Seventh International Conference on Emerging Security Information, Systems and Technologies, 7th edition, Barcelona. Proceedings IARIA: XPS Press, 2013. pp. 8 - 13.
- [2] KARP, Alan H.; HAURY, Harry; DAVIS, Michael H. From ABAC to ZBAC: the evolution of access control models. Hewlett-Packard Development Company, LP, v. 21, 2009.
- [3] JASON Program Office. Horizontal Integration: Broader Access Models for Realizing Information Dominance. [S.l.], 12 2004. [Online]. Available: <http://www.fas.org/irp/agency/dod/jason/classpol.pdf>.
- [4] MCGRAW, Robert W. Risk-Adaptable Access Control (RAdAC). In: NIST Privilege (Access) management Workshop, [s.e.], USA. NIST: Setembro 2009. [Online]. Available: [http://csrc.nist.gov/news\\_events/privilege-management-workshop/radac-Paper0001.pdf](http://csrc.nist.gov/news_events/privilege-management-workshop/radac-Paper0001.pdf).
- [5] GRUBER, T. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199-220, 1993.
- [6] BREITMAN, K. K.; LEITE, J. C. S. P.. Ontologias-Como e porque criá-las. Anais do Simpósio Brasileiro de Computação, XXIII JAI - Jornada de Atualização em Informática, 2004.
- [7] FININ, T.; JOSHI, A.; KAGAL, L.; NIU, J.; SANDHU, R.; WINSBOROUGH, W.; THURASINGHAM, B. ROWLBAC: representing role based access control in OWL. In: 13th ACM symposium on Access control models and technologies, 13th edition, Estados Unidos. Proceedings ACM: ACM Press, 2008. p 73-82.
- [8] DERSINGH, A.; LISCANO, R.; JOST, A.; FINNISON, J., "Dynamic Role Assignment Using Semantic Contexts," Advanced Information Networking and Applications Workshops, 2009. WAINA '09. International Conference on , vol., no., pp.1049,1054, 26-29 May 2009.
- [9] DIEP, N. N. et al. Contextual risk-based access control. In: Security and Management. [S.l.: s.n.], 2007. p. 406-412.
- [10] ARIAS-CABARCOS, P. et al. A metric-based approach to assess risk for "on cloud" federated identity management. *Journal of Network and Systems Management*, v. 20, n. 4, p. 513-533, 2012.
- [11] CHOUDHARY, R. A policy based architecture for NSA radac model. In: IEEE Information Assurance Workshop - IAW '05, 6th edition, Estados Unidos, Proceedings... IEEE: IEEE Press, 2005. p. 294-301.
- [12] BRITTON, D. W., BROWN, I. A. A Security Risk Measurement for The RAdAC Model. 2007. Master Thesis, Naval Postgraduate School, USA, 2007. [Online]. Available: <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA467180>.
- [13] FARROHA, B.; FARROHA, D. Challenges of operationalizing dynamic system access control: Transitioning from abac to radac. In: 2012 IEEE International Systems Conference (SysCon), [S.e.], Vancouver, Canada, Proceedings... IEEE: IEEE Press, 2012. p. 1-7.
- [14] SHARMA, M. et al. Using risk in access control for cloud-assisted ehealth. In: High Performance Computing and Communication 2012 - IEEE 9th International Conference on Embedded Software and Systems (HPCC-ICESSE), 9th edition, Estados Unidos. Proceedings... IEEE: IEEE Press, 2012. p. 1047-1052

- [15] SARIPALLI, P.; WALTERS, B. QUIRC: A Quantitative Impact and Risk Assessment Framework for Cloud Security. In: 2010 IEEE 3rd International Conference on Cloud Computing (CLOUD), 3 rd edition, USA. Proceedings... IEEE: IEEE Press, 2010. pp. 280-288. doi: 10.1109/CLOUD.2010.22.
- [16] PROTÉGÉ. Protégé-owl api. 2013. [Online]. Available: <http://protege.stanford.edu/plugins/owl/api/>

**SESSION**  
**NETWORK SECURITY + SECURITY**  
**MANAGEMENT**

**Chair(s)**

**Prof. Bon Sy**  
**City Univ. of New York - USA**  
**Dr. Rita Barrios**  
**Univ. of Detroit Mercy - USA**



# Malicious Device Inspection in the HAN Smart Grid

Eric McCary<sup>1</sup>, Yang Xiao<sup>1</sup>

<sup>1</sup>Department of Computer Science, The University of Alabama, Tuscaloosa, AL, US

**Abstract** - *Smart grid is an emerging power infrastructure and software solution that integrates the newest communication and information technology. The supporting infrastructure and networks extend and connect through every avenue of the grid. This includes networks resident in the consumer homes. In this paper, we explore extending the accountability established in the home area network (HAN). We propose several algorithms, which allow for grouping and inspection of the devices in a HAN in order to efficiently discover and pinpoint malicious devices in the HAN. With this, a higher level of fine-grained accountability can be achieved in the smart grid HAN.*

**Keywords:** malicious, accountability, inspection, witness, estimation, smart grid

## 1 Introduction

The status of the smart grid has progressed from an infrastructure where in depth research in security was scarce, to the end of its infancy, where there has been a notable amount of requirements which make the grid much more secure. Much attention has gone toward smart grid [1-3] and methods which secure the grid from obvious threats [4-6, 26].

The smart grid can be described as the current power delivery system with integrated bidirectional communication and real-time analysis on energy generation, transmission, and distribution data in order to create predictive and necessary recommendations for consumers [7]. The National Institute of Technology and Standards (NIST) defines six key areas which make up the grid below [8]: bulk generation domain, transmission domain, distribution domain, operations domain, service provider domain, and customer domain. These areas are expressed as domains which house several major components in the energy arena. Each domain has a unique distributed computing environment, sub-domains, and equipment to suit its mission-specific needs. It is also important to note that the domains of the grid are interconnected with adjacent domains which provide coordinated functionality.

Inaccuracy in estimation and malicious devices is one of a many problems that tend to plague many smart grid installations. Even throughout the networks in the home, it has been well-known to harbor legacy and insecure software on devices which can be actively targeted as entry points into a network. From there, the malicious individual or applications can freely take advantage of exploits on any of the devices

resident on the network, including devices which interface the smart grid networks including advanced metering infrastructure (AMI) which allows for automated measurement and communication of the metering data from the consumer to the utility. With this convention, there must be some form of trust between these two entities. In addition to trust, there should be correctness and truthfulness to the metering data being communicated. This can be provided upon adding sufficient accountability into networks which perform actions on the smart grid.

Since large amounts of data are generated on the consumer side, a more fine-grained process must be put in place to assure accountability as well as a technique to efficiently discover the devices in the HAN which are behaving in a manner that is outside of their expected correct requirements. Also, since the HAN interfaces other networks in the grid and is most influenced by the consumer, who normally has little to no knowledge in hardening their hardware and software for grid purposes, it should be equally regarded in terms of prominence in research and requirements. Within this deficiency resides the lack of a mechanism for inspecting devices in the HAN at a fine-grained level, which will further establish accountability in this domain and is the topic that this study will cover. This paper gives solutions to some of the HAN's needs in the form of algorithms which efficiently inspect the status of the devices connected to the network in order to find which are performing actions that are undesired or incorrect in the network. We propose several algorithms, which allow for grouping and inspection of the devices in a HAN in order to efficiently discover and pinpoint malicious devices therein.

The rest of the paper is organized as follows. Section II provides some background in smart grid and accountability, Section III will give an appropriate model for HAN energy consumption, Section IV will detail the varying consumption device estimation scheme. Section V will give an analysis of the scheme and performance, and finally, the paper is concluded in Section VI.

## 2 Background

Sensing and measurement are constantly occurring in the HAN and reported in the smart grid. This data is then analyzed and tendencies can be created to allow major and catastrophic events to be predicted and avoided based on past measurements. This type of wide situational awareness plays a large role in operational foresight, and when catastrophic events are avoided, price spikes will decrease.

Accountability can be viewed as a complement to the core principals of information security and the component that allows authorized individuals robust tracking and auditing history as well as establishing trust and confidence within the HAN among devices. Current accountability in the smart grid does not extend far beyond a single consumer (household, business, etc.) reporting monthly energy usage to the utility. This is not a sufficient method for grid operation as the lack of accountability in the HAN allows for false reporting before consumers send data to the utility.

Even in the case of data reporting, there is still much room for error and we cannot always expect for the record that the utility manages and what is recorded at the customer's end to be identical. Malicious actions, malfunction, miscalculation in estimation, or calibration may be the cause of such differences. Making the HAN accountable on a more fine-grained level can help alleviate problems such as these and provide us with a means of locating a compromised device which can immediately be disabled or serviced instead of canceling service to the residence until the problem is determined.

### 3 HAN in the Smart Grid

The smart grid HAN can be explained as a grid subsystem which is dedicated to demand-side management through energy efficiency and quality demand-response implementation. It can be further described as a dedicated network of devices which will inevitably become "smart". These devices include, load and control devices, along with some form of software applications which allow the consumer to control these devices normally on a detailed level in the context of energy usage. With the application of smart grid concepts, the home area will be transformed into intelligent nodes in the grid network where much of the energy that it uses is produced locally by renewable generators (likely photovoltaic generation).

The papers [16, 17] give several categories of technology that are normally found in smart homes. Intelligent management devices are the first type of devices found in the HAN. These devices include the control and monitoring capabilities locally required. These tasks require the data created and used inside of the HAN as well as external data generated from the utility of other intelligent nodes connected to the grid networks locally that work in tandem with the HAN. A simple example of how this data can be used is management of energy usage and discovering local producers and consumers. Management controls not only higher level functions, but also controls and manages individual devices. Scheduling and peak usage avoidance are keys in maintaining an optimum demand and supply relationship.

The HAN is the most important smart grid domain for the consumer as this is where they reside and also where the utility implements demand management systems to help regulate the energy usage and production in the grid. [12] provides some demand management applications as listed below: behavioral energy efficiency, technology-enabled dynamic pricing, and deterministic direct load control.

While the significance of the HAN lies in the distribution of electricity and demand response, none of this would be possible without AMI. AMI presents what was the first step in the transformation of homes into what is envisioned in the smart grid. This technology serves as the interface of the smart grid into the customer's domain and has been the focal point of much progress as well as scrutiny.

#### 3.1 Problem Details

Assured accountability allows for an audit trail for each device's actions. With implicit knowledge of the environment, the status of a network or device can be verified. Currently the smart grid landscape requires little accountability, and even less in the HAN (or customer domain). This means that locating malfunctioning or malicious devices in this area is very difficult, and all but impossible for entities outside of these networks to verify this data.

Given the dynamic nature of devices in these types of network, the solution must be scheduled to run at a specific interval which it will be most effective, or triggered by a specific action. This means that the efficiency of the solution will be important.

For power calculation purposes, disaggregation techniques have been well studied. This is insufficient for entities inside of the HAN as observation and troubleshooting at a more fine-grained level is necessary, as well as location of devices affected by malware and physical abnormalities. Solution to this problem is given in the following sections.

### 4 HAN Device Inspection

In this section, a method is proposed to search for devices in the smart grid HAN environment. We propose several algorithms, which allow for grouping and inspection of the devices in a HAN in order to efficiently discover and pinpoint malicious devices.

The application of accountability has been established in the neighborhood area network (NAN) in [9-11]. We will assume that the network has the ability to employ a device witnessing scheme similar to that proposed in these works, and that the HAN model under inspection is composed of many devices. The data from monitoring will be written to tamper evident logs which will help ensure accountability.

In the proposed environment, we will assume that each device has a set of witnesses ( $W_m$ ). We must also assume that any generation and energy storage devices are truthful of their potential capacity and how much energy they are storing when the system is commenced. The objective of the inspection is to efficiently search the active devices which are connected to the network in order to find malicious or malfunctioning devices. In the initial model, the smart meter resident in the house will be responsible for accountability management and maintaining a record of devices and their witnesses. As this will be constantly changing due to mobile devices moving in and out of the network, the management device will need to be aware of this. Smart devices can effortlessly broadcast their presence

on a network and the management device can simply observe the network for traffic from new hosts.

### 4.1 HAN Device Grouping

The first step of the inspection solution is to find a way to group the devices in the network. The papers [12-14] present similar problems in which they implement in order to solve their issue. As in the paper [12] the differences in the application of the scheme are very different from that of the paper [13]. The most prominent differences in the proposed algorithm are headed by the variable number of witness devices and their target selection process in addition to the fact that device's status in the house may change to malicious or fail at any time.

Each of the devices in the network will have a special "witnesses" attribute which keeps track of the details of each of its targets (which it witnesses), as well as the number of devices witnessing (monitoring) itself. Each device will need to have a minimum number of witnesses in order to satisfy the accountability requirements of the network. This is accomplished by each device querying the network for witnesses in the case that it does not meet the network required amount.

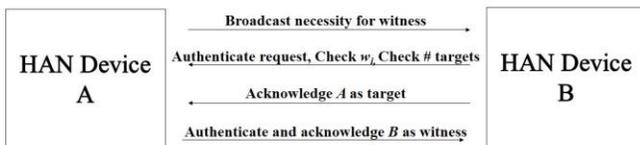


Fig. 1 Request for Witness

Fig. 1 shows that device A requests device B as witness. The request for a witness must first be authenticated and then the fielding device(s) must verify that the device requesting a witness has not already reached the number of required witnesses in the network. If this number of witnesses has been met, the fielding device will explore other options in the network before agreeing to witness said device. This is done to maintain efficiency in device usage in the network by keeping the number of witness devices in the network at the minimum. This will also achieve a more efficient scanning procedure.

A more detailed explanation of the selection process is given in Table 1.

Table 1: Selection Algorithm

<p><b>Input:</b> A device <math>dev_i</math> and its targets (<math>targets(dev_i)</math>);  <b>Begin at <math>dev_0</math> (smart meter)</b></p> <ol style="list-style-type: none"> <li><b>procedure:</b> <math>groupDevs(dev_i; targets(dev_i); \forall dev_j \in targets(dev_i))</math></li> <li><math>(\beta, newWitness) = receive();</math> // Read target/witness data</li> <li>Establish witness connection</li> <li><b>while</b> (<math>newWitness \neq NULL</math>) do //pool on witness request</li> <li>  <b>if</b> (<math>!witnessMax(newWitness)</math>) then</li> <li>    <math>newWitness = getNewWitness(targets(dev_i))</math>.</li> <li><b>end if</b></li> </ol>
---

```

8. end while
9. if ( $!targetReq(newTarget)$ ) then
10.   $newTarget = getNewTarget(targets(dev_i))$ .
11. end if
12. end procedure
    
```

Table 1 accomplishes a very necessary task which is maintaining ideal and required target and witness devices for each device. Understanding the fact that mobile devices will be constantly moving in and out of the network, as well as powering on and off, the witnesses and target attributes must be maintained and verified constantly instead of simply established and ignored. Devices which do not meet these requirements cannot be fully trusted in their actions as witness devices. Such devices may maintain status as target devices in the accountable scheme.

### 4.2 Inspection Algorithm

The aim of the inspection algorithm is to locate malicious or malfunctioning devices. Many actions that a device can take may not be explicitly labeled as malicious. In instances that a devices may act maliciously or innocently malfunction or falsify data in a logging method, there may be concessions for these types of actions to be flagged. Once these actions are observed, there must be a method to, without a doubt, deduce which device are participating in such acts.

In the typical smart grid HAN, the number of inspecting devices is defined by the number of devices resident in the HAN and the variable witness requirement of the accountable scheme. The dynamic nature of the scheme and the typical HAN are represented in the set of malicious devices  $Q$  and the set of inspecting devices  $I$ . the inspecting devices are taken from the set of witnesses  $W$ . While  $I \in W$ , the inverse is not true, as only a small subset of  $W$  is needed to have the necessary witnesses for each of the devices in  $D$ . Once  $I$  is established, it must only be changed when a device in the set leaves the network. The device inspection algorithm is given in Table 2:

Table 2: HAN Device Inspection

<p><b>Input:</b> (<math>I</math>) Inspecting Devices  <b>Output:</b> (<math>Q</math>) The set of possibly malicious nodes  <b>Initial:</b> <math>Q = \{ \}</math></p> <ol style="list-style-type: none"> <li><b>procedure:</b> <math>Detect(I)</math></li> <li>  <b>if</b> (<math>I</math> is empty)</li> <li>    <b>return</b> <math>Q</math>;</li> <li>    // retrieve the witness data <math>D</math> from the set of witnesses;</li> <li>    <b>while</b> (faulty or suspected <math>dev(s)</math> are reported)</li> <li>      <b>for each</b> (<math>dev \in D.s_0 \cap D.s_1 \cap D.s_2 \cap \dots D.s_n \cap</math>)</li> <li>        <b>if</b> (<math>dev \neq "clean"</math>)</li> <li>          <math>I := I - \{dev\}</math></li> <li>          <math>Q := Q \cup \{dev\}</math></li> <li>        <b>if</b> (all <math>devs</math> in <math>D == "clean"</math>)</li> <li>          <b>return</b> <math>Q</math>;</li> </ol>
--

```

else go to step 5;
10. end for
11. end procedure

```

We can view the set of inspecting devices and their targets to be a tree structures consisting of  $m$  parent nodes and  $n$  children nodes, with  $n$  representing the number of target devices required to fulfill the accountable scheme's witness requirement for each device. The set of inspecting nodes have edges directed to the root node which will either be the smart meter or an energy services interface (ESI), connecting the HAN to the NAN and the utility. Fig. 2 shows the view from the witness device down. In this manner, the root node will only have to inquire of the single witness device shown to find any possibilities of ill-reported or malicious data from any of the witness device's targets.

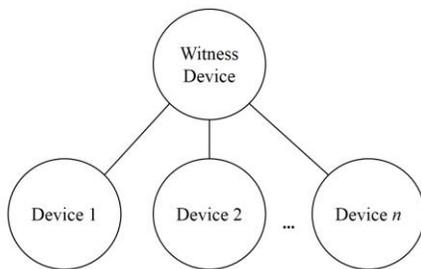


Fig. 2 Witness-Target Structure

This approach is basically a selective scanning procedure similar to the D-scanning method in [12] for utilizing the multiple witness scheme. In other words, we take advantage of the fact that each device has several witnesses, and in this, the inspector set can be reduced to a significantly small number in order to have sufficient coverage of each device on the network. Notice that here, the inspectors are inspecting home appliances/devices but in the paper [12], the inspectors are inspecting meters in the neighborhood area.

Inspection will be carried out based on the policy of the network and each instance of inspection will be completed in a single sweep based on the report from each of the inspectors. Since the devices in  $I$  are already contained in a devices witness set, there is no need to further monitor the inspectors. This means that the completion of the inspection will require the inspecting device(s) to only scan the devices in  $I$  to gain a full view of the entire network. Errors or suspect reports propagate up the chain and will reach the respective device in  $I$  which is witnessing the faulty/suspect device in question. In a tree-based scheme with many levels, there is inherent trust in the devices which are at the lower levels. In situations where there are only a single witness of a device, either device may be malicious and that data may not be propagated as it should due to the malicious device which the data travels through. In the proposed scheme, there will always be multiple devices witnessing a single target, and the witness devices will always be witnessed by other devices. Therefore, there is

accountability on many levels which is necessary in such environments which due to the nature of the networks, and may have malicious and/or faulty devices in it at any time.

## 5 Evaluation and Analysis

The average number of active internet connected devices in each home averages about 5.7 in the U.S. Based on the paper [21], the number of overall appliances can be expected to be about three of four times more. In a smart grid where all devices communicate via the network, each of these devices should be expected to participate in accountability processes and report their actions as the network or scheme requires.

The selection process is either automatically inclusive of all of the devices in the HAN, in the case of the most trivial scanning technique, which is the complete scanning of the entire set of devices  $D$ , or the set of  $I$  inspector devices described in the previously proposed algorithm. The algorithms proposed and described earlier are each simulated through software. The data is created based on common number of devices found in consumer homes to medium size businesses. Fig. 3 shows the number of devices required for scanning in the proposed algorithm for a specific number of devices as utilized in the several simulations.

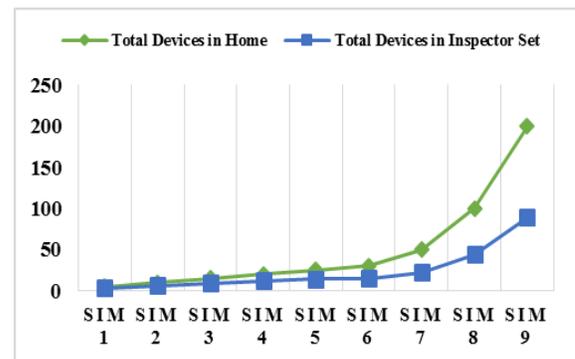


Fig. 3 Inspector Selection 3 Witnesses

The results in Fig. 3 requires three witnesses per connected device which will be input into the selection algorithm from Table 1. We observe that the set of inspectors  $I$  is roughly half of the complete set of HAN devices. If we require only two witnesses, we do not observe much change in these numbers as shown in Fig. 4.

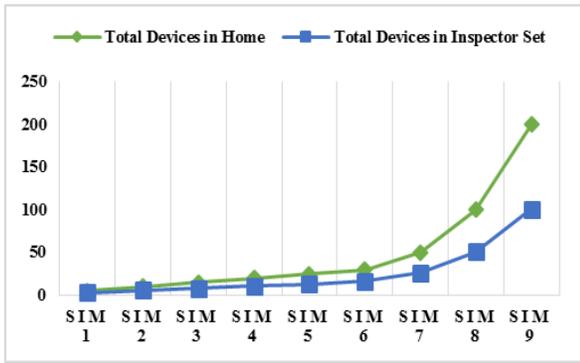


Fig. 4 Inspector Selection 2 Witnesses

With this application, one detail of note is that as the volume of devices trends upward as the inspector set ratio slowly trends down at an average of about 0.2 over the life of the simulations which span a minimum of 5 devices for small HAN environments to 200 for the larger environment. This factor is slightly more pronounced when the number of witnesses is increased. Also, the number of devices that a specific device is allowed to witness is a major factor. We will limit this to 4 devices in order to maintain a low level of computational and networking latency. This method of scanning is obviously more efficient than the scanning each of the devices in the HAN which must include all device in  $I$ .

[12] proposes a method for dynamic inspection of devices in the NAN which can be applied with few modifications in the HAN. This method makes the assumption that the malicious devices, once discovered, will be removed from the set which is being scanned and then proceeds with its scanning in a round-by-round fashion and queries a global monitor to check results. The problem in tree-based approaches that operate on the basis of scanning rounds is that in an extreme case it will repetitively scan the same set of devices assuming that devices with child nodes continue to infect or falsely report of the nodes below them at the scan time or the children devices continue to report falsely during scan time. The premise here is that the higher level nodes must be trusted inherently as the child nodes may not participate in the accountability of their parent nodes. This in essence makes networks such as this trusted networks from the view of those nodes looking up the tree.

### 5.1 Inspection Analysis

The scanning of the complete set of devices in the home  $D$ , can be accomplished by inquiring of all devices in the network. If the number of malicious devices in the network is much less than the total number of devices, the efficiency will suffer due to unnecessary operations. The extreme cases being  $D_s = 1$  and  $D_n = n$ , the complexity of such scanning is bounded by the number of devices in  $D$  although  $Q$  is unknown. Even in environments such as the HAN where  $Q$  is likely to be constantly changing, there is little benefit to inquiring select devices and not the entire set due to the possibility of run-time

infractions. The proposed protocol solves this problem receiving data on each device at run-time through  $W$ .

The inspector selection process is viewed as finding a Borel set of all singleton node sets of the nodes in set  $D$ . This produces a minimal set of nodes that contains all nodes in  $D$ . Each element in the  $D$  will be represented by their target nodes data as opposed to their own data as in Fig. 5.

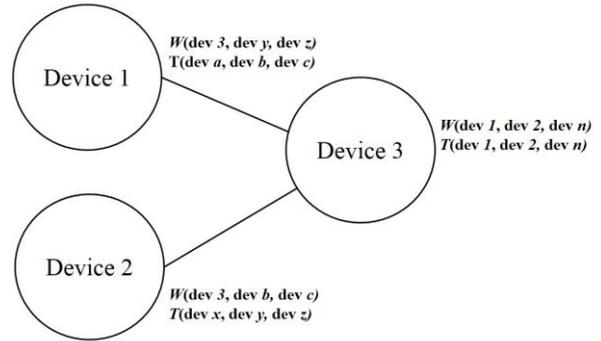


Fig. 5 Node Witness-Target Data

We can define the generating of the Borel set by iterating through the following which performs all countable intersections and all countable unions possible with the subsets in  $D$ :

$$D \rightarrow D_{\delta\sigma} \tag{1}$$

It is now necessary to remove the subsets which are non-singletons. This is achieved in Equation 2:

$$D_{\delta\sigma} \rightarrow I \text{ only if } |D_{\delta\sigma}| = 1 \tag{2}$$

This set of inspectors is essentially minimal while maintaining full view of the entire set  $D$  which allows a single sweep scan of  $I$  to determine the reported status of each device.

As the devices in  $I$  maintain the reported statuses of all devices in the network, the upper bound and lower bound of necessary scanning of the size of  $I$  remains the same while  $D$  does not change. Upon such a change, devices affected in the network will make the necessary changes due to network requirements. This shows that performance remains the same in either of the extreme cases of malicious node existence where  $n = 1$ , or  $n = n-1$ .

The accountable environment presented in [9-11] makes some assumptions such as activity patterns of varying consumption devices, generation, and storage devices. The papers [9-10] adopt the multiple witness concept for the HAN, and the paper [12] adopts the inspection procedure in the NAN, while the HAN and the NAN are quite different networks. The HAN is mostly defined by the appliances/devices in each home, while the NAN is defined mainly by the AMI in a neighborhood or service area. In the proposed algorithms, we use the multiple witness concept and

the inspection procedure in the HAN which are similar to those used in [9-10,12] though with key differences.

The proposed witness selection eliminates the inherent trust requirements of higher level nodes that tree-based methods maintain, as the trust of witnessing nodes is normally necessary. Our method leaves no nodes without witnesses that will monitor their actions. Other methods require more devices to actively witness targets on the network.

## 6 Conclusion

This paper has proposed algorithms to efficiently provide accountability in the smart grid HAN through multiple witness monitoring and malicious inspection of nodes. An inspector selection algorithm is proposed as well. These algorithms are employed in environments of a dynamic nature, where complete device inclusion is required and there is always a possibility of changing malicious status. There is still work to complete in are of accountable generation and storage. The analysis in the study shows that the proposed algorithms are effective in creating an accountable environment in a smart grid HAN.

## Acknowledgments

This work was supported in part by the National Science Foundation under Grant CNS-1059265.

## REFERENCES

- [1] M. Amin and B. F. Wollenberg, "Toward A Smart Grid: Power Delivery for the 21st Century," *IEEE Power and Energy Mag.*, vol. 3, no. 5, 2005, pp. 34-41.
- [2] X. Kai *et al.*, "The Vision of Future Smart Grid," *Electric Power*, vol. 41, no. 6, 2008, pp. 19-22.
- [3] National Institute of Standards and Technology. NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 2.0 [online] Available: [http://www.nist.gov/public\\_affairs/releases/upload/smartgrid\\_interoperability\\_final.pdf](http://www.nist.gov/public_affairs/releases/upload/smartgrid_interoperability_final.pdf)
- [4] M. A. Rahman, P. Bera, and E. Al-Shaer, "SmartAnalyzer: A noninvasive security threat analyzer for AMI smart grid," Proc. Of IEEE INFOCOM, 2012 pp.2255-2263, 25-30 March 2012
- [5] Z. Lu, X. Lu, W. Wang, and C. Wang, "Review and evaluation of security threats on the communication networks in the smart grid," Prof. of MILCOM 2010, pp.1830-1835, Oct. 31 2010-Nov. 3 2010
- [6] C. Neuman and K. Tan, "Mediating cyber and physical threat propagation in secure smart grid architectures," Proc. of 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm), pp.238-243, 17-20 Oct. 2011
- [7] Cisco Systems Inc., "Internet protocol architecture for smart grid" White Paper, Jul 2009. [Online]. Available: [http://www.cisco.com/web/strategy/docs/energy/CISCO\\_IP\\_INTEROP\\_STDS\\_PPR\\_TO\\_NIST\\_WP.pdf](http://www.cisco.com/web/strategy/docs/energy/CISCO_IP_INTEROP_STDS_PPR_TO_NIST_WP.pdf)
- [8] National Institute of Standards and Technology. NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 2.0 [online] Available: [http://www.nist.gov/public\\_affairs/releases/upload/smartgrid\\_interoperability\\_final.pdf](http://www.nist.gov/public_affairs/releases/upload/smartgrid_interoperability_final.pdf)
- [9] J. Liu, Y. Xiao, and J. Gao, "Achieving Accountability in Smart Grid," *Systems Journal*, IEEE , vol.PP, no.99, pp.1,16, 0
- [10] J. Liu; Y. Xiao, and J. Gao, "Accountability in smart grids," Proc. Of IEEE *Consumer Communications and Networking Conference (CCNC), 2011.*, pp.1166-1170, 9-12 Jan. 2011
- [11] Z. Xiao, Y. Xiao, and D. H. Du, "Non-repudiation in neighborhood area networks for smart grid," *IEEE Communications Magazine*, , vol.51, no.1, pp.18-26, January 2013.
- [12] Z. Xiao, Y. Xiao, and D. H. Du, "Exploring Malicious Meter Inspection in Neighborhood Area Smart Grids," *IEEE Transactions on Smart Grid*, vol.4, no.1, pp.214-226, March 2013
- [13] Du, D., Hwang, F., *Combinatorial Group Testing and its Applications*. Singapore: World Scientific, 1993.
- [14] R. Dorfman, "The detection of defective members of large populations," *Ann. Math. Statist.*, vol. 14, pp. 436-440, 1943.
- [15] India Smart Grid Forum. "Home Area Network" [Online] Available: <http://indiasmartgrid.org/en/technology/Pages/Home-Area-Network.aspx>
- [16] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. P. Hancke, G.P., "Smart Grid and Smart Homes: Key Players and Pilot Projects," *IEEE Industrial Electronics Magazine*, Vol.6, No.4, pp.18-34, Dec. 2012
- [17] K. Kok, S. Karnouskos, J. Ringelstein, A. Dimeas, A. Weidlich, C. Warmer, S. Drenkard, N. Hatzigiorgi, and V. Lioliou, "Fieldtesting smart houses for a smart grid," in Proc. 21st Int. Conf. Electricity Distribution (CIRED), Frankfurt, June 2011, pp. 1-4.
- [18] K. Kok, S. Karnouskos, J. Ringelstein, A. Dimeas, A. Weidlich, C. Warmer, S. Drenkard, N. Hatzigiorgi, and V. Lioliou, "Fieldtesting smart houses for a smart grid," in Proc. 21st Int. Conf. Electricity Distribution (CIRED), Frankfurt, June 2011, pp. 1-4.
- [19] E. McCary and Y. Xiao, (2014) "Smart Grid HAN Accountability with Varying Consumption Devices". Manuscript submitted for publication.
- [20] N. Parker "How Many Net Connected Devices are in Your Home". [Online] Available: <http://www.nbnco.com.au/blog/how-many-net-connected-gadgets-in-your-home.html>
- [21] J. Liu, Y. Xiao, S. Li, W. Liang, C. L. P. Chen, "Cyber Security and Privacy Issues in Smart Grids," *IEEE Communications Surveys & Tutorials*, Vol. 14, NO. 4, pp. 981 - 997, Fourth Quarter 2012.

# Developing and Assessing a Multi-Factor Authentication Protocol for Revocable Distributed Storage in a Mobile Wireless Network

Track: Security Management, Network Security

Scott Bell, Eugene Vasserman, *Member, IEEE* Dan Andresen, *Member, IEEE*

**Abstract**—The ability to share real-time data among soldiers provides a huge tactical advantage for modern military units. There is, however, significant risk involved in distributing this information across a mobile wireless network. An adversary could capture one or more of the mobile devices, potentially granting access to this data, and putting the entire unit at risk. While there are no feasible ways to completely eliminate this risk, we can effectively reduce the adversary's window of opportunity by requiring multi-factor, revocable authentication to access individual devices and files which are distributed across the mobile network.

While this new protocol does incur some costs, tests show that the costs for this improved security are more than acceptable. Cryptographic operations slow down the request-response process but response time is only increased by 61 milliseconds, which is more than acceptable given the improved security our protocol provides. Additionally, analysis of battery consumption shows that a tablet can send over 2000 requests or respond to over 800 requests with a 1% drop in battery power and a smart phone can make over 300 requests with a similar 1% drop in battery power.

**Keywords**—multi-factor authentication, mobile devices, security, file-sharing

## I. INTRODUCTION

The tactical advantage provided to today's military by secure communication channels transmitting real-time data is evident in "blue force tracking" (BFT) systems, used in military vehicles and command stations to display tactical information such as the locations of other friendly and enemy forces [7, 22]. This technology has been hereto confined to vehicles and ground stations due to the required power level and antenna size, but ideally this type of data sharing should be available to individual soldiers using small, low-power, devices.

One of the most versatile methods for disseminating and displaying information in real time, and the current tool of choice, is a network of hand-held wireless devices [19]. However, these can be easily lost or stolen, potentially giving adversaries access to all the information the owner would have been able to access. To prevent this, devices might encrypt their local storage and require users to enter a short PIN. However,

the complexity of PINs are limited, since they must be capable of being entered quickly in combat, making them vulnerable to brute-force cracking.

We introduce, implement, and analyze a protocol for practical secure storage and access of data using mobile devices in a battlefield, resilient to device capture by a powerful nation-state adversary, *without relying on strong hardware tamper resistance*. We do this without sacrificing usability in the field or incurring prohibitive computation, time, or battery consumption costs. Our protocol eliminates the previously-required assumption that any device (software or hardware) which connects to a military network is trusted to use the on-demand file access, and reduces the window of vulnerability during which captured devices may serve as network gateways for adversaries. Further, we allow devices and users to be revoked independently of each other, improving flexibility if devices are lost, stolen, or damaged.

We analyze the security of this system, and show its practicality by implementing a prototype and measuring battery consumption and latency costs. We show the relative probability that an adversary may gain unauthorized access to data is essentially reduced to having a malicious insider or capturing a legitimate user and device and coercing action on the adversary's behalf. Our prototype implementation, on mobile devices running Android, shows that the costs associated with these security gains (in terms of increased response time and battery consumption) are acceptable with a software-only implementation, and can be further reduced through specialized hardware [5, 6, 18, 20].

## II. RELATED WORK

MDFS is a distributed file system for mobile networks [10] which uses encrypted then erasure-coded files and Shamir secret sharing encoded keys [21]. Key and data fragments are stored in pairs spread among multiple devices. This system provides resilient data storage, and limited protection of the data in that an adversary is unable to recover information from the file fragments stored on a single captured device. However, it does not attempt to address the issue of restricting access to fragments stored on the system from a compromised device. Anyone possessing a device that is recognized as part of the network is able to obtain access to any of the shared files.

In both [4, 17], the authors present the idea of a wearable token which can be used to authenticate a user within the

---

The authors are with the Department of Computing and Information Sciences, Kansas State University, Manhattan, KS 66506.  
E-mail: rsbell@ksu.edu, eyv@ksu.edu, dan@ksu.edu

work place. The computer or mobile device is able to detect when the token has moved too far away and then disables sensitive operations. The National Institute of Standards and Technology (NIST) has published a standard that addresses proximity-based authentication for mobile devices, describing the use of a Bluetooth token for user authentication on a single mobile device [12]. In all of these examples, the mobile device and token are paired at initialization and form a long-term association, which may not be desirable in all situations. Our solution allows users to connect to different devices within the system as needed, and utilizes the authentication information provided by the token as one factor for authenticating file fragment requests made to remote devices.

### III. DESIGN

The goal of our system is to reduce the likelihood of an adversary gaining access to protected files via captured mobile devices. We recognize that fully preventing this type of attack is an impossible goal given the potential adversaries and the application environment. Therefore, we look to reduce this likelihood by limiting the window of opportunity during which such an attack might occur and increasing the difficulty of the tasks an adversary must accomplish within this window.

To prevent device hardware compromise from leaking information, files are dynamically reconstructed from the network while being accessed by a user (as in threshold secret sharing) then discarded and never stored locally. We therefore no longer rely on device tamper-resistance to prevent access to locally stored files. Latency and connectivity is a concern when using real-time file retrieval within a mobile network, but can be reduced using erasure coding combined with threshold cryptography [10]. A file is split into  $n$  fragments which are distributed across  $n$  devices, but only  $m < n$  fragments must be retrieved to reconstruct the file. This minimizes latency by requesting fragments in parallel and enhances security, since an adversary controlling fewer than  $m$  devices cannot learn any information about the stored content. It also serves to reduce the effect of denial of service attacks, where individual devices within the network are jammed.

Our system consists of four primary components: adversaries, users, wearable tokens, and the mobile devices participating in the wireless network. A token uniquely identifies a specific user, while mobile devices do not. This allows a user to login to any device within the system and gain access to the distributed data files. When logging in to a device, a user is required to enter a valid PIN that is unique for that user (or provide some other authentication input such as a biometric scan). The device will then attempt to locate that user's wireless token in order to obtain a second form of authentication. This authentication information is used to prove the user's identity both to the local device and to other devices on the network when making requests for file fragments. The details of the token and mobile device along with their communication protocol are described below.

#### A. Adversary Model

We focus on reducing the threat posed by an adversary capturing a limited number of mobile devices, and assume a

well-funded and motivated nation-state adversary attempting to access the secure files stored and shared between such devices. This adversary would be capable of defeating tamper-resistance on captured devices or tokens and cracking user PINs. The scope of this work is not concerned with networking protocols and issues such as denial of service attacks. We assume secure inter-device communication. Similarly, attacks via malicious software are beyond the scope of this work.

#### B. User

Before the start of a mission, each user is issued a unique key pair and PIN. The user's private key is stored on that user's token while the public key is distributed to all mobile devices participating in the mobile network for that mission. A user's PIN must be simple to be useful in combat, so *we do not assume that this PIN provides strong protection against a skilled adversary*, and treat it more like a user identifier. Other, potentially more secure, identifiers such as biometric scans could be utilized in place of the PIN.

#### C. Token

The token should be a self-contained tamper resistant wireless device. To provide the second form of authentication for a user, the token contains that user's private key which corresponds to the public key known by all mobile devices within the mission network, and associated with the user's PIN. The token provides a cryptographically secure means of authenticating a user without increasing the user's involvement in the authentication process. The only additional requirements for the user are to initially pair the token with a given mobile device and then keep the token within range of the mobile device while accessing data files. Minimizing user involvement in this process is a priority given the context in which this system will be used. Once a mobile device connects to the token, the device will issue a challenge message which the token signs using the private key and then returns to the device.

We envision that the token is a wearable device such as a watch, ring, dog tags or key fob. This makes it easier for users to keep the token with them at all times and also makes it more difficult for an adversary to acquire a token as opposed to a mobile device which may be dropped or forgotten somewhere. The token could utilize technologies such as standard Bluetooth, Bluetooth LE [2], ZigBee [23], Near Field Communication [16], or any other short range wireless protocol. The choice of technology is based on application specific requirements such as cost, operating environment, token concealment, durability, availability and battery life expectations. Transmission range for token communication should be limited to a few feet as the purpose is to ensure that the mobile device is being used by the soldier who is in possession of a specific token.

#### D. Mobile Device

1) *Initialization and Startup*: Each mobile device provides a user with access to files stored within the mobile network and responds to requests from other devices for file fragments

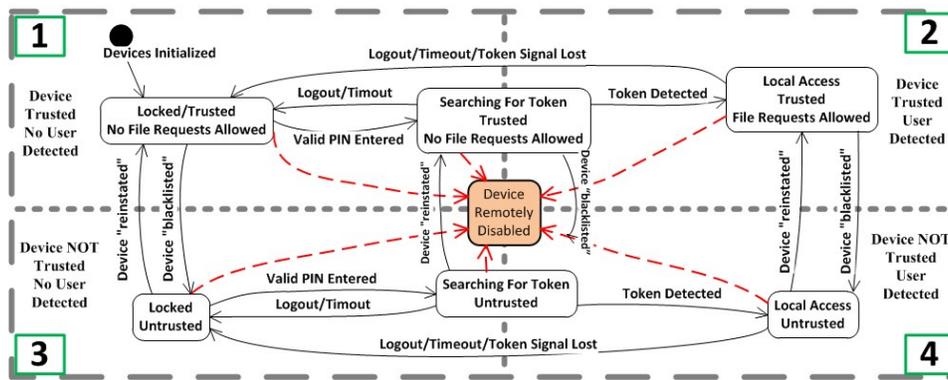


Fig. 1: State diagram for the mobile device. Note the four “trust quadrants.” Q1: Device is trusted but no user is connected. Q2: Device is trusted and a trusted user is connected. Q3: Device is NOT trusted and no user is connected and Q4: Device is not trusted and a trusted user is connected. In the central state, the device has been disabled and must be reinitialized.

which it possesses. All devices participating in the system are initialized at the start of *every* mission with the following information:

- A hash of the PIN (or other login information) assigned to each authorized user along with:
  - The user’s public key
  - The user’s token connection details
- The ID and public key for other devices
- A device-specific key pair
- Administrative account public key
- Device clocks are loosely synchronized
- Time-to-live for authentication messages
- File fragment IDs and where each is stored

The state diagram in Figure 1 shows the progression of states for a device once it has been initialized for a mission. The device starts in a locked state where no file requests are allowed to be sent out.

2) *User Authentication:* Figure 2 shows the communication process between the system components. In order to grant a user access to the file system, the device first requires a PIN (or another form of authentication) from the user. Longer PINs are more secure but can be more difficult to remember as well as more difficult to enter when under stress. To alleviate the need for an overly long PIN, we include the token as a second form of authentication. Once a user has entered a valid PIN, the device locates the token associated with that PIN and issues a challenge message which the token signs and returns. This challenge should contain a time stamp to prove that the signed message is fresh and the ID of the mobile device to show which device issued the challenge and reduce the possibility of replay attacks. After the device has received the signed challenge response, it can use the hashed value of the PIN to locate the corresponding public key and verify the signature, authenticating the user. If the signature cannot be verified the device will drop the connection to the token and return to the initial locked state.

3) *Logout/Timeout:* The device periodically sends new challenges to the token. Each response is verified and replaces the previous challenge response. If the token is not detected, or

responds incorrectly, OR if the user does not interact with the data files over a specified period of time, the device returns to the “locked” state and any files that are currently open are deleted from memory. These features limit the window of opportunity for an adversary to recover a device which is in a state that allows access to files on the system.

4) *Requesting Files:* As indicated in Figure 2, the next step is for the requesting device to send requests for fragments of the file to other devices within the mobile wireless network. It is assumed that this communication channel is secure although the details of that are beyond the scope of this work. Each request must contain enough information to prove the following:

- An authorized user possesses the device.
- The user authentication information is fresh, and was obtained by the requesting device.
- The requesting device is a currently authorized device.

Requests for file fragments are built using authentication information for both the user and the requesting device along with the identity of the file fragment being requested. Specifically, the challenge response and a hash of the user’s PIN are included to authenticate the user while a copy of the request signed using the requesting device’s secret key is used to authenticate that device. The responding device verifies the user and the requesting device, checks that the time stamp is fresh, and that the device ID in the challenge matches the ID of the requesting device. If all of these checks are found to be valid, the responding device will return the file fragment to the requesting device. If not, the request is ignored.

5) *Blacklisting vs. Disabling a Device:* Up to this point, the device has been in a trusted state (in the upper half of Figure 1) meaning that other devices on the network trust it and will respond to requests it makes as well as sending their own requests for fragments it possesses. If a device has not been detected for an extended period of time, or is thought to be compromised, the device is “blacklisted.” Other devices will ignore requests from the blacklisted device and will not request file fragments from it. This notification is expected to be initiated by an administrative account and broadcast over the network. In Figure 1, states below the horizontal dashed

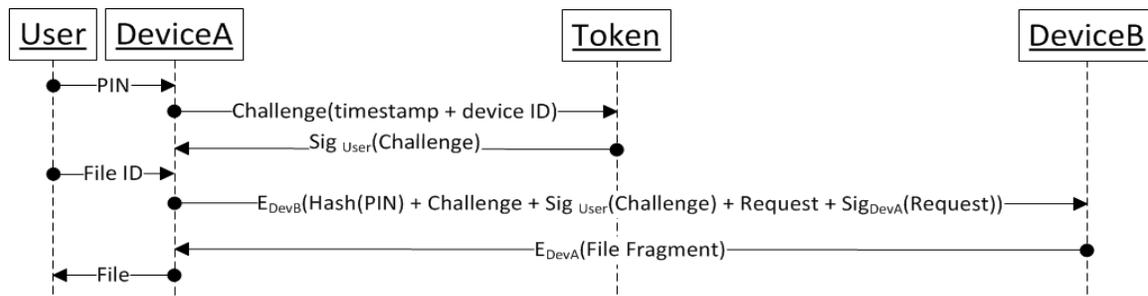


Fig. 2: File request protocol for mobile file access. Inter-device communication is authenticated and confidential.

line represent a device that has been “blacklisted” and is not currently trusted by the other devices in the system.

In a more severe case where a device is known to be compromised or lost, the administrative account holder can attempt to remotely disable the device (e.g. the iOS remote wipe feature [1]). This may require a long distance communication channel to send a signal that instructs the device to effectively destroy all of its contents. This is represented by the center state in Figure 1.

A device which has been disabled is in an unrecoverable state and no longer contains any of the initialization information listed in Section III-D1. A device in this state must be reinitialized to be used again whereas a “blacklisted” device could potentially be reauthorized once the administrator is certain the device can be trusted and would then be immediately able to participate in the wireless mobile network again. For example, the soldier with that device may have been separated from his unit and is now back with them.

6) *Blacklisting a User*: It is possible that a user’s information (both PIN and token) is compromised. This would allow the adversary to access the file system from any mobile device which is trusted by the network. The system can handle this threat in a manner similar to that discussed for a compromised device. A notice is sent out by the administrative account and each device within the system marks the PIN/public key pair for that user as “untrusted” and requests made using that user’s authentication information are ignored. If the user is later proven to once again be “trustworthy” then the devices on the network can be notified of this fact and they will once again respond to requests from that user.

#### IV. IMPLEMENTATION

We implement a prototype of our design using two Samsung Galaxy Tab tablets as the requesting and responding devices, while the token is simulated on a laptop computer. Bluetooth as well as IP over Wi-Fi are used to model inter-device communication. (Methods of managing wireless mobile networks have been studied extensively [3, 8, 14] and are outside of the scope of our work.)

##### A. Token

For token communication, we utilize standard Bluetooth because it is available on most mobile devices in the market

today. When launched, the application waits for a connection request over the Bluetooth socket. If this is the initial connection request from a device, the user must verify the device pairing procedure (for example, by pressing a button on the token), and this verification will cause the token to replace the previously stored device connection information. Once a connection is established, the process continuously loops through a (receive challenge)/(sign challenge)/(respond) sequence. If the connection is broken or fails, the device returns to the initial state, waiting for a new socket connection.

##### B. Device

1) *User Authentication*: As Figure 2 shows, when the user wants to access files, they are first required to enter a PIN. Once a valid PIN is entered, the mobile device retrieves the connection information for that user’s token and launches a separate thread which detects, and connects to, the token via a Bluetooth socket. It then sends a challenge message containing a current time stamp and the mobile device’s network ID to the token. Receiving the signed challenge back from the token, the mobile device is able to verify the signature using the public key associated with the PIN that was entered.

2) *File Fragment Requests*: In order to reconstruct the files a user wishes to access, the requesting device must retrieve file fragments from other devices in the network. We are using IP connections over Wi-Fi for inter-device communication. This is acceptable for our testing purposes since our work is not focused on routing or the reliability of the network communication system. Once a socket is created between the two mobile devices, the requesting device sends the request to the responding device. As can be seen in Figure 2, this request contains the information needed to verify that the user and requesting device are authorized to access the requested file and that the same device which obtained the challenge response is making the file fragment request.

The contents of the fragment request are described in III-D4. The responding device can verify the user, the requesting device, that the device issued the challenge and the freshness of the signed challenge response. If all of these checks are acceptable, the device returns the file fragment over the socket connection. If any of the checks fail, the device ignores the request.

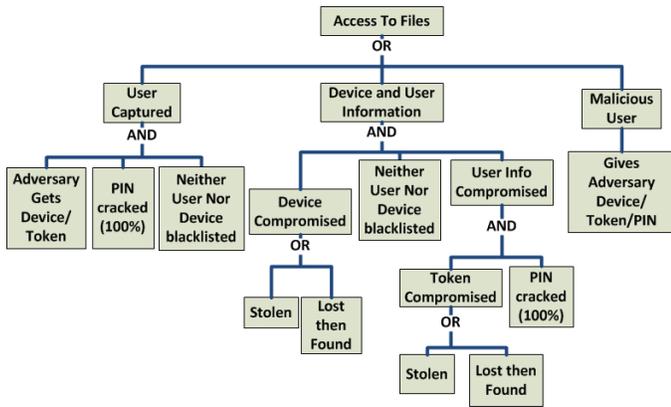


Fig. 3: Attack tree showing possible paths for an adversary to gain access to files within the network.

V. EVALUATION

We evaluate our approach in two ways: first we quantify the expected security improvements, then we demonstrate a prototype implementation and measure response time, computation overhead, and battery consumption.

A. Security Improvement

Given the context for expected deployment of this system and even conservative estimates of adversarial capabilities (a nation-state), we must expect that the adversary will gain access to the file system of at least one captured device. However, we wish to reduce the probability of data exposure, i.e. reduce the amount of data accessible to the adversary. We consider strong adversaries, who employ sophisticated techniques to crack encryption, bypass tamper resistance, and even capture targeted users to obtain devices, tokens, and login information.

Figure 3 shows an attack tree for file access against our system. Our protocol requires the adversary to provide fresh, authentic requests containing signatures from both a trusted device and a trusted user’s token in order to gain access to file fragments via the mobile network. There are three basic ways the adversary can gain access to a mobile device/ token combination: 1) capturing a soldier with a trusted device and token, 2) acquiring both a lost or stolen device and token, or 3) having a collaborator within the unit. The left branch shows the case wherein a user is captured. The adversary likely gains access to both a device and token, leaving the PIN as the only feature keeping the adversary from accessing the system. It takes a small amount of brute-force computation to crack the PIN, so we assume that the probability of cracking it is 1, and it happens almost instantly. The variable in this scenario is the time required to realize the user, device, and token have been captured, and blacklist them, denying access to the system even if the device and token are compromised.

The middle branch covers scenarios wherein the device and token are independently acquired by the adversary. In this situation, it is imperative that soldiers report lost or stolen devices and tokens quickly to limit the window of opportunity an adversary has to use these items to access the system. It should be noted that since devices are initialized at the start

TABLE I: Variables For Equations 1a and 1b.

Variable	Probability of:
$P[c]$	Soldier captured
$P[m]$	Malicious user in unit
$P[d]$	Device acquired/compromised
$P[p]$	PIN compromised
$P[t]$	Token acquired/compromised
$P[nbl]$	Components not blacklisted

of each mission, compromised tokens and devices can only be utilized on the mobile network for which they have been initialized and if they are possessed by the same adversary.

In the right branch, a user is collaborating with the enemy, and the adversary is assured access to the data. Our system has no effect on this attack, and insider attack protection is beyond the scope of this work.

1) *Probability of Compromise During a Mission:* We first consider the likelihood that the security of the system will be compromised over the course of a mission. Equation (1a) represents the likelihood that the adversary will gain access to data in a mobile wireless system which stores file fragments across devices and does not require multi-factor authentication or utilize revocation. Equation (1b) incorporates revocation and multi-factor authentication in addition to as-needed access. The variables are defined in Table I.

$$P[a] = P[c] + P[m] + P[d] * P[p] \tag{1a}$$

$$P[b] = P[c] + P[m] + P[d] * P[p] * P[t] * P[nbl] \tag{1b}$$

The factors in these equations are estimated, as the military does not make such values readily available. We conservatively assume that a malicious user will always successfully compromise the system and that a captured soldier also results in a compromised system before the token or device is blacklisted. This is the normal case when not implementing our protocol and the worst case when using our protocol. Given these assumptions, the two equations have equivalent values for the first two terms.

This leaves the middle branch of the attack tree, and the third term, as the difference between the two equations, and the measure of our security improvement. In general, we assume that  $P[d] > P[c] > P[m]$ , making this term dominant in both equations. We assume that the PIN will always be cracked ( $P[p] = 1$ ), so the terms are reduced to  $P[d]$  and  $P[d] * P[t] * P[nbl]$ . Thus, the difference in the dominant term in the two equations is the probabilities of a token being compromised and that a compromised token and device are not blacklisted. We can apply estimated values to these variables to understand their interaction. For example, if we say there is a 50% chance of each event, then the terms become  $P[d]$  and  $P[d] * .50 * .50$ , so we have reduced this term by 75%. Given that the token is a wearable item not easily lost, and that we assume that every lost token is recovered by the same adversary that has a

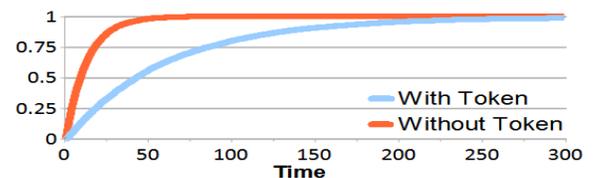


Fig. 4: Probability of compromise over time (CDF).



Fig. 5: Average time (a) to complete the number of requests; (b) per request for tests containing a given number of requests.

mobile device, these assumptions are likely very conservative. We therefore claim that our system has essentially reduced the probability of compromise to the first two terms in the equations.

2) *Probability of Compromise by a Given Time.*: Above, we calculate the probability of an adversary gaining access to the system at *some point* during a given mission. Here we evaluate how this risk changes with time and show how our protocol reduces this threat. Given enough time, the system will be compromised by a dedicated adversary, so we ask the question “How much of a difference is our system making at each point during the mission?” To answer this, we simulate token and device states through discrete time slices of a mission. The possible states for devices and tokens are: in a soldier’s possession, lost/stolen but not blacklisted, or blacklisted.

Transitions between these states during a given time slice are based on the probability of 1) a device or token being lost or stolen, or 2) a lost item’s absence being noted and reported (the item is blacklisted). Once a device is compromised, that instance of the simulation is halted and we record the time step when this occurs. For systems not implementing our protocol, the a device is compromised at the time slice when it is lost or stolen *or* a soldier is captured. Our system is considered to be compromised during a time slice if both a token and a device have been lost or stolen and neither has been blacklisted, *or* a soldier is captured (we assume conservatively that a missing token and device are in the possession of the same adversary). We performed 6,000 iterations of this simulation to the point of compromise recorded how often the system is compromised during each time slice. These values were divided by the total number of iterations to give the probability of compromise occurring during a given time slice. The cumulative value of all probabilities prior to a given time slice were added to that slice’s probability to give the likelihood that the system is compromised at or before that point in time. Figure 4 shows the simulation results – the difference in the two curves shows the increased security provided by our protocol.

TABLE II

Variable	Probability of:	Value
$N$	Number of devices/tokens	40
$P[c]$	Soldier captured	0.001
$P[d]$	Device lost/stolen	0.0008
$P[t]$	Token lost/stolen	0.0012
$P[dbl]$	Lost device blacklisted	0.50
$P[tbl]$	Lost token blacklisted	0.85

Probability of event occurring during a given time segment

Table II shows the probabilities we utilized in the simulation. We are tracking each individual item during multiple, shorter, time segments within a mission, so they answer questions such as: ‘*what is the probability of this item being lost during time slice 1.*’ These probabilities are our best conservative estimates of what the values are in a real-world scenario. We do not specify the units of time, as our concern is to determine how the configurations perform compared to one another. It is clear that during a reasonably large number of iterations, the system utilizing our protocol provides substantially stronger security than systems which do not.

### B. System Response

We measure response time due to cryptographic operations at the requesting device from request generation to response receipt. While the device typically sends one request to each device containing a needed file fragment, our tests consist of 5 to 100 request/response cycles. This ensures the times being measured are not dominated by other operations on the device. We used RSA with 1024-bit keys, but other algorithms may provide better security/performance metrics. Our implementation is software-only, and we expect that dedicated hardware will be used as part of the token and device, reducing time and power consumption [5, 6, 18, 20].

Figure 5a shows that elapsed time grows linearly with the number of request/response cycles in a given test run, though at different rates, reflecting the additional time per request for cryptographic operations. Figure 5b shows that the average response time per request is consistent regardless of the number of cycles in a given test. The average response time without cryptographic overhead is  $67 \pm 0.4$  milliseconds, while with cryptography this increases to  $128 \pm 0.5$  milliseconds. Although this *may* be noticeable for a user, delays will not increase linearly with the number of requests under normal operation. A portion of the delay occurs at the responding device, and since many devices will be queried simultaneously, these delays occur in parallel. Considering the expected transmission times for a mobile wireless network with multiple devices distributed throughout a military unit’s operating area, this delay does not contribute significantly to latency.

### C. Battery Consumption

Given the context in which our system is expected to be deployed, battery consumption must be a priority, so we evaluate power overhead using a methodology similar to the one above for measuring time overhead. We repeat the request/response cycles as for the system response tests, measuring the battery

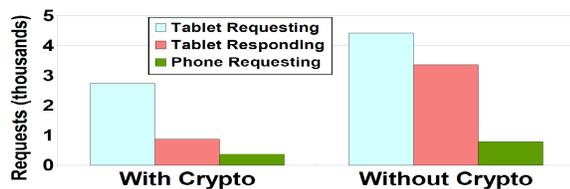


Fig. 6: Battery consumption comparison for tablet and phone.

drop on each device with and without cryptographic operations. Figure 6 shows the number of operations required to cause a 1% drop in battery power on each device tested.

1) *Requesting Device*: For the requesting device,  $4415 \pm 153$  requests cause a 1% drop in battery power without cryptography, and cryptography reduces this to  $2736 \pm 85$  requests. Tests were also performed using a Droid Pro mobile phone to more closely match the device a soldier might carry. With this device, the tests without cryptography averaged  $779 \pm 18$  requests per 1% drop in battery power and cryptography drops this to  $371 \pm 5$  requests.

2) *Responding Device*: For the responding device,  $3365 \pm 127$  requests cause a 1% drop in battery power without cryptography, and cryptography reduces this to  $867 \pm 14$  requests. This greater battery consumption reflects the extra verification performed by the responder.

3) *Token*: Assuming it is self-powered and generalizing from worst-case results (the mobile device request tests above) with a 30 second polling interval, 1% of token battery power drains in 3 hours, allowing over 12 days of use. Since the token performs a single operation (signing the received challenge), performance is expected to be much better than the requesting device. Further, given that the token is expected to be composed of optimized hardware, battery life can improve with little monetary investment [11, 13, 15].

## VI. SUMMARY

We show how to reduce the risk of an adversary accessing files stored within a mobile wireless network while keeping battery life and delay time costs reasonable. Multifactor authentication alleviates the problem of full network access from logged-in devices and provides revocation for users and devices independently. Our system provides a significant improvement in security over prior work without sacrificing ease of use. Our system is also more flexible, allowing users to operate any device in the system while limiting an adversary's window of opportunity to use an acquired device.

## REFERENCES

- [1] Apple. Bluetooth specification version 4.0, June 2012. <http://www.apple.com/ios/>.
- [2] Bluetooth-SIG. Bluetooth specification version 4.0, June 2010. <https://www.bluetooth.org/Technical/Specifications/adopted.htm>.
- [3] C. Candolin and H. Kari. A security architecture for wireless ad hoc networks. In *MILCOM 2002. Proceedings*, volume 2, pages 1095 – 1100 vol.2, oct. 2002.
- [4] Y. Chen and M. Sinclair. Tangible security for mobile devices. In *Proceedings of the 5th Annual International*

*Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services, Mobiquitous '08*, 2008.

- [5] H. Eberle, A. Wander, N. Gura, S. Chang-Shantz, and V. Gupta. Architectural extensions for elliptic curve cryptography over  $GF(2^m)$  on 8-bit microprocessors. In *ASAP*, 2005.
- [6] T. English, M. Keller, K. L. Man, E. Popovici, M. Schellekens, and W. Marnane. A low-power pairing-based cryptographic accelerator for embedded security applications. In *SOCC*, 2009.
- [7] S. Fox. JSTARS adds blue force tracking capability. Air Force Print News Today, 2006.
- [8] N. Garg and R. Mahapatra. Manet security issues. *International Journal of Computer Science and Network Security*, 9(8), 2009.
- [9] S. Huchton. Secure mobile distributed file system. Master's thesis, Naval Postgraduate School, 2011.
- [10] D. Hwang, B.-C. Lai, P. Schaumont, K. Sakiyama, Y. Fan, S. Yang, A. Hodjat, and I. Verbauwhede. Design flow for HW/SW acceleration transparency in the thumbpod secure embedded system. In *Design automation conference*, 2003.
- [11] W. Jensen, S. Gavrilu, and V. Korolev. Proximity-based authentication for mobile devices. In *Proceedings of The 2005 International Conference on Security and Management*, pages 398–404, June 2005.
- [12] Y. Matsuoka, P. Schaumont, K. Tiri, and I. Verbauwhede. Java cryptography on KVM and its performance and security optimization using HW/SW co-design techniques. In *CASES*, 2004.
- [13] N. Michalakis and D. Kalofonos. Designing an NFS-based mobile distributed file system for ephemeral sharing in proximity networks. In *Applications and Services in Wireless Networks, 2004. ASWN 2004. 2004 4th Workshop on*, pages 225 – 231, aug. 2004.
- [14] V. Nambiar, M. Khalil-Hani, and M. Zabidi. Accelerating the AES encryption function in OpenSSL for embedded systems. In *ICED*, 2008.
- [15] NFC-Forum. Near field communication specification, May 2012. <http://www.nfc-forum.org/specs/>.
- [16] B. D. Noble and M. D. Corner. The case for transient authentication. In *Proceedings of the 10th workshop on ACM SIGOPS European workshop*, EW 10, pages 24–29, New York, NY, USA, 2002. ACM.
- [17] L. Oliveira, D. Aranha, E. Morais, F. Daguano, J. Lopez, and R. Dahab. TinyTate: Computing the Tate pairing in resource-constrained sensor nodes. In *NCA*, 2007.
- [18] K. Osborn. Smart phones increase 'SPOT' reporting in Army evaluations, 2011.
- [19] M. Scott, N. Costigan, and W. Abdulwahab. Implementing cryptographic pairings on smartcards. In *CHES*, 2006.
- [20] A. Shamir. How to share a secret. *Commun. ACM*, 22(11):612–613, Nov. 1979.
- [21] ViaSat. Blue Force Tracking 2, 2012.
- [22] ZigBee-Alliance. ZigBee 2007 specification. <http://www.zigbee.org>.

# Smart Grid HAN Accountability with Varying Consumption Devices

Eric McCary<sup>1</sup>, Yang Xiao<sup>1</sup>

<sup>1</sup>Department of Computer Science, The University of Alabama, Tuscaloosa, AL, US

**Abstract** - Smart grid is an emerging power infrastructure that integrates the newest communication and information technology. Along with the advent of newer technologies, security challenges have never been encountered before. Among the principals for securing these infrastructures, accountability is one with few studies in smart grid literature. This paper will ensure greater accountability of devices in the home area network (HAN) by providing an algorithm to more accurately calculate and estimate the energy consumption of devices whose consumption varies while it is powered on. Analysis and simulations will show that the method is effective.

**Keywords:** varying consumption, accountability, smart grid

## 1 Introduction

With the current state of technological advancement today, the demand for energy has begun to outpace the growth of efficient production capability [1]. Inaccuracy in estimation and malicious devices are a few of the problems which are to blame. This of course, requires the energy sector to create a more efficient demand-response cycles. Solutions of many of these energy inefficiencies are addressed in the smart grid convention. The smart grid can be described as the current power delivery system with integrated bidirectional communication and real-time analysis on energy generation, transmission, and distribution data in order to create predictive and necessary recommendations for consumers [2]. The National Institute of Technology and Standards (NIST) defines six key areas which make up the grid below [3]: bulk generation domain, transmission domain, distribution domain, operations domain, service provider domain, and customer domain.

These areas are expressed as domains which house several major components in the energy arena. Each domain has a unique distributed computing environment, sub-domains, and equipments to suit its mission-specific needs. It is also important to note that the domains of the grid are interconnected with adjacent domains which provide coordinated functionality Which Figure 1 details.

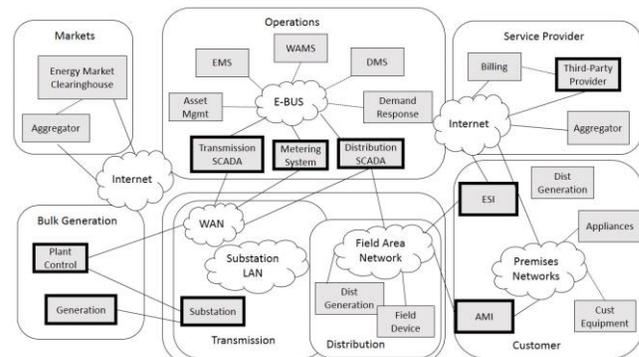


Figure 1: Smart Grid Overview [3]

To utilize maximum efficiency in a smart grid, the power utility must employ a customized energy demand management scheme which will allow for balancing of demand and supply in a manner which reduces costs for the customer and producer [4, 5]. For example, power usage can be scheduled to avoid peak load times which incur higher costs for both the utility and the consumer. This is one of the basic requirements for a smart grid and one of the concepts that make smart grid novel.

An unsolved problem in this arena lies in not only estimating power usage in general, but estimating power intricately in the home area network (HAN) and establishing accountability in this domain. An accountable environment has been researched and created in [6-8] while making some assumptions such as activity patterns of varying consumption devices, generation, and storage devices, which may or may not lead to sufficiently accurate estimations in the HAN. This paper will detail a scheme for more accurately estimating usage of varying consumption devices in order to create a more fully accountable and more accurate environment in the HAN.

The rest of the paper is organized as follows. Section II provides some background in smart grid and accountability, Section III describes the problem statement and Section VI details the varying consumption device estimation scheme. Section V gives some analysis of the scheme and finally, the paper is concluded in Section VII.

## 2 Background

The smart grid is an increasingly expanding network of networks and system of systems. Advanced metering infrastructure (AMI) plays a large role in the customer domain with two-way communications from the consumers to the system operators [9]. From the data, the consumer can receive pricing information, and schedule/modify their power usage.

Sensing and measurement are constantly occurring and reported throughout the smart grid. This type of wide situational awareness plays a large role in demand-response. The same action and consequence are often replicated in the HAN. If energy usage can be more accurately predicted and identified, peak prices will not be as high or occur as often. Achieving accountability in the HAN has rarely been studied in the past. Related areas such as disaggregation and load monitoring [10-12], are useful, but normally estimate device usage based on the aggregate amount from the residence as a whole and the normal consumption of a device. This type of estimation is more useful when identifying customer behavior patterns and device malfunctions instead of a more fine-grained approach of estimating and assuring that each device is accountable.

### 2.1 Accountability

Accountability serves as a complement to the core principals of information security and the component that allows authorized individuals more robust tracking and auditing history as well as establishing trust and confidence within the HAN between devices. Currently, accountability in the distribution domain of the grid only extends to a single residence which aggregates the appropriate data of all of the devices located therein. This is sufficient for the currently required duties of billing the customer based on total use of all appliances to be fulfilled, but in order for demand-response to be optimized on both sides of the equation, a more fine-grained approach must be utilized.

There are several requirements which can contribute to an effective accountable environment or mechanism. These include [13]: decentralization of accountability mechanisms, scalability, minimal impact, data collection, identity management. Inclusion of these elements in the accountable scheme can be very profitable and will help to cover the accountability requirements.

Sufficient and appropriate data must be extracted and archived for review and evaluation. It is imperative to discern the most effective location and type of data to archive and evaluate in each environment. Some devices may require a specific set of data to be analyzed, while others may require a set of data which includes a few of the parameters explicitly required by one, and a few not needed to satisfy the requirements of another. Although in most environments these parameters will be uniform

For smart grid applications, the main objective of accountability currently is to maintain record of and assure that a device acts as it says and/or is expected to. In other words,

the assurance that a device truthfully reports its power usage and other parameters required at a pre-specified time interval and/or when requested.

Even in the case of AMI, there is still much room for error and we cannot always expect for the record that the utility manages and what is recorded at the customer's end to be identical. Malicious action, malfunction, miscalculation in estimation, or calibration may be the cause of such differences. Making the HAN accountable on a more fine-grained level can help alleviate problems such as these and provide us with a means of locating a compromised device which can immediately be disabled or serviced instead of canceling service to the residence indefinitely.

## 3 Energy Consumption in the HAN

Currently, automation and energy management in the home has been well researched. In the smart home, the following offerings should be available and implemented [14]: *Information* – Graphically represented energy usage data, *Automation* – priority setting and scheduling, *Advanced Control* – Information and control locally of from third parties, *Integration* – Use of all previous offerings and forecasting information.

Implementation of the previous offerings allows for a much “smarter” home environment. In these smart homes, many new components are being added in order to complete its objective of minimizing the daily power costs and shaving the power consumption peak.

## 4 Problem Statement

The only accountability required in the distribution domain is or the power grid is basically the monthly reading of the power meter which records energy usage at each consumer residence. Inside the HAN, security measures can be put in place to introduce accountability to a specific entity such as an IDS (intrusion detection system), or the energy service interface (ESI), but with the amount of risk and the number of devices, the accountability mechanism should be distributed and implemented in many devices in the network.

Disaggregation techniques are wide ranging and have been well researched. The three major techniques include [15]: survey, single point sensing, and distributed direct sensing.

These techniques alone are simply not sufficient as more smart devices are being added to homes that are programmable. This means that the threat of malicious and/or malfunctioning devices is greater than ever and still increasing, and brings about the need for a more fine-grained model of accountability in the HAN.

In addition to creating an environment where the devices are accountable for their energy usage, the devices should also be made accountable for all of their actions with a scheme implemented in a distributed fashion. If a device takes an

action that causes an unexpected or unscheduled amount of energy use, that action and the energy should be verified and the device accountable for it.

Many devices in the HAN normally maintain energy consumption patterns which are not static in the sense that though operation may be scheduled, they may use varying amounts of energy at a given time. Such devices may not have a constant power capacity factor at any given time while it is active. Some examples of these types of devices include water heaters, boilers, even coffee makers. It is of great importance to implement a mechanism which can correctly assure accountability in environments utilizing such devices as well as detect any events that can be problematic. Once the events have been discovered in a timely manner, it should be categorized so that the appropriate actions can be taken to resolve the issue.

#### 4.1 Varying Energy Consumption in the HAN

There are hundreds of millions of devices which have the flexibility of varying consumption in the HAN. Much of this flexibility can be attributed to the time flexibility of the device usage due to management, or the upper and lower bounds of flexible energy usage amounts of that device. The paper [16] describes the aggregation and disaggregation of these flexible devices. Since the aim of this paper is accountability and accuracy in the HAN, aggregation and disaggregation is out of the scope of this discussion.

Energy consumption is considered to generally be constant in many devices. In fact, it is an estimation of the necessary energy the devices needs to function. This means that there will normally be some slight overuse or underuse. Therefore, for many reasons, including the previous, the measurement and estimation techniques are threshold-based. Scheduling helps to manage and estimate energy use to a certain extent, but the details of a varying consumer are normally partially defined by the habits of the user and cannot be completely estimated unless sample testing and forecasting is in place along with high level control of the varying consumer devices.

### 5 Estimation Scheme

This discussion will propose extension to the status reporting mechanisms that are required to be communicated between targets and witnesses. The basis of protocols is that devices called “witnesses” are used to monitor target devices in order to verify specific actions

We can categorize HAN devices as schedulable and non-schedulable. This means that for scheduled devices, we will likely have more insight into the specific times in which the target device will be using power. We assume that witnessing devices have the capability to sample the power of target devices in the HAN.

To identify any device as a “varying consumer” (VC), the consumption patterns must be identified. Some details may also be inferred from the devices rated power usage even

though this amount will likely not be constantly consumed during the duration of the devices active phases.

We can assign each device  $n$  in the network to a specific grade of rated power usage. Each of these devices possesses a certain number of attributes  $\mathbf{attr}$  defined by its hardware and software requirements and its environment as  $\{\mathbf{attr}_{n,1}, \mathbf{attr}_{n,2}, \mathbf{attr}_{n,3}, \dots, \mathbf{attr}_{n,n}\}$ . Once the attributes of the device are combined we find the rated power usage grade of that specific device,  $G_p \mid p \in \{1, \rho\}$ , where  $\rho$  denotes the number of grades in the HAN and  $p$  represents the grade of the device in question. Calculating the grade of the  $i^{\text{th}}$  device can be completed using this equation:

$$G_p(n_i) = \sum_{k=0}^n w_{i,k} \cdot \mathbf{attr}_{i,k} \quad (1)$$

where,  $w_{i,k}$  denotes the weight of the attribute represented in the summation while  $n_i$  will be assigned to one of the power usage grades by the following:

$$\text{if } (G_{min} < n^i < G_{max}), n^i \in G_p \quad (2)$$

$G_{min}$  and  $G_{max}$  denote the lower and upper limits of the specific usage grade  $G_p$ , respectively. The value of  $n_i$  depends heavily on the environment and the attributes considered with the device. With the usage grade of a specific device known, the range of its power consumption can more effectively be estimated. The VC algorithm is defined in Table.

Table 1: VC Algorithm

1. **Function Group** (*witness\_devs, target\_dev*)
2. discover\_VC\_tendencies(*target\_dev*) // function samples usage, infer from
3. // target\_dev power rating
4. target\_dev reports median and max power usage
5. **REPEAT UNTIL** usage\_group(*target\_dev*) is discovered
6. **IF** *target\_dev(max) && target\_dev(med) ∈ (f(x<sub>1</sub>), ..., f(x<sub>n</sub>))*
7. usage\_group(*target\_dev*) => Group(*f(x<sub>1</sub>), ..., f(x<sub>n</sub>)*)
8. **ELSE** redo group test for GROUP(*f(x<sub>n+1</sub>), ..., f(x<sub>n+Δ</sub>)*)
9. **END REPEAT**
10. Verify usage\_group(*target\_dev*) with sampled power usage
11. 12. **REPEAT** discover\_VC\_tendencies(*target\_dev*)
12. **IF** data sampled and calculated are differing above  $\Delta$ , mark *target\_dev* “suspect”
13. **END REPEAT**

## 5.1 Multiple Status Reporting

Any status report made by a target to a witness device, whether randomly requested or scheduled, will contain the target device's power status. Instead of defining this with only two values (on/off), we have at least 5 states for the target report. With these various reporting states, we can more closely estimate the power usage of the current device without sampling it. How these states are defined is based on the reported power input amount of the target device. While the power usage grade is known, we can more precisely estimate the device's energy usage at a specific time by having multiple status levels within that devices usage grade which are relative to the device. As  $G_p$  has an explicit minimum and maximum usage rates, we can further divide the state into 4 power ranges (the 5<sup>th</sup> state represents the "off" state). The power consumption can be even more accurately estimated as future power sampling observes energy usage and finds a common power level within a specific state. With this information, the witness can assume that the target device uses a specific amount of power (which has been constantly observed) while the target device reports the recently reported state.

## 5.2 Estimating Power Usage

In order to calculate the power usage of any device or to closely estimate it, the operational capacity  $P_i$  and the state of the device must be known [1]. In the case of VC devices, the power capacity is not necessarily known and is more difficult to estimate any time. Herein lies a major problem, and the solution lies within the estimation explained in this section and the next.

With the knowledge of the power usage grade  $G_p$ , and the status communicated from the target device, we can more accurately estimate the power usage at time  $t_i$  for any device  $i$ . For the group/witness-based accountability scheme in the HAN, trivially, we understand that we can derive the power usage of a particular device  $P_i$  with the function:

$$P_i = \int_{t_b}^{t_a} p(t) \cdot R_i(t) dt \quad (3)$$

Where  $p$  is the expected power consumption at time  $t$  and  $R$  is the running state that the device is in at time  $t$ . Understanding the rated power usage amount is not sufficient for the VC device, as the amount of power this device uses is not necessarily constant. Therefore, in the previous discussion we introduced the premise of a power usage status  $S_i$  which is relative to the power usage group  $G_p$ . These values are used to more accurately estimate the power usage at a specific time. It is understood that if  $i \in G_p$  then at any time  $p_i < \max(G_p)$ . In other words, the power usage for device  $i$  will never be greater than the upper bound on the range or the group it

resides among. If this does occur the device which reports this reading will be considered suspect or faulty and marked for further evaluation.

## 5.3 Varying Consumption Status Reporting

With the usage grade and the device power level status known, we can effectively construct a table for quick reference, or calculate the estimated amount in on-board the device at the necessary time. The protocol for building accountability into varying consumption devices is presented in the following steps.

1. The potential VC device is connected to the network. Recommended scheduling is verified as well as device attributes
2. The potential VC device informs proper authorities (e.g., witnesses, smart meter) of the minimum and maximum power usage/requirement. This value will be used to assign the VC into a usage grade  $G_p$ .
3. The VC device informs proper authorities of its initial usage state (level 0-4) while witness devices sample the VC device's power consumption. Witnessing devices calculate the expected power usage (from usage grade and status report) and compare it with the actual usage. If the difference is too large, label the device faulty.

At any time that a device's reported status, of time constraints are found to be incorrect or suspect, that device is labeled as suspect or faulty according to the reports of witnesses.

## 6 VC Protocol Evaluation

This section will analyze the algorithm detailed in earlier sections, while comparing to the most common and currently used method in academic literature. In accomplishing this, it must be understood that the purpose of this effort is to contribute to and enhance accountability in smart grid networks. Most publications in today's smart grid HAN energy academic arena focus on optimizing load scheduling in the customer domain. While this is important, establishing accountability is also an extremely important feature that the grid should focus on, which the proposed algorithm helps to accomplish. This accountability is on a per-device basis which differs from some other studies in accountability which focus on multi-user households [18], and understanding users and how to motivate savings [17].

The following will discuss the performance of the proposed algorithm. In the considered smart grid HAN, the simulation will encompass the devices connected in the HAN network which is also interconnected digitally and electrically with the smart grid. As a luxury, we assume that each device in the HAN is complete with an energy sampling unit in order to verify attributes and actions of peer devices in the network. This is a responsibility of each device which, in working together, will work to create a trusted and accountable environment. There are  $n=40$  devices in each home, and a

single home represented in each simulation. Attributes of each of the devices are all pseudo randomly generated at simulation runtime. We assume that the smart meter is able to discover the operational time of each device, as well as its power requirement information. This functionality can be programmed into the specific device so that this information is broadcasted upon connection, or especially for legacy devices, assumptions can be made and verified with the sampling unit. Flexibility in operational time should be available, but as discussed earlier, load management and estimation of future energy usage is not of concern in the assurance of HAN accountability. Renewable energy is also considered, which means that residences in the neighborhood may have some form of generation on their premises.

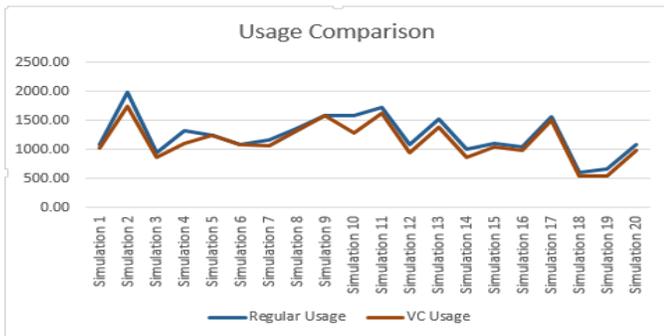


Fig. 2: Average Power Usage

Fig. 2 shows the results of energy usage of a HAN comprised of VC devices and devices that use a static amount of energy in kW-h format. The attributes for these devices are selected pseudorandomly, and therefore the number of VC devices in each home differs while the total number of devices is static. The measurements are both carried out under the same load, and the energy data is recorded over a months' time, meaning that the results are accumulated. The amount of energy usage difference visible here is fairly significant whether evaluating a single hour of device powered on activity or from the monthly view. We can also see that in the simulation there several instances where a home did not have any VC devices (simulations 6, 7, and 10) where the energy consumption is measured as being equal with the VC algorithm and with the currently utilized measurement algorithms.

Fig. 3 gives us more insights into the improved calculation of the VC devices in question.

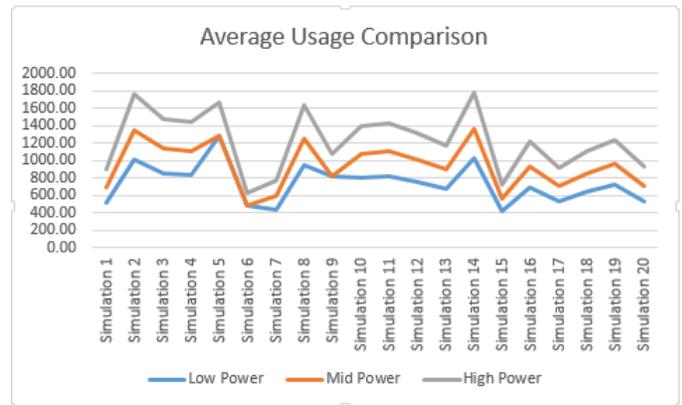


Fig. 3: Average Power Phase

As introduced in Section V, Fig. 3 details how much power the average device used in the simulation in watts. The VC algorithm recognizes the power usage of a device with a threshold-based method and categorizes the amount in one of several categories. For simplicity, the categories are broken down into three (high, medium, and low). The current common algorithm does not take this into account, and although any checks and balances of systems are also threshold based, they do not take into account VC devices, and therefore a larger deviation from what is estimated must be utilized; otherwise the number of false positives will increase dramatically. We can also see that from Fig. 3, that the power draw can be fairly dramatic between the high and low phases and must be accounted for even if the devices are in those phases for a short time in order to make estimation and efficiency more effective.

It is also necessary to understand and analyze the time that the devices are operating in these phases as its importance cannot be understated. Fig. 4 gives us insights into these measurements.

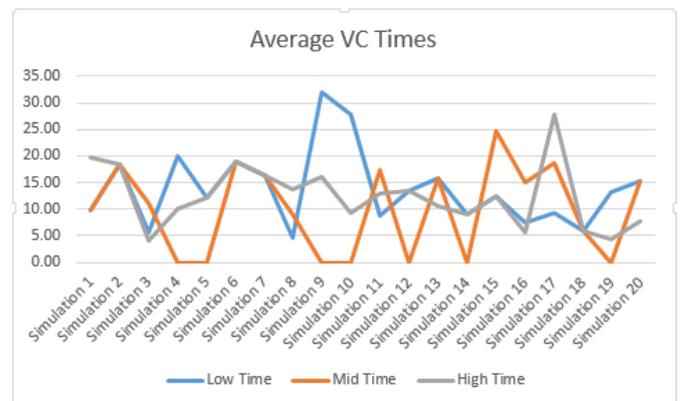


Fig. 4: Phase Time for VC Devices

As is visible, much time is spent outside of the median power usage or expected usage of the VC devices. This data must be utilized to gain more accurate measurements on these devices.

## 7 Conclusion

The state of smart grids will continue to evolve. This obviously means that technology and all other necessary components will need to advance also, and without accountability in the HAN, many deficiencies in security and accuracy can be exacerbated. This paper highlights the modern state of accountability in the HAN and typical home energy usage. The proposed VC energy usage calculation algorithm was given as a solution to more effectively provide accountability and enhance accuracy in the HAN energy calculation schemes.

## Acknowledgment

This work was supported in part by the National Science Foundation under Grant CNS-1059265.

## References

- [1] Xiao, J.; Li, J.; Boutaba, R.; Hong, J.W.-K., "Comfort-aware home energy management under market-based Demand-Response," *Network and service management (cnsm), 2012 8th international conference and 2012 workshop on systems virtualization management (svm)*, vol., no., pp.10,18, 22-26 Oct. 2012
- [2] Cisco Systems Inc., "Internet protocol architecture for smart grid" White Paper, Jul 2009. [Online]. Available: [http://www.cisco.com/web/strategy/docs/energy/CISCO\\_IP\\_INTEROP\\_STDS\\_PPR\\_TO\\_NIST\\_WP.pdf](http://www.cisco.com/web/strategy/docs/energy/CISCO_IP_INTEROP_STDS_PPR_TO_NIST_WP.pdf)
- [3] National Institute of Standards and Technology. NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 2.0 [online] Available: [http://www.nist.gov/public\\_affairs/releases/upload/smartgrid\\_interoperability\\_final.pdf](http://www.nist.gov/public_affairs/releases/upload/smartgrid_interoperability_final.pdf)
- [4] Fouda, M.M.; Fadlullah, Z.M.; Kato, N.; Takeuchi, A.; Nozaki, Y., "A novel demand control policy for improving quality of power usage in smart grid," *Global Communications Conference (GLOBECOM), 2012 IEEE*, vol., no., pp.5154,5159, 3-7 Dec. 2012
- [5] Bu Shengrong, F. R. Yu, and P. X. Liu, "Dynamic Pricing for Demandside Management in the Smart Grid," IEEE Online Conference on Green Communications (GreenCom'11), Sep. 2011.
- [6] J. Liu, Y. Xiao, and J. Gao, "Achieving Accountability in Smart Grid," IEEE Systems Journal, DOI: 10.1109/JSYST.2013.2260697, accepted, 2013.
- [7] J. Liu, Y. Xiao, and J. Gao, "Accountability in smart grids," *Consumer Communications and Networking Conference (CCNC), 2011 IEEE*, vol., no., pp.1166,1170, 9-12 Jan. 2011
- [8] Z. Xiao, Y. Xiao, and D. Du, "Non-repudiation in neighborhood area networks for smart grid," *IEEE Communications Magazine*, Vol.51, No.1, pp.18-26, Jan. 2013.
- [9] U.S. NETL, "Advanced Metering Infrastructure," White Paper, Feb. 2008. [Online]. Available: <http://www.smartgrid.gov/standards/roadmap>
- [10] Filippi, A.; Pandharipande, A.; Lelkens, A.; Rietman, R.; Schenk, T.; Ying Wang; Shrubsole, P., "Multi-appliance power disaggregation: An approach to energy monitoring," Energy Conference and Exhibition (EnergyCon), 2010 IEEE International, vol., no., pp.91,95, 18-22 Dec. 2010
- [11] J. Z. Kolter and M. J. Johnson, "Redd: A public data set for energy disaggregation research," in Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA, 2011.
- [12] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 1, pp. 76–84, february 2011.
- [13] Squicciarini, A.C.; Wonjun Lee; Bertino, E.; Song, C.X., "A Policy-Based Accountability Tool for Grid Computing Systems," *Asia-Pacific Services Computing Conference, 2008. APSCC '08. IEEE*, vol., no., pp.95,100, 9-12 Dec. 2008
- [14] Harkin, S. (2011, Autumn) "Home energy management in Europe, lots of solutions, but what's the problem", Delta Energy & Environment, [Online]. Available: UK [http://www.delta-ee.com/downloads/2011/Delta\\_Research\\_Paper\\_Home\\_Energy\\_Management](http://www.delta-ee.com/downloads/2011/Delta_Research_Paper_Home_Energy_Management)
- [15] J. Froehlich, E. Larson, S. Gupta, G. Cohn, M. S. Reynolds, and S. N. Patel, "Disaggregated End-Use Energy Sensing for the Smart Grid," *IEEE Pervasive Computing*, vol. 10, no. 1, pp. 28-39, 2011.
- [16] Siksnys, L., Khalefa, M.E., Pedersen, T.B.: Aggregating and Disaggregating Flexibility Objects. In: Ailamaki, A., Bowers, S. (eds.) SSDBM 2012. LNCS, vol. 7338, pp. 379–396. Springer, Heidelberg (2012)
- [17] T. Schwartz, G. Stevens, L. Ramirez, V. Wulf. "Uncovering practices of making energy consumption accountable: A phenomenological inquiry". TOCHI 2013, 20(2), article 9. (2013).
- [18] Guo, Y. Jones, M. Cowan, B. Beale, R. 2013. "Take it personally: personal accountability and energy consumption in domestic households". In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '13). ACM, New York, NY, USA

# Implementation of Oblivious Bloom Intersection in Private Set Intersection Protocol (PSI)

L. Ertaul<sup>1</sup>, A. M. Mehta<sup>2</sup>, and T. K. Wu<sup>2</sup>

<sup>1</sup>Math & Computer Science, California State University, East Bay, Hayward, CA, USA

<sup>2</sup>Math & Computer Science, California State University, East Bay, Hayward, CA, USA

**Abstract** - Today we are in the era of Big Data. The design of privacy preserving protocols in data processing is really challenging as the amount of data grows largely and complex. How to preserve privacy while meeting the requirements of speed and throughput have become critical criteria in the design. In this paper, we implement a practical use of Private Set Intersection (PSI) Protocol based on the new approach of oblivious Bloom intersection. The high scalability is achieved with parallel operations. We implemented the basic protocol and utilized Google Contact API to directly access the private contact information from two different Google accounts. The intersection of contact information could be found without disclosing any other private information from each account. We reported the result of the performance with respect to the number of contacts for different security levels. We only computed the intersections of two sets up to 25,000 contacts.

**Keywords:** Privacy, Private Set Intersections, Privacy Preserving Protocols.

## 1 Introduction

Recent controversies about the leakage of documents revealing how big data can be fatal even though it creates tremendous opportunities for the world in field of medical research and national security [26]. Privacy issues and collection of consumer information have also been hot topics in the political circles around the world like the Prism program of the National Security Agency (NSA) under the guise of anti-terrorism [27]. Everyone has the right to privacy, but in the case of big data computation it's necessary to maintain data protection and privacy so that it cannot be misused. Using someone's information without their consent is unethical and we need high security. But if Big Data analytics leads to a terrorist suspect then in this case security of the society is counted much higher than an individual's security and privacy.

According to a study by Wikibon [28], shows that the market for Big Data will reach \$50 billion mark in the next 5 years. According to results shown in Figure 1, in 2012 Big Data stood at just over a \$5 billion in terms of services, hardware and software revenue. The awareness and the interest in Big Data have increased in the recent years. The power and the capability of Big Data to improve the efficiency of operations together with its influence in

technological developments and services make Big Data's CAGR increase 58% from now and 2016.

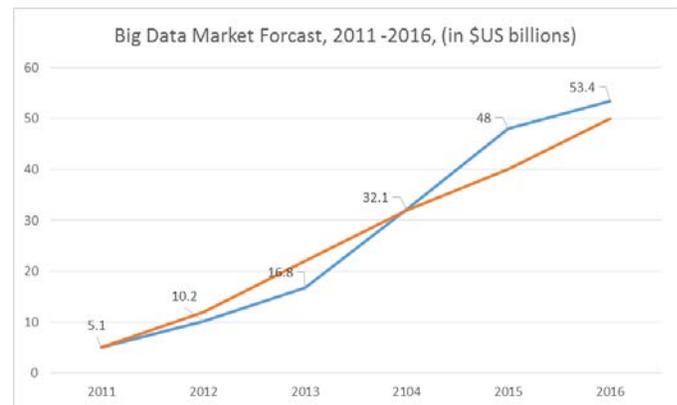


Figure 1 Big Data market Share

Privacy is often addressed as how the information in the application is kept secured and it's an essential issue with big data applications. Everyone has the right to be free from disturbances and intrusion in their respective personal life and also they are subject to right to privacy. Policy makers have therefore started addressing the most fundamental privacy laws, also "personally identifiable information" and role of consent were reviewed.

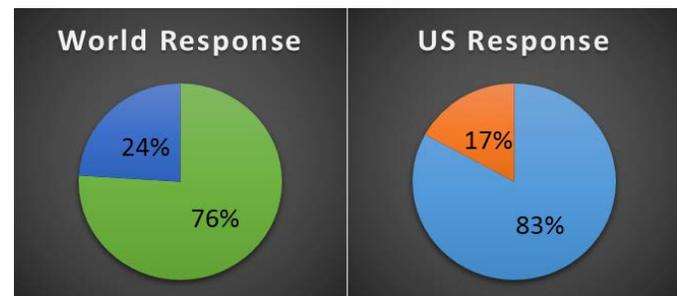


Figure 2 Privacy is the top preference according to World.

Figure 2 shows that trust plays a huge role for the success of Big Data. The survey was carried by Boston Consulting Group (BCG). The result of this survey shows that privacy is the most important preference for Big Data. Top issue according to 76% of consumers feel that the privacy is top issue with Big Data, but in the US 83% of the consumers feel the same. Big data allows organizations to boost their

chances for success by enhancing customer service, manufacturing and other technological aspects. Privacy will create a trust which will help these organizations to benefit themselves and the consumers with Big Data capabilities.

In this paper, we first discuss on the problem of Private Set Intersection (PSI). The scenario is this. There are two parties, a client and a server, who want to compute and find out the intersection of their private inputs. At the end, client learns the intersection and the server learns nothing. The value in this study is that there are many practical applications, such as homeland security, two different law enforcement entities who want to compare their respective databases of suspects [8], detection of online game cheating [21], and find tax evaders [14]. To solve this kind of problem, many proposed PSI protocols are proposed, such as [3, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19]. However, the performance becomes an issue and unacceptable as the required security parameter and the size of the input data are getting big. Based on the result in [3], we found out Changyu Dong's protocol with oblivious Bloom intersection has the best performance comparing with other existing protocols, RSA-OPRF-based protocol by De Cristofaro et al [8] and the garbled circuit protocol by Huang et al [9]. The computational time of two million-element sets with 80 bit security for Dong's protocol needs only 41 seconds while De Cristofaro's protocol needs 10.6 minutes and Huang's protocols needs 27 hours [3].

Next, we implement the basic protocol, proposed by Dong, using the approach of oblivious Bloom intersection with actual private information from Google Contact. The reason we chose this protocol over other existing protocols is not only due to its efficiency and scalability, but also its simple operations. The computational, memory, and communication complexities are all linear in the size of the input sets [3]. Two Google Accounts are created, one as a server and the other one as a client. We first uploaded 25,000 contacts to each account and jointly compute the intersection of their private contact lists. At the end, client learns the intersection and the server learns nothing. The result shows that our implementation can compute the intersection of two 25,000 element sets from both Google Account efficiently.

The rest of the paper is organized as follows: In section 2, we present the definition of the key components of the basic protocol. In section 3, we will discuss the implementation of the basic protocol. In section 4, we evaluate the result.

## 2 The Basic Protocol

In this section, we review the flow and algorithms used in the basic protocol of PSI. The concept is actually simple. First, the client encodes its set  $C$  by computing a Bloom Filter ( $BF_C$ ) and server encodes its set  $S$  by computing a Garbled Bloom Filter ( $GBF_S$ ). By running an oblivious transfer (OT) protocol, the client receives a Garbled Bloom Filter representing the intersection while server learns nothing. At the end, the client uses it to query and obtain the intersection. Figure 3 illustrates the basic PSI protocol.

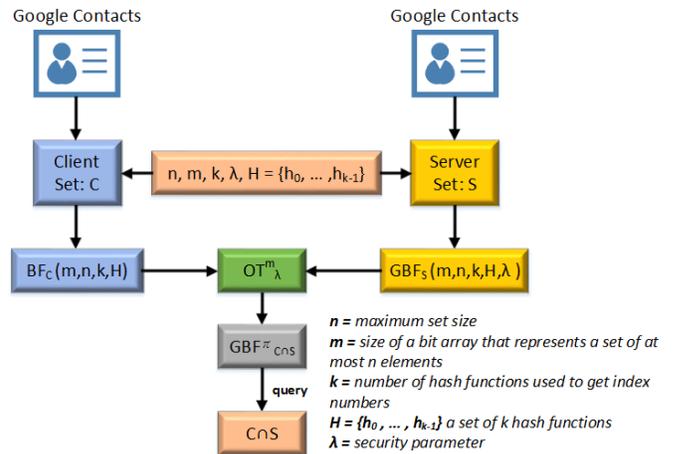


Figure 3: The basic PSI protocol

### A. Bloom Filter

A Bloom Filter [1], designed by Burton H. Bloom in 1970, is probabilistic data structure that is used to test whether an element is present in a set in a rapid and memory-efficient way. A Bloom Filter has a base data structure of bit vector, an array of  $m$  bits that presents a set of  $S$  with  $n$  elements at most. A Bloom Filter uses a set of  $k$  independent hash functions  $H = \{h_0, \dots, h_{k-1}\}$ . For each hash function  $h_i$ , the elements get mapped and uniformly distributed to the index numbers in the range of  $[0, m-1]$ . In this paper, we use  $BF(m, n, k, H)$  to denote a Bloom Filter with the parameters of  $(m, n, k, H)$ , use  $BF_S$  to denote the set  $S$  encoded by Bloom Filter, use  $BF_S[i]$  to denote the bit at the index  $i$  in  $BF_S$ .

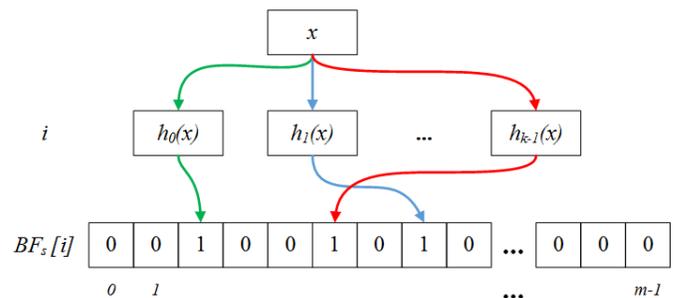


Figure 4: Add an element  $x$  to Bloom Filter

To create a Bloom Filter, as shown in Figure 4, for a set of  $S$ , all  $m$  bits in the array are first initialized to 0. Each element  $x$  that belongs to the set  $S$  is inserted into the filter by hashing  $x$  with  $k$  hash functions to get  $k$  index numbers and then setting all the bits at these indexes to 1, i.e. set  $BF_S[h_i(x)] = 1$ , where  $0 \leq i \leq k-1$ . We can also verify whether an element  $y$  is in the set  $S$  by hashing  $y$  with  $k$  hash functions to get  $k$  indexes and checking these indexes in the filter. If any of the bits at these index locations is 0,  $y$  is not in the set  $S$ . Otherwise, there is a probability that  $y$  is present in set  $S$ . Bloom Filter never yields a false negative due to the nature of hash functions being deterministic. However, it is possible to have false positive,

which means  $y$  is not actually in set  $S$  while all  $BF_s[h_i(x)]$  are set to be 1.

According to [2], the probability of a bit is still 0 in the Bloom Filter is

$$p' = (1 - 1/m)^{kn}$$

The probability of a certain bit is set to 1 is

$$p = 1 - p' = 1 - (1 - 1/m)^{kn}$$

And the upper bound of the false positive probability is:

$$\epsilon = p^k \times (1 + O\left(\frac{k}{p} \sqrt{\frac{\ln(m) - k \times \ln(p)}{m}}\right)) \quad (1)$$

which is negligible in  $k$ .

To be practical, it is necessary to build a Bloom Filter with a false positive probability that is capped. Based on [3], the efficiency of a Bloom Filter depends on the parameters of  $m$  and  $k$ . In our case, we assume that optimal  $m$  is used, which is  $kn \log_2 e$  [3].

### B. Oblivious Transfer

Oblivious Transfer (OT) [4] is a protocol that allows a sender to send part of its input to a receiver that protects both parties. The sender does not know which part of its input the receiver receives while the receiver does not know any information about other part of sender's input. A scenario that best explains the protocol is in the following: a server has a list of  $n$  strings  $x_1 \dots x_n$  and a client wants to learn  $x_i$ . The client does not want the server to know  $i$  and the server does not want the client knows  $x_j$  where  $j$  is not equal to  $i$ . The process of the server should transfer  $x_i$  to the client without knowing  $i$  is called oblivious transfer.

The operation of Oblivious Transfer protocols are actually costly and can be the bottleneck of efficiency in the design. However, Beaver has shown a solution to keep the oblivious transfer calls minimal [5]. In addition, efficient OT extensions were proposed in [6]. In our implementation, we kept the number of Oblivious Transfer calls at minimal.

### C. Google Contact API

The Google Contact API v3 [7] allows client applications to request service and access to a user's contacts. These contacts are stored in user's Google account. However, the user account is limited to a maximum of 25,000 personal contacts and 128KB per contact [25]. The requests to these private user data must be authorized by an authenticated user before the access is granted. Google uses OAuth 2.0 for this authorization process. By specifying the scope information and user's credential in the application, we can retrieve the contact list from the user's Google Account. The details of how to use the APIs are available at Google developers' website and Google's OAuth 2.0 Documentation [7].

### D. Garbled Bloom Filter

A Garbled Bloom Filter [3], introduced by Dong, is a garbled version of a standard Bloom Filter. Essentially, there is no difference between a Garbled Bloom Filter and a Bloom Filter from high level point of view. In the creation of these filters,  $k$  uniform and independent hash functions are used to map each element into  $k$  index numbers. The corresponding array locations are set or checked for adding or querying an element respectively. What makes a Garbled Bloom Filter

different than a standard Bloom Filter is the underlying data structure. To be specific, a Garbled Bloom Filter uses an array of  $\lambda$ -bit strings, where  $\lambda$  is a security parameter, and a standard Bloom Filter uses an array of bits.

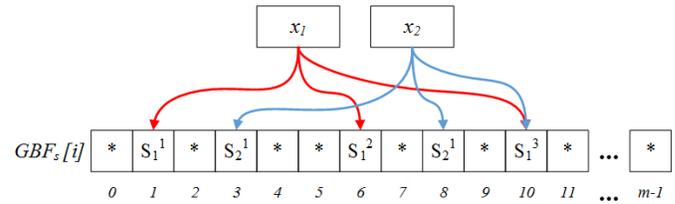


Figure 5: Add elements to Garbled Bloom Filter

Algorithms 1 and 2 [3] in the following are the pseudo codes for adding a set  $S$  into a Garbled Bloom Filter and for querying an element respectively.

**Algorithm 1: BuildGBF( $S, n, m, k, H, \lambda$ )** **E. input:** a set  $S$ ,  $n$ ,  $m$ ,  $k$ ,  $\lambda$ ,  $H = \{h_0, \dots, h_{k-1}\}$

**output:** a  $GBFs(m, n, k, H, \lambda)$

1  $GBFs =$  new  $m$ -element array of bit strings;

2 **for**  $i = 0$  **to**  $m-1$  **do**

3  $GBFs[i] = NULL$ ;

4 **end**

5 **for each**  $x \in S$  **do**

6  $emptySlot = -1$ ,  $finalShare = x$ ;

7 **for**  $i = 0$  **to**  $k-1$  **do**

8  $j = h_i(x)$ ;

9 **if**  $GBFs[j] == NULL$  **then**

10 **if**  $emptySlot == -1$  **then**

11  $emptySlot = j$ ;

12 **else**

13  $GBFs[j] \leftarrow \{0,1\}^\lambda$ ;

14  $finalShare = finalShare \oplus GBFs[j]$ ;

15 **end**

16 **else**

17  $finalShare = finalShare \oplus GBFs[j]$ ;

18 **end**

19 **end**

20  $GBFs[emptySlot] = finalShare$ ;

21 **end**

22 **for**  $i = 0$  **to**  $m-1$  **do**

23 **if**  $GBFs[i] == NULL$  **then**

24  $GBFs[i] \leftarrow \{0,1\}^\lambda$ ;

25 **end**

26 **end**

In Algorithm 1, first an empty Garbled Bloom Filter is created and initialized to NULL (line1-4). To add an element  $x \in S$  into a Garbled Bloom Filter, the element gets spitted into  $k$   $\lambda$ -bit shares using XOR-based Shamir's secret sharing scheme [20] and the shares gets stored in  $GBFs[h_i(x)]$  (line5-21). In this process, it might be possible that  $j = h_i(x)$  has been occupied by a previously added element. For this scenario, the existing share stored at  $GBFs[j]$  is reused (line16-18) as shown

in the Figure 5. The 3 shares of  $x_1, s_1^1, s_1^2, s_1^3$  are added to the  $GBF_s$  first. Then the 3 shares of  $x_2$  get added next. However,  $GBF_s[10]$  has been occupied by  $s_1^3$ .

To prevent  $x_j$  from becoming unrecoverable due to the replacement of  $s_j^3$  with another string, it is reasonable to reuse the string  $s_j^3$  as a share of  $x_2$ , where  $x_2 = s_2^1 \oplus s_2^2 \oplus s_j^3$ . After all the elements in  $S$  are added, the locations in filter that are still NULL will be filled with randomly generated  $\lambda$ -bit strings. According to [3], the reuse of shares will not cause security problems, and the probability of getting all shares of an element that is not in the intersection in this protocol is negligible. The detailed proofs and analysis are presented in [3].

**Algorithm 2: QueryGBF( $GBFs, x, k, H$ )**

**input** : a  $GBFs$ , an element  $x, k, H = \{h_0, \dots, h_{k-1}\}$

**output**: True if  $x \in S$ , False otherwise

```

1 recovered =  $\{0\}^\lambda$ ;
2 for  $i=0$  to  $k-1$  do
3    $j = h_i(x)$ ;
4    $recovered = recovered \oplus GBFs[j]$ ;
5 end
6 if  $recovered == x$  then
7   return True;
8 else
9   return False;
10 end

```

*E. Produce an Intersection GBF*

The idea of how to produce an intersection of Garbled Bloom Filter is based on performing the logic AND operation on two Bloom Filters. The resulting bits copied to a new filter that are set to 1 will be the intersection. The Algorithm 3 [3] in the following is the pseudo code used to build the intersection of Garbled Bloom Filter.

**Algorithm 3: GBFIntersection( $GBFs, BFc, m$ )**

**input**: a  $GBFs(m, n, k, H, \lambda)$ , a  $BFc(m, n, k, H)$ ,  $m$

**output**: a  $GBFcns(m, n, k, H, \lambda)$

```

1  $GBFcns =$  new  $m$ -element array of bit strings;
2 for  $i=0$  to  $m-1$  do
3   if  $BFc[i] == 1$  then
4      $GBFcns[i] = GBFs[i]$ ;
5   else
6      $GBFcns[i] \leftarrow \{0,1\}^\lambda$ ;
7   end
8 end

```

If an element  $x$  is in  $C \cap S$ , we know that  $BFc[i]$  must be a 1 bit and  $GBFs[i]$  must be a share of  $x$  for each location  $i$  it hashes to. By running this algorithm, all elements in  $C \cap S$  are preserved in a new Garbled Bloom Filter. The resulted intersection  $C \cap S$  is called Oblivious Bloom Intersection as shown in Figure 3. The detailed proofs and analysis are presented in [3].

### 3 Implementation

Based on the result presented in [3], the approach of oblivious Bloom intersection is very promising and more scalable and efficient than other existing PSI protocols. Our initial plan is to implement the protocol on mobile phones for practical use. However, the computation requires large amount of memory resources. Due to the fact of limited resources mobile phones have, we decided to implement on laptops.

We have implemented the basic PSI protocol of Oblivious Bloom Intersection in conjunction with Google Contact API in Java. Currently the hash function we used to build and query Bloom Filters and Garbled Bloom Filters is SHA1 [22, 23, 24]. We registered two Google Accounts, one is used as client and the other one is as server. For the initial account setup, we uploaded 25,000 randomly generated contacts with phone numbers to each account and intentionally made 15 contacts commonly exist in both accounts. The purpose is to be able to verify result later. To access the contact information from Google Account, we use Google Contact API v3 libraries to call the Contact Service.

The detailed specification of the implementation is shown in the following table.

Table 1: Specification of Implementation

Platform	Intel® i5 Quad-Core 2.5Gz, 16GB RAM
Operating System	Windows 7
Programming Language	Java
Runtime Environment	JRE 7
Network Model	TCP/IP Client/Server Model
IDE	Eclipse
Crypto Library	Java.Security
Hash Algorithm	SHA-1
Key Size and Security Parameter	80, 128 bit
Mode	Single Threaded, Parallel Mode
Input Set	Google Contacts: two 25,000-element sets

### 4 Results and Evaluation

In this section, we show the performance result of our implementation with Google Contacts. Both client and server programs run on the same laptop with an Intel® i5 quad-core 2.5Gz, 16GB RAM, Windows 7 platform and are developed in Eclipse IDE with JDK 1.7.0.45. In our implementation, we set  $k = \lambda$  to keep the false probability of a Bloom Filter to be as low as  $2^{-\lambda}$  and set  $m$  to be optimal value  $kn \log_e e$ . For example, at 80 bit security  $k = \lambda = 80$ , when  $n = 25,000$ ,  $m = 2,885,390$ . We measured the total running time of the protocol that starts from the client sending request and ends when client output the intersection. The time of fetching the contacts from the Google Accounts and the time of setting up sockets are excluded.

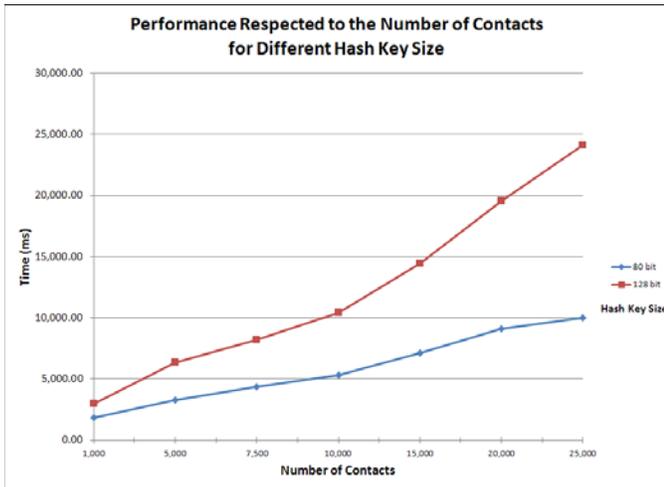


Figure 6: Performance of basic protocol respected to the number of contacts for different security key size

A. Performance

First, we show the performance in single threaded mode. We vary the size of contacts ( $n$ ) from 1,000 to 25,000 and the security ( $k = \lambda$ ) from 80 to 128 bit. The result is shown in Figure 6. As we can see, the running time increases almost linearly as the number of contacts increases at each level of security. For 25,000 contacts, it takes 10 seconds with 80 bit security and 24 seconds for 128 bit security.

Next, we show the comparison of performance between single-threaded and multi-threaded modes. We keep the key size to be 80 bit and vary the size of contacts ( $n$ ) from 10,000 to 25,000. The result is shown in the Figure 7. The total running time in multi-threaded mode is significantly less than in single-threaded mode as the number of contacts increases. For 80 bit security and 25,000 contacts, it takes 10 seconds in single-threaded mode while it only takes 6.3 seconds in multi-threaded mode.

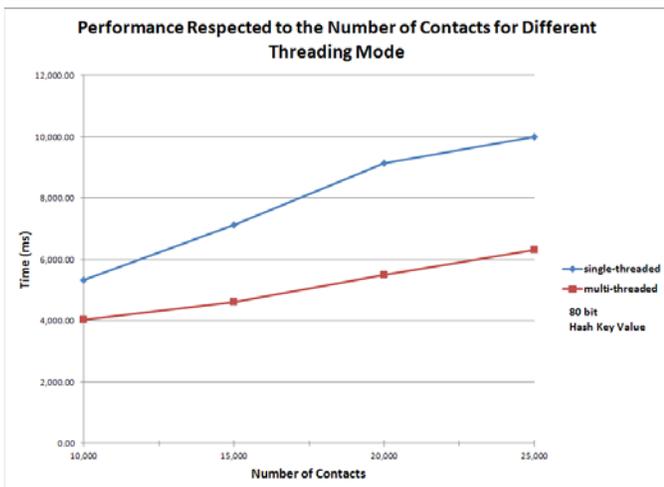


Figure 7: Performance of basic protocol respected to the number of contacts for different threading modes

In comparison to De Cristofarrot's RSA-OPRF protocol and Huang's Sort-Compare-Shuffle with Waksman Network protocol that are previously the fastest PSI protocols, Dong [3] showed that the approach of oblivious Bloom intersection is in orders of magnitude faster than these protocols.

B. Screenshot from Implementation

The Figures 8 and 9 demonstrate the user interface of our implementation in Oblivious Bloom Intersection. The interaction between client and server can be easily observed. Here is the process of computing the intersection:

1. The server and client connect to its corresponding Google Account we set up initially and get initialized to run in the environment.
2. The server will generate the symmetric key and send to the client.
3. The client and server will each encode their data set to Bloom Filter and Garbled Bloom Filter respectively.
4. The client and server then perform oblivious transfer and server will generate a new Garbled Bloom Filter for intersection for the client
5. The client will use the new GBF to query and compute the intersection
6. At the end, we allow client to send the set back to the server for verification purpose.

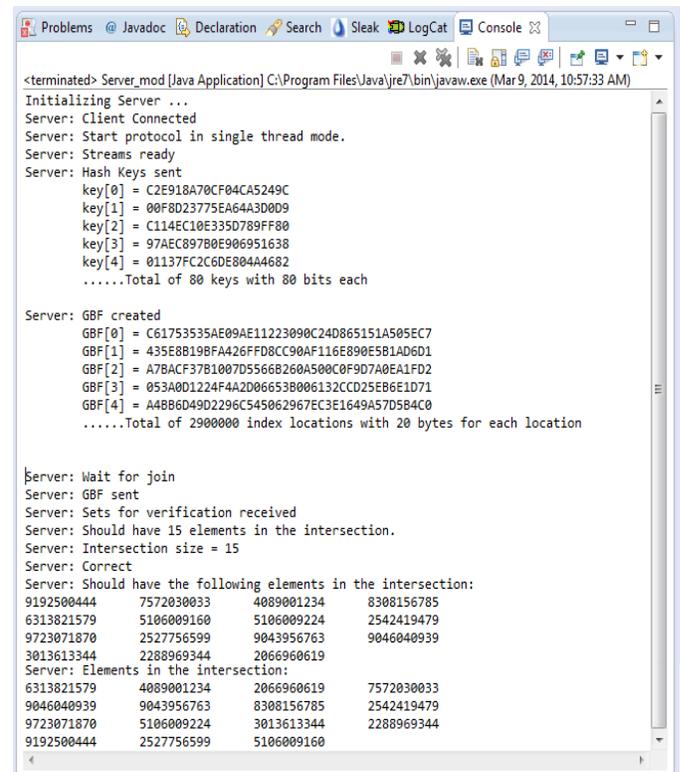
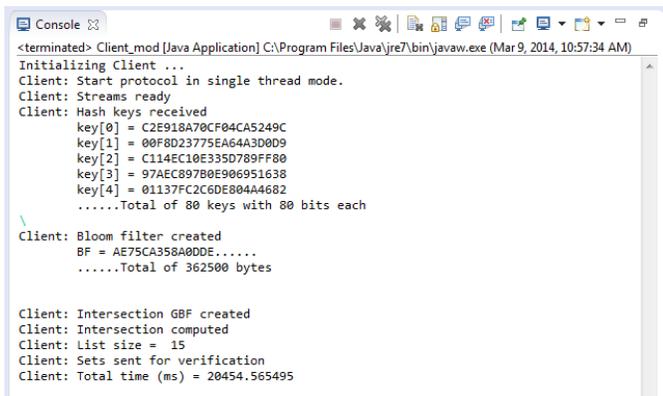


Figure 8: Interactive Server Interface



```

<terminated> Client_mod [Java Application] C:\Program Files\Java\jre7\bin\javaw.exe (Mar 9, 2014, 10:57:34 AM)
Initializing Client ...
Client: Start protocol in single thread mode.
Client: Streams ready
Client: Hash keys received
key[0] = C2E918A70CF04CA5249C
key[1] = 00F8D23775EA64A3D0D9
key[2] = C114EC10E335D789FF08
key[3] = 97AEC89780E906951638
key[4] = 01137FC2C6D0E804A4682
.....Total of 80 keys with 80 bits each

Client: Bloom filter created
BF = AE75CA358A0DDE.....
.....Total of 362500 bytes

Client: Intersection GBF created
Client: Intersection computed
Client: List size = 15
Client: Sets sent for verification
Client: Total time (ms) = 20454.565495

```

Figure 9: Interactive Client Interface

## 5 Conclusions

In this paper, we presented the practical use of a highly efficient and scalable PSI protocol based on the approach of oblivious Bloom intersection by implementing it in conjunction with Google Contacts. We also showed how this protocol can be easily integrated with cloud services like Google accounts to get contact information to be used as the input for both the client and the server. As explained by Dong, this protocol mainly depends on efficient symmetric key operations and these operations can be easily run in parallel. What makes the approach of oblivious Bloom intersection different than other protocols is mainly from its underlying data structure while other protocols are based on improving previous work with better algorithm. Its high performance is pretty encouraging and promising. In addition, it is suitable for large scale privacy preserving data processing. We hope that more applications can be developed with this protocol to provide secure and fast data processing

## 6 References

- [1] B. H. Bloom. *Space/time trade-offs in hash coding with allowable errors*. Commun. ACM, 13(7):422–426, 1970.
- [2] P. Bose, H. Guo, E. Kranakis, A. Maheshwari, P. Morin, J. Morrison, M. H. M. Smid, and Y. Tang. *On the false-positive rate of bloom filters*. Inf. Process. Lett., 108(4):210–213, 2008.
- [3] Changyu Dong, Liqun Chen, Zikai Wen, *When Private Set Intersection Meets Big Data: An Efficient and Scalable Protocol*, Page 4-15, 17-22. 2013
- [4] M.O. Rabin. *How to exchange secrets by oblivious transfer*. Technical Report TR-81, Harvard Aiken Computation Laboratory, 1981.
- [5] D. Beaver. *Correlated pseudorandomness and the complexity of private computations*. In STOC, pages 479–488, 1996.
- [6] Y. Ishai, J. Kilian, K. Nissim, and E. Petrank. *Extending oblivious transfers efficiently*. In CRYPTO, pages 145–161, 2003.

- [7] Google Contact API v3, Google.com. Retrieved Feb 18, 2014, from <https://developers.google.com/google-apps/contacts/v3/>
- [8] E. D. Cristofaro and G. Tsudik. *Practical private set intersection protocols with linear complexity*. In Financial Cryptography, pages 143–159, 2010.
- [9] Y. Huang, D. Evans, and J. Katz. *Private set intersection: Are garbled circuits better than custom protocols?* In NDSS, 2012.
- [10] M. J. Freedman, K. Nissim, and B. Pinkas. *Efficient private matching and set intersection*. In EUROCRYPT, pages 1–19, 2004.
- [11] L. Kissner and D. X. Song. *Privacy-preserving set operations*. In CRYPTO, pages 241–257, 2005.
- [12] J. Camenisch and G. M. Zaverucha. *Private intersection of certified sets*. In Financial Cryptography, pages 108–127, 2009.
- [13] C. Hazay and Y. Lindell. *Efficient protocols for set intersection and pattern matching with security against malicious and covert adversaries*. In TCC, pages 155–175, 2008.
- [14] E. D. Cristofaro, J. Kim, and G. Tsudik. *Linear-complexity private set intersection protocols secure in malicious model*. In ASIACRYPT, pages 213–231, 2010.
- [15] D. Dachman-Soled, T. Malkin, M. Raykova, and M. Yung. *Efficient robust private set intersection*. In ACNS, pages 125–142, 2009.
- [16] C. Hazay and K. Nissim. *Efficient set operations in the presence of malicious adversaries*. In Public Key Cryptography, pages 312–331, 2010.
- [17] S. Jarecki and X. Liu. *Fast secure computation of set intersection*. In SCN, pages 418–435, 2010.
- [18] S. Jarecki and X. Liu. *Efficient oblivious pseudorandom function with applications to adaptive OT and secure computation of set intersection*. In TCC, pages 577–594, 2009.
- [19] G. Ateniese, E. D. Cristofaro, and G. Tsudik. *(if) size matters: Size-hiding private set intersection*. In Public Key Cryptography, pages 156–173, 2011.
- [20] A. Shamir. *How to share a secret*. Commun. ACM, 22(11):612–613, 1979.
- [21] E. Bursztein, M. Hamburg, J. Lagarenne, and D. Boneh. *Openconflict: Preventing real time map hacks in online games*. In IEEE Symposium on Security and Privacy, pages 506–520, 2011.
- [22] Florent Chabaud, Antoine Joux. *Differential Collisions in SHA-0*. CRYPTO 1998. pp56–71
- [23] Eli Biham, Rafi Chen, *Near-Collisions of SHA-0*, Cryptology ePrint Archive, Report 2004/146, 2004 (appeared on CRYPTO 2004), IACR.org
- [24] Xiaoyun Wang, Hongbo Yu and Yiqun Lisa Yin, *Efficient Collision Search Attacks on SHA-0*, CRYPTO 2005
- [25] Google App Account Support, Google.com. Retrieved Feb 18, 2014, from <https://support.google.com/a/answer/1146409?hl=en>

- [26] Rudarakanchana, Nat. "*Big Data: Cat-And-Mouse Escalates On Privacy Concerns, As NRF Retail Conference Looms.*" International Business Times. [Http://www.ibtimes.com/](http://www.ibtimes.com/), 09 Jan. 2014. Web. 23 Feb. 2014.
- [27] Bloomberg, Jason. "Big Data Governance: 5 Lessons Learned From PRISM." *Big Data Governance: 5 Lessons Learned From PRISM.* [Http://www.baselinemag.com](http://www.baselinemag.com), 08 July 2013. Web. 23 Feb. 2014.
- [28] Kelly, Jeff. "Big Data Market Size And Vendor Revenues - Wikibon." *Big Data Market Size And Vendor Revenues - Wikibon.* Wikibon, n.d. Web. 23 Feb. 2014.

# A Comparative Evaluation of Intrusion-Alert Prediction Techniques

Kian-Moh Terence Tan, Neil C. Rowe, Christian J. Darken, and Farn-Wei J. Khong

DSO National Laboratories, Singapore, tkianmoh@dso.org.sg

U.S. Naval Postgraduate School, 1411 Cunningham Road, Monterey, CA 93943, United States, ncrowe@nps.edu (contact author)

U.S. Naval Postgraduate School, United States, cjdarken@nps.edu

Defence Science and Technology Agency, Singapore, kfarnwei@dsta.gov.sg

Track: Network Security

**Abstract**—Recognition of patterns of intrusion alerts can permit prediction of future alerts and thus earlier countermeasures. Previous work has focused on building attack models to enable prediction, but this approach cannot handle novel attacks. We tested six methods of predicting novel alerts in what appears to be the first systematic comparison of their relative merits. The techniques were evaluated on real non-simulated attacks, both deliberately staged ones and those recorded by a honeypot. The best performance was achieved by an approach which exploits partial structural matching between time-grouped sets of alerts and finds analogies in them. This approach is slow in its basic form, but we found several methods to improve its speed.

**Keywords:** intrusion detection, prediction, alerts, analogy, learning

## I. INTRODUCTION

Network intrusion-detection systems [1] such as Snort [2] screen incoming packets for suspicious activities. They often recognize attacks only after they occur because they need strong evidence to keep their false-alarm rate low. That means that damage may already have been done when serious alerts are generated. Contextual alert prediction could enable a more anticipatory network defense. Attacks often generate multiple alerts and early alerts can provide strong clues as to what will happen next. Often intrusion-detection experts can see an alert cluster and use their intuition. With automation of such early warnings, anticipated additional rule sets could be loaded into a signature-based intrusion-detection system, or prior probabilities modified for more accurate Bayesian inference in an anomaly-based system. In this paper we describe previous work in alert and time-series prediction, describe six approaches to online intrusion-alert prediction, and report comparative tests of their performance.

## II. PREVIOUS WORK

Some previous work [3-5] predicted intrusion-detection alerts by correlating already-observed ones and matching the correlated sets against libraries of known scenarios. This approach requires modeling attacks in advance, and fails with deliberately novel adversary behavior. [6-8] addressed

some limitations of this approach by using more general rules. [6] is probably the closest work to our approach to alert prediction with its approach of mining alert data to build attack graphs. [9] and [10] used Bayesian networks to learn correlations between alerts and conclusions about them using historical records, and [11] and [12] did something similar using Hidden-Markov models. Other work has addressed alert prediction for anomaly-based intrusion detection using state-transition models [13]. There appears to have been no work on reasoning by analogy for alert prediction, an idea important in artificial intelligence [14] that could inspire signature-based alert prediction that could learn dynamically as alerts come in.

Our previous work [15] developed a situation-learning approach to predicting relational time series in a role-playing game. Alert sequences can be seen as a time series of observations involving such properties as alert type, protocol, and IP addresses. Situation learning is an unsupervised online learning method that can flexibly learn with each new event.

## III. LEARNING AND PREDICTION OF RELATIONAL TIME SERIES

We give a summary of the situation-learning methods we tested for prediction of alerts. See [16] for more details.

### A. The Prediction Problem

The prediction problem is defined as  $a_1 a_2 \dots a_c \mid a_f$  where  $a_i$  are the alerts previously observed until the current alert  $a_c$ ,  $\mid$  is the prediction operator, and  $a_f$  is the next future alert.

Each intrusion alert (such as from Snort) contains a rule identifier (ID), a set of attributes, and a time of occurrence. A time series of intrusion alerts is a time-ordered sequence of intrusion alerts. The current intrusion alert is the most recent alert.

A prediction is an expected intrusion alert at a future time. A prediction may have restrictions on time, space, and object. In this paper we primarily use the time restriction that the prediction must be the next event and the object restriction that it be an alert for one particular site. A predicted intrusion alert is said to be correct if the next intrusion alert that arrives matches it exactly in everything except time.

A sample alert sequence is given in Table I. The prediction problem here is, given alert  $a_1$  to  $a_5$ , predict the next possible alert  $a_6$ .

TABLE I. A SEQUENCE OF FIVE ALERTS

Alert	time	Alert id	Source ip	Destination ip
$a_1$	17:56:07.236	384	80.135.185.28	63.205.26.73
$a_2$	17:56:07.251	384	80.135.185.28	63.205.26.70
$a_3$	17:56:07.253	384	80.135.185.28	63.205.26.69
$a_4$	17:56:09.18	1256	80.135.185.28	63.205.26.69
$a_5$	17:56:09.557	1002	80.135.185.28	63.205.26.69
$a_6$	?	?	?	?

### B. Situation Learning

In this work, a situation  $S$  is defined as a collection of intrusion alerts that occur within a fixed-duration time window. Other definitions have been explored but this is the simplest and enables better comparison of prediction techniques. The current situation  $S_c$  is the situation that contains the most current intrusion alert  $a_c$  as its last alert. A target alert  $a_t$  of a situation is the next alert that arrives after the situation. A situation-target tuple  $(S, \{a_t\})$  is a pair of a situation and a set of target alerts for that situation.

Situation learning is an unsupervised learning technique that learns a set of situation-target tuples from a time series by sliding the time window forward as each alert arrives. As an example, for Table I before alert  $a_5$  is observed where  $a_c = a_4$ , we form the current situation  $S_1 = \{a_1, a_2, a_3, a_4\}$  for a situation window duration of 2 seconds. When alert  $a_5$  is observed,  $a_5$  becomes the target of situation  $S_1$ . If the situation-target tuple  $(S_1, a_5)$  already exists in our knowledge base, we only need to update its statistics. The alert  $a_5$  can also form a new current situation with the previous alerts that are within the 2-second time window. Note that when we compare situations or targets, we ignore the relative time of occurrence of each alert except for the Markov model. Situation learning has low complexity and is capable of learning from a relational time-series in a high-entropy, non-stationary, or noisy environment. Variations on situation learning can infer more complex targets such as an attacker's intentions [17].

### C. Prediction Techniques

Given a set of situation-target tuples  $\{(S_i, t_i)\}$  where  $S_i \in S$ ,  $t_i \in T$ ,  $S$  is a set of all previously encountered situations,  $T$  is the set of distinct alerts, and  $S_c$  is the current situation, the prediction problem seeks a situation-target tuple such that some matching function  $M(S_i, S_c)$  is maximized. The  $t_i$  or an inferred  $t_i'$  can then be returned as the prediction. We tested six prediction techniques that cover reasonably well the space of possible prediction techniques.

Exact matching (EM) looks for a situation-target  $(S_i, t_i)$  such that  $S_i = S_c$ . Since there may be multiple matching tuples each with a different  $t_i$ , we select the target that has the highest probability given the situation  $S_i$ . In implementation, the EM technique represents the set of situation-target tuples  $\{(S_i, t_i)\}$  as a hash table, and searches the hash table for a situation that exactly matches

the current situation ignoring times. EM prediction works well in stationary environments in which the same entities are often encountered such as same rule ID, same IP addresses, etc. It has more difficulty improving its performance if there are many instances of alerts with different IP addresses, protocols, and rule ids, and cannot make prediction when all alerts have not been encountered before. EM is fast because it just requires a hash lookup.

EM performance can be improved by replacing all constants except the rule ID with variables to permit more frequent predictions. We call this the Variabilized Matching (VM) prediction technique. The matching of two situations becomes the problem of variable matching though unification. A unification is a set of variable bindings, e.g.  $\theta(\alpha, \beta) = \{a_1:b_1, a_2:b_2, \dots\}$  where a distinct variable  $a_i$  from situation  $\alpha$  is bound to a distinct variable  $b_j$  in situation  $\beta$ . Finding matches is equivalent to a graph isomorphism problem. Two situations match when there is a graph bijection (one to one unification) among the variables in both situations. VM prediction should find more matches than EM prediction when the number of new instances is higher because its matching criterion is less strict. The worst case time complexity of the variable matching technique is  $O(2^{n \log(n)})$  for two situations with same number of  $n$  constants [18]. But there exist many heuristics to speed up the graph isomorphism process by avoiding comparisons of two situations that are structurally different, such as number of constants (distinct IP addresses), predicates (rule ids), in/out degrees (source and destination), etc.

The structural matching problem encountered by VM's graph isomorphism process can be relaxed to partial matching. We describe four such methods. First, Bayesian methods are popular ways to predict events, and were previously applied to intrusion event prediction in [9] and [10]. One Bayesian approach computes the probability of occurrence for each encountered target alert, conditioned on the set of alerts in the current situation. During learning, it constructs one Bayesian network for each distinct target alert. Each distinct alert is the parent node in its respective Bayesian network, while the alerts that appeared in the preceding situations are the child nodes. Since we have one Bayesian network for each distinct target, we have a multiple simple Bayesian networks prediction technique (MSB). To formalize this, let  $T$  be a set of distinct alerts and  $S$  be a set of situations. We assume alerts are conditionally independent to allow partial matching through the naive Bayesian formalism. The probability of  $T$ , conditioned on  $S$  is:

$$\begin{aligned} P(T|S) &= P(T = t_i | S = s_c) \\ &= P(T = t_i | s_1 = p_1, s_2 = p_2, \dots, s_n = p_n) \\ &= \frac{P(T = t_i) \prod_{k=1}^n P(s_k = p_k | T = t_i)}{P(s_1 = p_1, s_2 = p_2, \dots, s_n = p_n)} \end{aligned}$$

where  $n$  is the number of alerts in the current situation,  $a_j$  are alerts in the current situation.

To allow partial matching with MSB, where only a subset of events in a situation match a situation seen before, Laplace smoothing is used by adding 1 to the counts. In

extreme cases, this does allow targets to be selected with no alerts in the current situation, but this should be rare. The time complexity of the MSB technique is  $O(n^2)$  where  $n$  is the number of distinct alert events, but this only occurs when each distinct percept is a child of very other distinct percepts, which would be very unusual.

Simple Bayesian mixture (SBM) is an improvement over MSB that can learn functions such as exclusive-OR. We were curious whether it could improve upon MSB. SBM contains probability mixture densities, constructed by normalizing a linear combination of two or more Bayesian networks probability densities having the same parent and child percepts. In MSB, we have one distribution for one parent-child network. In SBM, the same distribution for one parent-child network is divided into several weighted distributions. SBM is implemented using the Estimate and Maximize algorithm.

The abovementioned prediction techniques ignore the order of events in the situations. Variable-order Markov model (VOMM) are the most popular way to predict ordered sequences, and were previously applied to alert prediction in [11] and [12]. VOMM is an extension to the Markov chain models in which a variable order is used in place of a fixed order. This means that the prediction of the target alert depends on a varying number of the most recent alerts such that the matching function thinks is best.

While the Bayesian and the Markov techniques can handle partial situation matching, the abovementioned techniques except for VM can only predict target alerts whose exact combination of parameters have been encountered before. However, the VM technique requires full graph isomorphism, which is rare in non-stationary domains. Its generality can be improved with partial variabilized matching (PVM), implemented as a subgraph isomorphism process instead of a full graph isomorphism. This idea is related to reasoning by analogy, and our approach to implementing it was inspired by the psychological theory of single-scope blending [19]. Since solving a subgraph isomorphism problem is central in PVM, we explored heuristic algorithms to do it with a time complexity of  $O(n^2)$ . The idea is to generate a score for all possible pairings between attributes in two situations based on similarities in constants, in-degrees, and out-degrees. We do a lexical sort of the pairs and greedily choose the best remaining pairs where none of the constants in the pair have been selected previously.

#### IV. EXPERIMENTAL SETUP

##### A. Experiment Methodology

In the experiments that follow, we collected alerts sequences and ran the prediction algorithms on them. Each prediction technique returned their best prediction on the next alert to be triggered. The predicted alert is said to be correct if the next percept matches on rule ID, protocol, source IP address, and destination IP address. Note that time prediction is not included. A given situation can have multiple possible predictions, and their number varies considerably over techniques. To make comparison fair,

each technique was compared on their best guess for the next event.

We used two metrics to measure performance of prediction: accuracy and computation time. Prediction accuracy was computed as  $c/n$  where  $n$  is the number of alerts received and  $c$  is the count of correct predictions. Computation time was the time to predict all alerts in the relational time series.

We compared the metrics in two kinds of experiments: explicitly generated attacks and live network traffic on a honeypot.

##### B. Experiment 1: Explicitly Generated Attacks

Experiment 1 used the penetration-testing tool Pytbull (pytbull.sourceforge.net) to generate attacks on an isolated two-machine network [20]. Attacks went from a host machine to a victim machine, with Snort for intrusion detection on the victim with port mirroring, as described in [20]. We used VMWare Player 4 to implement three virtual machines on the attacker and three virtual machines on the victim, as well as to facilitate restoration of system state. The victim virtual machines had 512 MB of memory and 14 GB of disk space.

Pytbull provided about 300 tests in 11 testing modules of badTraffic, bruteForce, clientSideAttacks, denialOfService, evasionTechniques, fragmentedPackets, ipReputation, normalUsage, pcapReplay, shellCodes, and testRules. We excluded some client-side attacks and pcap-replay attacks that our previous experiments had shown could not be detected by Snort [21]. We used Snort 2.9 with the free public rules since they are a de facto standard, and ran it on a Linux machine. We also put on the victim the Apache2 Web services, Vsftpd, SSH, and the BackTrack security analysis tool.

##### C. Experiment 2: Live Network Traffic on a Honeypot

Tests with staged attacks may not be representative of real-world threats. Thus our primary experiments were with alert sequences obtained from the honeynet setup described in [22]. This honeypot was situated outside our school's firewall on a network provided by a commercial Internet service provider. Three sequences were collected in 2012 from Internet traffic trying to connect to the honeynet. The honeynet had five virtual machines. Some active responses to attackers were provided as described in [22], but they almost entirely affected the quantity and not the content of the traffic, as discussed there. A summary of the first two and primary datasets is given in Table II. The counts on distinct alerts ignore the time attribute of the alerts. The repetition rate is the ratio of distinct alerts to the entire alerts. We use entropy to measure the uncertainty of the occurrences of alerts, computed as:

$$-\sum_{i=1}^n p(x_i) \log_2(p(x_i))$$

where  $p(x_i)$  is the probability of alert  $x_i$ .

TABLE II. ALERT COUNTS OF DATASETS 1 &amp; 2

	Total Alerts	Distinct Alerts	Repetition Rate	Entropy	Duration
DataSet1	6482	1590	4.08	7.46	8 weeks
DataSet2	9619	4304	2.23	11.08	2 weeks

A sequence of alerts will have a lower entropy if it contains alerts that have a high frequency of reoccurrence (e.g., many ping alerts). A low-entropy alert sequence is expected to be easier for prediction task since its repetition rate is high. A high-entropy sequence will be harder to predict since it has high count of new alert encounters.

We extracted the rule ID, protocol, source IP address, and destination IP address. A sample sequence is given in Table III. To more easily apply subgraph isomorphism algorithms, each alert was converted into an arity-2 representation. The time window was fixed at 0.1sec to keep the situations small, although a bigger window may better model complicated attack profiles.

TABLE III. SAMPLE ALERT SEQUENCE FROM SNORT NETWORK INTRUSION-DETECTION SYSTEM

Data & Time	Alert ID	Protocol	Source IP	Destination IP
12/02/11-17:07:46.196272	402	ICMP	63.205.26.70	75.126.23.106
12/02/11-17:10:25.653574	449	ICMP	154.54.26.66	63.205.26.89
12/02/11-17:10:25.656963	449	ICMP	154.54.26.66	63.205.26.89
12/02/11-17:29:58.526864	2924	TCP	78.45.215.210	63.205.26.80
12/02/11-17:30:00.261283	2924	TCP	78.45.215.210	63.205.26.80
12/02/11-17:30:00.261756	2924	TCP	63.205.26.77	78.45.215.210
12/02/11-17:30:01.768373	2924	TCP	78.45.215.210	63.205.26.80
12/02/11-17:30:01.768855	2924	TCP	63.205.26.77	78.45.215.210
12/02/11-17:30:06.248677	2924	TCP	78.45.215.210	63.205.26.80
12/02/11-17:30:06.249809	2924	TCP	63.205.26.77	78.45.215.210
12/02/11-17:30:13.883712	2924	TCP	78.45.215.210	63.205.26.80
12/02/11-17:30:13.884162	2924	TCP	63.205.26.77	78.45.215.210

To give a sense of the data, Table IV shows the histogram of the full set of alert descriptions for alerts occurring in the datasets. We did nothing to encourage particular types of attacks, so mostly we saw quite-familiar Internet attacks. This was to demonstrate our prediction methods could help in defending ordinary computer systems.

## V. RESULTS

### A. Experiment 1: Explicitly Generated Attacks

In the first experiment, the overall average prediction accuracies for all six prediction techniques for three attackers, three victims, and four prediction methods were similar at around 80%. PVM lead in prediction accuracy, followed by VOMM, MSB, VM, SBM and EM [20]. (Results with one attacker and one victim were similar, as were results with staggered attacks.) We grouped alerts into groups by alert type and computed the entropy for each group. The alerts were then sorted into six ranges of entropy as shown in Figure 1. The F-score is the harmonic mean of the precision (fraction of correct predictions of all predictions) and recall (fraction of correct predictions of all alerts). EM and VM were omitted because they are special cases of PVM.

TABLE IV. HISTOGRAM OF ALERTS IN THE TWO DATASETS

Attack	Count
ICMP PING	3250
ICMP Destination Unreachable Host Unreachable	2488
POLICY Outbound Teredo traffic detected	1604
SHELLCODE x86 inc ecx NOOP	490
SHELLCODE x86 NOOP	396
SPECIFIC-THREATS ASN.1 constructed bit string	219
NETBIOS DCERPC NCACN-IP-TCP ISystemActivator RemoteCreateInstance attempt	79
ICMP traceroute	50
ICMP PING *NIX	44
RPC portmap mountd request UDP	38
ICMP Destination Unreachable Protocol Unreachable	19
NETBIOS DCERPC NCACN-IP-TCP umpnpgmgr PNP_QueryResConfList attempt	12
WEB-IIS WEBDAV nessus safe scan attempt	6
ICMP PING Flowpoint2200 or Network Management Software	2
POP3 PASS format string attempt	1
WEB-CLIENT Microsoft wmf metafile access	1
NETBIOS SMB-DS repeated logon failure	2786
ICMP Echo Reply	1977
ICMP Destination Unreachable Port Unreachable	1555
SHELLCODE x86 inc ebx NOOP	454
ICMP Time-To-Live Exceeded in Transit	282
NETBIOS DCERPC NCACN-IP-TCP srvsvc NetPathCanonicalize overflow attempt	144
NETBIOS DCERPC NCACN-IP-TCP lsassDsRolerUpgradeDownlevelServer overflow attempt	60
ICMP PING Windows	48
ICMP PING BSDtype	44
WEB-CGI awstats.pl configdir command execution attempt	22
DNS SPOOF query response with TTL of 1 min. and no authority	13
ICMP PING Sun Solaris	8
SHELLCODE base64 x86 NOOP	6
SPYWARE-PUT Trackware funwebproducts mywebsearchtoolbar-funtools runtime detection	2
NETBIOS Microsoft Windows SMB malformed process ID high field remote code execution attempt	1

We observe that when the entropy was less than 3, the F-scores for all prediction techniques are similar. This is because when the alerts repeat frequently, prediction is easier. However, when the frequency of occurrence reduces (or entropy increases), prediction performance began to differentiate, with PVM having the most accurate prediction accuracy. Paired T-tests comparing F-scores supported the hypothesis that PVM was the best with a 95% confidence level. However, the mean time to generate a prediction by PVM was 0.2207 seconds, whereas that for SBM was 0.0389 seconds, VOMM 0.0108 seconds, and MSB 0.0170

seconds. Similar results were obtained with one-at-a-time attacks and single-attacker single-victim setups.

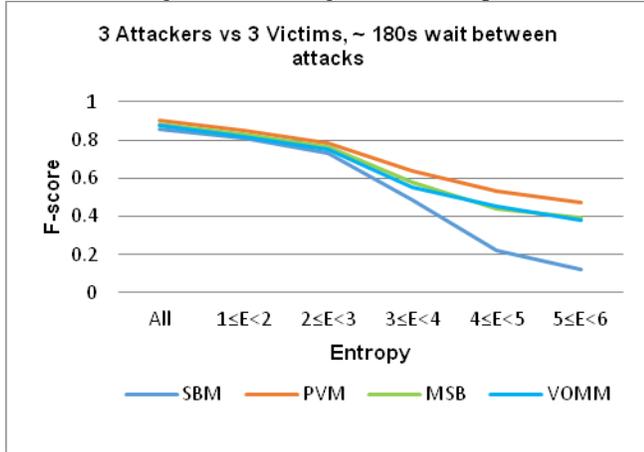


Figure 1. Results of Experiment 1 with a penetration-testing tool as a function of entropy range

**B. Experiment 2: Live Network Traffic on a Honeypot**

The engineered attacks in the previous experiment had many low-entropy alerts because there were only six IP addresses and limited types of attack. So we also tested data from a honeypot [16].

Figure 2 shows the variation of the average prediction accuracies over alert events for the honeypot data. After initial variations, predictors' performance reached steady states. For data 1 (two weeks of alerts) the PVM achieved a prediction accuracy of around 70% while most settled around 55% and SBM was around 35%. For dataset 2 (eight weeks of alerts), The PVM achieved a steady state prediction accuracy of around 48%, VM 37%, VOMM and MSB 28%, and EM 25%. SBM was excluded from this experiment onwards because its prediction accuracies were poor in the previous experiments and its run time was significantly longer than the other techniques. It appears that its poor performance is due to the non-stationary environment in which data is insufficient to create a mixture of more than one distribution.

These percentages are considerably less than those of intrusion-detection systems, as alert prediction is a harder problem. But even with imperfect success, prediction can tell intrusion detection what to anticipate and this can save it significant effort and make its conclusions more reliable.

We examined the types of errors made by each method and concluded that the success of PVM was due to its greater flexibility in making predictions with limited data. The limited numbers of identical alerts in our datasets did not provide VOMM and SBM with sufficient statistics on events to make reliable inferences.

To study the effects of entropy on prediction, we partitioned a third dataset that had over a year of data with 200,000 alerts into 10 sequence of 20,000 alerts. To study the effect of shorter alert sequence (higher entropy), we also partitioned the first part of the data into 150 sequences of 100 alerts (Figure 3). PVM performed the best on both short

and long sequences, but its advantage over the others was less on longer sequences.

Table V illustrates a key advantage of PVM: It can predict single-occurrence alerts that Markov and Bayesian methods cannot. For instance, row one of Table IV says of 2800 alerts that occurred once, PVM correctly predicted 618 even before seeing any instance of them. Table VI shows some examples of Snort alerts that occurred once and were correctly predicted by PVM but not the other methods.

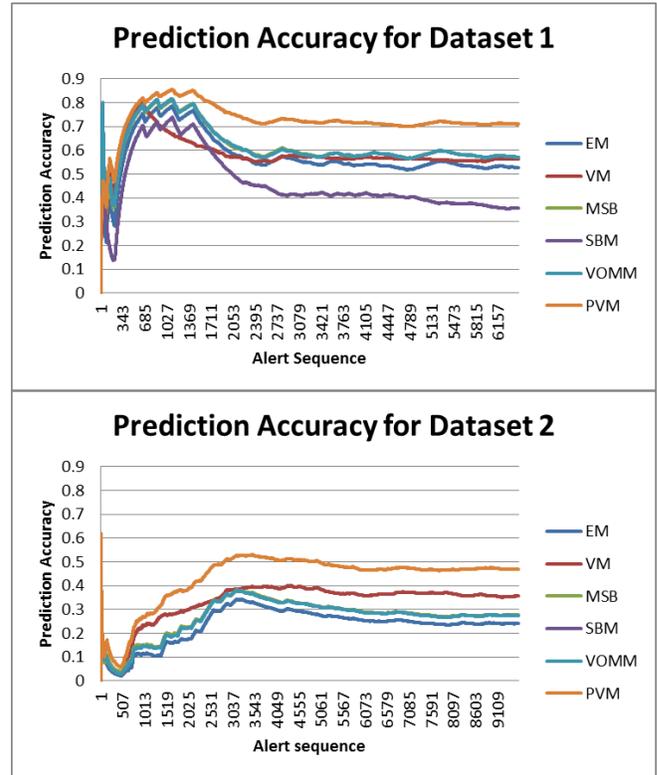


Figure 2. Dataset 1&2: Prediction accuracy with respect to the number of events (percepts). SBM for Dataset 2 is excluded.

As a more detailed example to show why this is so, suppose that our knowledge base contains only one situation target tuple of (alertID, source, destination) where  $S_1 = \{(384, 80.135.185.28, 63.205.26.69), (1256, 80.135.185.28, 63.205.26.69)\}$  and  $T_1 = \{(1002, 80.135.185.28, 63.205.26.69)\}$ . Let the current situation be  $S_c = \{(384, 80.135.185.29, 63.205.26.70), (1256, 80.135.185.29, 63.205.26.70)\}$ . Suppose that the system has encountered five alerts (384, 80.135.185.28, 63.205.26.69), (1256, 80.135.185.28, 63.205.26.69), (1002, 80.135.185.28, 63.205.26.69), (384, 80.135.185.29, 63.205.26.70) and (1256, 80.135.185.29, 63.205.26.70). Note that the  $S_1$  and  $S_c$  have some commonalities; both contain one alert of ID 384 and one alert of ID 1256. Furthermore, the source IP address 80.135.185.28 of  $S_1$  can be associated with the source IP address 135.185.29 of  $S_c$  since both IP addresses appear as source IP addresses in two alerts in their respective situations. Similarly, 63.205.26.69 can be associated with 63.205.26.70. We want to predict the next alert from the target alert by substituting the target IP

addresses with those found in the current situation after the association process. We want the prediction to be (1002, 80.135.185.29, 63.205.26.70), something not observed before.

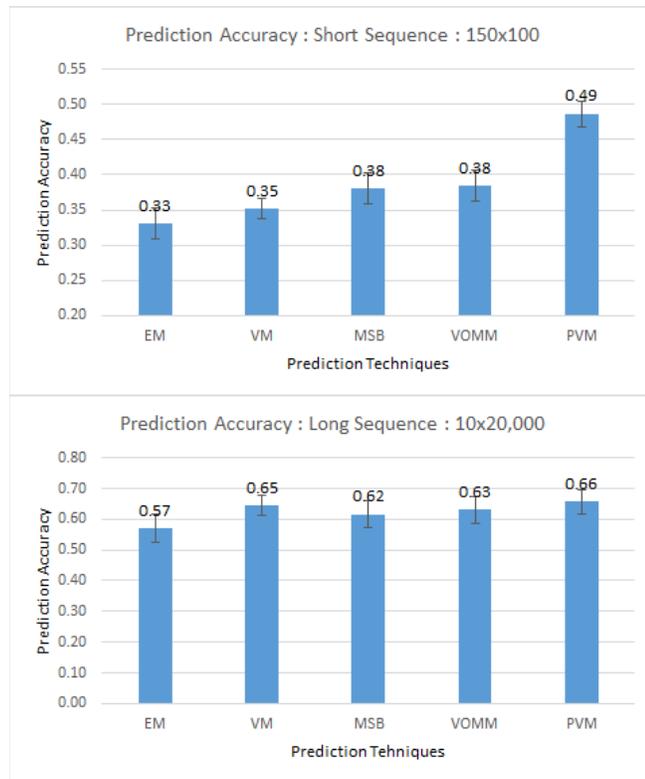


Figure 3. Average prediction accuracies for shorter versus longer sequences on the third dataset

TABLE V. COMPARATIVE PREDICTION PERFORMANCE AS A FUNCTION OF THE NUMBER OF OCCURRENCES OF AN ALERT

Frequency of occurrence	Number of Alerts	PVM Detects	MSB Detects	VOMM Detects
1	2800	618	0	0
2	2403	1847	707	574
3	174	107	107	62
4	285	250	218	250
5	29	10	11	12
6	56	46	35	38
7	22	12	6	9
8	18	14	10	12
9	11	8	7	8
10	11	8	8	8

The main disadvantage of PVM was that it was slower than the other methods since it requires subgraph matching, and its speed disadvantage increased with the number of situations it stores. Hence, it is imperative to keep either the situation size or the time window small. One simple way to improve computation time was to eliminate old or low-utility situation-target tuples. Figure 4 illustrates the effect of eliminating all but the best 10% of the number of possible situations from consideration, including only the most recent and occur more often. Experiments showed that

the prediction accuracy did not decrease significantly but the computation time was reduced significantly. Decreasing the maximum situation size from 2000 to 200, for instance, only decreased accuracy from 0.55 to 0.54 while increasing speed by a factor of 10.

TABLE VI. SNORT ALERTS PREDICTED BY SSB BUT NOT BY MSB, SBM, AND VOMM

	Protocol	Source	Destination
12710, ASN.1 constructed bit string	TCP	external	honeypot
402, Destination Unreachable Port Unreachable	ICMP	honeypot	external
3397, DCERPC NCACN-IP-TCP ISystemActivator RemoteCreateInstance attempt	TCP	external	honeypot
408, Echo Reply	ICMP	honeypot VM 1	external
402, Destination Unreachable Port Unreachable	ICMP	honeypot	external
1390, x86 inc ebx NOOP	TCP	honeypot VM 2	external
384, PING	ICMP	external	honeypot VM 3

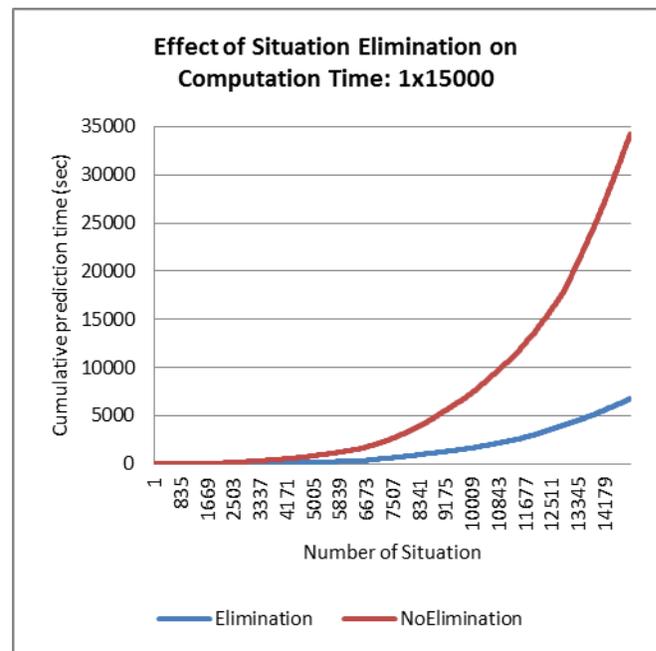


Figure 4. Effect of situation elimination on computation time with number of events encountered

There are also many heuristic methods for improving the speed of subgraph matching since it is needed in many applications of artificial intelligence. We got best results by a heuristic method of ranking match options by the strength of their similarity so that the most-similar ones were tried first, then conducted something close to a best-first search. It was not a strict best-first search because we got better performance when we included additional "second-best"

options on the best-first agenda. The final performance improvement using this idea alone was around a factor of 10, with significantly more improvement for high-entropy data.

The criterion for evaluating prediction accuracy so far has been stringent: Whether the next percept matches the prediction. If we relax the criterion to allow prediction of the percept within a time window, testing of the prediction would seem fairer in busy network traffic. We conducted experiments with a window of two standard deviations beyond the mean time of next occurrence of a percept. Results showed that prediction performance was surprisingly worse by around 1%, with VM 5% worse. This appeared to be due to the large variety of distinct alerts, for which it takes time to obtain sufficient occurrences of an alert to derive a good estimate of the statistical time interval of occurrences. Therefore, the intervals are usually zero, which require predicted alerts to occur exactly at the exact time of occurrence. As a result, prediction within an output time window becomes even more stringent than with next-event prediction.

## VI. CONCLUSIONS

Prediction of alerts can provide earlier responses to cyber-attacks and enable better recognition of attacks that are truly novel. Relational time-series learning and prediction methods provide useful new ideas for predicting intrusion alerts. They offer a special advantage in that they do not require an attack model in advance, which gives them flexibility in real-time response to traffic. In this work we tested several approaches with Partial Variable Matching (PVM) performing the best. On short alert sequences or those with high entropy, the superiority of PVM was more pronounced. PVM (and VM as well) also have the advantage of being able to predict new alerts, which can be useful in the dynamic world of frequent new cyber threats.

## ACKNOWLEDGMENT

This opinions expressed are those of the authors and do not represent the Singapore or U.S. governments.

## REFERENCES

- [1] Trost, R.: Practical Intrusion Analysis. Upper Saddle River, NJ: Addison-Wesley, 2010.
- [2] Roesch, M.: Snort—Lightweight Intrusion Detection for Networks. In: Proceedings of LISA '99: 13th Systems Administration Conference, 1999, pp. 229-238.
- [3] Ning, P., Cui, Y., Reeves, D.: Constructing Attack Scenarios through Correlation of Intrusion Alerts. In: Proceedings of the 9th ACM Conference on Computer & Communications Security, 2002, pp.245–254.
- [4] Cuppens, F., Mieke, A.: Alert Correlation in a Cooperative Intrusion Detection Framework. In: Proceedings of the IEEE Symposium on Security and Privacy, 2002, pp. 202–215.
- [5] Cheung, S., Lindqvist, U., Fong, M.: An Online Adaptive Approach to Alert Correlation. In: Proceedings of the 7th International

- Conference on Detection of Intrusions and Malware, and Vulnerability Assessment, 2010, pp.153–172.
- [6] Li, Z.-T., Wuhan, J., Wang, L., Li, D.: A Data Mining Approach to Generating Network Attack Graphs for Intrusion Prediction. Proc. Fourth Intl. Conf. on Fuzzy Systems and Knowledge Discovery, Haikou, CN, pp. 307-311, August 2007.
- [7] Sundaramurthy, S., Zomlot, L., Ou, X.: Practical IDS Alert Correlation in the Face of Dynamic Threats. In: Proceedings of the International Conference on Security and Management, 2011, available at <http://130.203.133.150/viewdoc/summary?doi=10.1.1.218.535>.
- [8] Templeton, S., Levitt, K.: A Requires/Provides Model for Computer Attacks. In: Proceedings of the 2000 Workshop on New Security Paradigms, 2000, pp. 31–38.
- [9] Qin, X.: A Probabilistic-Based Framework for INFOSEC Alert Correlation. Ph.D. dissertation, Georgia Institute of Technology, Atlanta, Georgia, USA, 2005.
- [10] Tabia, K. and Leray, P.: Handling IDS' Reliability in Alert Correlation. Proc. Intl. Conf. on Security and Cryptograph, Athens, GR, pp. 1-11, July 2010.
- [11] Zan, X., Gao, F., Han, J., and Sun, Y.: A Hidden Markov Model Based Framework for Tracking and Predicting of Attack Intention. Proc. Intl. Conf. on Multimedia Information Networking and Security, Hubei, CN, pp. 498-501, November 2009.
- [12] Fava, D., Byers, S., and Yan, S., Projecting Cyberattacks through Variable-Length Markov Models. IEEE Transactions on Information Forensics and Security, Vol, 3, No. 3, pp. 359-369, September 2008.
- [13] Soule, A., Salamatian, K., and Taft, N.: Combining Filtering and Statistical Methods for Anomaly Detection. Proc.5th ACM SIGCOMM Conference on the Internet, pp. 331-344, 2005.
- [14] Winston, P.: Learning and Reasoning by Analogy. Communications of the ACM, Vol. 23, No. 12, pp. 689-703, 1980.
- [15] Tan, K., Darken, C.: Learning and Prediction in Relational Time Series. In: Proceedings of Behavior Representation in Modeling & Simulation, 2012, pp.100-107.
- [16] Tan, T.: Learning and Prediction of Relational Time Series. Ph.D. dissertation, U.S. Naval Postgraduate School, Monterey, California, USA, March 2013.
- [17] Qin, X., and Lee, W.: Attack plan Recognition and Prediction Using Causal Networks. Proc. 20th Computer Security Applications Conference, pp. 370-379, December 2004.
- [18] Babai, L.; Codenotti, P.: Isomorphism of Hypergraphs of Low Rank in Moderately Exponential Time., FOCS '08: Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science, IEEE Computer Society, pp. 667–676, 2008.
- [19] Fauconnier, G., Turner, M.: The Way We Think: Conceptual Blending and the Mind's Hidden Complexities. New York: Basic Books, 2002.
- [20] Khong, F.: Performance Assessment of Network Intrusion-Alert Prediction. M.S. thesis, Naval Postgraduate School, Monterey, California, USA, 2012.
- [21] Albin, E., Rowe, N.: A Realistic Experimental Comparison of the Suricata and Snort Intrusion-Detection Systems. In: Proc. of Eighth International Symposium on the Frontiers of Information Systems and Network Applications, Fukuoka, Japan, March 2012.
- [22] Frederick, E., Rowe, N., Wong, A.: Testing Deception Tactics in Response to Cyberattacks. In: Proceedings of the National Symposium on Moving Target Research, Annapolis, June 2012, <http://cps-vo.org/node/3711>.

# Email Encryption

## Discovering Reasons Behind its Lack of Acceptance

Kendal Stephens LaFleur

Department of Computer Science  
Sam Houston State University  
Huntsville, TX, United States  
kks016@shsu.edu

Lei Chen

Department of Computer Science  
Sam Houston State University  
Huntsville, TX, United States  
shsu.chen@gmail.com

**Abstract**— Email encryption is a critical component of data security and privacy, yet many people fail to use it. Prior studies have been performed to find ways to improve existing encryption methods and to develop new ones, in the attempt of providing a program that more people will utilize. Despite these efforts, email encryption is still not widely adopted. Our study uses a survey to collect data about users' views of email encryption and provide us with insight on why many choose not to use it. After analyzing these results, we found that this is not only due to a lack of usability of the encryption programs, but it's also due to the fact that many people do not fully understand encryption. There is a substantial lack of knowledge of what exactly should be encrypted, how to operate encryption programs, and the many threats associated with unencrypted emails. Our results are based on the responses of thirty people answering a multiple-choice survey we designed, made up of ten questions dealing with email encryption. All survey participants came from a variety of different backgrounds and careers. To the best of our knowledge, this research study is unique and our findings represent the true viewpoints of our participants.

**Keywords**— *email encryption; security; privacy; education*

### I. INTRODUCTION

In today's fast-paced, technology-centered world, email encryption is becoming more important than ever. Incidences of data leakage and security breaches happen every day, and many of these originate from unencrypted emails. According to a recent study done by Cranfield University [6], an average person receives 63 emails per day and sends 34 emails per day. This demonstrates the significant role that email plays in people's daily lives, and the immense amount of information that is transported this way. People rely heavily on email as a form of communication for both personal and work-related matters. Because of this, email encryption needs to be used to ensure the security and privacy of any sensitive data or private information sent through emails. PGP, S/MIME, and other encryption programs exist, yet many people fail to use them. Because of this, we were motivated to perform this study to determine why exactly these programs are not being used when they play such a critical role in data security.

Our research study gathers data from common email users to gain insight on their encryption habits and their thoughts on the subject. We collected the data by creating a survey consisting of ten multiple-choice questions, and then we sent it out to thirty people for completion. We wanted to keep the survey concise and to the point. We wanted the questions to be

easy to understand so that even the more inexperienced technology users would be able to answer them truthfully and avoid confusion. We kept the survey at ten questions because we feared that making it too lengthy could cause participants to lose interest. We also felt that this study could be expanded further in the future, so for the information that we wanted to look into at this time, the ten questions provided us with the data we needed. This study stands out from prior research in the way that we focused completely on the views and opinions of users, in order to determine exactly why they choose not to use encryption techniques. We learned from our results that many participants feel that encryption programs are difficult and frustrating to use. Many of them are also very uneducated on email encryption, including how to use it and why it is important. This seems to have a major impact on the choice to send unencrypted emails. Our study first analyzes prior work in the area of email encryption, then discusses the methodology and details of our research, and then analyzes results and draws conclusions based on the findings.

### II. BACKGROUND

Many studies have been conducted in the area of email encryption to propose and try ways to improve the existing methods. We analyzed prior work in the field to gain more knowledge on what has been studied in the past, what conclusions have been made, and what areas still need further focus and investigation. We wanted to see what valuable findings other researchers had discovered, as well as where other studies had fallen short and needed to be extended upon further. We felt that a strong understanding of past research would help us better direct our own study and see what contributions we should aim to provide.

In a study by Poole et al. [8], the authors discuss how users employ different computer tools and which various characteristics of a technological tool or program affect its usage. They discuss in great detail the use of RFID technologies, and then move on to email encryption. They address how the lack of use of email encryption programs is commonly due to these applications having poor usability. They also discuss how many people feel that regular use of encryption in emails is abnormal and unnecessary. Then they explore how email encryption is usually associated with high importance of a message, and people don't use it or feel that it's necessary for smaller scale or less important emails. They

make the conclusion that the lack of email encryption being used today is more due to non-functional aspects than to actual technical difficulties affecting usability. However, they don't offer any solutions for this. This paper is weak in that it doesn't provide any proof or data to back up the assertions, weakening its impact. This influenced us to use a survey and gather data from actual email users in order to substantiate our findings and conclusions.

In another study, Gabrielson and Levkowitz [4] discuss the need for more user-friendly encryption tools. They offer a solution that involves a security pattern based upon existing technologies and ideas. Their primary goal is to create a trusted encrypted channel that is easy to use. They discuss their definition of "trust" and the requirements of their development. A main necessity is that minimal interaction be needed between the user and the application. They discuss their proposed solution in detail, covering functionality and technical aspects. Using their guide for future improvements at the conclusion of the paper, this study could be extended and work could be done to expand their application development. The authors only focus on two different use cases, so improving the proof-of-concept is definitely needed in future work. Payne and Edwards [7] also look at security applications and flaws in their design. However, they don't really take all of their conclusions about security designs and apply them to email encryption to show how it can be improved. This research could definitely be extended upon by looking at the successful security tools and what made them effective and usable, and then discussing how those same aspects could be used in email encryption applications to make them more popular among users.

In another study, Kainda et al. [5] develop a security and usability threat model. They identify main factors of usability and security by looking at prior studies, and categorize them into six different groups of security topics. One of these groups is email encryption. They discuss how users' understanding and knowledge of the application plays a huge role in email encryption. They explain how their threat model can be used to analyze different security scenarios. This is a unique study because it takes on a different approach to security usability by creating the threat model, and it provides a great deal of detail and clear explanation. One weakness is that while it does explain how this model can be applied to a scenario, it doesn't provide an example of actually doing so. It could be improved upon by actually applying this to a specific security scenario, and putting specific focus on how it can be used to analyze and improve email encryption methods.

Abdalla et al. [1] introduce a development called identity-based encryption with wildcards (WIBE) in their research study. This can be used to send encrypted emails to groups of recipients. The authors discuss the history of this concept, which was first introduced in 1984. They focus on providing an encryption method to be used when sending emails to multiple people of organizational hierarchies, rather than just one single person. They provide details about the syntax and security aspects of this encryption scheme. They go into

details on numerous other encryption schemes explored in prior studies that are the basis for WIBE. While this paper provides an immense amount of information, it can be hard to follow with all of the many algorithms given that can distract from the real meaning of the study. It is difficult to understand what these authors are actually contributing. However, this study influenced our study because we learned that we needed to make our work and its contributions clear and concise, so that other researchers in the field can use it to gain knowledge and expand upon in their own studies.

Another research study conducted by Dingledine and Mathewson [3] discusses the network effects of usability on privacy and security. These authors address how email encryption requires all participants, including the sender and any recipients, to work together and have an understanding of the process. They list the many ways that difficult to use programs can impair security. They also discuss the issue of privacy and data confidentiality, making usability even more critical when sending emails. This is where anonymizing network comes in, which is a technique that basically hides users among users so that they cannot be identified. The authors provide multiple case studies to help readers better understand this. They make conclusions that the success of any security application relies on the behavior of users, and work on network anonymity still needs further work and experimentation. Their study makes contributions by demonstrating the usefulness of anonymity and drawing attention to its need for better design and usability. This study could be taken a step further by exploring ways to improve anonymity based on its flaws found in these case studies, and by gathering data about user habits to really understand their behavior towards this security technique in order to improve its usability. Because of this, we knew it would be beneficial to gather data directly from users in our study, which led us to create the survey.

In another study we analyzed, Weisband and Reinig [9] first discuss user perceptions of privacy and address how people behave as though emails are private when in fact they have many vulnerabilities. Email privacy in organizations is complex and people often have false views of it. Numerous theoretical explanations are given for why users believe it is private. These are based on different things including technical factors, system design, corporate management policies, and social effects. Each of these areas is then discussed in more depth, providing details and examples. Conclusions are made that employers need to provide their employees with more information on their email policies and technology security. They also need to gain a better understanding of legal issues dealing with email privacy. This study could be very useful for organizations looking to improve the security of their employees' email and help them understand it better. The only weakness of this study is that it focuses mainly on company email, and not on personal/home email use. It could be improved by applying these same theoretical explanations and ideas about encryption usability and user perceptions of email security to using it for personal matters outside of the

workplace. We were sure to include questions in our survey about both workplace and personal use of email and encryption methods.

The final study we examined addresses the issue of email encryption techniques failing to be widely adopted, and authors Adida et al. [2] present a deployment and adoption process to help solve this problem. They begin by discussing previous key management strategies and then provide some information on their own development in a previous study, Lightweight Public Key Infrastructure (PKI) for email authentication. Then they explore how Lightweight PKI could be used for encryption. They address the two main goals of their solution, which are to protect honest domains and users. They go into details about their development, providing all of the technical aspects and algorithms. The authors also provide an example of how messages could be sent between two users, e.g. Alice and Bob, using this technique. They then discuss the flexible deployment options available with lightweight encryption, and give specifics of one scenario with naïve users, and another scenario with more advanced users. They also explore splitting IBE master keys, and what algorithms would be involved with this. Also they go over the ways that untrusted and malicious servers could damage security schemes, and how their method can protect against this. This study explores new ideas and contributes useful and meaningful information that can be used as a step towards making email encryption more widely accepted. The fact that they have already used PKI for email authentication and it has worked successfully also adds strength to the study, showing that the authors have a great deal of knowledge and background in working with this type of technique. One of their ideas for future work includes user interface considerations. While their research doesn't address this, it seems like an element that would have a large impact on the success of the method and could benefit from further research about user behavior and preferences to create an effective design.

While prior work is extensive in the area of email encryption, we believe that there are still many avenues to be explored. Our study aims to provide a closer insight on why users choose to encrypt emails, or why they don't, and what could be done to influence this. Our study departs from prior works because while they mention the fact that email encryption is not widely adopted by users, and address usability concerns with encryption programs, we actually gather information from real users about their specific dealings with encryption programs, and we then apply our findings to offer possible solutions.

### III. MAIN RESEARCH

#### A. Methodology

The basis for our method of data collection centered on wanting to gain honest and true views of average email users about their experience with encryption. To do this we created a survey made up of ten multiple-choice questions and then distributed it online to thirty participants. These participants

ranged in age from 23 to 56, and they were all employed by a variety of different companies. We did not want to limit our participants to those from a certain workplace or a certain age group, because we wanted contributors with various backgrounds and experiences. We asked them all to answer the questions as honestly as possible, and assured them that all results were to remain anonymous. We chose this approach of gathering data in order to gain answers from many different people to a variety of questions, and have organized results that we were able to analyze and draw conclusions from. While open-ended questions can provide more detailed answers, it can also make it difficult to measure the results logically and make accurate conclusions. With our multiple-choice survey, the results are more clear and conclusive and led to rewarding findings.

#### B. Data Collection

A critical part of our research method was determining the specific questions to ask on our survey. We wanted them to be simple yet still provide us with a good understanding of each participant's views on email encryption. We began by asking the following question:

*Do you use methods of email encryption?*

- Yes, only at work*
- Yes, only at home*
- Yes, at work and at home*
- No*

This question allowed us to determine from the very beginning how many of our participants actually utilized email encryption programs, and if that was for work or personal use. All ten questions we asked revolved around the topic of email encryption, discussing reasons for not utilizing encryption as well as discussing typical emailing habits of participants. Each question was multiple-choice, and answer choices varied from two to four different options. Our variety of questions allowed us to gain a great deal of insight on how users commonly interact with email and encryption applications, and how they feel about using encryption.

#### C. Benefits

Our research method differs from those in prior studies because it focuses more on the user perspective of email encryption and reasons why people are still failing to make use of encryption applications. PGP, S/MIME, and other encryption methods have been available for many years, and many studies have been done to look at new techniques and ways to improve them, but most of those studies have not given attention to users' opinions. We strived to focus solely on users' views and practices in order to gain the most accurate understanding of what influences their choice in using or not using email encryption. While a great deal of prior work has focused on improving technical operations of encryption applications, it won't matter how great a technical designer believes a program to be if users still fail to use it. Our study concentrates on this and tries to determine the main reasons why people are choosing not to encrypt emails, both in the work place and at home. Our research method provides us with sufficient results to determine this, allowing us to present new and unique information to the research community.

#### IV. RESULTS

Our results come from the data gatherings of surveys with thirty participants. We found that only 57% surveyed actually use email encryption, none of which use it “only at home.” This demonstrates the dire need to determine why people make the choice to not encrypt emails, since almost half of our participants fit into this category. With the great amount of sensitive data sent through emails, it is essential to understand why people aren’t encrypting and what can be done to change this. Figure 1 below shows results for the second question we asked participants.

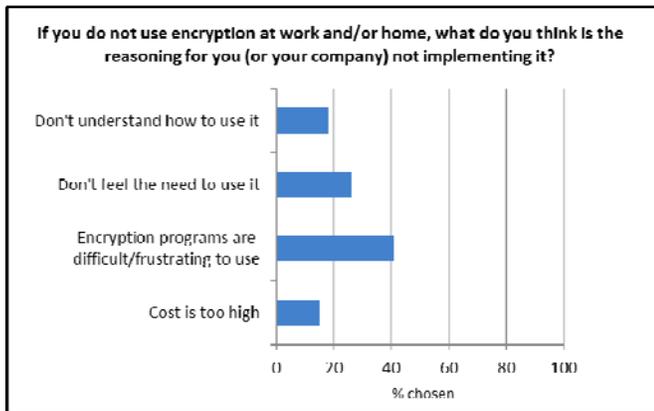


Figure 1. Survey results on reasons of not using email encryption.

From these results, we can see that difficulty of use is the top reason for encryption not being used, followed by users not feeling the need to use it and not understanding how to use it. Usability has always been a major issue with email encryption, and this data proves that it is in fact a heavy influence in people’s choice to use encryption techniques. Many people also seem to be uneducated on email encryption, since a total of 44% of those surveyed either don’t understand how to use encryption, or don’t feel the need to use it, meaning they aren’t aware of the serious risks with sending unencrypted emails. A small portion of survey participants felt that cost was the main reason for not using encryption. This also shows unawareness and the need for more education on the matter, since there are many cost-efficient encryption options for both personal and business use.

Other questions asked showed that a large majority of participants send personal or sensitive data in emails, and a majority also send personal emails from their company email server at work. We found that only 40% of participants said that management at the company where they work strongly enforces the use of email encryption. If it isn’t being enforced at work, then many people likely won’t see the need to use encryption at all. Managers need to understand the seriousness of data leakage and security breaches that happen so often, and realize that enforcing the use of email encryption can help prevent this. There are also many types of data confidentiality laws, some differing by state and some based on the type of sensitive information being sent, such as health records, that requires email encryption to be used. Some of the companies choosing not to enforce it may be violating laws and regulations. Only 7% of survey participants feel very informed

of policies and regulations concerning their company using email encryption. This proves that there is definitely a need for education on this subject so that employees understand what is required of them to be in accordance with policies and laws.

When looking at satisfaction with the usability of encryption programs, we found that very few people are completely satisfied. Results demonstrating this are displayed in Figure 2 below.

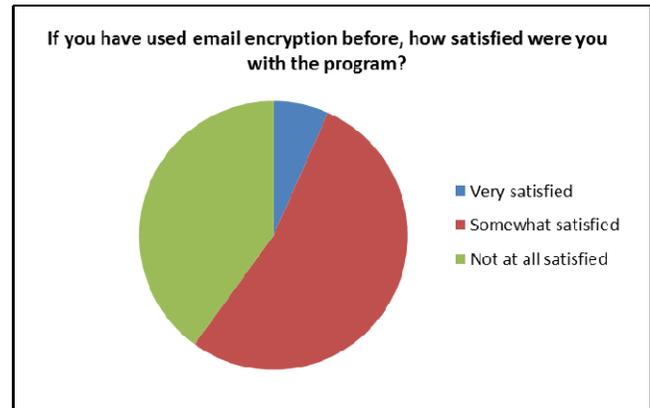


Figure 2. Survey results on user satisfaction of email encryption program.

For a program to be successful, users need to feel very satisfied, which is obviously not the case with encryption methods. This seems to be the trend in our results, since many users also named difficulty of use as the top reason for not using encryption. We also found that 37% of those surveyed have tried to open an encrypted email on their smartphone. Since smartphones and other mobile devices have become increasingly popular in recent years and many people rely on them to perform work-related tasks, it means that encryption programs will also need to be compatible with these devices. If usability is even more difficult on mobile devices, then users are likely to become more frustrated and reluctant to use encryption methods. While email encryption does have many advantages such as ensuring the security and privacy of data, it seems that users believe its disadvantages outweigh those. The lack of an easy-to-use encryption program is definitely a drawback and a large factor in people commonly sending unencrypted emails containing sensitive information.

#### V. CONCLUSIONS AND FUTURE WORK

From our study and data analysis, we can conclude that the main reasons for people failing to use the available email encryption methods is that they lack simple usability, and people lack knowledge on the topic of email encryption. A large majority of our participants don’t know exactly what should be encrypted in an email, and many of them don’t understand how to use encryption programs. This highlights the need for education on the subject. We believe a solution to the lack of encryption use might be to provide people with more information on the risks associated with sending unencrypted emails, and on the available encryption programs and how they operate. Email servers could send out information about this to its users, or companies could make it

a priority for management to become more educated on the issue and then conduct workshops for its employees to teach them all about how to use encryption techniques. If more people were actually taught how to use it then they would feel more comfortable with it and understand what needs to be encrypted, making them more likely to actually use encryption on a daily basis. Employers should also work harder at strongly enforcing the use of encryption methods and informing employees of the laws and regulations relating to it. This could lead people to finally comprehend the critical need for encryption, which may also drive them to use it at home.

When looking at the usability issue, many researchers have already known that encryption programs are difficult to use and work has been done trying to improve them. However, these attempts have not proved very successful since it is still a major issue with users. We believe this could be solved by performing extensive evaluations and surveying users, to determine what exactly they don't like about their current encryption programs. Researchers could also try to learn which specific characteristics users do like about other computer security programs they commonly use. This would be a good avenue of exploration for a future study done in this area. After collecting all of the information from users and having a better understanding of what it is that they precisely need and want in a program, then a technical designer would be more capable of creating a successful encryption program suited to the needs of users.

Future studies could also extend upon ours by trying to educate users on email encryption through some of our suggested methods, and then observing how that actually impacted their use of encryption. Our study led to useful findings and conclusions but there is always room for further exploration on the critical topic of email encryption.

#### REFERENCES

- [1] M. Abdalla, J. Birkett, D. Catalano, A. Dent, J. Malone-Lee, G. Neven, J. Schuldt, and N. Smart, "Wildcarded Identity-Based Encryption," in *Journal of Cryptography*, 2011, pp. 42 – 82.
- [2] B. Adida, S. Hohenberger, and R. Rivest, "Lightweight Encryption for Email," in *USENIX SRUTI '05: Steps to Reducing Unwanted Traffic on the Internet Workshop*, 2005, pp. 93 – 99.
- [3] R. Dingledine and N. Mathewson, "Anonymity Loves Company: Usability and the Network Effect," in *Proceedings of the Fifth Workshop on the Economics of Information Security*, 2006, pp. 100 – 112.
- [4] A. Gabrielson and H. Levkowitz, "Reducing Error by Establishing Encryption Patterns," in *PATTERNS 2011, The Third International Conferences on Pervasive Patterns and Applications*, 2011, pp. 133 – 137.
- [5] R. Kainda, I. Flechais, and A. Roscoe, "Security and Usability: Analysis and Evaluation," in *ARES '10 International Conference on Availability, Reliability, and Security*, 2010, pp. 275 – 282.
- [6] C. Moore, "You Are What You Email @ Your Inbox," in *Cranfield University School of Management Research Briefings*, 2011, pp. 1 – 4.
- [7] B. Payne and W. Edwards, "A Brief Introduction to Usable Security," in *IEEE Internet Computing*, 2008, pp. 30 – 38.
- [8] E. Poole, C. Le Dantec, J. Eagan, and W. Edwards, "Reflecting on the Invisible: Understanding End-User Perceptions of Ubiquitous

Computing," in *Proceedings of the 10<sup>th</sup> International Conference on Ubiquitous Computing*, 2008, pp. 192 – 201.

- [9] S. Weisband and B. Reinig, "Managing User Perceptions of Email Privacy," in *Communications of the ACM*, 1995, pp. 40 – 47.

#### APPENDIX

Below is the survey we designed and conducted in this research study.

#### Survey

Please answer all following questions as honestly as possible. All results will remain anonymous.

1. Do you use methods of email encryption?
  - Yes, only at work.
  - Yes, only at home.
  - Yes, at work and at home.
  - No.
2. If you do not use encryption at work and/or home, what do you think is the reason for you (or your company) not implementing it?
  - Cost is too high.
  - Encryption programs are difficult/frustrating to use.
  - Don't feel the need to use it.
  - Don't understand how to use it.
3. Do you ever worry about the privacy and security of your emails?
  - Yes, frequently.
  - Yes, sometimes.
  - No, never.
4. Do you ever send personal information or sensitive data in emails?
  - Yes, frequently.
  - Yes, sometimes.
  - No, never.
5. Do you ever send personal emails from work using your company email server?
  - Yes, frequently.
  - Yes, sometimes.
  - No, never.
6. Does management at your company strongly enforce the use of email encryption?
  - Yes.
  - No.
7. Are you aware of policies and regulations concerning your company using email encryption?
  - Yes, very informed of them.
  - Yes, somewhat informed of them.
  - No, not at all informed of them.

8. If you have used email encryption before, how satisfied were you with the program?
  - Very satisfied.
  - Somewhat satisfied.
  - Not at all satisfied.
  
9. Have you ever tried to open an encrypted email on your smartphone?
  - Yes.
  - No.
  
10. Do you know exactly what should be encrypted in an email?
  - Yes.
  - No.

# IT Security Policies and Employee Compliance: The Effects of Organizational Environment

Kendal Stephens LaFleur, Narasimha Shashidhar  
 Department of Computer Science  
 Sam Houston State University  
 Huntsville, TX  
 {kks016, karpoor}@shsu.edu

**Abstract**— A major threat to IT security in today's business world is the simple problem of careless employees choosing not to comply with security policies and guidelines. Many studies have been done to track the reasoning behind this problem and try to find a solution. To address this issue, our study proposes to look at the effects of company culture and environment, including relationships among co-workers as well as their feelings toward upper management, in order to determine if a correlation exists between this relationship and an employee's compliance with IT security policies. In order to do so, we designed a survey comprised of various questions dealing with workplace environment and thoughts on IT security policies in order to gain insight on the topic and gather useful data. We administered the survey to an anonymous group of people we assembled, ranging in age and employed by a variety of companies. The survey results were then analyzed and data trends were uncovered in order to see if a correlation does in fact exist. We found that there is a positive correlation between employees' organizational environment and their compliance with IT security policies. We also discovered that there appears to be a lack of employee education on security policies in the workplace, which needs to be studied further in the future. To the best of our knowledge, our method is unique and stands apart from prior work in that the data gathered was not limited to employees from a specific company or of equal job status. We believe that the wide variety of our chosen participants will provide a more comprehensive look at this issue. Additionally, our study does not focus on any specific behavioral theories or try to implement any new methods in the workplace as was done earlier. We merely look at the employees' daily, ordinary feelings and actions, giving a special "real" and "true" quality to our results.

**Keywords**—*policy; compliance; security; education*

## I. INTRODUCTION

Now more than ever, IT security issues pose a major threat to companies everywhere. Security breaches, computer infections, system failures, and data loss are all serious dangers. One way to help combat these risks is by implementing IT security policies, which provide employees with a set of rules and guidelines to follow to help lessen the likelihood of security issues. A major problem, however, involves employees choosing not to comply with all of these policies, which puts the organization at risk. Our study explores a possible connection between the level of employee compliance with IT security policy and the state of their organizational environment. We gather data from employees at various jobs in several companies in order to gain insight on

this correlation, studying how they function in their ordinary workplace. Our method enables one to take a look at the daily life of employees, without altering anything about the way the workplace normally operates. Using this approach rather than studying behavioral methods and then implementing those in the workplace makes our study unique and provides us with useful results distinctive from the findings of prior studies.

## II. LITERATURE REVIEW

A great number of studies have been done in the area of employee compliance with security policy. It has been shown that many IT issues and security breaches are merely the result of an employee failing to adhere to the company's rules and procedures, which indicate that these should be easily preventable. This leads researchers to question what exactly influences this type of employee behavior and what could be done to increase compliance with IT security policies and suggested guidelines.

Many of the prior research studies focus on trying to implement different behavioral methods to control the behavior of employees. Tyler's [5] study uses a command-and-control model where managers implement sanctions and punish undesired behavior. While this can provide useful insights, it does not uncover how employees' behavior is impacted on a normal daily basis. We believe that it would be more beneficial to study the current company culture, relationships among employees, and other factors of this type that affect their behavior, rather than having managers make sudden changes and test different theories that are not part of the usual workplace. Forgas et al. [13] conducted a study focusing on the Affect Infusion Model to delve into experimental social psychology and organizational behavior to determine the influence that affective states have on decision making and behavior in organizations. Dillon et al. [2] analyzed how different behavioral models affect the ways that employees accept new technologies and the consequences of those innovations. Our study will differ from these earlier studies in the way that it gathers data by having participants answer questions about their daily routine workplace, not about the effects of some foreign new behavioral method.

Vroom et al. [6] studied the human factor in IT security from many different angles. *Organizational culture* is defined and explained so that its meaning is clearly understood, and conclusions are drawn that the culture of an organization could have a huge impact on the security of information, either in a negative or a positive way. The subject of organizational

behavior is then discussed, showing how it affects the actions of employees. Final conclusions state that it will be beneficial to study a combination of both organizational culture and behavior in order to determine a way to successfully change an organization's culture one portion at a time. This perspective is interesting, as the researchers look at how various characteristics of the organization come into play with employee behavior and compliance with policy, and then conclude that they should be changed a small bit at a time to eventually get the organization where it needs to be. This approach seems more practical and beneficial than using a specific behavioral theory to completely change the way the workplace operates and then recording the results of that experiment. Our study approach will build on this paradigm, observing how the workplace functions on a daily basis rather than examining how it changes with the different methods being implemented.

Chan et al. [1] showed that information security is a very complex matter and its study should incorporate not only technical factors, but also social factors as well. They introduce a concept called *organizational climate*, which they distinguish from organizational culture since it refers to the way employees view both formal and informal policies, procedures, and practices of an organization. The study uses organizational climate to provide researchers with some insight on the more subtle, less apparent aspects of an organization's culture. They show that emerging evidence exists to show that specific climates are predictive of specific outcomes. For example, employees in a workplace with a strong safety climate seemed to comply with more safety guidelines. They claim that this is important because safety programs and information security programs share many characteristics and are both critical components of an organization. The authors then discuss how organizational climate can be influenced by socialization with coworkers and peers, and by practices of supervisors and upper management. This is definitely an important observation; however, this article fails to explore it further and see how these relationships actually affect the climate and the behavior of employees. One primary goal of our research study is to look into this further and see what connections might be uncovered in order to fill this gap. We like to stress that our study is the first of its kind, to the best of our knowledge, to undertake this approach on examining this correlation.

A study done by Herath et al. [3] brings up organizational commitment and how it has influenced organizational behavior literature. They define *organizational commitment* as an individual's feelings of dedication and contribution to their work organization. The study discusses how an employee's commitment to an organization is likely to play a role in his or her engagement in security behaviors. The conclusion that is drawn is that the strength of employees' organizational commitment will positively affect their intentions to follow policies in the workplace. But does an employee's *intention* to follow a policy mean that they actually do it? We believe that this is a fundamental weakness in this work. Although they present several interesting ideas revolving around

organizational commitment, it still remains uncertain as to how this directly affects an employee's real actions. There are several questions that remain unanswered - Just because an employee thinks that something is the right choice, or feels more obligated to do a certain thing, does that mean that they actually follow through and do that? We examine this question further. In contrast to this study and many others, our study will focus on actual employee actions, and not just what they are "likely" to do because of feelings of obligation.

Pahnila et al. [4] conducted a study of employee behavior towards IS security policy compliance in which they suggest that the behavior of employees is generated by the way that they interact with each other and prove that this is true through data collection. This is an important deduction and raises the following questions in this area. Does positive interaction between employees increase their compliance with security policies? Does interaction between an employee and their supervisor or top management members have a greater or lesser effect on compliance than their interaction with same-level co-workers? We aim to address these questions in this study.

While prior work is extensive in this area, we have chosen to survey only the most directly relevant work and it is not to be treated as an exhaustive survey. We found many informative and significant pieces of research regarding employees' compliance with IT-related security policies. However, we believe that there are still many avenues and ideas that need to be explored. Our study aims to provide a closer insight on what exactly affects an employee's compliance with security policies in their typical workplace by examining organizational culture and climate as well as the interactions among different co-workers. The ultimate goal is to determine if and how these factors positively impact an employee's adherence to security policies. Our study departs from earlier work by not looking only at the "intentions" to comply with policies and also by not implementing different behavioral models to try and alter employees' behavior.

### III. MAIN RESEARCH

#### A. Methodology

The reasoning behind our theory about employees' compliance with security policies being related to their work environment is well justified. A great deal of scientific research has been done in the area of employee compliance with rules and different behavioral influences. For instance, Tyler [5] discusses some of these behavioral theories including the "command-and-control model" and the "self-regulatory model." The first of these models looks at how employee behavior is controlled by managers or bosses and the way they punish undesired behavior. The latter looks at how the ethical values of employees motivate them to follow rules. A study done by Cardona et al. [12] discusses how the social exchange relationship between employees and their organization affects their attitude and feelings. The more positive they view the relationship, the more connected they feel to the organization, leading to increased feelings of obligation and commitment. This leads us to question if this increased commitment holds

true in the area of IT security policies, and if these positive feelings affect employees' compliance with them. Although studies such as this one give the impression that such a connection exists, there is no definite evidence. We wanted to look deeper into assessments of employee behavior and relationships and see how compliance with IT security policies is connected.

We were motivated to study this topic because of the increasing importance of IT security policies in the workplace, and issues with non-compliance. Employees failing to follow IT policies can lead to major consequences for a company, such as data breaches, viruses and infections, unauthorized access to information, and many other issues. Because of the serious impact of these threats, we felt the need to study what might affect compliance with policies. Our goal was to establish conclusions that could be studied further and could be used as valuable information by employers to help increase employee compliance with IT security policies in the workplace. In the recent years, organizational culture and environment have become increasingly popular topics, with many companies striving to promote a friendly and caring atmosphere while providing great amenities for employees, and with national awards being given for "Great Place to Work" and "Best Corporate Culture." This led us to ask questions about the possible correlation between organizational environment and employee compliance with security policies. Could these two things be connected? We began to think about the different types of relationships and interactions among people in the workplace and explored possible links between employee behavior and following rules and policies.

Psychological studies have been criticized for focusing only on the negatives of situations and not the positives. In a study on organizational behavior, it has been shown that studies in psychology, as well as in business and management, need to use a more positive approach [8]. Thus, our motivation in this study is to place emphasis on a positive correlation, rather than a negative one. We chose to differ from the norm by not looking at how negative work relationships impact non-compliance with policies, but rather focus on how positive relationships increase compliance with policies.

A study conducted by Bishop et al. [7] looks at how commitment to an organization and commitment within work teams leads to desired employee outcomes. We used this idea of positive commitment leading to positive job performance in developing our theory. If an employee is committed to their organization and holds themselves and their workplace to high standards, then will this positively influence whether or not they follow policies? This is one of the many questions we aim to answer in this study. We show that our theory is rational and in the next section outline how we collect the requisite data to substantiate our claim.

### B. Beginning Processes

First, we designed a survey consisting of eleven multiple-choice questions relating to employees' feelings towards co-workers and management, as well as their thoughts on IT security policies in the workplace. This online survey was then

sent out to 40 individuals for completion. Participants were told to be as candid and honest as possible and all results were to remain anonymous. The participants ranged in age from 22-55 and included both men and women, providing us with a good amount of variation. A study by Morris et al. [11] found that age can play a part in perceptions of technology and its use in the workplace, so we chose to use participants with a fairly large age deviation in order to get accurate results from the general working population, and not narrow it down to the outlooks of one specific group. Also, participants were all employed by a variety of different companies. For example, one participant was an elementary school teacher, one worked for a large accounting firm, and one was a field supervisor for an oil and gas company. A great amount of diversity was present in this group of people surveyed. We wanted to test our hypothesis in a wide variety of work settings and not limit it to one specific company or type of employee.

### C. Study Details

A critical component of our research method was determining the specific questions to be included in the survey. One important goal was to keep the number of questions below 15 so that participants would not get bored, which might drive them to answer quickly and carelessly in an attempt to finish. We also required the questions to provide us with insight on how employees get along with others in the workplace, how they feel about IT security policies, and what possible associations might exist between these two areas. We created multiple questions revolving around these subjects. The survey was led by the following question:

*How informed are you of your workplace's IT security policies (computer passwords, data protection, web monitoring, rules and legal issues, etc.)?*

- a. *Very well informed*
- b. *Well informed, familiar with most policies*
- c. *Partly informed, not familiar with numerous policies*
- d. *Not informed at all*

Since the survey participants had no idea beforehand what the survey would be concerning, this question was asked first in order to get them thinking about IT security policies and enable us to gauge how well-educated they were on the subject. Questions were also asked concerning their relationships with others in the workplace. Here are examples of two such questions: *How would you describe the company culture/environment in your workplace? How would you describe your relationship with top management and bosses at your workplace?* The goal was to get an idea of participants' feelings on both topics of workplace relationships and IT security policies, design questions combining the two ideas.

*Would you be more likely to violate an IT security policy if the reasoning was to help out a co-worker you're friends with?*

- a. *Definitely*
- b. *Maybe*
- c. *Not sure*
- d. *No*

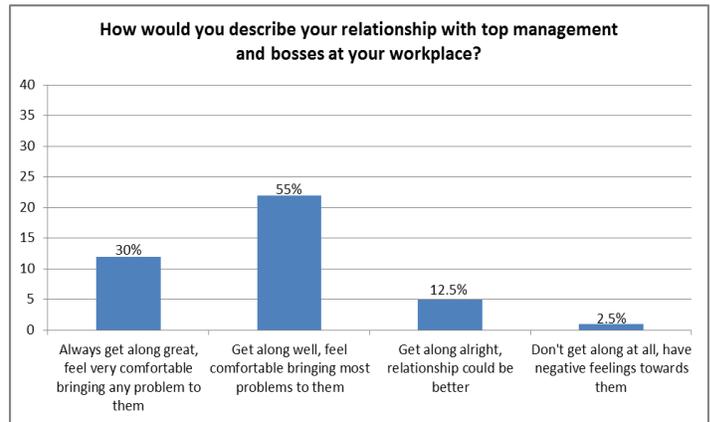
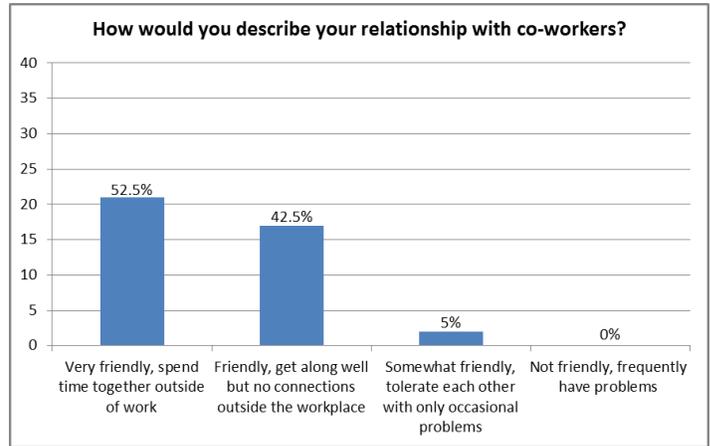
Responses to a question like this provide us with concrete answers as to how positive work relationships impact compliance with security policies. While our theory is that this correlation is positive, it also seems plausible that it might be just the opposite. Employees might be more willing to violate a policy if doing so would help out a fellow employee they're friends with, and they feel a higher obligation to that friendship than to the company. Prior studies done by Truckenboldt [9] and by Wayne et al. [10] examine this concept, looking more closely at different levels of employee commitment and social exchanges. We asked questions like the one above in order to test out multiple possibilities and look at the theory from all perspectives.

Recall that we had a total of eleven questions on the survey. Some questions were only concerned with IT security policies, some were only concerned with workplace relationships, and some comprised a combination of the two topics in order to see what effects they had on each other. We chose to perform the research and data gathering in this manner for several reasons. First, we wanted to get the most honest and accurate answers from participants as possible. By hosting the survey online and having responses remain anonymous, employees would not be inhibited to share their true views and not worry about feeling embarrassed to admit something, if for example they don't know very much about security policies or they have violated them in the past. Secondly, a multiple-choice survey makes it easier to compare and analyze results than other survey options. Had we elected to include open-ended questions, this might have forced us to analyze many different responses for each question and possibly not have been conducive to data mining and analysis. Our survey method was also less time-consuming and provided quicker results than doing workplace observations of employees. It also allowed us to gather responses from individuals in many different professions. We also wanted to collect information that reflected how employees operate in their daily routine workplace. Many prior studies have tried implementing different behavioral methods and then studying employees' reactions. For example, having a boss punish certain behaviors and reward others to see how it affects employees' compliance with policies. We wanted our study to differ from those by not introducing any type of changes in the workplace before we gathered our data. Our primary goal was to ensure that the results reflected real world circumstance.

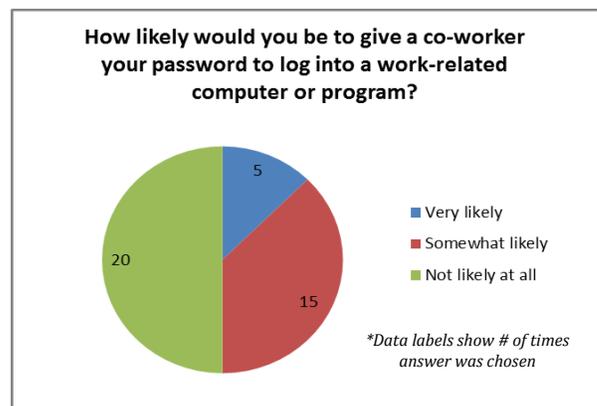
#### IV. RESULTS

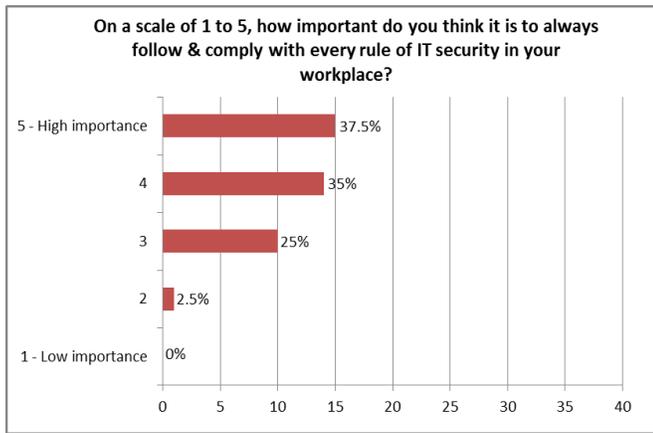
##### A. Data Observations

We observed from our data analysis that 45% of participants feel that they are only partly informed of their workplace's security policies and feel unfamiliar with numerous policies. This leads us to believe that one major cause of employees not complying with policies might be simply because they are unaware of them. Regardless of the social environment in the workplace, it appears that increasing user education would likely have a positive impact on policy compliance. The following two graphs display the results of questions dealing with relationships in the workplace.



From this data, we conclude that the majority of participants have a very positive relationship with both co-workers and upper management. A survey question asking participants to describe the company culture/environment in their workplace also produced positive answers, with 38% choosing "very comfortable and friendly environment where everyone feels relaxed" and the remaining 62% choosing "somewhat comfortable environment where everyone usually gets along fine." No participants chose slightly tense or very tense environment. Keeping this in mind, we now take a look at their usual inclinations towards IT security policies.





From these responses, we can gather that the majority of employees surveyed take security policy seriously and understand that compliance is extremely important. Results showed that 50% of employees answered “not likely at all” when asked how likely they would be to give a co-worker their password. We also asked if participants would be more likely to violate a policy if the reasoning was to help out a co-worker who is their friend, and 41% answered “no” and only 10% answered “definitely”. From these responses we can draw the conclusion that a positive relationship with co-workers seems to positively impact compliance with IT security policies. For most participants, they would not put their friendship above all else and violate a policy because of obligation to the co-worker.

It also seems that participants very rarely violate a policy, or if they do they are unaware of it, since 82% of participants chose “hardly ever” or “never” when asked how many times they have violated or not complied with a security policy or guideline. We also found that 59% of participants are very highly concerned with pleasing their boss and making others in their workplace happy by always doing an outstanding job at work. This makes it no surprise that 62% feel they have an extremely high sense of responsibility and accountability in the workplace.

When looking overall at the questions asked and the data collected, a majority shows positive results when dealing with both environment of the workplace and feelings towards IT security policies. The figure below displays a summary of all eleven survey questions and their most commonly chosen answer, along with the percentage of participants that selected that response.

How informed are you of your workplace’s IT security policies?	<i>Partly informed, not familiar with numerous policies – 45%</i>
How likely would you be to give a co-worker your password to log into a work-related computer or program?	<i>Not likely at all – 50%</i>
How likely would you be to violate or not comply with an IT security policy if it required you to do extra work or you felt it wasn’t all that important?	<i>Not likely at all – 56%</i>
How would you describe your relationship with co-workers?	<i>Very friendly, spend time together outside of work – 52.5%</i>

How would you describe your relationship with top management and bosses at your workplace?	<i>Get along well, feel comfortable bringing most problems to them – 55%</i>
How would you describe the company culture/environment at your workplace?	<i>Somewhat comfortable, everyone usually gets along fine – 62%</i>
On a scale of 1 to 5 (with 5 being highly important), how important do you think it is to always follow and comply with every rule of IT security in your workplace?	<i>5 – 37.5%</i>
Would you be more likely to violate an IT security policy if the reasoning was to help out a co-worker you’re friends with?	<i>No – 41%</i>
On a scale of 1 to 5, how concerned are you with pleasing your boss and making others in your workplace happy by always doing an outstanding job at work?	<i>5 – 59%</i>
On a scale of 1 to 5, how high is your sense of responsibility and accountability in your workplace?	<i>5 – 62%</i>
To your knowledge, how many times have you violated or not complied with a security policy or guideline at your workplace?	<i>Hardly ever – 46%</i>

**B. How our Results Compare to Prior Research**

When comparing our results to those found in prior related studies, we proved Pahnla et. al’s conjecture that “individuals create their behavior based on the interaction with each other” [4], and discovered how this is true in the area of IT security policies. We found that a good relationship between employees and co-workers, as well as that between employees and upper management causes them to have a more positive attitude towards following security policies. From our research it is not apparent if either of these relationships has a stronger effect on compliance than the other, but it is an interesting question to explore and we hope that our present work will motivate future studies on this topic.

Our results also relate to Herath et al.’s [3] study on organizational commitment. They claimed that an individual’s feelings of organizational commitment have a positive influence on their intentions to follow policies. We questioned whether an employee’s *intentions* are actually carried out and shown in their actions. Looking at our survey results, we can see that a great majority of employees have high levels of organizational commitment, and they also very rarely violate security policies and clearly understand their importance. We can conclude from this that organizational commitment not only influences the *intentions* to follow policy, but it also influences the actual actions of employees and leads them to follow policies more often.

Our findings are valuable to the research community because we have shown an important link between organizational environment and employee compliance with IT security policies. This correlation exists in the ordinary workplace in a variety of different job settings. We did not implement behavioral methods and try to alter the behavior of employees as many researchers have done in the past, but rather chose to focus on daily, typical relationships and actions. This has led us to significant research findings that will be meaningful and helpful in future studies because they exhibit the behavior of actual employees in their natural workplace interactions.

## V. CONCLUSIONS

From our data analysis, we conclude that a positive relationship does in fact exist between workplace environment/relationships and employee compliance with IT security policies. However, since a majority of study participants already have a positive work environment as well as high inclinations to follow security policies, this makes us question why so many security breaches and IT problems still occur due to employee negligence? The only hint at this answer that can be gathered from our study is lack of employee education when it comes to policies. A total of 53% believed they were only partly informed or not at all informed in regards to their company's IT security policies. Since a majority also believed that following the policies is important and that they hardly ever violate them, it seems that a possible explanation is that they simply aren't familiar with the specifics of all policies. Perhaps future research done in this area could focus on employee education and knowledge of rules and policies, and possibly find that to be a reason behind the lack of compliance. Future research is also needed in the area of company culture and relationships affecting policy compliance, testing larger groups of participants and possibly separating companies into different categories to get a better idea of specific variations among different types of employees.

## VI. REFERENCES

- [1] M. Chan et al., "Perceptions of Information Security in the Workplace: Linking Information Security Climate to Compliant Behavior," *Journal of Information Privacy & Security*, 2005, pp 18-41.
- [2] A. Dillon and M. Morris, "User acceptance of new information technology: theories and models," *Annual Review of Information Science and Technology*, Vol. 31, 1996, pp 3-32.
- [3] T. Herath and H. Rao, "Protection motivation and deterrence: a framework for security policy compliance in organizations," *European Journal of Information Systems*, 2009, pp 106-125.
- [4] S. Pahnla M. Siponen, and A. Mahmood, "Employees' Behavior towards IS Security Policy Compliance," *Proceedings of the 40th Hawaii International Conference on System Sciences*, 2007, pp 1-10.
- [5] T. Tyler, "Promoting Employee Policy Adherence and Rule Following in Work Settings: The Value of Self-Regulatory Approaches," *Brooklyn Law Review*, Vol. 70:4, 2005, pp 1287-1312.
- [6] C. Vroom and R. Solms, "Towards information security behavioral compliance," *Computers & Security*, 2004, pp 191-198.
- [7] J. Bishop, K. Scott and S. Burroughs, "Support, commitment, and employee outcomes in a team environment," *Journal of Management*, Vol. 26, 2000, pp 1113 – 1132.
- [8] A. Baker and W. Schaufeli, "Positive organizational behavior: Engaged employees in flourishing organizations." *Journal of Organizational Behavior*, 2008, pp 147 – 154.
- [9] Y. Truckenbrodt, "The relationship between leader-member exchange and commitment and organizational citizenship behavior." *Acquisition Review Quarterly*, 2000, pp 233 – 244.
- [10] S. Wayne, L. Shore and R. Liden, "Perceived organizational support and leader-member exchange: a social exchange perspective." *Academy of Management Journal*, Vol. 40, 1997, pp 82 – 111.
- [11] M. Morris, V. Venkatesh and P. Ackerman, "Gender and age differences in employee decisions about new technology: an extension to the theory of planned behavior." *IEEE Transactions on Engineering Management*, Vol. 52, 2005, pp 69 – 84.
- [12] P. Cardona, B. Lawrence and P. Bentler, "The influence of social and work exchange relationships on organizational citizenship behavior." *IESE Business School*, 2003, pp 1 – 27.
- [13] J. Forgas and J. George, "Affective influences on judgements and behavior in organizations: an information processing perspective." *Organizational Behavior and Human Decision Processes*, Vol. 86, 2001, pp 3-34.

## VII. APPENDIX A

Below is the survey we designed for our study.

### Survey

*Please answer all questions and be as honest as possible. All results are anonymous.*

**1. How informed are you of your workplace's IT security policies (computer passwords, data protection, web monitoring, rules and legal issues, etc.)?**

- Very well informed
- Well informed, familiar with most policies
- Partly informed, not familiar with numerous policies
- Not informed at all

**2. How likely would you be to give a co-worker your password to log into a work-related computer or program?**

- Very likely
- Somewhat likely
- Not likely at all

**3. How likely would you be to violate or not comply with an IT security policy if it required you to do extra work or you felt it wasn't all that important?**

- Very likely
- Somewhat likely
- Not likely at all

**4. How would you describe your relationship with co-workers?**

- Very friendly, spend time together outside of work
- Friendly, get along well but no connections outside the workplace
- Somewhat friendly, tolerate each other with only occasional problems
- Not friendly, frequently have problems

**5. How would you describe your relationship with top management and bosses at your workplace?**

- Always get along great, feel very comfortable bringing any problem to them
- Get along well, feel comfortable bringing most problems to them
- Get along alright, relationship could be better
- Don't get along at all, have negative feelings towards them

**6. How would you describe the company culture/environment in your workplace?**

- Comfortable & friendly environment, everyone feels relaxed
- Somewhat comfortable, everyone usually gets along fine
- Somewhat tense, not usually very friendly
- Very tense, not a healthy environment

**7. On a scale of 1 to 5 (with 5 being highly important), how important do you think it is to always follow and comply with every rule of IT security in your workplace?**

- 1
- 2
- 3
- 4
- 5

**8. Would you be more likely to violate an IT security policy if the reasoning was to help out a co-worker you're friends with?**

- Definitely
- Maybe
- Not sure
- No

**9. On a scale from 1 to 5 (with 5 being highly concerned), how concerned are you with pleasing your boss and making others in your workplace happy by always doing an outstanding job at work?**

- 1
- 2
- 3
- 4
- 5

**10. On a scale of 1 to 5, how high is your sense of responsibility and accountability in your workplace?**

- 1
- 2
- 3
- 4
- 5

**11. To your knowledge, how many times have you violated or not complied with a security policy or guideline at your workplace?**

- Many times
- A few times
- Hardly ever
- Never

**SESSION**

**BIOMETRICS AND FORENSICS I +  
CRYPTOGRAPHIC TECHNOLOGIES**

**Chair(s)**

**Dr. Rita Barrios**  
**Univ. of Detroit Mercy - USA**  
**Dr. Victor Gayoso Martinez**  
**CSIC - Spain**



# State of the Art in Similarity Preserving Hashing Functions

V. Gayoso Martínez, F. Hernández Álvarez, and L. Hernández Encinas

Information Processing and Cryptography (TIC), Institute of Physical and Information Technologies (ITEFI)  
Spanish National Research Council (CSIC), Madrid, Spain

**Abstract**—*One of the goals of digital forensics is to analyse the content of digital devices by reducing its size and complexity. Similarity preserving hashing functions help to accomplish that mission through a resemblance comparison between different files. Some of the best-known functions of this type are the context-triggered piecewise hashing functions, which create a signature formed by several hashes of the initial file. In this contribution, we present the state of the art of the most important similarity preserving hashing functions, analysing their main features. We conclude our work listing the most relevant properties that such type of functions should satisfy in order to improve their efficiency.*

**Keywords:** Forensics, Hash Functions, Similarity Preserving

## 1. Introduction

In modern society, the amount of information has increased in an incommensurable way and therefore the management of big quantities of data represents a major challenge. Digital forensics is the branch of Computer Science which, through investigation and analysis techniques, gathers evidence from the content of a particular electronic device in a way that is suitable for presentation in a court of law, for example. When inspecting the content of a computer, digital forensics experts need to reduce the large amount of data available to them to information that can be analysed in an easier way.

An initial approach to reach that reduction is using cryptographic hash functions. Hashing algorithms like MD5 [22] and the family SHA [17], [19], [20], among others, have been traditionally used in computer forensics to determine if two files were the same. Given the importance of this topic, NIST (National Institute of Standards and Technology) developed a database, called NSRL (National Software Reference Library), which contains hash values of files of several trusted operating systems [18]. With this public service, NIST contributes to reduce the search time of known files and to detect content forgery on the devices. However, the main limitation of cryptographic hash functions when comparing files is that, if one of the files is modified, the outcome of the comparison is negative, even if the two files are identical except in one byte.

In contrast to cryptographic hash functions, Similarity Preserving Hashing Functions (SPHF), also known as Piecewise Hashing Functions (PHF) or fuzzy hashing functions, aim to detect the resemblance between two files by mapping similar inputs to similar hash values. These functions, which compare files at byte level, are useful in order to compare a broader range of input data and detect not only similar texts, but also embedded objects (e.g. a JPEG image in a Word document) or binary fragments (e.g. a data packet in a network connection or a virus inside an executable file).

The technique behind SPHF was originally devised by Harbour [12], and consists in creating a signature formed by several hashes of the initial file, instead of only one. In this way, even if part of the content is modified, only the hashes related to that updated parts would change, allowing to detect if the rest of the file is related or similar to the original one. There are four types of SPHF:

- *Block-Based Hashing (BBH)*: functions that produce a hash after a fixed amount of bytes have been handled, so the number of hashes depends directly on the data object size and the length of the hash input.
- *Context-Triggered Piecewise Hashing (CTPH)*: functions where the number of hashes is determined by the existence of special points, called *trigger points*, within the data object. A point is considered to be a trigger point if it matches a certain property, defined in a way so that the number of expected trigger points falls within a range.
- *Statistically-Improbable Features (SIF)*: the basic aim of this functions is to identify a set of features (sequence of bits) which are least likely to occur in each of the data objects by chance and then compare the features themselves to obtain the similarity level.
- *Block-Based Rebuilding (BBR)*: these functions make use of external auxiliary data, such as binary blocks, to compare the bytes of the original file and calculate the differences between them (e.g. using the Hamming distance). Then, these differences are used as a base to find possible similar data objects.

In this paper, we present a study about the most important similarity preserving hashing functions, including a study of their main properties, and we conclude our work listing the main properties that such type of functions should satisfy in

order to improve their efficiency. The rest of this paper is organized as follows: Section 2 summarizes the block-based hashing functions, whereas in Section 3 context triggered piecewise hashing functions are presented. In Section 4 functions based on statistically-improbable features are analyzed. Block-based rebuilding functions are studied in Section 5. Finally, Section 6 summarizes our conclusions in this topic.

## 2. Block-Based Hashing

The most basic scheme that can be used for determining similarity of binary data is Block-Based Hashing (BBH). In short, using this method cryptographic hashes are generated and stored for every block of a chosen fixed size (e.g. 512 bytes). Later, the block-level hashes from two different sources can be compared and, by counting the number of blocks in common, a measure of similarity can be determined.

An example of this kind of similarity hashing functions was performed by Harbour, who developed a program called `dconfldd` [12]. This software splits the input data into sectors or blocks of a fixed length and computes the corresponding cryptographic hash value for each of these blocks.

The main advantage of this scheme is that it is already supported by existing hashing tools and it is computationally efficient. The disadvantages become fairly obvious when block-level hashing is applied to files: success heavily depends on the intrinsic layout of the files being very similar. For example, if we search for versions of a given text document, a simple character insertion/deletion towards the beginning of the file could render all block hashes different. This means that `dconfldd` is not alignment robust.

Similarly, block-based hashes will not tell us if an object, such as a JPEG image, is embedded in a compound document, such as a Microsoft Word document. In short, the scheme is too fragile and a negative result does not reveal any useful information.

## 3. Context Triggered Piecewise Hashing

The second type of piecewise hashing functions, which are usually known as Context Triggered Piecewise Hashing (CTPH) functions, were originally proposed by Tridgell [30]. Later, Tridgell developed a context triggered piecewise hashing based algorithm to identify mails which are similar to known spam mails. He called his software `spamsum` [31]. The basic idea is to identify content markers, called *contexts*, within a binary data object and to store the sequences of hashes for each of the pieces, also called *chunks*, in between contexts. In other words, the boundaries of the chunk hashes are not determined by an arbitrary fixed block size but are based on the content of the object.

Nowadays, `ssdeep` is the best known CTPH application, but another algorithms based in the same concepts or improvements to the original `ssdeep` algorithm have been proposed: `FKSum`, `SimFD`, `MRS`, etc.

### 3.1 `ssdeep`

In 2006, Kornblum released `ssdeep` [15], one of the first programs for computing context triggered piecewise hashes. In this algorithm, blocks (chunks or segments) are not determined by an arbitrary fixed block size but are based on the content of the object.

The algorithm's core is a rolling hash very similar to the rolling hash used in `rsync` [30] and `spamsum` [31]. The rolling hash is used to identify a set of reset points (also known as distinguished points or triggered points) in the plaintext that depend on the content of a sliding window of seven bytes. The algorithm reaches a reset point whenever the rolling hash (which is based on the *Adler32* function) generates a value which meets a predefined criteria. Let

$$BS_p = B_{p-s+1}B_{p-s+2} \dots B_p$$

denote the byte sequence in the current window of size  $s$ , which is 7 by default, at position  $p$  within the file, and let  $PRF(BS_p)$  be the corresponding rolling hash value. If  $PRF(BS_p)$  hits a certain value, the end of the current chunk is identified. So, the byte  $B_p$  is a trigger point and the current byte sequence  $BS_p$  a trigger sequence. The subsequent chunk starts at byte  $B_{p+1}$  and ends at the next trigger point or the end of the file. As there are only low-level operations, Kornblum's  $PRF$  is very fast in practice.

In order to define a hit for  $PRF(BS_p)$ , Kornblum introduced a modulus,  $b$ , called block size, which determines the reset frequency. The byte  $B_p$  is a trigger point if and only if  $PRF(BS_p) \equiv -1 \pmod{b}$ . If  $PRF$  outputs are equally distributed values, then the probability of a hit is reciprocally proportional to  $b$ . Thus if  $b$  is too small, we have too many trigger points and vice versa.

As Kornblum aims at having 64 chunks, the block size depends on the file size as given in Eq. (1), where  $b_{min}$  is the minimum block size with a default value of 3,  $S$  is the desired number of chunks with a default value of 64, and  $N$  is the file size in bytes (for a complete explanation of the formula and the chosen default values, please see [15]).

$$b = b_{min} \cdot 2^{\lceil \log_2 \left( \frac{N}{S \cdot b_{min}} \right) \rceil} \quad (1)$$

Given that  $b \approx N/S$ , the procedure generates as a result around  $S$  chunks. Once a chunk is identified, a second hash based on the *FNV* algorithm is then used to produce hash values of the content between two consecutive trigger points. Using its last 6 bits, each of those hash values is translated into a Base64 character, so the resulting signature is the concatenation of the single characters generated at all the trigger points (with a maximum of 64 characters per signature). In this way, if a new version of the object is created by localized insertions and deletions, some of the original chunk hashes will be modified, reordered, or deleted, but enough will remain in the new composite hash to identify the similarity.

As the frequency of the trigger points strongly determines how many characters will appear in the signature, at the beginning of its execution the algorithm estimates the value of the block size, which would theoretically produce a signature of around 64 characters. Once the signature is produced, if its length is less than 32 characters *ssdeep* adjusts the block size ( $b \leftarrow b/2$ ) and the algorithm is executed one more time, which generates a new signature. This procedure continues until a signature of at least 32 characters is produced.

In the comparison process, *ssdeep* computes how similar are two files based on their signatures. The similarity measurement that *ssdeep* uses is an edit distance algorithm based on the Damerau-Levenshtein distance [11], [16], which compares the two strings and counts the minimum number of operations needed to transform one string into the other, where the allowed operations are insertions, deletions, and substitutions of a single character, and transpositions of two adjacent characters [13], [33].

In *ssdeep*, insertions and deletions are given a weight of 1, while substitutions are given a weight of 3, and transpositions a weight of 5. As an example, using *ssdeep*'s algorithm the distance between the strings "Saturday" and "Sundays" is 5, as it can be checked with the following steps and the computations of Table 1.

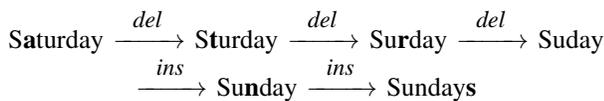


Table 1: *ssdeep* edit distance example.

	S	a	t	u	r	d	a	y	
0	0	1	2	3	4	5	6	7	8
S	1	0	1	2	3	4	5	6	7
u	2	1	2	3	2	3	4	5	6
n	3	2	3	4	3	4	5	6	7
d	4	3	4	5	4	5	4	5	6
a	5	4	3	4	5	6	5	4	5
y	6	5	4	5	6	7	6	5	4
s	7	6	5	6	7	8	7	6	5

A consequence of assigning the weights 3 and 5 to the substitution and transposition operations is that, in practice, the edit distance computed by *ssdeep* only takes into consideration insertions and deletions. In this way, a substitution has a cost of 2 (a deletion plus an insertion) instead of 3, and a transposition has also a weight of 2 (again an insertion and a deletion) instead of 5.

One of the limitations derived from this design is that, given a string, a rotated version of the initial string is credited with many insertion and deletion operations, when in its nature it is basically the same string (i.e., the content is the same, although the order of the substrings is different). Consider for example the strings "1234abcd" and "abcd1234".

The resulting distance is then scaled to produce a score in the range 0-100, where a value of 100 indicates a perfect match and a score of 0 indicates a complete mismatch. There are two conditions that have to be taken into consideration at this point: if the two signatures have a different block size, then the score is automatically set to 0 without performing any additional calculation. Besides, if the minimum length of the longest common substring in the comparison is less than the windows size (7), then *ssdeep* provides a score of 0.

In retrospect, *ssdeep* represented a cornerstone in similarity detection techniques. Its source code is freely available, and there are implementations for Windows and Linux [21], [32]. The latest version of *ssdeep* is 2.10, which was released in July 2013. Even though *ssdeep* is not a multi-threaded program, the author states that the library on which its based can be used in multi-threaded applications [14].

Despite the benefits brought by the release of this program, during the last years some limitations have been brought to attention by different researchers, proposing improvements or even different theoretical approaches (for example, see [1], [2], [4], [5], [8], [10], [24]).

### 3.2 FKSum

An improvement to Kornblum's algorithm was proposed by Chen *et al.* [10]. In their algorithm, *FKSum*, they showed that it is possible to improve the efficiency of *ssdeep*, since until the very last step it does not examine the signature and, if it is too short (i.e., shorter than 32 characters), the file has to be processed again using an adapted block size  $b \leftarrow b/2$ .

As this fact is very likely to happen (they showed that it happens in 38% of the cases), the goal of their modification to the original algorithm was to generate intermediate hashes using numbers in the geometric progression with factor 4 as block size. If the current block size is  $b$ , they perform the same algorithm and compute the hashes with block sizes  $b$  and  $4b$ , counting the trigger points for block sizes  $b$ ,  $2b$ ,  $4b$ , and  $8b$ . As the authors used *FNV* as a homomorphic hash function, it is possible to create the hashes for  $2b$  by using the hashes of  $b$  and hence the process runs more efficiently. The drawback of this approach is that the combination of the hashes might be only possible with the *FNV* hash. If we use any cryptographic hash function such as MD5, we would have to do more computations and the efficiency advantage would no longer be available.

### 3.3 New version of *ssdeep*

Breitinger and Baier [5] discussed the efficiency of *ssdeep*, presenting some enhancements that, in their opinion, would increase the performance of his algorithm by 55% if applied to a real life scenario:

- Each file should be processed only once. As it was proposed by Chen *et al.* [10], they use four different block size values:  $2b$ ,  $b$ ,  $b/2$ , and  $b/4$ .

- Implementation should be flexible in order to be able to change the *PRF* and chunk functions.
- It should be able to determine untypical behaviour of trigger sequences (which may be caused by an active adversary), in order to mitigate the security concerns regarding adversary attacks detected in [2].

Their main idea is to process the file once and count the trigger sequences for all reasonable block sizes (according to Kornblum approach). In the next step, the file is read again and the block size  $b$  is set to the largest value that yields at least 32 signature characters. This fact is the main disadvantage of this proposal, since the file has to be read twice.

An important point related to security is the restriction of the signature length of *ssdeep*. Kornblum forces the resulting hash length to be between 32 and 64, but he does not give any justification for those limits. Even though Breitingner and Baier considered that the upper boundary was a weakness, and as such it was exploited in [2], they maintained both limits.

### 3.4 SimFD

Seo *et al.* [29] developed the SimFD algorithm by combining *ssdeep* with other statistical analysis and improved the false positive rate, but at the cost of efficiency.

Statistical analysis uses byte frequency analysis to detect if a file is similar to the original one. This method generates a result through byte frequency analysis for the original file. As a result of it, the process is performed before any similar file detection task, and some *reference values* are established from the original file for comparing other files. These reference values consist of three types, where each type has a different purpose for detection.

The first reference values are computed for comparing similar files, and they are determined from features of the original file through numerical values obtained from data distribution. The second reference values are computed using metadata, such as file signature, header/footer, null values etc, which was eliminated during the computation of the first reference values. Finally, the third reference values are calculated as a clustering value for all binaries. The clustering scheme is divided into increase, decrease and stagnation for accumulated frequency. The clustering results have the advantage of grasping the distribution type for a file in a character string.

SimFD consists of four modules. First the input module, used by selecting the original copy and the target object. Then the analysis modules, that consists of the *CTPH* analysis module (mainly *ssdeep*) and the statistical analysis module. And finally, the detection result module, that judges final similarity by checking the results of *CTPH* and statistical analysis through the reference values.

### 3.5 md5bloom

Roussev *et al.* [26] proposed a new tool, *md5bloom*, which uses Bloom filters as an efficient tool for fast comparisons. Bloom filters are a space-efficient probabilistic data structure, first introduced by Bloom in [3], and widely used in areas such as network routing and traffic filtering. They allow to test whether an element is a member of a set.

A Bloom filter  $B$  is a representation of a set of  $n$  elements,  $S = \{s_1, \dots, s_n\}$ , taken from a universe  $U$ . The filter consists of an array of  $m$  bits, initially all set to 0. To represent the set of elements, the filter uses  $k$  independent hash functions,  $h_0, \dots, h_{k-1}$ , that produce values in the range of 0 to  $m - 1$ . All hashes are assumed to be independent and to map elements from  $U$  uniformly over the range of the function.

To insert an element  $x$  from  $S$ , each hash function is applied to it, which gives  $k$  values. For each value,  $h_1(x), \dots, h_k(x)$ , the bit with the corresponding number to one is set (setting a bit twice has the same effect as setting it once). To verify if an element  $x$  is in  $S$ , we must hash it with all the hash functions and check the corresponding bits: if all of them are set to one, we return *yes*; otherwise, *no*. The filter will never return a false negative; that is, if the element was inserted, the answer will always be *yes*. However, we could have a false positive for an element that has never been inserted but whose bits have been set by chance by other element insertions. False positives are the price we pay for the compression gains.

As it turns out, the routine use of cryptographic hashes in digital forensics makes it easy to introduce Bloom filters into the process. Instead of computing  $k$  separate hashes, we can take an object's cryptographic hash, split it into several nonoverlapping subhashes, and use them as if different hash functions had produced them. This is the way the *md5bloom* application works: the MD5 function returns 128 bits, any individual bit of the hash value can be viewed as an independent random variable and, by extension, any subset of the 128 hash bits can be selected to produce a value within a desired range. Let the bits in  $h^{md5}$  be numbered 0:127 (we use the notation  $h_{d_1:d_2}$  and the term subhash to denote the selection of bits numbered  $d_1$  through  $d_2$ , inclusively), thus,  $h^{md5} = h_{0:127}$  and can also be expressed as the concatenation of subhashes, for example:

$$h^{md5} = h_{0:15}h_{16:31}h_{32:47}h_{48:63}h_{64:79}h_{80:95}h_{96:111}h_{112:127}$$

### 3.6 MRSH—Multi-Resolution Similarity Hashing

In [27] Roussev *et al.* applied three main changes to the *ssdeep* algorithm:

- 1) First, they stated that it is not necessary to use a cryptographic hash function for the *PRF*. Therefore, instead of using the *PRF* based on *Adler32*, they used the following polynomial hash function, *djb2*:

$$h_0 = 5381; h_{k+1} = 33h_k + c_k \pmod{232}; \text{ for } k \geq 0,$$

where  $c_k$  denotes the  $k^{\text{th}}$  character of the input.

Given that `djb2` has the disadvantage that each window has to be processed from scratch, this change influences negatively the efficiency.

- 2) Second, they changed the hash function for processing each chunk. Instead of the *FNV* hash, they used MD5. Then, the least significant 11 bits of the MD5 output are used as input for a Bloom filter to represent the final signature.
- 3) The next step in the design process is to determine whether the composite hash will be of fixed or variable size. Fixed-size hashes have an obvious appeal (minimum storage requirements and simple management). However, they also have some scalability issues as they limit the ability of the hashing scheme to compare files of varying sizes. `md5bloom`, on its own, has a very similar problem if the attempt is to produce a composite hash which consists of a single filter. Moreover, to compare two filters, they must be of the same size and use the same hash functions. This analysis points out the need to devise a variable-sized hashing scheme that scales with the object size but also maintains a low relative overhead. In this sense, to enable universal comparison of filters Rousev standardized a set of Bloom filters of 256 bytes, 8 bits per element, using four hash functions. To obtain the four hashes, they take the MD5 chunk hash, split it into four 32-bit numbers and take the least significant 11 bits from each part.

With all these design changes, the process of this new algorithm works in the following way:

- 1) A 32-bit `djb2` hash is computed on a sliding window of size 7. At each step, the least significant  $t$  bits of the hash (the trigger) are examined, and if they are all set to 1, a context discovery is declared;  $t$  is the essential parameter that distinguishes the different levels of resolution. For the lowest level 0, the default value is 8.
- 2) Context discovery triggers the computation of the MD5 chunk hash between the previous context and the current one.
- 3) The chunk hash is split into four pieces and four corresponding 11-bit hashes are obtained and inserted into the current Bloom filter. If the number of elements in the current filter reaches the maximum allowed (256), a new filter is added at the end of the list and becomes the current one.
- 4) The hash consists of the concatenation of all the Bloom filters, preceded by their total count.

Even though this modification slows down `ssdeep`, it increases the security aspects, therefore this change is considered to be very useful.

### 3.7 MRSH v2—Multi-Resolution Similarity Hashing, version 2

In [8] Breitinger *et al.* reviewed in terms of efficiency and performance the parameters used in the Multi-Resolution Similarity Hashing function (MRSH) proposed by Rousev (see §3.6) and developed a new version, MRSH v2, which recovers some of the original `ssdeep` parameters, such as *Adler32* and *FNV*.

In order to be more efficient, they decided to use again the original rolling hash (*Adler32*) instead of `djb2`, since it computes the hash value over the 7-byte window in an easier way just by removing the last byte and adding the new one instead of doing seven loops per window. Moreover, as collision resistance is not necessary, the new version makes use of *FNV*, as the original `ssdeep`, instead of MD5. For performance reasons they stated that the minimum block size should be  $b/4$ , which is in line with *FKSum* (see §3.2). Finally, the maximum number of elements has been changed to 160 and 5 subhashes are used. The maximum is therefore 800 bits, so one Bloom filter could represent approximately 40,960 bytes. In order to insert the chunk hash value into a Bloom filter, they used the least significant  $k \cdot \log_2(m)$  bits (MRSH divides the chunk hash values).

Additionally, they demonstrated that the algorithm is compliant with the five properties that in their opinion a SPHF should have, namely: compression, ease of computation, similarity score, coverage, and obfuscation resistance.

## 4. Statistically-Improbable Features

This approach is based on the idea that finding similarities between two objects can be understood as identifying a set of features in each of the objects and then comparing the features themselves. A feature in this context is simply a sequence of consecutive bits selected by some criterion from the object.

Rousev [23] uses entropy as the way of finding statistically-improbable features and measures the false positive range for different kind of files (doc, xls, txt, html, pdf, etc.). With this idea, he proposed a new algorithm, called `sdhash`, whose goal is to pick object features that are least likely to occur in other data objects by chance.

Instead of dividing an input into pieces, `sdhash` identifies statistically-improbable features using an entropy calculation. These characteristic features, forming a sequence of length 64 bytes, are then hashed using the cryptographic hash function SHA-1 and inserted into a Bloom filter. Hence, files are similar if they share identical features.

A security analysis was performed by Breitinger *et al.* in [9], finding some bugs in the implementation as well as showing some possible attacks to circumvent the algorithm and analyzing the resistance of the parameters designed. Another analysis [7] in terms of measuring the compression, ease of computation, coverage and similarity score

showed different weaknesses of `sdfhash`, e.g. there is no full coverage (a change up to 20% of the input does not alter the fingerprint) and that the chosen design of the comparison function is made for fragment detection but not for comparing two files.

Moreover, in [24] Roussev performs a comparison between `sdfhash` and `ssdeep`, analyzing two different experiments (random files and real files) with three different scenarios (embedded object detection, single-common-block file correlation, and multiple-common-blocks file correlation) concluding that `sdfhash`'s accuracy and scalability outperforms `ssdeep`.

Finally, in [25] the `sdfhash` basic algorithm was made scalable by parallelizing it. The new modification was called `sdfhash-dd` and to reach this objective, some chain dependencies among the Bloom filter component filters were moved away in order to allow concurrent generation. The idea was to split the target into blocks of fixed size and run the signature generation in block-parallel fashion.

## 5. Block-Based Rebuilding

There are mainly three algorithms, `SimHash`, `mvHash-B`, and `bbHash`, which make use of external auxiliary data or blocks that can be chosen randomly, uniformly or as a fixed base, in order to rebuild a file. The process compares the bytes of the original file to the auxiliary data and calculates the differences between them (e.g., using the Hamming distance).

### 5.1 SimHash

Sadowski *et al.* presented an algorithm, called `SimHash` [28], which preselects 16 blocks of 8 bits each in order to find matches by scanning and comparing the original file to these blocks. When a match is found, it is stored in a sum table and then the hash key is computed as a function of the sum entries. Another function, called `SimFind`, identifies the files with key values within a certain threshold of a particular file, then performs a pairwise comparison among the sum table entries to return a filtered selection of similar files.

They performed some experiments using a Uniform Key which has all the 16 blocks weighted equally and a Skew Key which has uneven weights in 4 of the blocks.

### 5.2 mvhash-B

The `mvhash-B` function was described by Breitinger *et al.* [4], having three phases to create the fingerprint:

- 1) First, majority voting is used to map every byte of the input file to either 0x00 or 0xFF. Majority voting in this case means counting the amount of 0s/1s in the neighborhood of the currently processed input byte. If the neighborhood is crowded by 1s, the majority vote yields an output 0xFF and vice versa.
- 2) Next, Run Length Encoding (RLE) compresses these sequences of 0x00s or 0xFFs bytes.

- 3) Finally, the RLE sequence is inserted into Bloom filters to represent the actual fingerprint.

By design, `mvHash-B` aims at having a fingerprint length of 0.5% of the input length, but a drawback of this implementation is the dependence on the file type: each file type requires its own configuration (no standard configuration works for all file types). In other words, although `mvhash-B` works on the byte level, it needs different configurations.

### 5.3 bbHash

Another example of this way of finding similar files is the `bbHash` function, designed by Breitinger *et al.* in [6]. Their new fuzzy hashing technique is based on two concepts:

- *Deduplication*: is a backup scheme for saving files efficiently because instead of saving it completely, it makes use of small pieces. If two files share a common piece, it is only saved once, but referenced for both files.
- *Eigenfaces*: they are used in biometrics for face recognition, for example by representing any face as a combination of a set of  $N$  eigenfaces previously selected.

In this algorithm, they use a fixed set of  $N$  random byte sequences called building blocks of 128 bytes. The process is to slide through the file byte-by-byte and compute the Hamming distance of all building blocks against the current input sequence. If the building block with the smallest Hamming distance is smaller than a certain threshold, its index contributes to the files's hashing result. The disadvantage of this algorithm is that its runtime is high.

## 6. Conclusions

Breitinger and Baier presented a list of four general properties for SPHF [7], which they later extended to the following five general characteristics [8]:

- 1) *Compression*: the output must be much smaller than the input for space-saving and performance reasons.
- 2) *Ease of computation*: generating the hash value of a given file and making comparisons between files must be a fast procedure.
- 3) *Similarity score*: the comparison function must provide a number which represents a matching percentage value.
- 4) *Coverage*: every byte of the input must be used to calculate the hash value.
- 5) *Obfuscation resistance*: it must be difficult to obtain a false negative/false positive result, even after manipulating the input data.

Nevertheless, after analysing the characteristics of these functions, we have been able to identify a list of additional specific features that any SPHF should provide, either by improving a feature already existing or implementing it for the first time. These additional features are:

- *Generate more realistic results consistent with the content of the files compared:* this requirement implies the existence of a simple and clear definition of the concept of similarity, and how to express it as a number.
- *Detect content rotation:* by rotation we mean moving some part of the end of the document to the beginning (or vice versa). A visual examination of a pair of such rotated files would provide a result close to 100, as both files have the same content.
- *Detect content swapping:* by content swapping we mean taking some portion of the document and moving it into a different location, so graphically it could be seen as moving data blocks inside the document. For a file where several swaps have been made, the result should be close to 100, as again the content of both files is basically the same.
- *Compare files of different sizes without limit:* in some cases, it is necessary to compare files of very dissimilar sizes (e.g., considering a book, this could be seen as comparing one chapter with ten chapters in order to detect plagiarism).
- *Avoid insertion attacks:* there are two ways by which a user could alter the comparison results. He could repeatedly insert a specific byte string at the beginning of the file, or could insert a specific byte string scattered along the document (not necessarily at the beginning). Any new piecewise hashing application should try to provide countermeasures to ameliorate the effects of this type of attacks, at least to some extent.

## Acknowledgements

This work has been partially supported by the Ministerio de Ciencia e Innovación (España) under the project TIN2011-22668.

## References

- [1] K. Astebøl, "mvHash—A new approach for fuzzy hashing," Master's thesis, Gjøvik University College, 2012.
- [2] H. Baier and F. Breitingner, "Security aspects of piecewise hashing in computer forensics," in *Sixth International Conference on IT Security Incident Management and IT Forensics (IMF 2001)*, 2011, pp. 21–36.
- [3] B. H. Bloom, "Space time tradeoffs in hash coding with allowable errors," *Communications of the ACM*, vol. 13, no. 7, 1970.
- [4] F. Breitingner, K. Astebøl, H. Baier, and C. Busch, "mvhash-b - A new approach for similarity preserving hashing," in *Seventh International Conference on IT Security Incident Management and IT Forensics (IMF 2013)*, 2013, pp. 33–44.
- [5] F. Breitingner and H. Baier, "Performance issues about context-triggered piecewise hashing," in *Proc. of 3rd ICST Conference on Digital Forensics & Cyber Crime (ICDF2C)*, vol. 3, 2011.
- [6] —, "A fuzzy hashing approach based on random sequences and hamming distance," in *7th annual Conference on Digital Forensics, Security and Law (ADFSL 2012)*, 2012.
- [7] —, "Properties of a similarity preserving hash function and their realization in sdhash," in *Information Security for South Africa (ISSA 2012)*, 2012, pp. 1–8.
- [8] —, "Similarity preserving hashing: Eligible properties and a new algorithm mrsh-v2," in *Digital Forensics and Cyber Crime*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, M. Rogers and K. Seigfried-Spellar, Eds. Springer Berlin Heidelberg, 2013, vol. 114, pp. 167–182.
- [9] F. Breitingner, H. Baier, and J. Beckingham, "Security and implementation analysis of the similarity digest sdhash," in *1st International Baltic Conference on Network Security & Forensics (NeSeFo 2012)*, 2012.
- [10] L. Chen and G. Wang, "An efficient piecewise hashing method for computer forensic," in *Proc. of Workshop on knowledge discovery and data mining*, IEEE, Ed., 2008, pp. 635–638.
- [11] F. J. Damerau, "A technique for computer detection and correction of spelling errors," *Communications of the ACM*, vol. 7, no. 3, pp. 171–176, 1964.
- [12] N. Harbour, "Dcfddd. defense computer forensics lab," 2002. [Online]. Available: <http://dcfddd.sourceforge.net>
- [13] M. Karpinski, "On approximate string matching," *Lecture Notes in Computer Science*, vol. 158, pp. 487–495, 1983.
- [14] J. Kornblum, "ssdeep 2.10 released." [Online]. Available: <http://jessekornblum.livejournal.com/293679.html>
- [15] —, "Identifying almost identical files using context trigger piecewise hashing," *Digital Investigation*, vol. 3(S1), pp. 91–97, 2006.
- [16] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Soviet Physics Doklady*, vol. 10, no. 8, pp. 707 – 710, 1966. [Online]. Available: <http://profs.sci.univr.it/~liptak/ALBioinfo/files/levenshtein66.pdf>
- [17] A. Menezes, P. van Oorschot, and S. Vanstone, *Handbook of applied Cryptography*. Boca Raton, FL: CRC Press, 1997.
- [18] NIST. National Software Reference Library. [Online]. Available: <http://www.nsl.nist.gov>
- [19] —, "SHA-3 competition," National Institute of Standards and Technology, 2012, <http://csrc.nist.gov/groups/ST/hash/sha-3/index.html>.
- [20] —, "The Keccak sponge function family," 2013, [http://keccak.noekoon.org/specs\\_summary.html](http://keccak.noekoon.org/specs_summary.html).
- [21] Python Software Foundation. (2013) ssdeep python wrapper. [Online]. Available: <https://pypi.python.org/pypi/ssdeep>
- [22] R. L. Rivest, "The MD5 message digest algorithm," 1992, request for comments (RFC 1321), Internet Activity Boards, Internet Privacy Task Force.
- [23] V. Roussev, "Building a better similarity drap with statistically improbable features," in *Proc. of 42 Hawaii International Conference on System Science*, 2009, pp. 1–10.
- [24] —, "An evaluation of forensic similarity hashes," *Digital Investigation*, vol. 8, Supplement, no. 0, pp. 34 – 41, 2011.
- [25] —, "Scalable data correlation," in *Proc. of International Conference on Digital Forensics (IFIP WG 11.9)*, 2012.
- [26] V. Roussev, Y. Chen, T. Bourg, and G. Richard, "Md5bloom: Forensic filesystem hashing revisited," *Digital Investigation*, vol. 3, pp. 82–90, 2006.
- [27] V. Roussev, G. Richard, and L. Marziale, "Multi-resolution similarity hashing," *Digital Investigation*, vol. 4, Supplement, no. 0, pp. 105 – 113, 2007.
- [28] C. Sadowsky and G. Levin, "Simhash: Hash-based similarity detection," Tech. Rep., 2007. [Online]. Available: <http://simhash.googlecode.com/svn/trunk/paper/SimHashWithBib.pdf>
- [29] K. Seo, K. Lim, J. Choi, K. Chang, and S. Lee, "Detecting similar files based on hash and statistical analysis for digital forensic investigation," in *2nd International Conference on Computer Science and its Applications (CSA '09)*, 2009, pp. 1–6.
- [30] A. Tridgell, "Efficient algorithms for sorting and synchronization," Master's thesis, The Australian National University. Department of Computer Science, Canberra, Australia, 1999.
- [31] —, "Spamsun readme," 1999. [Online]. Available: <http://samba.org/ftp/unpacked/junkcode/spamsun/README>
- [32] —. (2013) Getting started with ssdeep. [Online]. Available: <http://ssdeep.sourceforge.net/usage.html>
- [33] R. A. Wagner and M. J. Fischer, "The string-to-string correction problem," *Journal of the ACM*, vol. 21, no. 1, pp. 168–173, 1974.

# VoIP Forgery Detection

Satish Tummala, Yanxin Liu and Qingzhong Liu

Department of Computer Science  
Sam Houston State University  
Huntsville, TX, USA

Emails: [sct137@shsu.edu](mailto:sct137@shsu.edu); [yanxin@shsu.edu](mailto:yanxin@shsu.edu); [liu@shsu.edu](mailto:liu@shsu.edu)

**Abstract-** With the rapid increase in low-cost and sophisticated digital technology the need for techniques to authenticate digital material will become more urgent. Inspired by the image forgery detection, we develop a method to detect the forgery in VoIP audio streams by checking the offset differential features with learning machine. Our experimental results on Speex VoIP streams show that our approach is promising to expose the forgery manipulation in VoIP audio streams.

**Keywords-** Multimedia forensics, forgery detection, SVM classifier, speex, VoIP

## 1. Introduction

Recently, there has been a tremendous increase in the use of VoIP applications. In some cases, they may be submitted as digital evidence. If these files are being submitted as digital evidence, then there is a need to authenticate such material. In order to authenticate such material, there is an urgent need to develop authentication methods for VoIP files. Digital watermarking and signature are the two typical technologies for digital multimedia authentication. These two techniques need some side information such as a signature or digital watermark at the time of detection. Since, it is impossible to retrieve any available side information from the questionable data; these methods are useless in many real applications.

In the last several years, multimedia forensics has been an emerging research field of information security, because it doesn't need any side information like digital watermark or signature during detection. What it needs is the features within the multimedia. These features can further be analyzed so as to provide forensics information on how this data is acquired and processed.

The paper is organized as follows. First, a brief overview of VoIP and the tools used in this project has been presented. Next to it, the procedure of creating a forgery database of VoIP has been described. After that, the feature extraction module and the machine learning module have been explained. The experimental results have been discussed. Finally, a conclusion along with future works has been presented.

## 2. Related Work

There are few works on authentication for digital audio. A technique for detecting digital audio forgeries by checking frame offsets [1, 16] has been proposed, based on the assumption that forgeries lead to the broken frame grids. Many Farid [2] used bispectral analysis to detect digital forgery in speech signals, based on an idea that forgery in speech would introduce unnatural correlations. Recently, the detection of doubly compressed MP3 audio streams has been in-depth investigated [17, 18].

Grigoras [3] reported that the Electronic Network Frequency Criterion can be used as a means of assessing the integrity of digital audio evidence. It could be used to verify the exact time when a digital audio was created. This could be done by comparing the ENF of audio recording with a reference frequency database from the electric company or the laboratory. Dittmann et al. [4] proposed that the authenticity of the speaker's environment could be determined by extracting the background features of an audio stream. These features provide information for determining its origin location and the used microphone.

There are still no passive authentication methods focusing especially on VoIP. This issue needs to be addressed, because a lot of applications now-a-days are based on VoIP. The audio forgery detection methods described above cannot be directly applied to VoIP, because there is a vast difference in the process of encoding between audio and VoIP. But the ideas of these methods can be improved so as to apply them to VoIP files.

## 3. Speex VoIP Codec

Speex is an open source audio compression format, based on CELP and is designed to compress voice at bitrates ranging from 2 to 44 kbps. [9]. It is part of the GNU Project [10] and is available under the revised BSD license [11]. The best part of using speex is that it is free compared to some other expensive proprietary speech codecs. It has a lot of useful features that are not present in many other speech codecs, which makes it well adapted to internet applications especially VoIP. Some of Speex's features include:

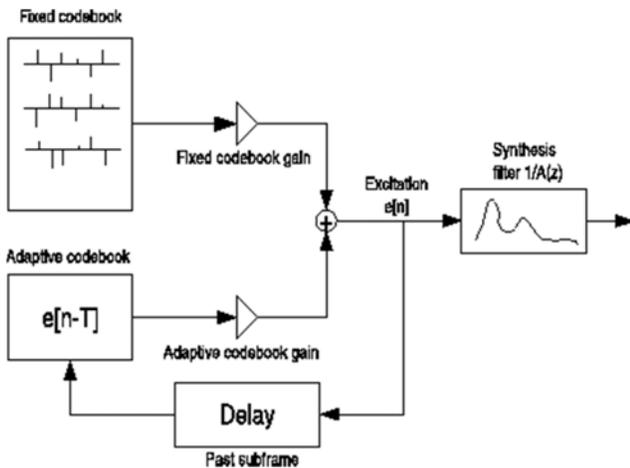
- Narrowband (8 kHz), wideband (16 kHz), and ultra-wideband (32 kHz) compression in the same bitstream
- Intensity stereo encoding
- Packet loss concealment
- Variable bitrate operation (VBR)

- Voice Activity Detection (VAD)
- Discontinuous Transmission (DTX)
- Fixed-point port
- Acoustic echo canceller
- Noise suppression

Speex is based on CELP (Code Excited Linear Prediction) [12]. The techniques are based on the following ideas:

1. Use of a Linear Prediction Model
2. Use of codebook entries as input of the Linear Prediction Model
3. Search performed in a closed loop in a perceptually weighted domain

CELP also utilizes the source-filter model for speech prediction. It assumes that the vocal cords are the source of spectrally flat sound, and that the vocal tract acts as a filter to spectrally shape the various sounds of speech i.e., the source and filter are totally independent of each other. This model is mainly used because of its simplicity. Also, this model is usually tied with the use of linear prediction, illustrated by Figure 1.



**Figure 1. Illustration of CELP**

In what follows we briefly describe the techniques used in CELP.

### Linear Prediction (LPC)

It is at the base of CELP. The basic idea behind it is that it predicts the signal  $x[n]$  using a linear combination of its past samples.

$$y[n] = \sum_{i=1}^N a_i x[n - i]$$

where  $y[n]$  is the linear prediction of  $x[n]$ . The prediction error is thus given by:

$$e[n] = x[n] - y[n] = x[n] - \sum_{i=1}^N a_i x[n - i]$$

The goal of the LPC analysis is to find the best prediction coefficients  $a_i$  which minimizes the quadratic error function.

$$E = \sum_{n=0}^{L-1} [e[n]]^2 = \sum_{n=0}^{L-1} [x[n] - \sum_{i=1}^N a_i x[n - i]]^2$$

That can be done by making all derivatives equal to zero

$$\frac{\partial E}{\partial a_i} = \frac{\partial}{\partial a_i} \sum_{n=0}^{L-1} [x[n] - \sum_{i=1}^N a_i x[n - i]]^2 = 0$$

For an order N filter, the filter coefficients  $a_i$  are found by solving the system  $N \times N$  linear system  $Ra = r$ , where

$$R = \begin{matrix} R(0) & R(1) & \dots & R(N-1) \\ R(1) & R(0) & \dots & R(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ R(N-1) & R(N-2) & \dots & R(0) \end{matrix}$$

$$r = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(N) \end{bmatrix}$$

with  $R(m)$ , the auto-correlation of the signal  $x[n]$ , computed as

$$R(m) = \sum_{i=0}^{N-1} x[i]x[i - m]$$

### Pitch Prediction

Pitch prediction is used in most speech codecs. Here, we find a period (because the signal is periodic) that looks similar to the current frame i.e., the pitch predictor look for similar patterns outside the current frame. The pitch period is encoded along with a prediction gain.

$$e[n] \cong p[n] = \beta e[n - T]$$

where  $T$  is the pitch period,  $\beta$  is the pitch gain.

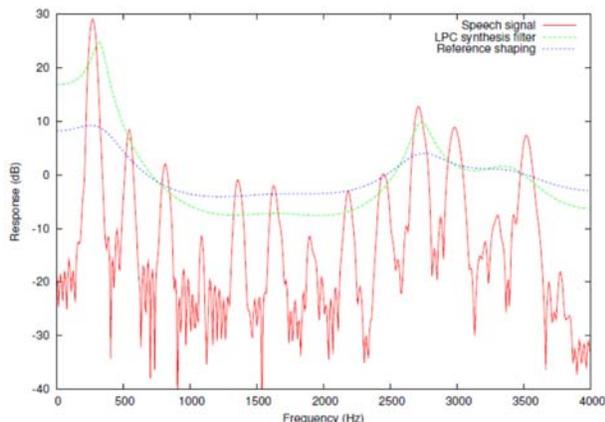
### Innovation Codebook

The final excitation will be the sum of the pitch prediction and an innovation signal taken from a fixed codebook, hence the name Code Excited Linear Prediction.

$$e[n] = p[n] + c[n] = \beta e[n - T] + c[n]$$

*Noise Weighting*

Most of the modern speech codecs shape the noise so that it appears mostly in the frequency regions where the ear cannot detect it. In order to maximize the speech quality, CELP codecs minimize the mean square error in the perceptually weighted domain.



**Figure 2. Standard Noise Shaping in CELP**

Analysis-by-Synthesis

This principle means that the encoding is performed by perceptually optimizing the decoded signal in a closed loop. The best CELP stream would be produced by trying all possible bit combinations and selecting the one that produces the best-sounded decoding signal.

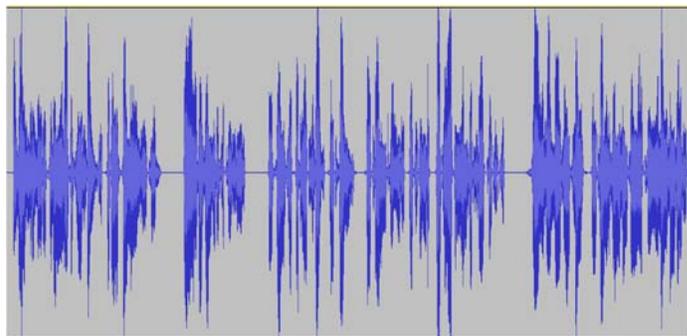
**4. Speex VoIP Tampering**

We have created a VoIP audio database containing 1000 files by using audacity [13] and speex [9]. While we create the forgery, we decode the VoIP file into temporal domain, and manipulate the file in temporal domain, and then encoded the doctored file to speex format. Here we show an original file and the tampering, shown in Figures 3 and 4.

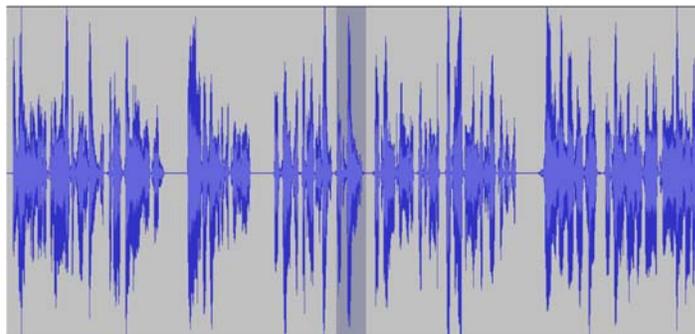
The original audio file contains the voice, “A Rose is a woody perennial of the genus Rosa, within the family Rosaceae. There are over 100 species. They form a group of erect shrubs, and climbing or trailing plants, with stems that are often armed with sharp prickles. Flowers are large and showy, in colors ranging from white through yellows and reds.”

The doctored audio file means: “A Rose is a woody perennial of the genus Rosa, within the family Rosaceae. There are over 100 species. They form a group of erect shrubs or trailing plants, with stems that are often armed with sharp prickles. Flowers are

large and showy, in colors ranging from white through yellows and reds.”



**Figure 3. The original audio stream**



**Figure 4. The audio tampering with the shading removed**

If you observe the above forged sample, the words ‘and climbing’ have been trimmed from the original VoIP file. In the above figure, the highlighted part has been trimmed.

**5. VoIP Forgery Detection**

5.1. Feature Mining Based on Shift-Recompression

To detect the VoIP forgery, inspired by the previous work in detecting image forgery and image steganography in JPEG format [5, 6, 7, 8], we design an algorithm to extract the differential features based on shift recompression, described as follows:

***Shift-Recompression-based Differential Feature Extraction***

1. Decode the examined VoIP audio stream to temporal domain, denoted by a vector  $S(i)$  ( $i=0,1, 2, \dots, M$ );
2. Shift the matrix  $S(i)$  by  $t$  samples in the temporal domain,  $t \in \{1, 2 \dots, N - 1\}$ , here  $N$  stands for the number of samples in a frame/block. For speex VoIP audio signal, a block consists of 160 samples ( $N = 160$ ). A shifted temporal WAV signal  $S'(i, t)$  is produced.  $S'(i, t) = S(i-t)$ ,  $i= t, t+1, t+2, \dots, M$ ;
3. For  $t=1:159$

- 3.1 Encode the shifted temporal signal  $S'(i, t)$  to speex VoIP audio stream at the same bit rate;
- 3.2. Decode the encoded audio signal from the above step to temporal domain, denoted by  $S''(i, t)$ ;
- 3.3. Calculate the difference  $D(i, t) = S'(i, t) - S''(i, t)$ ;
- 3.4. Shift-recompression based reshuffle characteristic features are given by:

$$SRSC(t) = \frac{\sum_i |D(i, t)|}{\sum_i |S'(i, t)|} \quad (1)$$

Where  $t = 1, 2, \dots, 159$ . There are 159 features for a speex VoIP audio file.

### 5.2. SVM

LibSVM [14] is being uses for support vector machine classification in this project. Support vector machines are a relatively new learning method used for binary classification [15]. The basic idea is to find a hyperplane which separates the d-dimensional data perfectly into its two classes. However, since example data is not often linearly separable, SVM's introduce the notion of a "kernel induced feature space" which casts the data into a higher dimensional space where the data is separable. Typically casting into such a space would cause problems computationally, and with overfitting. The key insight used in SVM's is that the higher dimensional space doesn't need to be dealt with directly, which eliminates the above concerns.

Suppose we are given data points each of which belong to one of two classes.

$$D = \{(X_i, C_i) | X_i \in \mathbb{R}^p, C_i \in \{-1, 1\}\}_{i=1}^n$$

In this method, the differential values are considered as vector  $x_i$ .  $c_i$  represents whether the given file is forged or not. Under this environment, the SVM classifier attempts to maximize the geometric margin which is the distance from the hyperplane to the closest instances on either side. The hyperplane can be written as the set of points  $x$  satisfying  $w \cdot x - b = 0$ . To maximize the margin of separation, two hyperplanes are represented by the following equations

$$W \cdot X_i - b \geq 1, \text{ for } X_i \text{ of the first class}$$

$$W \cdot X_i - b \leq 1, \text{ for } X_i \text{ of the second class}$$

Or equivalently

$$c_i(W \cdot X_i - b) \geq 1, \text{ for all } 1 \leq i \leq n$$

The figure below demonstrates the SVM classifier briefly. The points on the line  $w \cdot x - b = -1$  and  $w \cdot x - b = 1$  are called the support vectors.

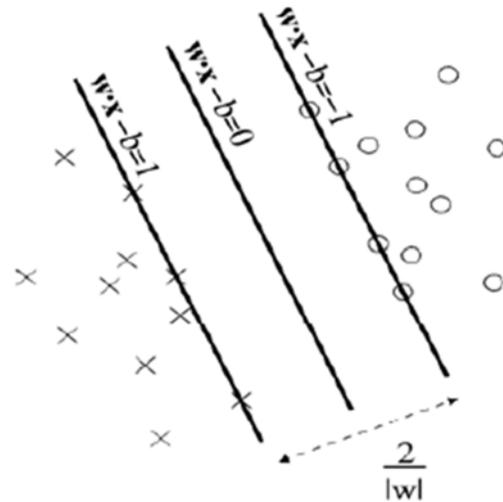


Figure 5. SVM illustration

$\|w\|$  should be minimized because the distance between the two hyperplanes is  $2/\|w\|$ . Therefore, the optimization problem is

$$\text{chose } X, b \text{ to minimize } \|W\|$$

$$\text{subject to } c_i(W \cdot X_i - b) \geq 1, \text{ for all } 1 \leq i \leq n$$

By substituting  $\frac{1}{2}\|w\|^2$  in place of  $\|w\|$ , the problem to find  $w$  and  $b$  becomes a quadratic programming optimization problem as follows

$$\text{minimize } \frac{1}{2} \|W\|^2, \text{ subject to } c_i(W \cdot X_i - b) \geq 1, \text{ for all } 1 \leq i \leq n$$

## 6. Experiments

In each experiments we randomly select 500 untouched VoIP audio files and 500 tampered VoIP files for training, and other 500 untouched and 500 tampered VoIP files for testing. The experiment has been repeated for 100 times and the detection accuracy is the mean of 100 experiments. Due to the computational cost, we only test the forgery manipulations misplaced by 30, 70, and 125 samples

The results to demonstrate the performance of this forgery detection method based on the SVM classifier and the differential values are given in Table 1.

Table 1. The mean detection accuracy

OFFSET (in samples)	DETECTION RATE
30	82.1%
70	79.8%
125	82.1%

## 7. Conclusion

While VoIP-based services are favorably disseminated in our real-life, there is an increasing challenge to detect the tampering in VoIP audio streams. To this date, there is no such an effective detection of the tampering. In this paper, inspired by the success in image forgery detection and steganalysis, a shift-recompression differential feature analysis is designed to detect VoIP audio forgery with the aid of learning machine. Our preliminary experimental results demonstrate the effectiveness of proposed method.

The future study will be focused on the improvement of proposed method and applied the improvement to other formats of VoIP forgery detection.

### REFERENCES

- [1] Yang R, Qu Z and Huang J, "Detecting digital audio forgeries by checking frame offsets". *Proc. 10th ACM Workshop on Multimedia and Security*, pages: 21-26, 2008.
- [2] Farid, Hany, "Detecting Digital Forgeries Using Bispectral Analysis," in *MIT AI Memo AIM-1657*, MIT, 1999.
- [3] Grigoras C, "Digital Audio Recording Analysis: The Electric Network Frequency (ENF) Criterion," *The International Journal of Speech Language and the Law*, vol. 12, no. 1, pp. 63-76, 2005.
- [4] Kraetzer C, Oermann A, Dittmann J and Lang A, "Digital Audio Forensics: A First Practical Evaluation on microphone and Environment Classification," in *ACM MMSEC*, pp 63-74, Dallas, 2007.
- [5] Liu Q, Sung AH and Qiao M, "Neighboring joint density based JPEG steganalysis". *ACM Transactions on Intelligent Systems and Technology*, 2(2), 16:1-16, 2011.
- [6] Liu Q, Sung AH and Qiao M, "Derivative based audio steganalysis". *ACM Transactions on Multimedia Computing, Communications and Application*, 7(3), 18:1-19, 2011.
- [7] Liu Q, "Steganalysis of DCT-embedding-based Adaptive Steganography and YASS", *Proc. 13th ACM Workshop on Multimedia and Security*, pp. 76-85, 2011.
- [8] Liu Q, "Detection of misaligned cropping and recompression with the same quantization matrix and relevant forgery", *Proc. 3rd ACM Workshop on Multimedia in Forensics and Intelligence*, pp. 25-30, 2011.
- [9] "Speex," [Online]. Available: <http://www.speex.org/>.
- [10] "GNU Project," [Online]. Available: <http://www.gnu.org/>.
- [11] "BSD License," [Online]. Available: <http://www.xiph.org/licenses/bsd/speex/>.
- [12] "CELP," [Online]. Available: <http://www.speex.org/docs/manual/speex-manual/node9.html>.
- [13] "Audacity," [Online]. Available: <http://audacity.sourceforge.net/>.
- [14] "LibSVM," [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/#matlab>.
- [15] "Support Vector Machine," [Online]. Available: [http://en.wikipedia.org/wiki/Support\\_vector\\_machine](http://en.wikipedia.org/wiki/Support_vector_machine)
- [16] Yang R, Qu Z and Huang J, "Exposing MP3 audio forgeries using frame offsets", *ACM Transactions on Multimedia Computing, Communications and Applications*, vol 8, issue 2S, September 2012.
- [17] Liu Q, Sung AH and Qiao M, "Detection of double MP3 compression", *Cognitive Computation* 2(4): 291-296, 2010.
- [18] Qiao M, Sung AH and Liu Q, "Revealing real quality of double compressed MP3 audio", *Proc. ACM Multimedia 2010*, pages 1011-1014, 2010.

# A Low-Complexity Procedure for Pupil and Iris Detection Suitable for Biometric Identification

V. Gayoso Martínez, F. Hernández Álvarez, and L. Hernández Encinas

Information Processing and Cryptography (TIC), Institute of Physical and Information Technologies (ITEFI)  
Spanish National Research Council (CSIC), Madrid, Spain

**Abstract**—*The goal of any biometric system consists in identifying individuals based on a certain characteristic possessed by the persons under examination. Among them, iris recognition is regarded as one of the most reliable and accurate biometric identification systems currently available. Most commercial iris recognition products use patented algorithms, which forces open source developers to design and use alternate algorithms. In this contribution, we propose two low-complexity methods for detecting and isolating the pupil and the iris using a greyscale image as the input data. The proposed algorithms can be easily implemented in any device, as they do not use complex operations or image transforms. In addition to that, we show a performance comparison which includes an implementation of our proposed algorithms and two other open source solutions.*

**Keywords:** Biometric Identification, Iris, Pupil, Java, Security

## 1. Introduction

With the ever-growing need for reliable authentication mechanisms, biometric systems have experienced a great development in recent years. Due to their capability to perform the automatic and instant verification of an individual based on some specific physical characteristic, biometric security is now in a privileged position regarding other authentication solutions [1], [2].

There are several biometric technologies, based on different physical features: fingerprints, hand geometry, retina or iris scan, face recognition, voice analysis, etc. Among them, iris recognition has become one of the leading biometric technologies in our society, as the physical patterns of the iris are unique and can be obtained from some distance using the proper equipment [3].

John Daugman is credited as the developer of the first algorithms for iris recognition [4], and in 1994 he patented some of the algorithms that conform its foundations [5]. Even though Daugman's patent expired in 2011 [6], most commercial biometric systems use other similarly patented algorithms, making necessary to develop alternative methods when implementing iris recognition applications.

In this contribution, we present a low-complexity method for locating the pupil and the iris in an image, so it can be processed afterwards for obtaining its associated code (also

known as the iris template). We have implemented the proposed algorithms in Java and have compared its performance using two already existing open source solutions. The results from our tests allow us to state that our algorithms are faster and provide better identification rates than the other solutions considered.

This paper is organized as follows: Section 2 offers a brief introduction to iris identification. Section 3 presents two open source solutions available on the internet. Section 4 includes a complete description of our algorithms for pupil and iris location. In Section 5, we describe the Java application developed by us that implements the proposed algorithms. Section 6 contains the tests results of the three applications with the same iris database. Finally, Section 7 includes our conclusions about the new algorithms.

## 2. Foundations of iris identification

In the first layers of the eye there are different elements: cornea, iris, pupil, sclera, etc. The purpose of the iris is to constrict or enlarge the aperture of the pupil. By doing this, it determines the amount of light that enters the pupil [7]. Figure 1 shows the elements present in a front view of the eye.



Fig. 1: View of the human eye (source: [8]).

The algorithms developed by Daugman for locating the iris and generating its template have greatly influenced the developments in this field, to the point that most implementations follow the scheme devised by him, and which consists in the following steps [9]:

- 1) Image acquisition: capture of an image of the user's eye.

- 2) Pre-processing: detection of the iris prior to obtaining its normalized version, which is the result of transforming the image from polar to Cartesian coordinates.
- 3) Template generation: computation of the iris code based on the characteristics of the iris.
- 4) Feature comparison: calculation of a similarity score by comparing the user's iris code to other templates.

One of the most important prerequisites for performing authentications based on the iris consists in developing and using robust methods for the automatic detection and isolation of the pupil and the iris. Due to its regular size and uniform dark shade, the pupil is relatively easy to locate. However, locating the iris is not a trivial task due to its irregular pattern, the obstruction of the iris by the eyelids presented in some images, and the relative similarity of the iris and the sclera near the outer iris boundary [10].

### 3. Iris recognition applications

In this section we describe two iris recognition applications whose source code can be freely downloaded and inspected by developers.

#### 3.1 Imperial College

The *Iris Recognition Application* from Project Iris is a free, open-source, cross-platform application developed using C++ and the Qt framework [11]. The application is the work of five Computing Science students at Imperial College London working under the supervision of Professor Duncan Gillies, and the source code is freely available under the GNU General Public Licence (GPL). Figure 2 shows a screenshot of the application after it has processed the image of an eye and located the pupil, the iris, and the eyelids.

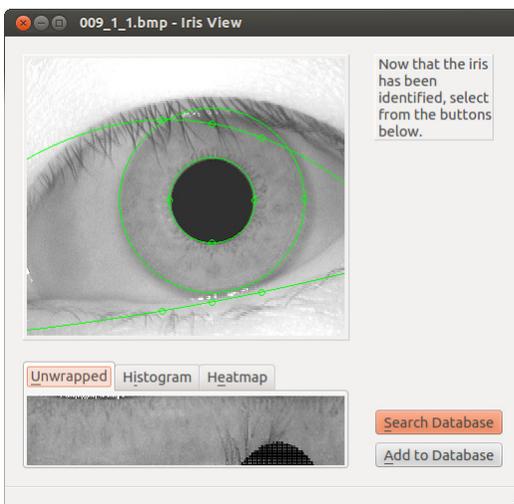


Fig. 2: Iris Recognition Application.

As it is stated in the project's final report [10], its aim was to implement a working prototype of the techniques and

methods used for iris recognition, and to test these methods on one of the databases of irides provided by the Chinese Academy of Sciences' Institute of Automation (CASIA) [12].

The application automatically detects the pupil and the iris by first removing noise with a median filter, and then applying the Sobel operator and the Hough transform for edge detection. In addition to that, it allows users to manually locate the pupil and the iris using the mouse.

The authors state that in virtually any case where the iris boundary is correctly located by the program, the iris is subsequently identified, being a solid evidence the lack of false matches after 35,000 comparisons using the CASIA database [10].

As a disadvantage, their implementation provides an estimation of the location of the iris based on concentricity with the pupil which, among other factors, lowers the detection rate to around 70% [10].

#### 3.2 Warsaw University of Technology

The *Iris Recognition* software, developed by Bernard Kobos and Piotr Zaborowski at the Warsaw University of Technology, is a freely available Java application [13] based on the work of Libor Masek presented in his Thesis [14].

The application is able to locate the iris and the pupil, excluding eyelids, eyelashes, and reflections. In order to do that, it uses the Hough transform before normalizing and filtering the iris region using a 1D Log-Gabor filter. The phase data associated to the iris patterns is then extracted and quantised with two bits per working point [13].

Regarding the comparison capabilities, this software uses the Minkowski distance. Figure 3 shows a screenshot of the application after detecting and isolating the iris and the pupil of one of CASIA's test subjects.

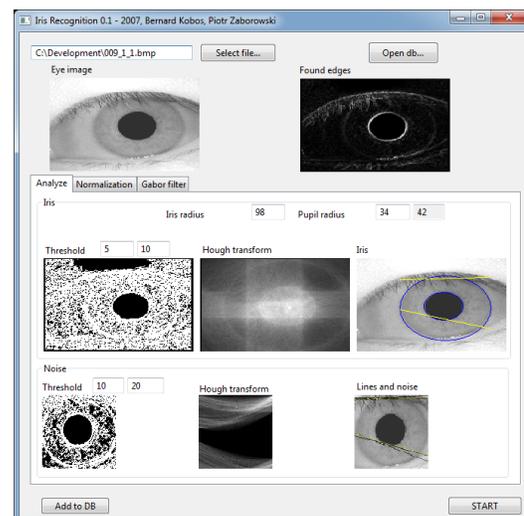


Fig. 3: Iris Recognition software.

## 4. Iris and pupil detection algorithms

After using the two applications commented in Section 3, we decided to create new algorithms for locating the pupil and the iris. As our detection procedures do not use image transforms or complex operations, they can be implemented in a broad range of devices with different computing capabilities.

Before presenting the algorithms, it is necessary to take into account the differences between the Cartesian system of coordinates and the system typically used in computers graphics, both represented in Figure 4. In the Cartesian system, coordinates increase to the right and up along the  $x$  and  $y$  axis, respectively. In comparison, computer graphics use by convention a coordinate system where the origin is in the upper-left corner, and the direction of the positive  $y$ -axis is downwards. Using this scheme, distances are measured in pixels, which are always integer values.

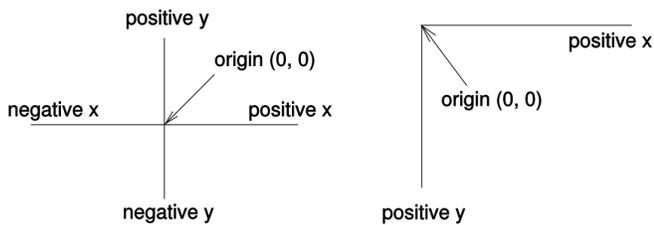


Fig. 4: Cartesian and computer graphics coordinate systems (source: [15]).

Another technical characteristic that must be taken into consideration is the format of the images. CASIA images are stored as 8-bit greyscale BMP (Windows bitmap) images. In that format, every pixel is encoded as a positive number ranging from 0 to 255, with 0 being pure black and 255 being pure white.

### 4.1 Pupil detection

Our pupil detection algorithm is a five-step procedure that basically computes a point acting as the centre of the circumference that represents the pupil and its associate radius, and then proceeds to adjust those values through a series of left-right and up-down movements.

Algorithm 1 provides the first approximation to the centre and radius of the pupil. The algorithm vertically scans the image and locates the longer sequence of at least 25 contiguous pixels whose value is below a certain threshold. The algorithm performs a loop until such a valid sequence is found, where each iteration uses a different limit value (starting at 5, and incremented in 5 units in each loop pass). Then, the algorithm identifies the middle point of the sequence in the scope of the image, which represents the first estimation for the pupil centre.

The following four steps of the procedure aim to improve the location of the pupil centre and the estimation of the

---

### Algorithm 1 Pupil detection (phase 1).

---

```

Require: image, height, width
1: count, pupilX, pupilY, pupilDiam  $\leftarrow$  0
2: limit  $\leftarrow$  5
3: repeat
4:   for all  $i$  such that height  $< i < 3 \cdot \text{height}/4$  do
5:     black  $\leftarrow$  false
6:     count, first, last, longFirst, longLast  $\leftarrow$  0
7:     for all  $j$  such that width  $< j < 3 \cdot \text{width}/4$  do
8:       val = image[ $i \cdot \text{width} + j$ ];
9:       if val  $<$  limit then
10:        count  $\leftarrow$  count + 1
11:        if black then
12:          last  $\leftarrow$  j
13:        else
14:          first  $\leftarrow$  j
15:          black  $\leftarrow$  true
16:        end if
17:        if count  $\geq$  25 then
18:          longFirst  $\leftarrow$  first
19:          longLast  $\leftarrow$  last
20:        end if
21:      else
22:        black  $\leftarrow$  false
23:        count  $\leftarrow$  0
24:      end if
25:    end for
26:    dif  $\leftarrow$  (longLast - longFirst)
27:    if dif  $>$  0 then
28:      if pupilDiam  $\leq$  dif then
29:        pupilDiam  $\leftarrow$  dif
30:        pupilX  $\leftarrow$  (longLast + longFirst)/2
31:        pupilY  $\leftarrow$  i
32:      end if
33:    end if
34:  end for
35:  limit  $\leftarrow$  limit + 5
36: until pupilDiam  $>$  0
37: return pupilX, pupilY, pupilDiam, limit

```

---

radius. In order to do so, Algorithms 2 and 3 identify the values such that the four circumference points located at the far left, right, up, and down (from the point of view of the circumference centre) all have values smaller than the limit found in Algorithm 1, which means that the circumference is circumscribed in the pupil. In Algorithm 2, the one-directional movements proceed along the  $x$  axis (i.e., only left and right movements and reductions in the circumference radius are allowed). In comparison, the circumference can move along the  $y$  axis in Algorithm 3, so up and down movements and circumference reductions are the only operations allowed.

Algorithms 4 and 5 are quite similar to Algorithms 2 and 3. The difference consists in replacing the reduction operations with enlargement ones. The goal for those algorithms is to locate the circumference whose left, right, up, and down extreme points all have values bigger than the limit, which means that the circumference is acting as the external border of the pupil.

Once Algorithm 5 finishes its execution, the centre of the pupil circumference is represented as the point whose coordinates are `pupilX` and `pupilY`, while the diameter of the circumference is computed as `pupilDiam`.

**Algorithm 2** Pupil detection (phase 2).

```

Require: image, height, width, pupilX, pupilY, pupilDiam, limit
1: finished ← false
2: left ← false
3: right ← false
4: while not finished do
5:   radius ← pupilDiam/2
6:   west = image[pupilY*width+pupilX-radius];
7:   east = image[pupilY*width+pupilX+radius];
8:   if ((east>limit) or (west > limit)) then
9:     if ((east>limit) and (west > limit)) or (left and right) then
10:      pupilDiam ← pupilDiam-2
11:      left ← false
12:      right ← false
13:     else
14:       if (west > limit) and (east ≤ limit) then
15:         pupilX ← pupilX +1
16:         right ← true
17:       else
18:         if (east > limit) and (west ≤ limit) then
19:           pupilX ← pupilX -1
20:           left ← true
21:         end if
22:       end if
23:     end if
24:   else
25:     finished ← true
26:   end if
27: end while
28: return pupilX, pupilY, pupilDiam

```

**Algorithm 3** Pupil detection (phase 3).

```

Require: image, height, width, pupilX, pupilY, pupilDiam, limit
1: finished ← false
2: up ← false
3: down ← false
4: while not finished do
5:   radius ← pupilDiam/2
6:   north = image[pupilY*width+posX-radius];
7:   south = image[pupilY*width+posX+radius];
8:   if (south>limit) or (north > limit) then
9:     if ((south>limit) and (north > limit)) or (up and down) then
10:      pupilDiam ← pupilDiam-2
11:      up ← false
12:      down ← false
13:     else
14:       if (north > limit) and (south ≤ limit) then
15:         pupilY ← pupilY +1
16:         down ← true
17:       else
18:         if (south > limit) and (north ≤ limit) then
19:           pupilY ← pupilY -1
20:           up ← true
21:         end if
22:       end if
23:     end if
24:   else
25:     finished ← true
26:   end if
27: end while
28: return pupilX, pupilY, pupilDiam

```

## 4.2 Iris detection

Algorithm 6 contains all the logic needed to detect the outer border of the iris taking as input data the pupil centre and diameter provided by the previous phase, and the image. The core of the algorithm consists in working along a certain row (the one corresponding to the  $y$  value of the pupil's centre, the one immediately before and the one immediately after), trying to locate the point in which a certain condition is fulfilled. When working with six consecutive groups of 8 pixels each, the condition is fulfilled when the average value of each of the three inner pixel groups (i.e., those closer to the pupil centre) is lower than the average value of each of the three outer pixel groups. The algorithm measures the distance from the pupil centre where that condition occurs when moving the groups of pixels first to the left and then to the right. The value of 8 pixels per group was the one that provided better results in our experiments.

Once the left and right distances have been taken in the three aforementioned rows, the bigger distance is taken, with the particularities described in Algorithm 6. If the left distance is not equal to the right distance, that means that the iris centre is not concentric to the pupil centre.

## 5. Iris Recognition Program

In order to test the practicability of the proposed algorithms and to compare the results to other solutions, we have developed a Java application using the software Java Development Kit (JDK) 1.6 update 27, with default parameters both for compiling and running the application. Figure 5 displays the main screen of our application, the Iris Recognition Program (IRP-CSIC).

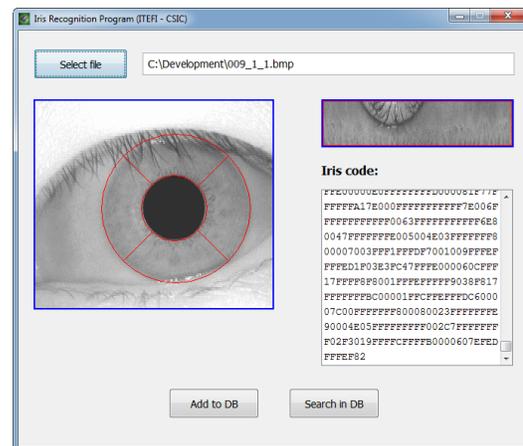


Fig. 5: Iris Recognition Program.

This software allows the user to select a greyscale BMP image using the *Select File* button. Once the file is loaded, the application automatically tries to detect the pupil and iris boundaries, generating a code from the normalized image of the iris.

The application also allows to store the generated template in a database and to compare the current code to the ones stored in the database, though in the scope of this contribution that functionality is not used. With regards to that, the lines that intersect the circumferences in Figure 5 are used to mark the two lateral sectors which contain the iris data employed during the computation of the template.

**Algorithm 4** Pupil detection (phase 4).

---

```

Require: image, height, width, pupilX, pupilY, pupilDiam, limit
1: finished ← false
2: left ← false
3: right ← false
4: while not finished do
5:   radius ← diam/2
6:   west = image[pupilY*width+posX-radius];
7:   east = image[pupilY*width+posX+radius];
8:   if (east ≤ limit) or (west ≤ limit) then
9:     if ((east ≤ limit) and (west ≤ limit)) or (left and right) then
10:      pupilDiam ← pupilDiam+2
11:      left ← false
12:      right ← false
13:     else
14:       if (west ≤ limit) and (east > limit) then
15:         pupilX ← pupilX -1
16:         left ← true
17:       else
18:         if (east ≤ limit) and (west > limit) then
19:           pupilX ← pupilX +1
20:           right ← true
21:         end if
22:       end if
23:     end if
24:   else
25:     finished ← true
26:   end if
27: end while
28: return pupilX, pupilY, pupilDiam

```

---

**Algorithm 5** Pupil detection (phase 5).

---

```

Require: image, height, width, pupilX, pupilY, pupilDiam limit
1: finished ← false
2: up ← false
3: down ← false
4: while not finished do
5:   radius ← pupilDiam/2
6:   north = image[pupilY*width+posX-radius];
7:   south = image[pupilY*width+posX+radius];
8:   if (south ≤ limit) or north ≤ limit) then
9:     if ((south ≤ limit) and (north ≤ limit)) or (up and down) then
10:      pupilDiam ← pupilDiam+2
11:      up ← false
12:      down ← false
13:     else
14:       if (north ≤ limit) and (south > limit) then
15:         pupilY ← pupilY -1
16:         down ← true
17:       else
18:         if (south ≤ limit) and (north > limit) then
19:           pupilY ← pupilY +1
20:           up ← true
21:         end if
22:       end if
23:     end if
24:   else
25:     finished ← true
26:   end if
27: end while
28: return pupilX, pupilY, pupilDiam

```

---

## 6. Tests and results

The tests whose results are presented in this section were completed using a PC with Windows 7 Professional OS and an Intel Core i7 processor at 3.40 GHz.

### 6.1 Detection comparison

Table 1 includes the results of the tests performed with the three applications. The CASIA database used in those tests, called *CASIA Iris Image Database Version 1.0* [12], contains 756 images from 108 different users. For each user, two iris recording sessions are provided (the first one producing 3 images, and the second one 4 images).

Table 1: Comparison of detection capabilities.

Application	Pupil	Iris
IRP-CSIC	754 (99,74 %)	683 (90,34 %)
Imperial College	745 (98,54 %)	538 (71,16 %)
Warsaw Tech. Univ.	411 (54,37 %)	381 (50,40 %)

In the pupil tests we have allowed a 5% error margin, whilst in the iris tests we have permitted an error margin of 10%. In the scope of these tests, the error is defined as the distance in pixels between the ideal circumference that represents the outer limit of the pupil (or, respectively, the iris), and the circumference drawn by the applications, divided by the diameter of the pupil (or the iris). That distance, which has been measured with the help of the Greenshot program [16], is taken in at most four points (left and right in all the cases, and up and down whenever those points were not covered by the eyelids).

Figures 6, 7, and 8 show three examples where our software outperforms the other open source solutions considered in this comparison in terms of iris and pupil detection. In those figures, the first image corresponds to IRP-CSIC, the second one to the Imperial College's application, and the third one to the application developed at the Warsaw Technical University.

### 6.2 Running time

In order to measure the performance of our solution, we conducted a new test with the following common characteristics for the three cases:

- 1) For each element to be detected (pupil and iris), we have used the same batch of CASIA images belonging to 10 different users, so in each case a total of 70 images have been processed.
- 2) The time displayed for each element represents the average time of the 70 images processed.
- 3) For each individual test, the corresponding application has been started and closed, so every time a new test has been passed a fresh instance of the application has been executed.

In the case of IRP-CSIC, the tests include the following features:

- 1) The Java function used for obtaining the timing is `System.nanoTime()`.
- 2) In the pupil case, the start time has been taken exactly before calling the application method that implements Algorithm 1, while the finish time has been taken after the application retrieves the last candidate values provided by Algorithm 5.

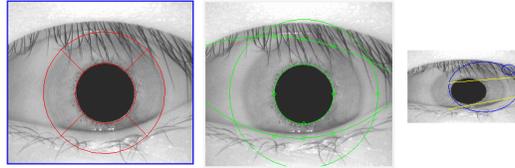


Fig. 6: Example of iris detection (input image 010\_1\_3.bmp).

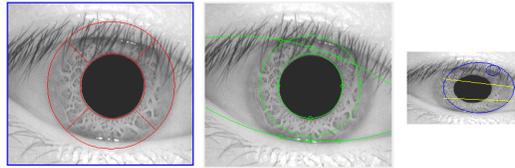


Fig. 7: Example of iris detection (input image 078\_2\_1.bmp).

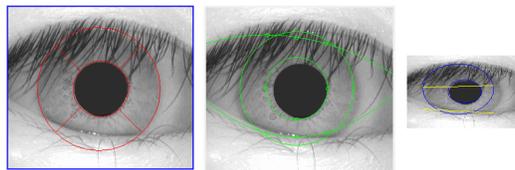


Fig. 8: Example of pupil detection (input image 082\_2\_2.bmp).

- 3) In the iris case, the start time and the stop time has been taken exactly before and after calling the method which implements Algorithm 6, respectively.

With regards to the Imperial College application, the most important characteristics of the tests are the following:

- 1) The C++ method used for obtaining the timing is `clock()`. The difference between the finish and start values given by that function has been divided by the value `CLOCKS_PER_SEC`, provided by the system, in order to obtain the time elapsed.
- 2) In the pupil case, the start time has been taken exactly before calling the method `findPupil()` that implements the pupil location algorithm, while the finish time has been taken immediately after that method returns.
- 3) In the iris case, the start time has been taken exactly before calling the method `findCircle()` that implements the iris location algorithm, while the finish time has been taken immediately after that method returns its estimation.

Finally, regarding the application from the Warsaw Technical University, the most relevant features are the following:

- 1) The Java function used for obtaining the timing is `System.nanoTime()`.
- 2) We have measured the time used by the Sobel and Hough methods and added half that time to both the pupil and the iris running time, as the location of both elements take advantage of the Sobel and Hough computations.

- 3) The Sobel and Hough start time has been taken before calling the method `sobelObject.init()`, while the stop time has been taken after the methods `sobelObject.process()`, `sobel()`, and `hough()` are executed.
- 4) In the pupil case, the initial running time comprises the execution of the `pupil()` method.
- 5) In the iris case, the initial running time comprises the execution of the `iris()` method.
- 6) In all the portions of code involved in the calculations, we have removed the printing methods which outputs information on the console in order to remove the delays added by the printing of information.

Table 2 includes the results obtained after completing the performance tests with the three applications. As it can be observed, our proposal is faster than the other solutions.

Table 2: Running time.

Application	Pupil	Iris
IRP-CSIC	526.6 $\mu$ s	454.0 $\mu$ s
Imperial College	51.0 ms	29.3 ms
Warsaw Tech. Univ.	388.0 ms	359.9 ms

## 7. Conclusions

Iris identification is one of the most interesting applications of biometric techniques. As the algorithms used in most commercial applications are patented, developers must either pay for the corresponding licences or create new algorithms.

**Algorithm 6** Iris detection.

---

```

Require: image, height, width, pupilX, pupilY, pupilDiam
1: irisL1X, irisL1Y, irisL1Diam, irisL2X, irisL2Y, irisL2Diam  $\leftarrow$  0
2: for all  $k$  such that  $-1 \leq k \leq 1$  do
3:   good  $\leftarrow$  true, finished  $\leftarrow$  false
4:   irisY  $\leftarrow$  pupilY +  $k$ ,  $x \leftarrow$  pupilX - pupilDiam/2
5:   irisDiam  $\leftarrow$  0, diam  $\leftarrow$  0
6:   repeat
7:     for all  $i$  such that  $0 \leq i < 8$  do
8:       val1  $\leftarrow$  val1 + image[irisY*width+x-0-i]
9:       val2  $\leftarrow$  val2 + image[irisY*width+x-8-i]
10:      val3  $\leftarrow$  val3 + image[irisY*width+x-16-i]
11:      val4  $\leftarrow$  val4 + image[irisY*width+x-24-i]
12:      val5  $\leftarrow$  val5 + image[irisY*width+x-32-i]
13:      val6  $\leftarrow$  val6 + image[irisY*width+x-40-i]
14:     end for
15:     val1  $\leftarrow$  val1/8, val2  $\leftarrow$  val2/8, val3  $\leftarrow$  val3/8
16:     val4  $\leftarrow$  val4/8, val5  $\leftarrow$  val5/8, val6  $\leftarrow$  val6/8
17:     if ((val1 < val4) and (val1 < val5) and (val1 < val6) and (val2 < val4) and (val2 < val6) and (val2 < val6) and (val3 < val4) and (val3 < val5) and (val3 < val6)) then
18:       distLeft = ((pupilX-x) + 3*length) + length/2;
19:       finished  $\leftarrow$  true
20:     end if
21:     x  $\leftarrow$  x-1
22:   until finished or (x-48) < 0
23:   finished  $\leftarrow$  false
24:   x  $\leftarrow$  pupilX + diam/2
25:   repeat
26:     for all  $i$  such that  $0 \leq i < \text{length}$  do
27:       val1  $\leftarrow$  val1 + image[irisY*width+x+0+i]
28:       val2  $\leftarrow$  val2 + image[irisY*width+x+8+i]
29:       val3  $\leftarrow$  val3 + image[irisY*width+x+16+i]
30:       val4  $\leftarrow$  val4 + image[irisY*width+x+24+i]
31:       val5  $\leftarrow$  val5 + image[irisY*width+x+32+i]
32:       val6  $\leftarrow$  val6 + image[irisY*width+x+40+i]
33:     end for
34:     val1  $\leftarrow$  val1/length, val2  $\leftarrow$  val2/length, val3  $\leftarrow$  val3/length
35:     val4  $\leftarrow$  val4/length, val5  $\leftarrow$  val5/length, val6  $\leftarrow$  val6/length
36:     if ((val1 < val4) and (val1 < val5) and (val1 < val6) and (val2 < val4) and (val2 < val6) and (val2 < val6) and (val3 < val4) and (val3 < val5) and (val3 < val6)) then
37:       distRight = ((x-pupilX) + 3*length) + length/2;
38:       finished  $\leftarrow$  true
39:     end if
40:     x  $\leftarrow$  x+1
41:   until finished or (x+48)  $\geq$  width
42:   if (distLeft = 0) and (distRight > 0) then
43:     distLeft  $\leftarrow$  distRight, good  $\leftarrow$  false
44:   end if
45:   if (distRight = 0) and (distLeft > 0) then
46:     distRight  $\leftarrow$  distLeft, good  $\leftarrow$  false
47:   end if
48:   if (distLeft > distRight) and ((distLeft-distRight)>18) then
49:     distRight  $\leftarrow$  distLeft, good  $\leftarrow$  false
50:   end if
51:   if (distRight > distLeft) and ((distRight-distLeft)>18) then
52:     distLeft  $\leftarrow$  distRight, good  $\leftarrow$  false
53:   end if
54:   if (distLeft > distRight) then
55:     irisX = pupilX - ((distLeft-distRight)/2);
56:   else
57:     irisX = pupilX + ((distRight-distLeft)/2);
58:   end if
59:   irisDiam  $\leftarrow$  distLeft + distRight
60:   if good and (irisDiam > bestL1Diam) then
61:     irisL1X  $\leftarrow$  irisX, irisL1Y  $\leftarrow$  irisY, irisL1Diam  $\leftarrow$  diam
62:   end if
63:   if (not good) and (irisDiam > irisL2Diam) then
64:     irisL2X  $\leftarrow$  irisX, irisL2Y  $\leftarrow$  irisY, irisL2Diam  $\leftarrow$  diam
65:   end if
66: end for
67: if irisL1Diam > 0 then
68:   return irisL1X, irisL1Y, irisL1Diam
69: else if irisL2Diam > 0 then
70:   return irisL2X, irisL2Y, irisL2Diam
71: else
72:   return null
73: end if

```

---

In this contribution we have described in detail two low-complexity methods for locating the pupil and the iris which are patent-free and that, due to their simplicity, can be implemented in a broad range of devices with a good performance. After comparing our proposal to two other open source solutions, we can state that the application that implements our algorithms is faster and more accurate than the other solutions considered in the comparison.

Even though the iris detection is a complex issue, we consider that the lower average running time (when compared to the average running time for pupil detection) in our application is due to the following facts:

- The pupil identification procedure generates five calls to other Java methods, whilst the iris identification procedure generates three calls. Calling a Java method increments the overall running time.
- The iris identification procedure takes advantage of the work done by the pupil identification procedure, as it starts its execution with the knowledge of the point which represents the center of the circle which simulates the pupil.

## Acknowledgment

This work has been partially supported by Ministerio de Ciencia e Innovación (Spain) under the grant TIN2011-22668.

## References

- [1] M. J. Burge and K. W. Bowyer, *Handbook of Iris Recognition*. New York, NY, USA: Springer, 2012.
- [2] CASIA. Biometrics Ideal Test. <http://biometrics.idealtest.org>.
- [3] C. Rathgeb, A. Uhl, and P. Wild, *Iris Biometrics*. New York, NY, USA: Springer, 2013.
- [4] J. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 12, pp. 1148–1161, 1993.
- [5] —, *Biometric Identification Based on Iris Analysis*, U.S. Patent No. 5,291,560, 1994.
- [6] Wall Street Journal. Iris Recognition: New Fingerprinting. <http://blogs.wsj.com/digits/2011/07/13/iris-recognition-the-new-fingerprinting/>.
- [7] D. H. Gold and R. A. Lewis, *Clinical Eye Atlas*. Oxford, UK: Oxford University Press, 2010.
- [8] Webvision. The Organization of the Retina and Visual System. [webvision.med.utah.edu/](http://webvision.med.utah.edu/).
- [9] J. Daugman, "How iris recognition works," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 21–30, 2004.
- [10] M. Boyd, D. Carmaciu, F. Giannaros, T. Payne, and W. Snell. (2010) Iris Recognition. <http://projectiris.co.uk/final.pdf>.
- [11] I. C. London. Project Iris. <http://projectiris.co.uk/>.
- [12] CASIA. CASIA Iris Image Database Version 1.0. <http://biometrics.idealtest.org/dbDetailForUser.do?id=1>.
- [13] GitHub. Iris Recognition repository. <https://github.com/bernii/IrisRecognition>.
- [14] L. Masek, *Recognition of Human Iris Patterns for Biometric Identification*, 2003, <http://www.csse.uwa.edu.au/~pk/studentprojects/libor/LiborMasekThesis.pdf>.
- [15] A. Downey. thinkajava - Java 2D Graphics. <http://fpl.cs.depaul.edu/jriely/thinkajava/thinkajava.2d.html>.
- [16] T. Braun, J. Klingen, and R. Krom. Greenshot. <http://sourceforge.net/projects/greenshot/>.

# Towards Designing a Greener Advanced Encryption Standard (AES)

S. Raghu Talluri  
 School of Computing  
 University of North Florida  
 Jacksonville, Florida 32224  
 Email: n00926109@ospreys.unf.edu

Swapnoneel Roy  
 School of Computing  
 University of North Florida  
 Jacksonville, Florida 32224  
 Email: s.roy@unf.edu

**Abstract**—In this work we study the energy consumption by Advanced Encryption Standard (AES), a symmetric key encryption protocol from the *algorithmic* perspective. Our work is motivated by the frequent use of AES as a specification for the encryption of electronic data established by the U.S. National Institute of Standards and Technology (NIST) in 2001.

We use a generic energy complexity model designed by Roy et. al. to analyze the energy consumed by AES. We then show how to reduce the energy consumption by AES by performing the processing of the blocks of AES encryption (of size 16 bytes) in parallel.

## I. INTRODUCTION

Motivation to consider energy efficiency in delivering information technology solutions comes from: 1. Data centers with strong focus on energy management for server class systems. 2. Personal computing devices such as smartphones, handhelds, and notebooks, which run on batteries and perform a significant amount of computation and data transfer. 3. Telecom providers expecting to invest in equipment that will form an integral part of the global network infrastructure. On the other hand, information security has become a natural component of all technology solutions. Security protocols consuming additional energy are often incorporated in these solutions. Thus, the impact of security protocols on energy consumption needs to be studied. Ongoing research in this context has been mainly focused on energy efficiency/consumption on specific hardware and/or different systems/platforms. Very little is known or has been explored regarding energy consumption or efficiency from an *applications* perspective, although apps for smartphones and handhelds abound.

**Our Contributions:** Our work makes two key contributions. Since energy or power has become a first class component in computing now a days, this could be very expensive especially for the battery driven devices like laptops and PDAs. As a conclusion to this observation, we found a lot of work done to reduce energy consumption in network communication and security protocols in the hardware, virtual machines, operating systems, and the system software levels [1], [2], [3], [4], [5]. But not much work has been done from the application or algorithmic perspective to minimize energy or power consumption in such protocols. In other words, the problem which we try to investigate is can we design energy aware security protocols? Or can we modify existing protocols to make them energy optimal, without compromising on the level of security they provide?

We next analyze the energy consumption by AES. Specifically we estimate the energy consumed by the AES algorithm using the energy model of [6]. Finally we modify the AES algorithm to lower the level of energy consumption pertaining to the energy model. Specifically we *parallelize* the input by accessing blocks of size 16 bytes in parallel for AES and observe it lowers the energy consumption of the protocol.

AES is based on a design principle known as a substitution-permutation network, and is fast in both software and hardware. Unlike its predecessor DES, AES does not use a Feistel network. AES is a variant of Rijndael which has a fixed block size of 128 bits (or 16 bytes), and a key size of 128, 192, or 256 bits. By contrast, the Rijndael specification per se is specified with block and key sizes that may be any multiple of 32 bits, both with a minimum of 128 and a maximum of 256 bits.

The rest of the paper is structured in the following manner. Section II describes the AES algorithm in detail. The energy complexity model we use, and the techniques we use to optimize energy consumption by AES is described in Section III. Section IV describes our experimental set and presents our key experimental results on energy consumption of AES. Finally, Section V summarizes the results and discusses future research directions.

## II. ADVANCED ENCRYPTION STANDARD (AES)

The AES algorithm on each block of 16 bytes can be described as follows:

- 1) **KeyExpansion.** Round keys are derived from the cipher key using Rijndael's key schedule. AES requires a separate 128-bit round key block for each round plus one more.
- 2) **InitialRound.**
  - a) **AddRoundKey:** Each byte of the state is combined with a block of the round key using bitwise xor.
- 3) **Rounds**
  - a) **SubBytes:** A non-linear substitution step where each byte is replaced with another according to a lookup table.
  - b) **ShiftRows:** A transposition step where the last three rows of the state are shifted cyclically a certain number of steps.

```

Data: A Plaintext block of 16 bytes byte  $in[4 * Nb]$ 
Result: A Ciphertext block of 16 bytes byte  $out[4 * Nb]$ 
word  $w[Nb * (Nr + 1)];$ 
state = in;
AddRoundKey(state, w[0, Nb - 1]);
for round = 1 step 1 to Nr1 do
  SubBytes(state);
  ShiftRows(state);
  MixColumns(state);
  AddRoundKey(state, w[round * Nb, (round + 1) * Nb - 1]);
end
SubBytes(state);
ShiftRows(state);
AddRoundKey(state, w[Nr * Nb, (Nr + 1) * Nb - 1]);
out = state;

```

**Algorithm 1:** The AES Algorithm

- c) MixColumns: A mixing operation which operates on the columns of the state, combining the four bytes in each column.
  - d) AddRoundKey
- 4) **Final Round (no MixColumns).**
- a) SubBytes
  - b) ShiftRows
  - c) AddRoundKey

AES is a symmetric key encryption cipher. That is the same set of keys are used for both encryption and decryption. For a detailed description of each of the operations, please see [7], [8], [9], [10].

### III. ENGINEERING AES FOR ENERGY EFFICIENCY

#### A. Energy Complexity Model

An asymptotic energy complexity model for algorithms was proposed in [6]. Inspired by the popular DDR3 architecture, the model assumes that the memory is divided into  $P$  banks each of which can store multiple blocks of size  $B$ . In particular,  $P$  blocks in  $P$  different memory banks can be accessed in parallel (Figure 1). The main contribution of the model in [6] was to highlight the effect of parallelizability of the memory accesses in energy consumption. In particular, the energy consumption of an algorithm was derived as the weighted sum  $T + (PB) \cdot I$ , where  $T$  is the total time taken and  $I$  is the number of parallel I/Os made by the algorithm.

#### B. $P$ -way Parallelism for AES input

Energy optimal algorithms proposed in [6] require data to be laid out in memory with a controlled degree of parallelism. We first propose a way to ensure desired memory parallelism for a given input  $M$  to the AES algorithm. We ensure memory parallelism for the processing of 16-byte blocks by the AES algorithm.

We treat the AES algorithm as a black box. Given an input  $M$ , AES divides  $M$  into blocks of 16 bytes and processes each block for encryption to produce the ciphertext (Figure 2).

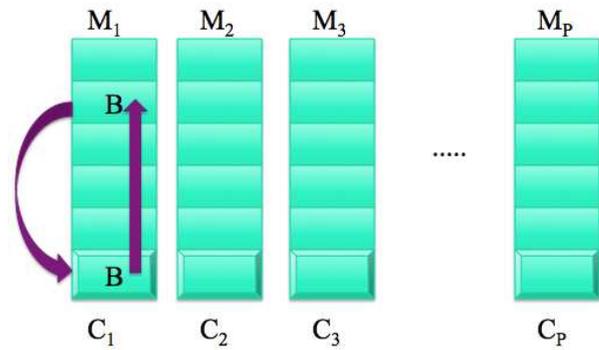


Figure 1. Memory divided in banks.

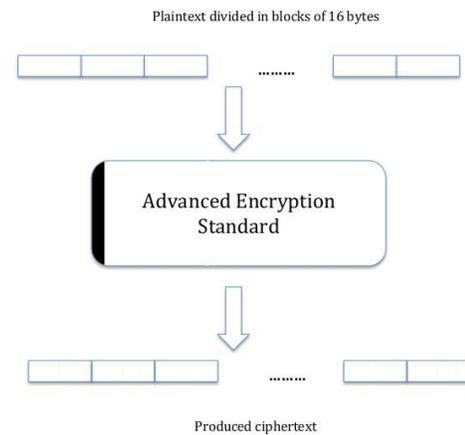


Figure 2. AES input in blocks of size 16 bytes.

For the message  $M$  which is a multiple of blocks of 16 bytes, we created a logical mapping which ensures access to the blocks in  $P$ -way parallel fashion, where  $P$  ranges from 1 to 8. More specifically, when  $P = 1$ , (almost) all the blocks of 16 bytes are clustered in a single bank. While for  $P = 8$ , the blocks are evenly spread across all 8 banks to ensure the maximum degree of parallelism of the access to the input ( $M$ ) to AES. We also experiment for  $P = 2$ , and  $P = 4$ .

To achieve the above, we create a mapping function which maps the physical input  $M$  into the logical input which defines the degree of parallelism. In other words, we define an *ordering* among the blocks of 16 bytes which defines the logical input (and the degree of parallelism).

### IV. EXPERIMENTAL RESULTS

We next evaluate the energy consumed by AES algorithm for  $\{1, 2, 4, \text{ and } 8\}$ -way parallel data. Again, for a recap, a  $k$ -way parallelism in the input data for AES suggests that the 16 bytes blocks in the input to AES algorithm are spread across  $k$  banks in the memory. So a 8-way parallelism in an 8 bank memory means full parallelism (optimal case), and a 1-way parallelism is the worst case. According to the energy complexity model of [6], an 8-way parallelism should account for lower energy consumption in AES.

The experiments were performed on a PC machine. The machine has an Intel i5-3427U processor with inbuilt graphics

with 4-GB ram running Windows 7 SP1. The machine was run on battery during the experiments.

We measured the power drawn by an application using the Joulemeter tool [11] developed by Microsoft Research. Joulemeter runs on Windows XP and Windows 7 currently. It is a software tool that estimates the power consumption of your computer. This software gives the user the ability to tag a CPU process and measure the power consumed by the application in real-time. The data over a period of time is logged and plotted to give us a visualization of the power and energy requirements of the software. It also tracks computer resources, such as CPU utilization and screen brightness, and estimates power usage.

We calculate energy by the product of the time to execute and the (average) power consumed during the time of execution. All the experiments were repeated a hundred times, and the mean value has been reported. The benchmark code was written in C and was compiled using `gcc`. We note that DDR3 has 8 banks.

The first results reported measures the energy consumed by AES with input sizes of 8MB, 16MB, 32MB, and 64MB. These numbers have been obtained for the best case (8-way parallelism). A key size of 128 bits was used for all the experiments.

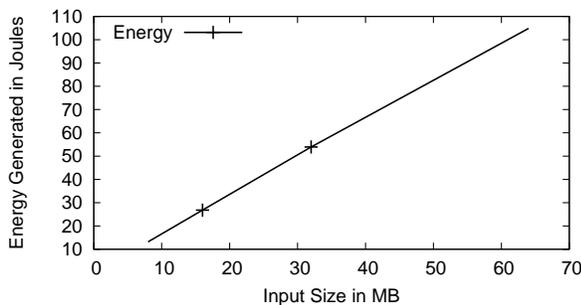


Figure 3. Energy consumption in joules by AES with various sizes of inputs.

We do not see any surprises in Figure 3. The energy consumption for AES varies linearly with respect to the input size.

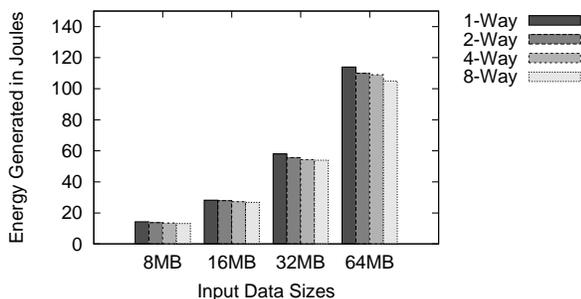


Figure 4. Energy consumption in joules by AES with various degrees of parallelism.

Figure 4 shows us interesting numbers. We measure the energy consumed by AES over fixed size input, varying the degree of parallelism of the input. We see change in energy consumption based in the degree of parallelism of the input. Higher degree of parallelism lowers the energy consumption. We note the difference is not very high between the best and

worst cases, and that is partly due to other factors like the code overhead, noise due to other processes, etc. The numbers indicate the applicability of the energy model of [6] on the AES algorithm. It signifies that the energy consumption of AES can be lowered by parallelizing the input data.

## V. CONCLUSION

In this paper, we have made two key contributions. We first call for designing algorithmic techniques to bring down the energy consumption of security protocols which build them. We next experiment on the applicability of the generic energy complexity model [6] on AES algorithm. We observe the model to be applicable to AES. Our numbers show a reduction in the energy consumption of AES by increasing the degree of parallelism in the input.

It would be interesting to compute the energy consumption for other security protocols like RSA (public key), or the advanced hash functions like MD4, and MD5. We conjecture the applicability of our techniques to lower the level of energy consumption in them.

## REFERENCES

- [1] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on*. IEEE, 2000, pp. 10–pp.
- [2] S. Lindsey and C. S. Raghavendra, "Pegasis: Power-efficient gathering in sensor information systems," in *Aerospace conference proceedings, 2002. IEEE*, vol. 3. IEEE, 2002, pp. 3–1125.
- [3] M. Handy, M. Haase, and D. Timmermann, "Low energy adaptive clustering hierarchy with deterministic cluster-head selection," in *Mobile and Wireless Communications Network, 2002. 4th International Workshop on*. IEEE, 2002, pp. 368–372.
- [4] W. Ye, J. Heidemann, and D. Estrin, "An energy-efficient mac protocol for wireless sensor networks," in *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3. IEEE, 2002, pp. 1567–1576.
- [5] V. Raghunathan, C. Schurgers, S. Park, and M. B. Srivastava, "Energy-aware wireless microsensor networks," *Signal Processing Magazine, IEEE*, vol. 19, no. 2, pp. 40–50, 2002.
- [6] S. Roy, A. Rudra, and A. Verma, "An energy complexity model for algorithms," in *ITCS 2013*.
- [7] "Announcing the advanced encryption standard (aes)," .
- [8] J. Daemen and V. Rijmen, *The design of Rijndael: AES-the advanced encryption standard*. Springer, 2002.
- [9] E. Conrad, "Advanced encryption standard," *White Paper*, 1997.
- [10] D. Selent, "Advanced encryption standard," *Rivier Academic Journal*, vol. 6, no. 2, 2010.
- [11] "Joulemeter: Computational energy measurement and optimization," <http://research.microsoft.com/en-us/projects/joulemeter/>.

**SESSION**  
**COMPUTER AND HARDWARE SECURITY**

**Chair(s)**

**Dr. Nicolas Sklavos**  
**Technological Educational Institute of Western Greece**



# Modeling and Attack for 4-MUXs based PUF

S. Kiryu<sup>1</sup>, K. Asahi<sup>1</sup>, and M. Yoshikawa<sup>1</sup>

<sup>1</sup>Department of Information Engineering, Meijo University, Nagoya, Aichi, Japan

**Abstract** - Physical unclonable function (PUF) is one of the technique to prevent forgery circuits. PUF uses analog characteristics of each device which are accidentally generated due to dispersion during Large Scale Integrated circuit (LSI) manufacturing as a measure of individual identification. Arbiter PUF, which uses the difference in signal propagation delay between selectors, is typical methods of composing PUF using delay characteristics. 4-MUXs based PUF, which is improved the performance of conventional arbiter 2-MUXs PUF is proposed. This paper proposes a modeling method of the 4-MUXs based PUF for machine learning attacks and discusses the vulnerability of the proposed 4-MUXs based PUF.

**Keywords:** PUF, Machine learning attack, Modeling

## 1 Introduction

Recently, the semiconductor counterfeiting problem has become a serious problem. This counterfeit problem causes not only financial damage, but also the safety of human life. Therefore, techniques to prevent forgery using random characteristic patterns which are difficult to artificially control have attracted attention. Physical unclonable function (PUF) is one of these techniques. PUF uses analog characteristics of each device which are accidentally generated due to dispersion during Large Scale Integrated circuit (LSI) manufacturing as a measure of individual identification. The basic operations of PUF are to give an input called challenge and to output a characteristic value (ID) called response. Since PUF is composed of a challenge-response function, its manufacturing and authentication methods are not necessary to conceal.

Several studies have been performed on PUF [1]-[14]. Arbiter PUF, which uses the difference in signal propagation delay between selectors, is typical methods of composing PUF using delay characteristics [4]-[7]. The advantage of arbiter PUF is that the response can be acquired at any time. However, the vulnerability of the arbiter PUF to machine learning attacks has been pointed out [1],[2].

On the other hand, we have developed a new 4-MUXs based PUF which is improved the performance of conventional arbiter PUF [4]. This paper proposes a modeling method of the 4-MUXs based PUF for machine learning

attacks and discusses the vulnerability of the proposed 4-MUXs based PUF.

## 2 Previous Studies

The operation of a conventional arbiter PUF can be expressed by using a model formula. Conventional PUFs are known to be attacked by forcing them to learn a model formula in reference [3]. This section describes a method of modeling a conventional PUF. In reference [3], the model formulae for the delay time and the challenge can be expressed as formulae (1) and (2), respectively.

$$\left. \begin{aligned} w^1 &= \frac{1}{2}(\delta_n^0 + \delta_n^1), & w^i &= \frac{1}{2}(\delta_{i-1}^0 - \delta_{i-1}^1 + \delta_i^0 + \delta_i^1) \\ w^{n+1} &= \frac{1}{2}(\delta_n^0 - \delta_n^1) & (i &= 2, \dots, n) \end{aligned} \right\} \quad (1)$$

$$\varphi^i = \prod_{l=1}^n (1 - 2b_l), \quad \varphi^{n+1} = 1 \quad (i = 2, \dots, n) \quad (2)$$

The response is obtained using  $\text{sgn}(w^T, \varphi)$ . Here,  $\text{sgn}$  is a function that outputs 0 and 1 when the values of the inner product  $(w^T, \varphi)$  are positive and negative, respectively. Here, the case of performing machine learning attacks is examined.

Since the delay time is generated by dispersion during semiconductor manufacturing, an attacker cannot know the difference in the delay time. However, when the challenge can be obtained, the challenge model can be created. When an attacker performs machine learning attacks against an arbiter PUF, the attacker uses only the challenge model and the corresponding response data. When the challenge is  $n$ -bit, the challenge model is  $(n + 1)$  bit. However, all the challenge models at the  $(n + 1)$  th are 1. Therefore, challenge models from the first to  $n$ th are used for machine learning attacks.

## 3 Proposed Modeling Method

Figure 1 shows the structure of the  $k$ -th step of a 4xPUF to be proposed.

In this figure, the delay time of each selector wiring is expressed as  $A_k, B_k, \dots, H_k$ , and the sum of the delay times is expressed as  $\delta$ . Here,  $\delta_1(k)$  is examined. When challenge  $C = (c_1, c_2, \dots, c_n)$  is assumed,  $\delta_1(k)$  can be expressed using the following formula.

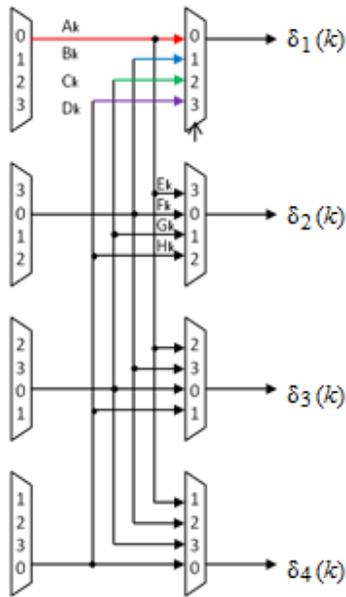


Fig. 1. Example of the proposed PUF

$$\begin{aligned}
 \delta_1(k) &= (1 - c_{2k-1})(1 - c_{2k})(A_k + \delta_1(k-1)) \\
 &+ c_{2k-1} \quad c_{2k} (B_k + \delta_2(k-1)) \\
 &+ c_{2k-1} \quad (1 - c_{2k})(C_k + \delta_3(k-1)) \\
 &+ (1 - c_{2k-1}) \quad c_{2k} (D_k + \delta_4(k-1)) \\
 \delta_0(k) &= 0
 \end{aligned} \quad (3)$$

Here, the difference in the delay time  $\Delta_{12}(k) = \delta_{1k} - \delta_{2k}$  can be expressed as follows:

$$\begin{aligned}
 \Delta_{12}(k) &= (1 - c_{2k-1})(1 - c_{2k})(A_k - F_k + \Delta_{12}(k-1)) \\
 &+ c_{2k-1} \quad c_{2k} (B_k - G_k + \Delta_{23}(k-1)) \\
 &+ c_{2k-1} \quad (1 - c_{2k})(C_k - H_k + \Delta_{34}(k-1)) \\
 &+ (1 - c_{2k-1}) \quad (D_k - E_k + \Delta_{41}(k-1)) \\
 \Delta_{12}(0) &= 0
 \end{aligned} \quad (4)$$

In the case of a conventional PUF, the difference in the delay time can be expressed as

$$\begin{aligned}
 \Delta(k) &= (1 - c_k)(p_k - q_k + \Delta(k-1)) \\
 &+ c_k(p_k - q_k + \Delta(k-1))
 \end{aligned} \quad (5)$$

Unlike a conventional PUF, the proposed PUF needs to obtain  $\Delta_{12}$ ,  $\Delta_{23}$ ,  $\Delta_{34}$ , and  $\Delta_{41}$ . Therefore, the model is computationally difficult to derive from formulae. Compared with a conventional PUF, the number of product terms of  $c_k$  is larger in the proposed PUF. This large number prevents the proposed PUF from machine learning attacks. The present study creates a new challenge model based on the challenge model for a conventional PUF.

Figure 2 shows the operation obtained when challenge 1101 is input into a conventional PUF. At this moment, the challenge model can be expressed as  $(-1, 1, -1, -1, 1)$  from formula (8). In this figure, the wirings on the upper and lower parts are expressed as classes 1 and 2, respectively. Here, the flow of signals input into the input signal side D is examined. The flow of signals between selectors changes from class 1 or 2 when the challenge model is 1 or -1, respectively.

The present study uses four classes and creates a challenge model that can express the operation of the proposed PUF in a manner similar to that of a conventional PUF.

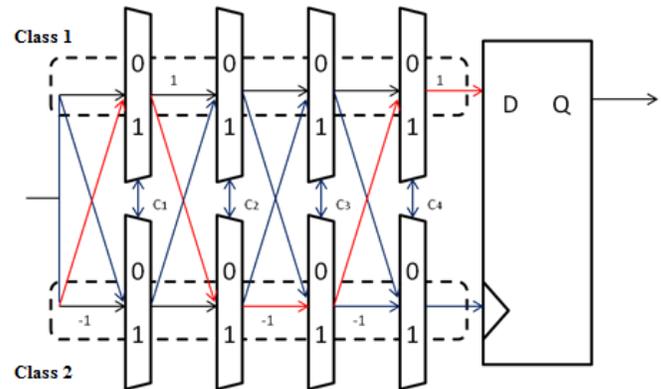


Fig. 2. Example of an operation by a conventional PUF

Here, signals to be noticed are assumed to pass through selector 1. Identification numbers (00, 01, 10, and 11) are given to the four classes and defined as the components of the challenge model. In the case where the challenge is n-bit ( $n = 2m$ ), since the number of steps of the selector is  $n/2$ , the challenge model is assumed to be n-bit.

Figure 3 shows the operation obtained when challenge 0011101 is input into the proposed PUF. At the first step, class 4 is used. At the second step, since challenge 11 is input into the selector, class 1 is used.

Therefore, the class when inputting the challenge changes in such a manner as 4, 4, 1, 2, and 1. Since signals to be noticed are assumed to pass through selector 1, the final

class is always 1. Identification numbers are given to all the classes except the final class and an 8-bit challenge model is obtained. In the case of Figure 3, the challenge model is 11110001.

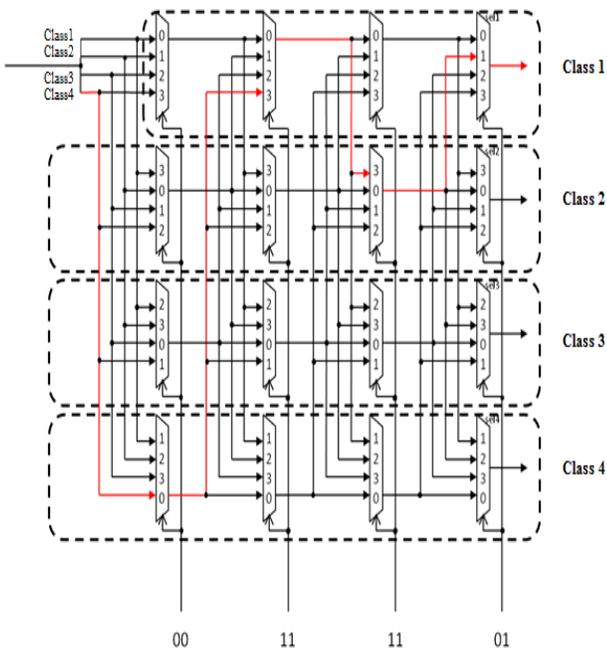


Fig. 3. Example of an operation by the proposed PUF

## 4 Experiments

### 4.1 Experimental Conditions

Unlike the SVM, a neural network (NN) does not handle the response as a label, but it handles as a numerical value. Therefore, for the response of the proposed PUF, it is easier to perform learning every one bit.

In an experiment, machine learning attacks using the NN were performed against the conventional and proposed PUFs and the correct answer rate was examined. The experimental procedure was as follows:

- (1) A learning data set for the conventional PUF and that for the proposed PUF, in which the response to the challenge at every one bit had been recorded, were read into a program for back propagation.
- (2) From the program, the learning results of the synaptic weight and threshold value of the intermediate and output layers were obtained.

- (3) Using the obtained learning results, the test data set, and the program for the NN, the correct answer rate was examined.

As the optimal values of the parameters used for the NN, the learning coefficient  $\alpha$  was set at 0.02 and the number of the intermediate layers was set at 2, based on the results obtained by performing a preparatory experiment. In the experiment, the NN was determined to be terminated when the square-sum of the error between the output from the network and the response of the learning data became below 0.001.

However, when the above termination condition was not satisfied after performing the learning algorithm one million times, the PUFs were determined not to be able to learn.

### 4.2 Results

Figure 4 shows the learning results. In this figure, the results obtained using the SVM were also exhibited. Although the conventional PUF could learn using the NN, the proposed PUF could not satisfy the termination condition in any case; i.e., the proposed PUF could not learn using the NN. Therefore, the results obtained using the proposed PUF were not shown in Figure 4.

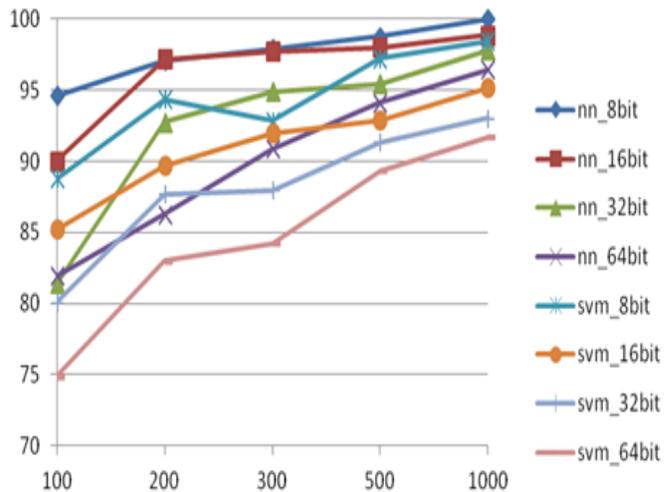


Fig. 4. Results of attack

The reason for this failure was probably that the generalization capacity was smaller in the NN than in the SVM. Since the data processing was more complicated and the number of response types was larger in the proposed PUF than in the conventional PUF, the responses could not be correctly identified. Therefore, the resistance to machine learning attacks using the NN was higher in the proposed PUF than in the conventional PUF.

The correct answer rate of the conventional PUF was higher in the NN than in the SVM. However, the time required for learning was longer in the NN than in the SVM.

The actual times required for learning were 697.1 and 0.5 seconds in the NN and the SVM, respectively, when the challenge was 64 bits and the number of learning data sets was 1000. In the experiment, the NN was set to be terminated when the square-sum of the error between the output from the network and the response of the learning data became below 0.001.

The time required for learning could be shortened by easing the termination condition. Therefore, the correct answer rate and the time required for learning was considered to be in the trade-off relationship in the NN.

## 5 Conclusions

This paper proposed a modeling method of the 4-MUXs based PUF for machine learning attacks and discussed the vulnerability of c 4-MUXs based PUF. Experiments proved the resistance against illegal attack using NN.

In the future, we will apply the method proposed in the present study to simulations of resistance verification to other illegal attacks.

## 6 References

- [1] Jae W. Lee, D. Lim, B. Gassend, G. E. Suh, M. vanDijk, and S. Debadas, "A Technique to Build a SecretKey in Integrated Circuits for Identification and Authentication Applications", in Proceedings of the IEEE VLSI Circuits Symposium, pp.176-179, 2004.
- [2] U. Ruhrmair, F. Sehnke, J. Solter, G. Dror, S. Devadas, J. Schmidhuber, "Modeling Attacks on Physical Unclonable Functions", in Proceedings of ACM Conference on Computer and Communications Security, pp.237-249, 2010.
- [3] Daihyun Lim. "Extracting Secret Keys from Integrated Circuits.", Msc thesis, MIT, 2004.
- [4] M.Yoshikiawa, T.Asai, "Multiplexing Aware Arbiter Physical Unclonable Function", Proc. of IEEE IRI 2012, pp.639-644, 2012.
- [5] Mitsuru Shiozaki, Kota Furuhashi, Takahiko Murayama, Akitaka Fukushima, Masaya Yoshikawa, Takeshi Fujino, "High Uniqueness Arbiter-Based PUF Circuit Utilizing RG-DTM Scheme for Identification and Authentication Applications", IEICE Trans. on Electronics Vol.E95-C No.4 pp.468-477, 2012.
- [6] Yohei Hori, Takahiro Yoshida, Toshihiro Katashita, Akashi Satoh, "Quantitative Performance Evaluation of Arbiter PUFs on FPGAs", IEICE Technical Reports, vol.110, no.204, RECONF2010-37, pp.115-120, 2010.
- [7] Takanori Machida, Toshiki Nakasone, Kazuo Sakiyama, "Evaluation Method for Arbiter PUF on FPGA and Its Vulnerability", IEICE Technical Reports, vol.113, no.135, ISEC2013-18, pp.53-58, 2013.
- [8] M.Majzooobi, F.Koushanfar, S.Devadas, "FPGA PUF using programmable delay lines", Proc. of IEEE International Workshop on Information Forensics and Security, pp.1-6, doi:10.1109/WIFS.2010.5711471, 2010.
- [9] Zouha Cherif, Jean-Luc Danger, Florent Lozac'h, Yves Mathieu, Lilian Bossuet, "Evaluation of Delay PUFs on CMOS 65 nm Technology: ASIC vs FPGA", Proc. of the 2nd International Workshop on Hardware and Architectural Support for Security and Privacy, doi:10.1145/2487726.2487730, 2013.
- [10] Kota Furuhashi, Mitsuru Shiozaki, Akitaka Fukushima, Takeshi Fujino, "The arbiter-PUF with high uniqueness utilizing novel arbiter circuit with Delay-Time Measurement", Proc. of IEEE International Symposium on Circuits and Systems, pp.2325-2328, 2011.
- [11] R.Kumar, S.N.Dhanuskodi, S.Kundu, "On Manufacturing Aware Physical Design to Improve the Uniqueness of Silicon-Based Physically Unclonable Functions", Proc. of 27th International Conference on VLSI Design and 13th International Conference on Embedded Systems, pp.381-386, 2014.
- [12] R.Kumar, V.C.Patil, S.Kundu, "Design of Unique and Reliable Physically Unclonable Functions Based on Current Starved Inverter Chain", Proc. of IEEE Computer Society Annual Symposium on VLSI, pp.224-229, 2011.
- [13] Hyunho Kang Y.Hori, T.Katashita, M.Hagiwara, K.Iwamura, "Cryptographie key generation from PUF data using efficient fuzzy extractors", Proc. of 16th International Conference on Advanced Communication Technology, pp. 23-26, 2014.
- [14] Lang Lin, S.Srivathsa, D.K.Krishnappa, P.Shabadi, W.Burleson, "Design and Validation of Arbiter-Based PUFs for Sub-45-nm Low-Power Security Applications", IEEE Trans. on Information Forensics and Security, Vol.7, No.4, pp.1394-1403, 2012.

## Acknowledgements

This study was supported by Japan Science and Technology Agency (JST), Core Research for Evolutional Science and Technology (CREST).

# Relationship Between Number of Stages in ROPUF and CRP Generation on FPGA

\*Muslim Mustapa, Mohammed Niamat

Electrical Engineering and Computer Science, University of Toledo  
muslim.mustapa@rockets.utoledo.edu, mohammed.niamat@utoledo.edu

**Abstract**—Physical Unclonable Function (PUF) is commonly used to prevent hackers from stealing information from semiconductor chips. The PUFs utilize the process variations on the chip to create an irreversible function that generates unique response bits for each challenge. A good response bit can be generated by comparing two Ring Oscillators (RO) frequencies, which have a significant amount of difference. An insignificant amount of frequency difference can cause bit flip in the response bit generated. A higher threshold for the frequency difference is preferred to dismiss the bit flip occurrence. As the frequency difference threshold (FDT) increased, the numbers of challenge and response pairs (CRP) were reduced. In this paper we proposed new parameter (diverseness) to measure the ROs frequencies range. High diverseness can compensate the higher FDT. We used our Full Scan Technique (FST) on different number of RO stages to determine the number of stages that have the highest diverseness of ROs frequencies. Our experimental results showed that the diverseness of ROs frequencies increased as the number of stages reduced. We also showed that by reducing the number of stages, we still obtained good uniqueness, reliability, bit-aliasing, and uniformity.

**Keywords**—PUF; Ring Oscillator; Hardware security; FPGA

## I. INTRODUCTION

Physical Unclonable Function (PUF) is an irreversible function that is derived from the process variation of a silicon chip. Until now, there was no technology that could measure the process variation with high accuracy which makes harder for the adversary to model PUF. An adversary that is trying to tamper PUF will change the properties of the process variation in the silicon chip and thus the tampering effort will fail [1]. The tamper resistance feature of PUF is important as hardware tampering is one of the unsolved hardware security issues. PUF can be applied as secret bits generator (which is known as response in PUF application) where it can generate  $n$  bits of response for the authentication purpose. PUF can also be applied as a cryptography key generator to encode and decode secure information.

There are various types of PUF that can be implemented on FPGA such as Arbiter PUF (APUF), Butterfly PUF (BPUF), and Ring Oscillator PUF (ROPUF). Out of these three PUFs, ROPUF has the most advantage to be implemented on FPGA because it does not need to have a mirror symmetry requirement, which is a must for APUF and BPUF [2].

Fixed routing on FPGA has made the mirror symmetry structure for APUF and BPUF impossible to be implemented. ROPUF features a simple circuit with strong security response bits generation. Because of these features, ROPUF has much to offer in solving the FPGA hardware security issues such as hardware cloning and IP protection [3].

Despite the promising solution offered by ROPUF, there are still challenges that need to be overcome for ROPUF to become a practical solution. Making the ROPUF response better in uniqueness and increasing in reliability are among the challenges. Uniqueness refers to the ability of similar ROPUF circuits to generate unique responses on different chips. Reliability refers to the generation of same response under various environmental conditions such as temperature and humidity.

ROPUF research areas can be divided into four main categories [1]: fabrication variation extraction, secret selection, error correction, and tests for security and reliability. Fabrication variation extraction is the study on the physical behavior of the silicon chip. This is the most fundamental research area in ROPUF which interacts directly with the process variation. The uniqueness and reliability parameters of the ROPUF are studied thoroughly in this part to take full advantage of the process variation [4][5].

Secret selection is the study of the algorithm to select the comparison pairs that is known as challenge. The randomness parameter of ROPUF is studied in this research area. Error correction research is focused on the algorithm that will correct any flipped bit. This is important for ROPUF implementation as a cryptography technique, where zero bit flipped occurrence is expected [7]. Finally the tests for security and reliability research area looks into the diverseness, bit-aliasing, and probability of misidentification parameters of ROPUF.

This paper focused on the fabrication variation extraction for ROPUF on FPGA. The main objective in this research was to increase the diverseness of ROs frequencies on FPGA. Three different RO stages were tested and compared in term of the diverseness, uniqueness, reliability, bit aliasing, and uniformity. The three different RO stages were tested using our new proposed Full Scan Technique (FST) which will record frequencies from all CLBs available on the FPGA. To the best of our knowledge this is the novel experiment comparing number of stages used in RO.

\*Author currently pursuing Ph.D. study at the EECS Department of the University of Toledo under the fellowship program sponsored by the University of Malaysia Perlis.

## II. BACKGROUND

RO frequency is generated from the inverted signal that travels through the RO loop as shown in Figure 1. The presence of process variation inside logic gates and wires causes an uneven delay across the chip, hence a pair of ROs will produce two different frequencies:  $f_a$  and  $f_b$ .  $f_a$  and  $f_b$  then will be compared to see if  $f_a$  is greater than  $f_b$ . If  $f_a$  is greater than  $f_b$ , response bit 1 will be generated; otherwise it will be 0 as shown in Equation 1.

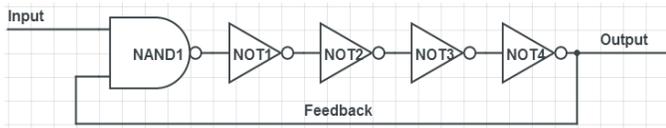


Fig. 1. 5-stage RO

$$\text{Response bit} = \begin{cases} 1 & \text{if } f_a > f_b \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

### A. RO number of stages

In this experiment there are three different number of stages used. Figure 1 shows the 5-stage RO where each component in the RO counts as one stage. The 5-stage RO consists of one NAND gate and 4 inverter gates. The NAND gate is used to control the switching on and off of the RO. The RO will be activated (start to produce an oscillation) when the input is set to high. Figure 2 shows the 4-stage RO. The 4-stage RO consists of one NAND gate, one buffer gate and two inverter gates. The reason of using 4-stage RO is explained Section IV. One buffer gate is used instead of inverter gate because the inverting components need to be odd in number to produce an oscillation. The buffer gate is added to increase the total delay in the RO therefore will reduce the RO frequency. Finally Figure 3 shows the 7-stage RO. The 7-stage RO consist of one NAND gate and 6 inverter gates.

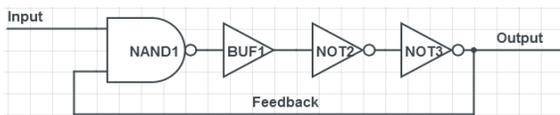


Fig. 2. 4-stage RO

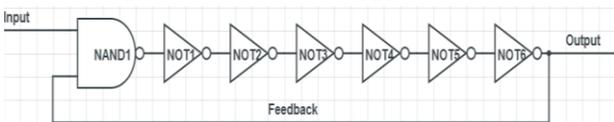


Fig. 3. 7-stage RO

### B. RO Parameters

There are numbers of parameters proposed to measure PUF performance such as uniformity, reliability, steadiness, uniqueness, diverseness, bit-aliasing and probability of misidentification [12][13][14][15][16]. In this research, 4 existing parameters and one newly proposed parameter are used. The 4 existing parameters are chose based on the suitability to measure the performance of different number of

stages used in ROPUF. The 4 parameters are uniqueness, reliability, uniformity and bit-aliasing. One newly parameter proposed in this research is the diverseness. The uniqueness represents the ability of a PUF to uniquely differentiate a particular chip among a group of chips of the same type [12]. Uniqueness can be measured by calculating the Inter-chip HD as shown in Equation 2.  $m$  is the number of chips used,  $u$  and  $v$  are the two chips being compared, and  $n$  is the number of response bits generated.  $R_u$  and  $R_v$  are the response bits from the same challenge  $C$  for chip  $u$  and  $v$ . HD is the hamming distance between response bits generated from chip  $u$  and  $v$ . The good uniqueness value is around 50%. This means that at least 50% of the responses generated from chip  $u$  and  $v$  differ from each other (responses obtained by given the same challenge to chip  $u$  and  $v$ ).

$$\text{Uniqueness} = \frac{2}{m(m-1)} \sum_{u=1}^{m-1} \sum_{v=u+1}^m \frac{HD(R_u, R_v)}{n} \times 100\% \quad (2)$$

The reliability is referring to how efficient is a PUF in reproducing the response bits. Reliability can be measured by using Equation 3 and 4.  $R_s$  is the response from chip  $i$  at normal operating condition (at room temperature).  $R_{s,t}$  is  $t$ -th sample of  $R_s$  response from chip  $i$  at different operating condition such as different temperature setting. The good reliability value is 100%. As can be seen in Equation 4, if the HD intra (comparison of response under normal operating condition and different operating condition) is low or zero, then the reliability will be around 100%.

$$\text{Intra - chip HD} = \frac{1}{k} \sum_{t=1}^k \frac{HD(R_s, R'_{s,t})}{n} \times 100\% \quad (3)$$

$$\text{Reliability} = 100\% - \text{HD Intra} \quad (4)$$

The uniformity estimates how uniform is the ratio of '0's and '1's in the response bits of a PUF. Uniformity can be measured by calculating the Intra-Chip Hamming Weight HW as shown in Equation 5 where  $r_{s,l}$  is the  $l$ -th binary bit. The good value for uniformity is around 50% which means the response from RO is well distributed between '0's and '1's.

$$\text{Uniformity} = \frac{1}{n} \sum_{l=1}^n r_{s,l} \times 100\% \quad (5)$$

The bit-aliasing is to estimates the uniformity of '1's and '0's in each bit in the responses across a group of chips of the same type. Bit-aliasing can be measured by calculating the Inter-chip HW as shown in Equation 6. The good value for bit-aliasing is 50% which means each bit in the responses across a group of chip is well distributed between '0's and '1's. Uniformity and bit aliasing are the parameters that could measure one of the randomness features in the responses generated.

$$\text{Bit - aliasing} = \frac{1}{m} \sum_{i=1}^m r_{s,i} \times 100\% \quad (6)$$

Finally the new parameter, diverseness, will measure the range of frequencies in different number of stages used in ROPUF. Diverseness can be measured by calculating the standard deviation SD of the frequencies from each stage used in ROPUF as shown in Equation 7, 8, and 9.  $h$  is the number of ROs used on a chip.  $f_{i,j}$  is individual frequency for each RO.  $f_{i,j,q}$  is the  $q$ -th frequency sample of the  $j$ -th RO in the  $i$ -th chip.  $f_{avg}$  is the average frequency on a chip.

$$Diverseness = \sqrt{\frac{1}{h-1} \sum_{j=1}^h (f_{i,j} - f_{avg})^2} \quad (7)$$

$$f_{i,j} = \frac{1}{q} \sum_{q=1}^q f_{i,j,q} \quad (8)$$

$$f_{avg} = \frac{1}{h} \sum_{j=1}^h f_{i,j} \quad (9)$$

### III. EXPERIMENTAL SETUP

In this experiment three Xilinx Spartan 2 XSA-100 boards are used. There are 600 CLBs on each chip as shown in Figure 4 [11]. Each CLB contains two slices and each slice contains two Lookup Tables (LUTs). One stage in RO occupied one LUT. One CLB is used for the 4-stage RO and two CLBs are used for 5-stage and 7-stage RO. Six hundred 4-stage ROs and 300 5-stage and 7-stage ROs are mapped on each chip.

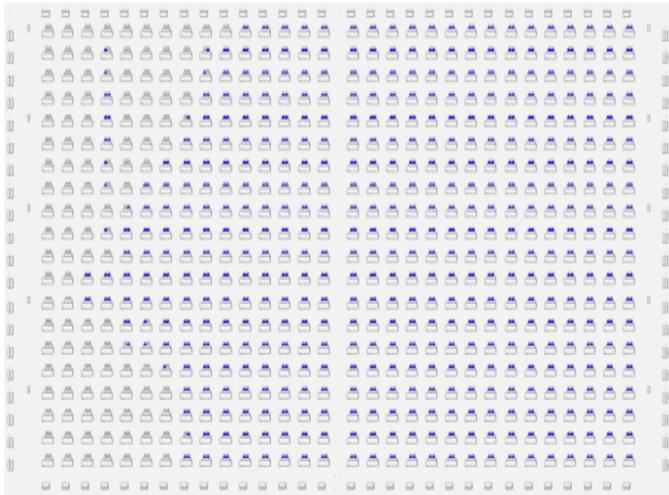


Fig. 4. Xilinx Spartan 2 CLBs layout

The FPGA area is divided into two areas left and right (three hundred CLBs on each area). The experiment will be run two times for each chip and RO stage. The first run occupied the right area with ROs and the left area with other circuits needed such as MUX and counters. The blue color boxes in Figure 4 show the occupied CLBs. It can be seen on the right side of Figure 4, 300 ROs occupied half of the FPGA. The other half of the FPGA is partially occupied by the other logics used in FST. The second run will just swap the left area for other logics and right area for ROs. For each RO, the frequency was recorded 10 times. Overall, 18000

frequencies for 4-stage ROs and 9000 frequencies for each 5-stage and 7-stage ROs were recorded.

Figure 5 shows the logic blocks for the FST test circuit. The challenge generator will produce the inputs to MUX which will activate one RO at a time. Each RO will be activated for 0.4 ms and there will be a 0.1 ms gap before the next RO is activated; this is to reduce the noise in the form of heat that generated from the adjacent CLB [8]. RO is activated from the top and moves down to the bottom of each column of the CLBs. A 0.2 ms gap is given between the RO and counter activation for the signal to be stabilized before the measurement starts. The timing controller will control all time intervals involved such as the time interval for each RO being activated and the time interval for the counter to measure each RO.

Frequency is computed using Equation 10 where  $x$  is the cycle counts from each RO and  $y$  is the cycle counts for the 50 MHz reference clock. The preset value for  $y$  is set to be 7000 cycles. That means the RO cycles will be measured within a 0.14 ms period. The accuracy of the measurement is 0.007 MHz/cycle which is good enough to notice the differences between frequencies generated from ROs.

$$x \times \frac{50}{y} \text{ MHz} \quad (10)$$

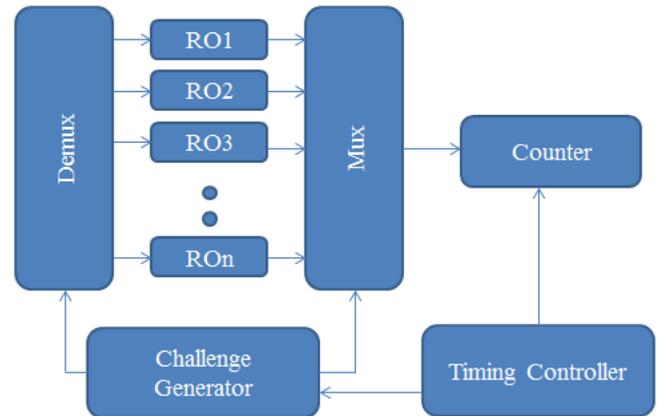


Fig. 5. FST circuit diagram

### IV. RESULTS AND ANALYSIS

Response bits from 4, 5, and 7-stage ROs are generated to calculate the diverseness, uniformity, uniqueness, bit-aliasing, and reliability. The response bits are generated using neighbor coding method where the neighboring ROs are compared [4]. The first response bit will be generated from the comparison of RO1 which is mapped in row 1 and column 1 of the CLB with RO2 which is mapped in row 2 and column 1 of the CLB. The comparison equation used is shown in Equation 1.

Table I shows the diverseness, uniformity, uniqueness, and bit-aliasing for 4, 5, and 7-stage ROs. The diverseness of frequencies for 4-stage is the highest compared to 5 and 7-stage. The results in Table I clearly show that as the number of stages used in ROs is reduced, the diverseness of ROs frequencies is increased. However, there is a limitation on

how low the number of stages can be used because each FPGA chip has its maximum operating frequency. Spartan 2 FPGA family has the maximum operating frequency of 200 MHz [11]. The lowest number of RO stages that can be used on Spartan 2 FPGA is 4 because the average frequency generated from the 4-stage RO on Spartan 2 is 182.77 MHz. The average frequency generated from the 3-stage RO on Spartan 2 is 220 MHz which has exceeded the maximum operating frequency for Spartan 2.

If an RO is producing frequency beyond the operating frequency of an FPGA, the other logics such as counter cannot measure the frequency from the RO correctly. Frequencies generated from the RO for all stages were verified using the Agilent Logic Analyzer, where the RO output is connected directly to the output pin of the FPGA board [9].

A high diverseness of ROs frequencies is good for ROPUF because it indicates that there are high amounts of frequency variations and important for generating higher number of good CRPs which will be discussed later in this section. For the authentication in ROPUF application, the challenge cannot be reused because this will reduce the security level of ROPUF as the response bits traverse the open domain for verification and is susceptible to the adversary attack [6]. This means that to make ROPUF practical, ample numbers of CRPs are needed.

A good RO is not just relying on the diverseness of ROs frequencies thus it needs to be proved that it has good uniformity, uniqueness, and bit-aliasing. As mentioned in Section II, good uniformity and uniqueness average should be around 50%. For the uniformity, the 5-stage ROs have the highest value and the 7-stage ROs have the lowest but still the difference is just 0.87%. The average uniformity results for all stages used can be considered good as the values are close to 50%. High uniformity value means the secret bits generated are uniformly distributed between 1s and 0s which is one of the good randomness characteristics.

TABLE I. DIVERSENESS, UNIFORMITY AND UNIQUENESS FOR 4,5 AND 7 STAGE ROS

Stage	Diverseness (MHz)	Uniformity (%)	Uniqueness (%)	Bit-aliasing (%)
4	1.9469	47.0228	40.1780	47.0228
5	1.2375	47.7146	34.5596	47.7146
7	0.7360	46.1538	40.5797	46.1539

For the uniqueness, the 4-stage and the 7-stage have better values compared to the 5-stage as can be seen in Table I. This shows that 4-stage and 7-stage have better uniqueness for the inter-chip comparison. Average uniqueness is obtained by comparing the responses generated from all three FPGA chips. The higher the differences between responses from each chip the higher value of the uniqueness. It is important to make sure the uniqueness is high because this represents that ROPUF could generate unique response from mass number of FPGA chips by given the same challenge.

For the bit aliasing, the 5-stage has the highest percentage that is 47.71% and the 7-stage has the lowest percentage that

is 46.15%. Nevertheless, all stages have good bit aliasing percentages that are close to 50%.

Table 2 shows the diverseness of ROs frequencies for 4, 5 and 7-stage for each FPGA chip used. The 4-stage ROs have the highest diverseness of ROs frequencies value for all three chips used compared to other stages. These results are still consistent with the previous results presented in [9] where we showed that as the number of stages used in RO reduced, the diverseness of ROs frequencies obtained will be higher. All three different FPGA chips are showing the same pattern where the diverseness of ROs frequencies increase as the number of stage in ROs is reduced. The same pattern obtained across the three different FPGA chips prove the consistency of our claim that the diverseness of ROs frequencies increases as the number of stages used in RO is reduced.

TABLE II. DIVERSENESS FOR CHIP1, CHIP2 AND CHIP3

	Diverseness (MHz)		
	CHIP 1	CHIP 2	CHIP 3
4 Stage	2.117440	2.586673	1.136851
5 Stage	0.938292	1.878756	0.895488
7 Stage	0.828066	0.821414	0.558649

The final check to ensure that ROs with lower number of stages can be a good ROPUF is the reliability. To calculate the reliability, responses need to be generated at different environment conditions. In this experiment, responses from 4, 5 and 7-stage ROs are generated at four different temperature settings as shown in Table III. The experiment was conducted in a controlled temperature test chamber. The frequencies from each RO will be recorded 10 times at each temperature setting. The responses are generated by comparing the average RO frequencies obtained. The bit generation equation used is shown in Equation 1.

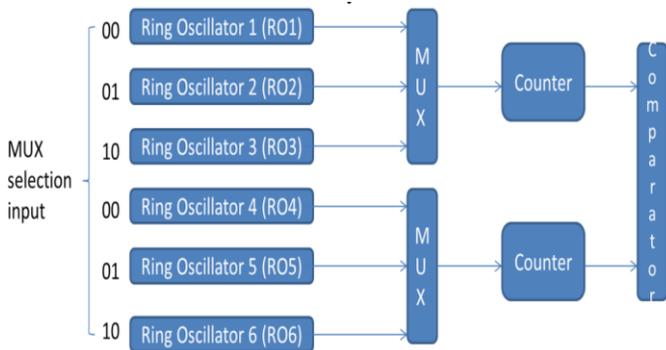
All responses obtained at various temperature settings are compared with the responses generated at room temperature. The results obtained are shown in Table III. The lowest reliability is 97.32% at 0°C for 4-stage ROs which mean 8 bits flipped out of 299 bits. The highest reliability is 99.332% at 20°C for 4 and 5-stage ROs which are 4 bits flipped out of 599 bits for 4-stage ROs and 2 bits flipped out of 299 bits for 5-stage ROs. From Table III it could be observed that reducing the number stages in ROs has no relationship with the reliability as there are no patterns can be observed.

TABLE III. RELIABILITY ON CHIP 3

ROs stage	Reliability %			
	0°C	20°C	45°C	70°C
4	98.1636	99.3322	98.9983	98.9983
5	97.3244	98.9967	98.9967	97.9933
7	98.3278	99.3311	98.9967	98.6622

The final step is to show the relationship between the number of stages used in ROs and CRP generation. To do this, all possible comparison pairs need to be generated. Note here that there are differences between the challenge and comparison pairs as shown in Figure 6. Challenge is selection of the comparison pairs to form a response bitstream. One challenge can consist of many comparison pairs depending on the design of the challenge and the length of the response. Figure 6 shows the example of ROPUF circuit that will generate three bit of response. In the left bottom of Figure 6 shows the list of all available comparison pairs with the MUX inputs on the right. In total there are only 9 comparison pairs available for this ROPUF circuit.

Figure 6 bottom right shows the list of possible challenges formation. The first three response bits will be generated from Pair 1, Pair 2, and Pair 3. Assume that comparison result for Pair 1, Pair 2, and Pair 3 are 1,0, and 1, then response bits are 101. The challenge for this response are the combination of the MUX inputs for Pair 1, Pair 2, and Pair 3 that are 0000 0001 0010. The number of possible challenges can be measure by  $n!/(n-r)!(r!)$ .  $n$  is the number of available comparison pairs and  $r$  is the number of response bits. As the number of available comparison pairs increase, the number possible challenges will also increase.



List of available Comparison Pairs:	List of possible Challenges that produce 3 response bits:
Pair 1: RO1-RO4 0000	Pair 1, Pair 2, Pair 3 = 0000,0001,0010
Pair 2: RO1-RO5 0001	Pair 1, Pair 2, Pair 4 = 0000,0001,0100
Pair 3: RO1-RO6 0010	Pair 1, Pair 2, Pair 5 = 0000,0001,0101
Pair 4: RO2-RO4 0100	Pair 1, Pair 2, Pair 6 = 0000,0001,0110
Pair 5: RO2-RO5 0101	Pair 1, Pair 2, Pair 7 = 0000,0001,1000
Pair 6: RO2-RO6 0110	Pair 1, Pair 2, Pair 8 = 0000,0001,1001
Pair 7: RO3-RO4 1000	Pair 1, Pair 2, Pair 9 = 0000,0001,1010
Pair 8: RO3-RO5 1001	Pair 1, Pair 3, Pair 2 = 0000,0010,0001
Pair 9: RO3-RO6 1010	

Fig. 6. Difference between comparison pairs and CRPs

The easiest way to generate all possible comparison pairs is by using selecting sort algorithm that has  $O(n^2)$  complexity. But as mentioned earlier the comparison pairs generated need to be good, which mean each comparison pair needs to pass certain FDT. To determine the FDT, the frequencies differences at all bit flips occurrence on all FPGA chips are checked. Then the maximum frequencies difference that caused the bit flip will be set as FDT. The majority bit flips

occurred when the frequency difference between ROs was 1 MHz and below. The maximum frequency difference where bit flip can occur is 3.5 MHz which will be set as the FDT in this experiment.

The algorithm used to generate the comparison pair is as shown below. The input of the algorithm will be all RO frequencies generated at room temperature. Then the algorithm will compare the frequency difference between one RO with the rest of ROs available based on the  $O(n^2)$  complexity. If the frequency difference pass the FDT, then those ROs will be selected as the comparison pair.

**Comparison Pair Generation in pseudocode**

**Input:**

- 1) 600 frequencies for 4-stage ROs and 300 frequencies for 5 and 7-stage ROs represented as ROs frequencies(i).
- 2) n is equal to the number of ROs.

**Output:** The list of all possible ROs comparison pairs that passed the FDT represented as ROs comparison pair(k,i).

**Algorithm**

```

1. i <- 0, j <- 0, k <- 1
2. for i = 1 to n-1
3.   for j = i + 1 to n
4.     frequency difference = absolute (ROs
       frequencies(i)-ROs frequencies(j))
5.     if frequency difference > FDT
6.       comparison pair(k,1) = i
7.       comparison pair(k,2) = j
8.       k++
9.     end if
10.  end for
11. End for
    
```

TABLE IV. NUMBER OF COMPARISON PAIRS GENERATED

FPGA Chip	ROs stage	FDT (MHz)			
		2	2.5	3	3.5
1	4	45757	28746	16606	8685
	5	5955	2749	1116	381
	7	1221	167	8	1
2	4	102800	95161	86713	75831
	5	22287	18422	14036	9776
	7	1171	525	342	302
3	4	37932	23095	12769	7122
	5	3811	2790	843	150
	7	154	44	1	1

Table IV shows the results obtained for good comparison pair generation. It can be seen that the highest number of comparison pairs generated from 4-stage ROs on FPGA chip 2 at FDT value is equal to 2 MHz. The lowest number of comparison pairs are generated from 7-stage ROs on FPGA chip 1 and 3 at FDT value equal to 3 and 3.5 MHz. There is a pattern in Table IV where the number of comparison pairs generated are higher when the number of stages used in RO is reduced. As the FDT increased by 0.5 MHz step, the number of comparison pairs reduced enormously.

As mentioned earlier the FDT used to filter all the bit flip occurrences is 3.5MHz. In Table IV it could be observed that the 7-stage is badly affected by the higher value of FDT. The comparison pairs that could be generated from 7-stage ROs is 302 on chip 2 and only 1 comparison pair on chip 1 and 3. This shows that 7-stage ROs cannot be used in ROPUF as the lower number of comparison pair generated really diminish the ROPUF application. For 5-stage ROs, the comparison pairs generated at FDT 3.5 MHz are very low for chip 1 and 3 (381 and 150). Except for 5-stage on chip 2 the comparison pairs that could be generated are 9776. The 4-stage ROs is showing the highest number of comparison pairs that could be generated at FDT equal to 3.5 MHz.

## V. CONCLUSION AND FUTURE WORK

This experiment was run on Xilinx Spartan 2 FPGA chip which uses 180 nm semiconductor process technology. This conclusion is based on Xilinx Spartan 2 FPGA and cannot be generalized on different FPGA technology. For the Spartan 2 FPGA chips, it can be concluded that the diverseness of ROs frequencies is increased as the number of stages used in RO is reduced. The lowest number of stages that can be used in RO is dependent on the operating frequency of the FPGA chip. For Spartan 2 FPGA the maximum operating frequency is 200 MHz, therefore the lowest number of RO stages that can be used is 4-stage RO as the frequency produced from 3-stage RO exceeded the maximum operating frequency. This paper shows that the lower number of stages used in RO will not compromise the uniqueness, uniformity, bit-aliasing, and reliability. The relationship between the number of stages used in ROs and CRPs has also been proven experimentally as the comparison pairs generation improved tremendously when a lower number of stages in RO is used.

In the future, the same experiment will be conducted on other FPGA technologies so that the conclusion can be applied across the semiconductor technology used on different FPGA chips.

## ACKNOWLEDGMENT

Muslim Mustapa is being supported for his Ph.D. through a fellowship program sponsored by the University of Malaysia Perlis.

## REFERENCES

- [1] C.E. Yin and Q.Gang, "Improving PUF Security with Regression-based Distiller," Design Automation Conference (DAC), pp. 1–6, Jun 2013.
- [2] S. Morozov, A. Maiti, P. Schaumont, "An Analysis of Delay Based PUF Implementations of FPGA," 6<sup>th</sup> International Symposium ARC 2010, Bangkok, Thailand, pp.382-387, March 17-19 2010.
- [3] R. Kastne and T. Huffmire, "Threats and Challenges in Reconfigurable Hardware Security," International Conference on Engineering of Reconfigurable Systems & Algorithms (ERSA), CSREA Press, July 2008.
- [4] A. Maiti, J. Casarona, L. McHale, P. Schaumont, "A large characterization of RO-PUF," HOST 2010, pp. 66-71, 2010.
- [5] H. Yu, P.H.-W. Leong, H. Hinkelmann, L. Moller, M. Glesner, P. Zipf, "Towards a unique FPGA-based identification circuit using process variations," International Conference on Field Programmable Logic and Application, pp.397,402, Aug. 31 2009-Sept. 2 2009.
- [6] G.E. Suh, S. Devadas, "Physical Unclonable Functions for Device Authentication and Secret Key Generation," Design Automation Conference DAC '07 44th ACM/IEEE, pp.9,14, 4-8 June 2007.
- [7] Y. Meng-Day, S. Devadas, "Secure and Robust Error Correction for Physical Unclonable Functions," Design & Test of Computers, IEEE, vol.27, no.1, pp.48,65, Jan.-Feb. 2010.
- [8] S. Lopez-Buedo, J. Garrido, E. Boemo, "Thermal testing on reconfigurable computers," Design & Test of Computers, IEEE, vol.17, no.1, pp.84,91, Jan-Mar 2000.
- [9] M. Mustapa, M. Niamat, M. Alam and T. Killian, "Frequency Uniqueness in Ring Oscillator Physical Unclonable Functions on FPGAs," MWSCAS 2013, pp. 465-468, Aug. 2013.
- [10] P. Sedcole and P.Y.K. Cheung, "Within-die delay variability in 90nm FPGAs and beyond," Field Programmable Technology FPT 2006, pp. 97-104, 2006.
- [11] Xilinx, "Spartan-II FPGA Family Data Sheet," DS001 June 13 2008
- [12] A. Maiti, V. Gunreddy, P. Schaumont, "A Systematic Method to Evaluate and Compare the Performance of Physical Unclonable Functions", Eds. P. Athanas, D. Pnevmatikatos, N. Sklavos, Springer 2012, ISBN 978-1-4614-1361-5.
- [13] Hori Y., Yoshida T., Katashita T., Satoh A., "Quantitative and statistical performance evaluation of arbiter physical unclonable functions on fpgas," International conference on reconfigurable computing and FPGAs (ReConFig) 2010, pp 298–303, Dec 2010.
- [14] Majzoobi M., Koushanfar F., Potkonjak M., "Testing techniques for hardware security," IEEE international test conference, ITC 2008, pp 1–10.
- [15] Su Y, Holleman J, Otis B., "A digital 1.6 pj/bit chip identification circuit using process variations," IEEE J Solid-State Circ 43(1):69–77.
- [16] Yamamoto D, Sakiyama K, Iwamoto M, Ohta K, Ochiai T, Takenaka M, Itoh K, "Uniqueness enhancement of puf responses based on the locations of random outputting rs latches," Proceedings of the 13th international conference on Cryptographic hardware and embedded systems, CHES 2011. Springer, Berlin, Heidelberg, pp 390–406.

# Autonomic Intrusion Detection System in Cloud Computing with Big Data

Kleber M.M. Vieira, Fernando Schubert, Guilherme A. Geronimo,  
Rafael de Souza Mendes, Carlos B. Westphal  
{kleber, schubert, r2, mendes, westphal}@inf.ufsc.br  
LRG - INE - UFSC - Florianopolis - SC - Brazil

**Abstract**—This paper analyzes real-time intrusion response systems in order to mitigate attacks that compromise integrity, confidentiality and availability in cloud computing platforms. Our work proposes an autonomic intrusion response technique enabling self-awareness, self-optimization and self-healing properties. To achieve this goal, we propose IRAS, an Intrusion Response Autonomic System, using Big Data techniques for data analytics and expected utility function for decision taking.

## I. INTRODUCTION

The quickly expansion in the volume of data generated in the Internet, with the consequent diffusion of personal, financial, legal and other data in the Web, has created a very valuable content for hackers, crackers and other cyber-criminals.

In this context, the need for a highly effective and quickly reactive security system gains importance. The growing number of attacks and vulnerabilities exploitation techniques requires preventive measures by system administrators. These measures are getting more complex with the growth of data heterogeneity and the increasing complexity of the attacks. In addition, slow reaction time from human agents and the huge amount of data and information generated makes the decision making process an arduous task. In response to this, there is an increase in the usage of IDS (Intrusion Detection Systems) [1], as a way to identify attacks patterns, malicious actions and unauthorized access to an environment [2].

The need for IDS is growing due to limitations from IPS (Intrusion Preventing Systems) - which focus on alerting administrators when vulnerability is detected, connectivity and threat evolution, as well as the financial appeal of cybercrime [3].

Despite its growing importance, current IDS solutions available have limited response mechanisms. While the research focus is on better intrusion detection techniques, response and effective reaction to threats are still mostly manual and rely on human agents to take effect [4].

Recently, some intrusion detection tools started to provide some limited set of automated responses, but with the intrusions growing complexity, the need for more effective response system strategies have raised. Due to implementation limitations, research works on intrusion detection techniques advance in a faster rate than intrusion response systems [2].

The development of reliable and quickly responsive systems is even more important for cloud computing, where the

elasticity increases the risk and costs of an attack [5].

## A. Motivation

The number of computer attacks has grown in quantity and complexity, making defense an increasingly arduous task. Each computer that suffers an attack has very limited information on who initiated the attack, and the origin of it. Current systems for intrusion detection and response do not follow the growing number of threats [4].

The focus on manual processes creates a delay between detection and response, leaving a time window for attackers [6]. Research findings by [4] indicate that if a skilled attacker has a range of 10 hours between intrusion and response, the attack has 80% chance of being successful, if the attacker has 20 hours, the attack has 95% of chances of being successful, and if the attacker has 30 hours the attack becomes virtually foolproof. In this situation, the system administrators skills become irrelevant. On the other hand, if the response is instant to the intrusion, the chance of a successful attack is almost zero. [4] says that statistics have shown that the number of pro-rated intrusion is growing. The high cost of the contract indicates serious financial commitment made by the Pentagon to prevent and secure their infrastructure from being attacked by another country.

An automated intrusion response system that combines the best techniques of intrusion detection would provide the best defense possible in short time, giving more time to the system administrator to develop a permanent solution to avoid further attacks or fix the vulnerability exploited [6][4].

According to Buyya, the Cloud [7] is complex, large-scale and heterogeneous and its management is challenging. This environment requires an automated and intelligent system to provide security services with efficient cost. Thus, cloud systems represent a distinct structure with several layers of abstraction that requires specific IDS and response techniques to address its complexity.

The paper is organized as follows: Section 2 describes the proposal underlying concepts and key technologies, section 3 presents an overview of the related work, section 4 details the proposal and section 5 concludes the paper with future directions and open challenges.

## II. BACKGROUND

Autonomic computing can overcome heterogeneity and complexity of computing systems being considered a new and

effective approach to implement complex systems, addressing several issues where humans are losing control due to complexity and slow reactions, such as security systems [8].

The autonomic computing model is based on the so called self properties. The self is inspired by the autonomic nervous system of the human body, which can manage multiple key functions through involuntary control. The autonomic computing system is the adjustment of software and hardware resources to manage its operation, driven by changes in the internal and external demands. It has four key features, including self-configuration, self-healing, self-optimization and self-protection.

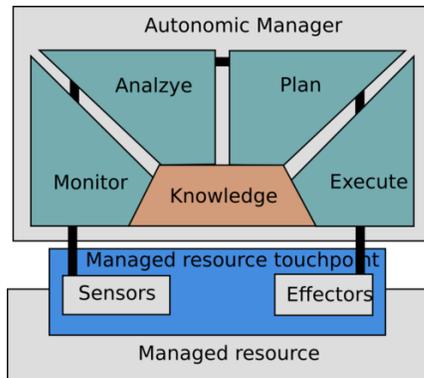


Fig. 1. An autonomous system.

Figure 1 shows the structure of an autonomic system and its MAPE-K cycle [9], composed by the monitoring, analysis, planning and executing modules. All the management of the autonomic component is performed by a meta-management element, which make decisions based on the knowledge-base built.

Sensors are responsible for collecting information from the managed element. Information collected by the sensors is sent to the monitors where they are interpreted, pre-processed, aggregated and presented in a higher level of abstraction. After this, the analysis phase is executed and planning takes place. At this stage, a work plan is resulted, which consists on a set of actions to be performed by the executor. Only the sensors and executors have direct access to the managed element. Through the autonomic management cycle, there may be a need for decision-making, thus it is also necessary the presence of knowledge base [10].

#### A. Autonomic Systems Properties

The essence of autonomic computing is self- management. To implement it, the system must be self-aware as well as environment-aware. Thus, the system must precisely know its current situation and be aware of the operational environment in which it operates. From a practical standpoint, according to [10], the term autonomic computing has been used to denote systems that have the following properties:

- Self-awareness: the system knows itself: its components, their state and behavior;

- Context-awareness: the system must be aware of the context of its execution environment and be able to react to changes in its environment;
- Self-configuring: the system must dynamically adjust its resources based on its status and the state of the execution environment;
- Self-optimizing: the system is able to detect performance degradations and functions to perform self-optimization;
- Self-protecting: the system is able to detect and protect its resources from external and internal attackers, keeping its overall security and integrity;
- Self-healing: the system must have the ability to identify potential problems and to reconfigure itself in order to continue operating normally;

### III. RELATED WORK

In this section, four related works that we considered important to our research were selected. To evaluate these works, five topics were chosen to analyze them. The chosen topics are: The chosen topics are:

- Does it propose IDS?
- Is it suitable for the Cloud scenario?
- Does it respond against attacks?
- Does it have a Self-Healing method?
- Which kind of algorithm is used?

#### A. Proposal of Wu 2013

[11] propose an autonomous manager which introduces a mechanism for multi-attribute auction. Its architecture has a layer of managed resources covering generically all physical devices like routers, servers or software applications. These resources should be manageable, observable, and adjustable. The state of resources refers to all data (events) that reflect the state of existing resources, including logging and real-time events. This architecture also has an autonomous agent as a detection engine, optimization strategy, autonomic response, and knowledge base module.

The architecture has agents responsible for MR information capture, preprocessing and redundancy removal before final submission to AM agents.

The multi-attribute auction model is defined as follows:

The auction model:  $M = A, B, S, V, C, Res$

$A$  refers to attributes, each auction has  $n$  attributes,  $(A_1, \dots, A_n)$ .

$B$  refers to the auction participants that needs to buy (win) an event.

$S$  refers to a seller (auctioneer) which includes  $n$  buyers.

Buyers can provide events with different attributes.  $V$ :  $V_i R$  refers to a buyer function.  $C$  refers to seller costs.  $Res$  refers to the transaction  $Res = P$ ,  $P$  is one of the members of  $A$ .

The buyer benefit  $B$  is determined by  $U = V(a) - P$  and the seller benefit is  $S_i$  is  $U_i = P - C_i(a)$

The auction process is performed in four steps: 1) the buyer publishes the evaluation  $V(a)$ . 2) each seller  $i$  makes a  $B_i, 0$  proposal. 3) the transaction is committed. 4) the transaction is processed.

Wu says that the autonomic response depends on knowledge base of possible actions. It is necessary to form knowledge base with attributes and valuations [11].

#### B. Proposal of Kholidy 2013

Kholidy approach describes how to extend the current technology and IDS systems. His proposal is based on hierarchical IDS [12]) to experimentally detect DDoS, host-based, network based and masquerade attacks. It provides capabilities for self-resilience preventing illegal security event updates on data storage and avoiding single point of failure across multiple instances of intrusion detection components.

His proposal consists on a hierarchical structure, autonomic and cloud-based, extending his earlier work [12] with features such as autonomic response and prediction. In particular, it assesses vulnerabilities and risks in the system through a mechanism that builds a security model based on risk assessment and security event policies criticality. It also provides the possibility of automatic response to actions based on a set of policies defined by the system administrator. However, a black box format does not clarify possible answers or makes clear how to choose the best answer leaving that decision to a system administrator. Finally, the architecture offers some predictive capabilities based on Holt-Winters algorithm [13], which predicts and detects abnormal behavior of network traffic when the amount of collected network traffic is either too high or too low, compared to normal network traffic. Predictive capabilities improve detection accuracy of both decision making and automated response [14].

#### C. Proposal of Vollmer 2013

This article describes new architecture that uses concepts of autonomic computing based on SOA and external communication layer to create a network security sensor. This approach simplifies the integration of legacy applications and supports a safe, scalable, self-managed structure.

The contribution of this work is a flexible two level communication layer, based on autonomic computing and SOA. A module uses clustering and fuzzy logic to monitor traffic for abnormal behavior. Another module passively monitors network traffic and deploys deceptive hosts in the virtual network.

This work also presents the possibility of an automatic response but it does not address this topic in detail, leaving it for future works [15].

#### D. Proposal of Sperotto 2012

It presents an autonomic approach to adjust the parameters of intrusion detection systems based on SSH traffic anomaly.

This paper proposes a procedure which aims to automatically tune system parameters and, in doing so, to optimize system performance. It validates their approach by testing it on a probabilistic-based detection test environment for attack detection on system running SSH [16].

#### E. About the related works

Related works representing the state of the art attempt to solve the problem of cyber-attacks by proposing intrusion detection mechanisms and increasing detection techniques. Although many of them show the need of automatic responses, none of them go deeper in this direction. The works of [11] and [15] mention the possibility of response to attacks; however, both works leave this point open not going deep into the issue.

Table I shows a brief comparison between the related works, based on the previous described topics.

### IV. PROPOSAL

In this work, we propose a model for autonomic intrusion detection system based on the autonomic loop, commonly referenced as MAPE-K (Monitor, Analyze, Plan, Execute and Knowledge Base). To monitor and analyze, we use sensors to collect data from IDS logs, network traffic, system logs, and data communication. For storage and further analytics, a distributed storage is used, for instance we chose Apache Hadoop as storage engine because its performance, scalability and further capabilities to be extended and suffer Map Reduce jobs.

For analysis, planning and execution we present a model based on expected utility function [17].

#### A. Proposed system: IRAS Intrusion Responsive Autonomic System

The approach of IRAS follows the line of an autonomic system for intrusion response. The sensors collect log data from network IDS and host systems. This information is compiled in a Big Data environment [18], preprocessed and placed on a higher level of abstraction, ready to be sent to analysis and planning cycles of the autonomic loop.

Based on the MAPE-K autonomic loop, the phases of IRAS are:

*M* data collection from sensors, storage on Big Data infrastructure.

*A* preprocessing (filtering, aggregation) and analysis.

*P* calculation of utility.

*E* execution, this means, based on results of utility function, effective measures will be taken in the system.

*K* the knowledge base built from the monitored and analyzed data is used to feedback the utility based function, weighting the utilities.

#### B. Monitoring

The first phase of the MAPE-K autonomic cycle corresponds to monitoring. In this step, sensors are used in order to obtain data reflecting changes in behavior of the managed element or information from the execution environment that are relevant to the self-management process.

The concept of sensor is a little generic, but it is possible to consider that a sensor is a component of the system that makes the connection between the external world and the management system.

Author	IDS	Cloud	Response	Self-healing	BigData	Algorithm
[11]	yes	no	yes	no	no	Auction
[14]	yes	yes	yes	no	no	Holt- Winters
[15]	yes	no	yes	no	no	Fuzzy
[16]	yes	no	no	no	no	Flor-based

TABLE I  
RELATED WORKS

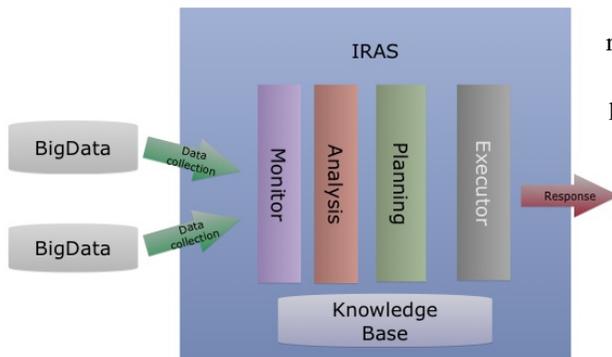


Fig. 2. IRAS Intrusion Responsive Autonomic System

However, the important nuance to observe in data monitoring for security in Cloud Computing, is that the data will be intrinsically temporal. This characteristic impose some peculiarities in the data structure to storage temporal information, as well as, in the queries to be executed in the sensor data base to retrieve useful information.

As defined in [19], Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze. [20] defines the three data characteristics of Big Data sets: volume, variety and velocity. We have a large volume of data from various sources like logs, IDS alerts, network traffic scans, where processing and analysis speed is necessary to extract meaningful information from these sources. Based on work by Suthaharan [18] where it was decided to use a structure with BigData tools, in this case Hadoop to organize the collected data in the cloud and perform the monitoring. However, Suthaharan use Machine Learning (ML) to find attacks and in this work we propose to use a technical knowledge based on intrusion detection systems [1] making it possible to detect attacks like Stuxnet or Duqu ones. Thus we make a map & reduce over the collected data to identify signatures of known attacks extracting some significant data such as origin, destination of attack, type, signature and timestamp.

### C. Analysis

There is a really resourceful set of analytics methods that correlates data in order to discover causality relationship, or events association. However, it is possible to think in three types of analytical methods which are useful for Cloud Computing security:

diagnostic the method means to synthesize a temporal flow of events arising from sensors in a *security state* of the cloud – it

is common represent the state as a dashboard;  
root-cause the goal of this type of analysis is to determine what events are the main causes of the actual cloud state;  
prediction the prediction methods aims to suggest forecast projections to cloud state.

It is possible to consider that the analysis phase in Cloud Computing security management must have some characteristics:

- there must exist evaluation methods able to supply a set of security metrics for parts and for the whole cloud;
- it must consider the uncertain – uncertainty of the diagnostics provided by analytics methods must carry some fuzzy or probabilistic measure to represent uncertainty;
- it must consider temporality – generally based on time series;
- it must be multi-criteria – may there exists multiples, seemingly uncorrelated, events that articulated, constitute an attack;
- it must be real-time – an fundamental characteristic of the events in Cloud Computing is the need to provide real-time evaluations of the cloud state;
- it must learn – the measures in a real world Cloud changes their statistical distribution, variance and behavior – in this context, an analytical method to security in Cloud Computing must be adaptive to follow this changes.

In this way, our proposal to proceed with the diagnostic is composed by two steps:

- a security state set, where each machine  $m \in M$  have a vector  $Al_m = (al_1, t_1, al_2, t_2, \dots, al_n, t_n)$  that contains all security alarms coming from IDs sensors  $al_x$  and the time  $t_x$ , since the alarm was triggered. The set of all possible cloud security states can be obtained by the product of arbitrary sets  $(Al_m)_{m \in M}$ , such that  $S = \prod_{m \in M} Al_m$ , where  $M$  is the set of all cloud machines (VMs and PMs);
- two utility functions, to evaluate the security value of machines, and to evaluate the total cloud state, respectively (1) and (2), given:

$$u(m) = - \sum_{k=1}^{n_m} w(al_k) \cdot (t_{now} - t_k) \quad (1)$$

, where  $n_m$  is the number of alarms triggered to machine  $m$ ,  $w(al_k)$  is the weight of alarm  $k$ , and  $(t_{now} - t_k)$  is the amount of time that the alarm  $k$  is triggered.

$$\sum_{m \in M} w_m \cdot u(m) \quad (2)$$

, where  $w_m$  is the weight that each machine  $m$  have in the cloud security, and  $u(m)$  is the security evaluation of  $m$ .

The root-cause analysis will not be addressed in this work. But, it may be important to correlate and determine the *what* and *how* of some configuration states (e.g. a blocked ip address in the firewall) influence the occurrence of security incidents. In this way, a sensor component that reads the data from logs, IDS agents, VM and Hypervisor [21] data collectors, network traffic sniffers, SNMP agents and alarms. This analysis will be important to determine and discover possible security actions.

The prediction will be important to establish the consequences of an action  $a \in A$  execution, where  $A$  is the set of all possible actions, over a state  $s \in S$ . So, the prediction of action consequences must provide a probability function  $p(s^{t+1}|a, s^t)$ , read as: the probability of action  $a$ , executed over a state  $s^t$  in time  $t$  conduce to a state  $s^{t+1}$  in time  $t+1$ .

#### D. Planning

In the planning phase, we will use a simple utility maximization method. However, it is interesting to study the Markov Decision Process (MDP) mathematical framework. It will supply a interesting set of elements to guide our decision function.

MDP is a framework generally described in the follow way:

- a set  $S$  of system states – here, product of the diagnostic method of analysis phase;
- a set  $A$  of possible actions to be taken in the system;
- a probability transition function  $P : S \times A \times S \rightarrow \mathbb{R}$  that express the probability of the system in state  $s$ , given an action  $a$  be conduced to a state  $s'$  – here, the probability function will be product of the forecasts provided analysis method;
- a reward function  $R : S \times A \times S \rightarrow \mathbb{R}$  that evaluate the reward of take the action  $a$  in state  $s$  and conduce the system to state  $s'$ .

There exists another ways to describe an MDP, however, this is the most useful to our objectives, that are use MDP for Cloud Computing security management.

It is important to observe that MDP is known as it does not work under incomplete information systems. To use this approach, we consider that:

- 1) Cloud Computing security monitoring will provide a big data environment that can supply the information needed by MDP;
- 2) the set of possible states will be finite and treatable;
- 3) there is an enough number of analytical methods to supply the forecasting needs to support probability function;
- 4) there are an enough number of analytical methods to supply the needs of state evaluation to support the reward function;

So, considering that there is an utility and a probability function to evaluate the current state, predict the future state and after execute an action, evaluating the future actions, we

will propose to select the action with the maximum reward function since it runs in state  $s^t$  (3).

$$r(a|s^t) = \left( \sum_{s^{t+1} \in S} p(s^{t+1}|a, s^t) \cdot u(s^{t+1}) \right) - u(s^t) \quad (3)$$

The reward function establish the difference between the utility of the each possible future state  $p(s^{t+1}|a, s^t) \cdot u(s^{t+1})$  and the utility of the current state  $u(s^t)$ .

So to choose an action, the function must be (4).

$$a = \arg \max_{a \in A} r(a|s^t) \quad (4)$$

#### E. Executor

With the calculation of the response with the highest expected utility, it is possible to forward the response to an executing agent in the cloud. The hypervisor is responsible for executing the response getting transperance for each virtual machine.

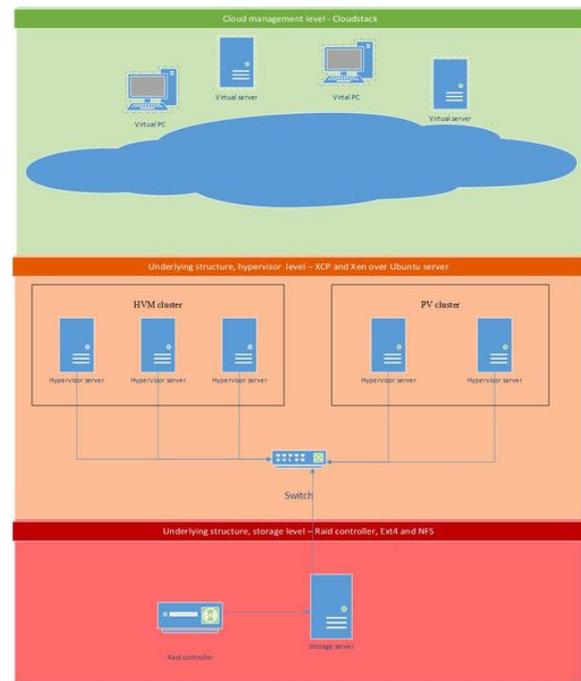


Fig. 3. Cloud environment for validation

#### F. Discussion

As shown on table 1, our work presents an increment in art when you use Big Data to locate attack occurrences and be able to provide a response that takes into consideration the impacts of the attack across the Cloud environment. Regarding the authors Wu et al. (2013), Vollmer et al. (2013) and Idss et al. (2012) the contribution of our research was to consider the environment Cloud and its peculiarities as the hypervisor, the complexity of providing an answer without being invasive to customers. Our work also considers self-healing and uses statistical function in expected utility to achieve the most efficient response and thereby, block the attacks.

## V. CONCLUSION

This paper suggests the use of autonomic computing to provide response to attacks on cloud computing environments. Thus, it is possible to provide self-awareness, self-configuration and self-healing in the cloud. An architecture that uses the expected utility function for choosing an appropriate response is a statistical model to adjust the answers given in order to provide more results. Furthermore, the work proposes the use of Big Data infrastructure using Hadoop to organize the large volume of data and extract information using the Map- Reduce framework. Thus, we could provide intrusion detection, response and self-healing in cloud environment.

## REFERENCES

- [1] K. Vieira, A. Schulter, C. Westphall, and C. M. Westphall, "Intrusion detection for grid and cloud computing," *It Professional*, vol. 12, no. 4, pp. 38–43, 2010.
- [2] N. Stakhanova, S. Basu, and J. Wong, "A taxonomy of intrusion response systems," *International Journal of Information and Computer Security*, vol. 1, no. 1, pp. 169–184, 2007.
- [3] C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, and M. Rajarajan, "A survey of intrusion detection techniques in Cloud," *Journal of Network and Computer Applications*, vol. 36, pp. 42–57, Jan. 2013.
- [4] K. Lumpur, "An investigation and survey of response options for Intrusion Response Systems ( IRSs )," 2010.
- [5] P. Mell and T. Grance, "The nist definition of cloud computing (draft)," *NIST special publication*, vol. 800, no. 145, p. 7, 2011.
- [6] C. A. Carver, "Intrusion response systems: A survey," *Department of Computer Science, Texas A&M University, College Station, TX*, pp. 77843–3112, 2000.
- [7] R. Buyya, R. Calheiros, and X. Li, "Autonomic Cloud computing: Open challenges and architectural elements," *Emerging Applications of ...*, pp. 3–10, Nov. 2012.
- [8] J. Kephart and D. Chess, "The vision of autonomic computing," *Computer*, vol. 36, pp. 41–50, Jan. 2003.
- [9] M. C. Huebscher and J. A. McCann, "A survey of autonomic computingdegrees, models, and applications," *ACM Computing Surveys (CSUR)*, vol. 40, no. 3, p. 7, 2008.
- [10] S. Hariri, B. Khargharia, H. Chen, J. Yang, Y. Zhang, M. Parashar, and H. Liu, "The autonomic computing paradigm," *Cluster Computing*, vol. 9, no. 1, pp. 5–17, 2006.
- [11] Q. Wu, X. Zhang, R. Zheng, and M. Zhang, "An Autonomic Intrusion Detection Model with Multi-Attribute Auction Mechanism," vol. 10, no. 1, pp. 56–61, 2013.
- [12] H. A. Kholidy, A. Erradi, S. Abdelwahed, and F. Baiardi, "Ha-cids: A hierarchical and autonomous ids for cloud systems," in *Computational Intelligence, Communication Systems and Networks (CICSyN), 2013 Fifth International Conference on*, pp. 179–184, IEEE, 2013.
- [13] C. Chatfield, "The holt-winters forecasting procedure," *Applied Statistics*, pp. 264–279, 1978.
- [14] H. Kholidy, A. Erradi, S. Abdelwahed, and F. Baiardi, "A hierarchical, autonomous, and forecasting cloud IDS," pp. 213–220, 2013.
- [15] D. Vollmer, M. Manic, and O. Linda, "Autonomic Intelligent Cyber Sensor to Support Industrial Control Network Awareness," *IEEE Transactions on Industrial Informatics*, vol. PP, no. 99, pp. 1–1, 2013.
- [16] A.-b. Idss, S. S. H. Case, A. Sperotto, M. Mandjes, R. Sadre, P.-t. D. Boer, A. Pras, and P.-T. de Boer, "Autonomic Parameter Tuning of Anomaly-Based IDSs: an SSH Case Study," *IEEE Transactions on Network and Service Management*, vol. 9, pp. 128–141, June 2012.
- [17] R. F. Bordley and S. M. Pollock, "A decision-analytic approach to reliability-based design optimization," *Operations research*, vol. 57, no. 5, pp. 1262–1270, 2009.
- [18] S. Suthaharan, "Big data classification: Problems and challenges in network intrusion prediction with machine learning," in *Big Data Analytics workshop, in conjunction with ACM Sigmetrics*, 2013.
- [19] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data: The next frontier for innovation, competition, and productivity," May 2011.
- [20] P. Zikopoulos, C. Eaton, et al., *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media, 2011.
- [21] C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, and M. Rajarajan, "A survey of intrusion detection techniques in cloud," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 42–57, 2013.

# Fourier Transform as a Feature Extraction Method for Malware Classification

Stanislav Ponomarev, Nathan Wallace, and Travis Atkison

Louisiana Tech University Ruston, LA, 71270

{spo013, nsw004, atkison}@latech.edu

**Abstract**—*Research efforts to develop malicious application detection algorithms have been a priority ever since the discovery of the first “viruses”. In this research effort Fourier transform is used to extract features from binary files. Each byte in these files is treated as a value of a discrete function. Discrete Fourier Transform then transforms binary files into frequency domain. Each frequency is used as a feature for malware classification. These features are then reduced by random projection algorithm to create a set of low-dimensional features that are used to classify whether the application is malicious or not. A 99.6% accuracy was reached by Random Forest classifier, while processing various worms, trojan horses, viruses, and backdoors.*

**Keywords:** Computer security, Virus issues

## 1. Introduction

Any Turing-complete machine can run malicious code that is designed to “harm or subvert a system’s intended functionality”. Applications utilizing such code are known as “malware” [1], [2]. Turing-complete machines include a vast set of devices - from personal computers and cell phones, to machinery automation and utility distribution controllers. According to CTIA - The Wireless Association, at the end of 2012, there were 326.4 million cell phone subscribers in the US alone [3]. Add in personal computers, laptops, and tablets and the list of possible malware carriers greatly expands. All of these devices now communicate over a network, which means they can be infected by malicious code without any user interaction. Non-networked devices can also be compromised by such code through user interaction - connecting it to computer, inserting a flash drive that contains infected files, or transferring data in any other means.

There have been many studies of detection and protection against malicious applications initiated by both industry and university labs [4]–[7]. Many classes of malicious applications have been defined. The most common ones include “viruses” - “a program that can ‘infect’ other programs by modifying them to include a possibly evolved copy of itself” [6], “worms” - “a program that self-propagates across a network exploiting security or policy flaws in widely-used services” [8], and “Trojan horse” - a term derived from

Greek mythology describing a software that masks itself as having useful features for the user [9].

Cohen was the one of the first researchers to study the defense against malicious software [6]. He also recognized the risks of a widespread “infection”, which was much harder in the time his publication was written due to low network connection count of the computers in that time period. Currently, many types of anti-virus software exist. They utilize static and dynamic analysis, neither of which are perfect [5], [10].

Static analysis refers to a set of algorithms that can determine whether a code is benign or malicious by looking at its signatures [5]. These algorithms typically compare the signatures of the scanned files with a database of known malicious code signatures. If the signatures match, the file is marked as malicious. This approach results in two major problems. Only malicious code that has already been “captured” and proven to be malicious will be in the database. Which means “zero day viruses” (viruses that were just discovered) have had a potential to stay hidden on users’ machines for years or until the security experts find a copy of such a virus.

Dynamic analysis typically runs an executable inside a virtual environment to determine whether it is malicious or not [10]. However, dynamic analysis can be obfuscated by certain conditional statements such as malicious code execution only on a certain date. Furthermore, it is possible to determine if a program is running inside the virtual environment and to write code that can escape virtual environment into the host system [11].

According to the Symantec 2013 Internet Threat Report [12], one in 291 emails sent in the year of 2012 contained a malicious application. One of the primary reasons for such an abundance of malware are “attack kits” - a set of tools that allow almost anyone with some computer knowledge to create or modify new viruses almost instantly. Attack kits allow the class of malware writers to grow past the people with computer penetration knowledge, and include the average users, who might try writing viruses for various reasons from unintentional to active attacks. Combining such tools with simple execution obfuscation techniques allows attackers to create a new strand of a virus by simply morphing the old one. Christodorescu refers to it as a “game between malicious code writers and researchers working on

malicious code detection” [5].

In this article, the principles of static analysis and data mining are used in this ongoing research effort to create a set of trained classifiers that are more robust in detecting malicious applications than signature based detection methods commonly used in modern anti-virus applications. By doing so, it is possible not only to detect malicious software that has already been known, but a malicious software that has been obfuscated by various methods.

Atkison was first to suggest the use of random projection in combination with  $n$ -gram analysis and data mining algorithms to classify computer applications [13]. Durand then further analyzed different parameters of  $n$ -gram analysis and the target feature count of random projection algorithm in combination with several common classifiers [14]. He determined that the 4-gram analysis, using 1500 features as a target feature count, and Support Vector Machines classifier have the highest accuracy of classification. This research extends the use of these algorithms to create a code that can be efficiently run on common computers. Fourier transform is evaluated as a possible replacement for  $n$ -gram analysis.

## 2. Background

The problem of malicious application detection is very popular, well-studied, and has gathered a significant body of research [2], [8], [9], [13]. All research ventures can be categorized as either static analysis or dynamic analysis. Static analysis refers to the process of determining whether an application is malicious without actually running the program in question. Dynamic analysis describes the process of determining whether a program is malicious by monitoring the behavior of a suspect program by executing it, usually within a virtual environment. Neither one of these approaches is a complete solution in itself, but each has a part to play in producing better malware detection systems.

### 2.1 $n$ -gram Analysis

When dealing with information retrieval or data mining, the features extracted from the data set play a pivotal role in the success of the prediction process. The information retrieval technique of  $n$ -gram analysis has proven to be a valuable tool for feature extraction in several research efforts which focus on the detection and/or classification of malicious applications [2], [4], [15]–[17]. An  $n$ -gram is any substring of length  $n$  [18]. Since  $n$ -grams overlap, they do not just capture statistics about sub-strings of length  $n$ , but also implicitly capture frequencies of longer sub-strings [19]. However, due to the high dimensionality of  $n$ -gram feature sets, the gathered data is a subject to the “curse of dimensionality” [20]. Many of these research efforts use some form of dimensionality reduction to curb these large feature sets in order to mitigate the effects. For this particular experiment,  $n$ -gram of size 4 is used, as it is shown to yield higher classification accuracy as shown in [21].

### 2.2 Fourier Transform

As a possible replacement for the  $n$ -gram analysis technique, Fourier transforms can be used to provide the information about frequency of byte patterns in malicious applications. Fourier transform is commonly used in digital signal processing to transform the signal from time domain to frequency domain by following the equation 1, where  $f(x)$  is the incoming signal, and  $\hat{f}(\sigma)$  is the signal in frequency domain.

$$\hat{f}(\sigma) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i x \sigma} dx \quad (1)$$

A discrete Fourier Transform can also be applied to a sequence of  $N$  complex numbers as defined in equation 2. Here,  $x_n$  is the value of the signal at a discrete time offset  $n$ , and  $X_k$  is the complex number that represents the amplitude and phase of a sinusoidal component in  $x_n$  at a frequency  $k/N$ . Since the analyzed frequency range depends only on the size of analysed data,  $N$ , a comparison of different length signals can be achieved by padding shorter signal with zeros, and maintaining the same sample size,  $N$  for both signals.

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-\frac{i2\pi kn}{N}}, k \in \mathbb{Z} \quad (2)$$

### 2.3 Random Projection

The feature selection technique known as random projection has been recently applied to the field of malware detection [14]. Random projection is a feature extraction technique which embeds a high dimensional feature set into a “low-dimensional subspace using a random matrix whose columns have unit length” [22], thus creating a completely new set of features. Random projection feature extraction technique was first introduced to the realm of malicious application detection in [23]. Similarly to Kolter, in [14], a vector space model was used with  $n$ -gram analysis to produce weighted feature vectors from binary executables [7]. Every dimension of these vectors represented a unique  $n$ -gram which could be extracted from the corresponding executable. Generated feature vectors were then used as input to random projection algorithms in order to produce feature vectors of a reduced dimension. Random projection used Achlioptas’ matrix multiplication with a random matrix of values of 0, +1, or -1 following a probability distribution of 2/3, 1/6 and 1/6 respectively to reduce the feature vectors [24]. Previous findings have shown the use of random projection to reduce the feature set to 1500 features to result in higher accuracy [21].

### 2.4 Classification algorithms

Nine classification algorithms were chosen for this research: Naïve Bayes, SVM, Simple Logistic, Bagging, Ridor,

Decision Stump, J48, LMT, and Random Forest. Previous research results used some of these methods to classify malicious applications [21]. A new set of bagging type classifiers was chosen because they perform well with training sets containing large noise [25].

#### 2.4.1 Naïve Bayes

Naïve Bayes classifier assumes that all the features are independent of each other and follows a Bayesian probabilistic model:

$$p(C|F_0, F_1, \dots, F_n) = \frac{p(C) \cdot p(F_0, F_1, \dots, F_n|C)}{p(F_0, F_1, \dots, F_n)}$$

where  $C$  is the class of dataset,  $F_0, \dots, F_n$  is a set of features, and  $p()$  is a probability function.

#### 2.4.2 SVM

Support Vector Machines constructs a hyperplane in a multidimensional space that maximizes the separation of classes. This hyperplane can then be used to classify new features. SVM also supports non-linear classification, where a hyper-surface is constructed that allows for better classification of statistical outliers.

#### 2.4.3 Simple Logistic

Simple Logistic classifier utilizes a binary logistic regression to describe an outcome in only two possible classes. It takes a set of features and applies regression analysis to create a classification parameters.

#### 2.4.4 Bagging

Given a training set, bagging algorithms derive  $m$  new sets by sampling from the original set. New datasets are fitted individually. The final result of bagging-type classifier is then the average of  $m$  fits.

#### 2.4.5 Ridor

Ripple Down Rules Learner utilizes a decision-tree like structure to compile a set of rules that either result in a classification of data, or passing of the parameters to another decision tree. The first rule is generated based on the dataset, then the rules are iteratively modified to account for all the exceptions of the original tree.

#### 2.4.6 Decision Stump

Decision stump is a decision tree which contains only one node - the root node. Root node's leafs are the classes when the root node's condition is met or not. The condition is based on union of numerical conditions which compare features in the original dataset.

#### 2.4.7 J48

J48 is Java's implementation of C4.5 algorithm. J48 builds a decision tree by finding features in the dataset that split the dataset evenly into separate classes. Branches of the decision tree are then used to further improve the classification results.

#### 2.4.8 LMT

Logistic Model Tree is a classifier that combines logistic regression analysis and a decision tree classification. Conditions of Logistic Model Tree are based on the logistic regression similar to Simple Logistics classifier. But instead of computing regression for the whole dataset, it is first split using C4.5 algorithm, which creates a tree structure for decisions.

#### 2.4.9 Random Forest

Random Forest classification is an ensemble learning type classifier that builds multiple decision trees. The classification of data is then passed to all the trees, and the output class is the mode class from all the decision trees.

### 3. Methodology

While developing experiments for the previous research effort [21],  $n$ -gram presence matrix was noted to require a significant amount of memory. To analyse a corpus of 15,000 files, a 240GB matrix had to be used. To overcome a large memory footprint requirement, Fourier Transform method was chosen as a frequency analysis tool.

Unlike  $n$ -gram analysis, Fourier transform does not require a built a set of all the  $n$ -grams, nor the creation of sparse matrices. which means a large decrease in RAM usage of the software. Fourier transform also simplifies the comparison of the features in files of different length - same frequencies are reported for different lengths of data.

To transform binary data from offset space to frequency space, a discrete Fourier Transform (equation 2) was used. A sequence size,  $N$ , was determined to be  $N = 2^p$  such that  $2^{p-1} < \max(\text{fileSize}) < 2^p$ . As having sequence size equal to the powers of 2 decreases the processing required. Every file was then appended with zeroes to reach a file size of  $N$ . Values of each byte in an executable were used as  $x_n$ , where  $n$  specified an offset of the byte from the beginning of file.

### 4. Experiment

This research effort targets the accuracy aspects of  $n$ -gram analysis, random projection, and Fourier transfer methods. The data set described in the next section was first processed by Fourier analysis or  $n$ -gram analysis on an xServe G-5 cluster (PPC970FX cpu, 2GB RAM per node), then the extracted features were reduced by random projection and the result was uploaded to a test machine running on Intel

Core i7-3770K, 16 GB RAM, 1TB HDD that ran machine learning algorithms to classify the data.

Methodology of the previous research effort [21] was used to generate a control result. The dataset of malicious and benign executables described below was combined into a single corpus. Using 4-gram analysis of the dataset, random projection was applied to create a 1500 feature embedding. This new low-dimensional dataset was analyzed using the SVM classifier in Waikato Environment for Knowledge Analysis (WEKA). This classifier was then used to determine whether a given application was malicious or benign. [17]. 10-fold cross validation was used by the trained classifier to determine average classifier accuracy.

After the control data generated, the algorithm was modified to use Fourier transform instead of 4-gram analysis. The rest of the algorithm remained untouched. The results of the Fourier transform were randomly projected to 1500 features and WEKA was used to train the classifiers. By treating the content of binary executables as a raw waveform, this research was able to transform the data from byte offset to byte pattern frequency, which allowed for an easy comparison of features in files of various sizes.

#### 4.1 Data set

The data set for this experiment consisted of 5124 windows executables in a PE format, with .exe extension. 4270 executables in the data set were malicious applications that break into 854 different Viruses, Backdoors, Trojan Horses, Worms, as well as 854 different instances of Zeus Trojan binary. These executables were obtained from various web-sites online, such as <http://www.trojanfrance.com>, <http://vx.netlux.org>, and <http://zeustracker.abuse.ch>. Previous research efforts used some of the same malicious applications [21].

854 of the executables in this data set were benign. They were obtained by installing an instance of Windows operating systems as well as Office environments in the virtual machine with a disconnected network adapter. The host computer was behind a NAT firewall, as well as the firewall of Louisiana Tech University. As executables were extracted from virtual machines, their MD5 sums were also recorded to make sure they did not get infected during the transfer process. All the executables and their MD5 sums were compressed in an archive, and copied to the research server. Once on the research server, the files were extracted into a dataset folder, given a read only access, and verified versus their MD5 checksums.

### 5. Results

The figures (Fig. 1 - 10) show classifier accuracies based on the amount of features generated by processing the data with Fourier transform, and random projection. Malicious files were separated into worms, trojans, viruses, and backdoors sections. Zeus trojan section was separately created

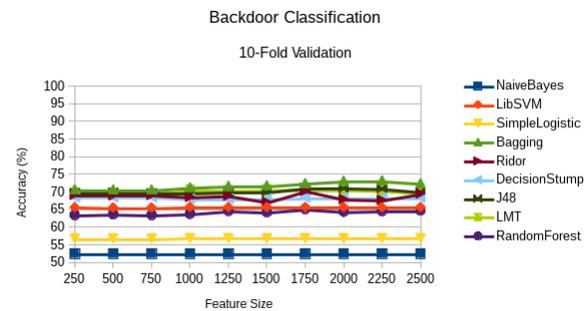


Fig. 1: Backdoors 10-fold classification

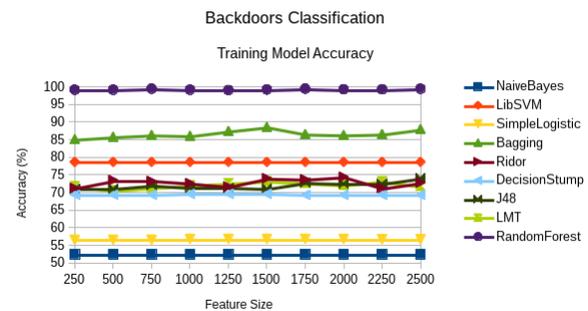


Fig. 2: Backdoors Training Model classification

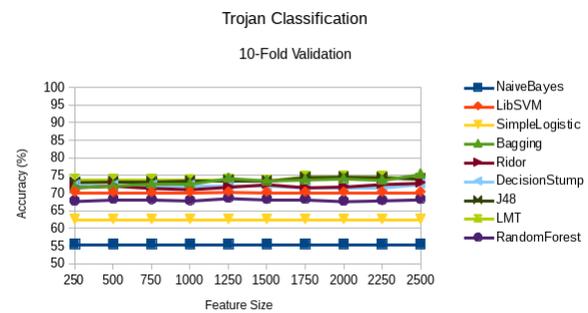


Fig. 3: Trojans 10-fold classification

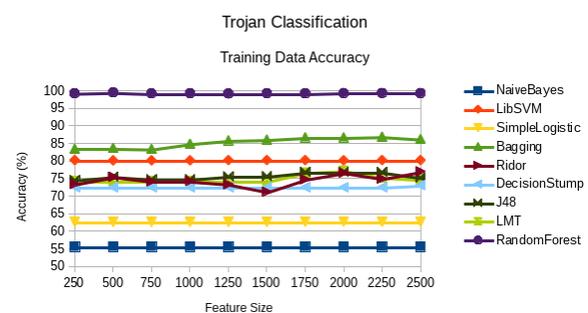


Fig. 4: Trojans Training Model classification

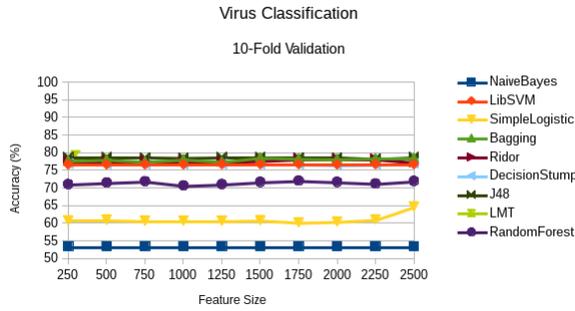


Fig. 5: Viruses 10-fold classification

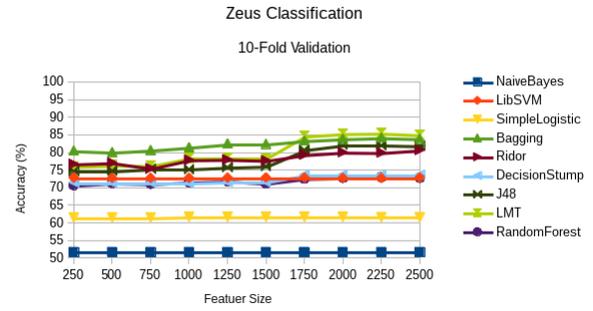


Fig. 9: Zeus Trojan 10-fold classification

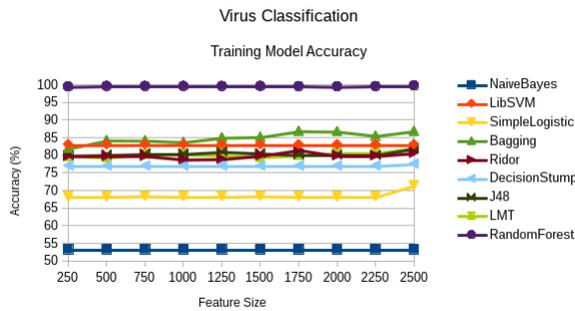


Fig. 6: Viruses Training Model classification

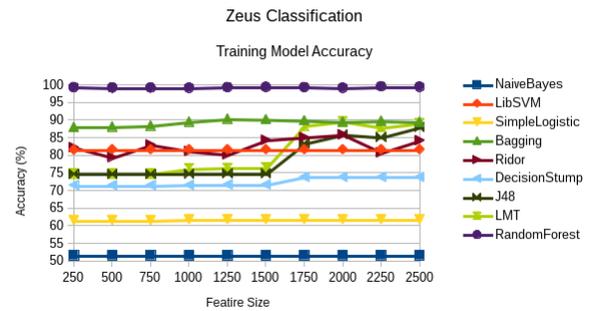


Fig. 10: Zeus Trojan Training Model classification

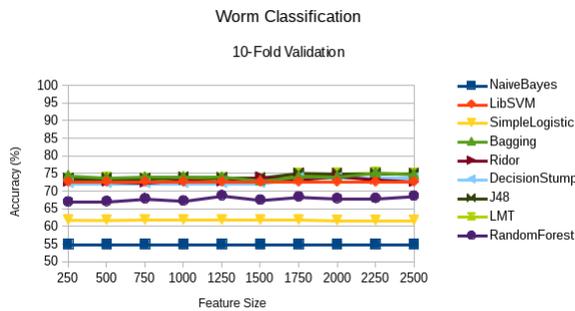


Fig. 7: Worms 10-fold classification

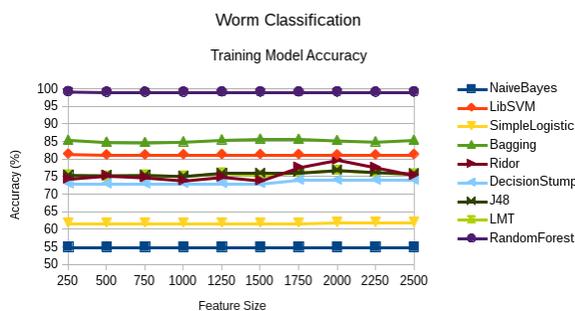


Fig. 8: Worms Training Model classification

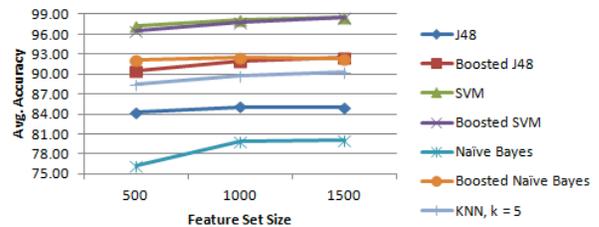


Fig. 11: Classification accuracy using  $n$ -gram analysis

from each individual instance of Zeus Trojan horse. The accuracy of each classifier was graphed in relation to the amount of features generated by the random projection feature reduction technique.

Random forest classifier was able to achieve more than 99% accuracy with every type of malicious application while evaluating the training model. During cross validation, bagging and LMT classifiers performed the best with a set of Zeus trojan horse malware, reaching 85% accuracy. Bagging also reached the highest accuracy while cross validating backdoor classification, reaching 74%. J48 reached the highest accuracy 80% while classifying viruses, and 75% while classifying trojans and worms.

## 6. Conclusions

Random projection has been proven to work well in reducing the amount of features in a dataset for malicious

application detection. In conjunction with Fourier transform, these algorithms allow for an accurate classification of malicious applications in various categories, without relying on specific signatures.

An added benefit of using Fourier transform instead of  $n$ -gram analysis is much lower memory footprint, and an ability to process new files without restructuring the feature set.  $N$ -gram analysis has to use large sparse matrices to generate features, which can take gigabytes to store. Processing files with Fourier transform and random projection for use with machine learning classifiers allows security researchers to detect zero day malicious applications before they have time to damage any critical infrastructure.

## 7. Acknowledgments

This material is based upon work supported by the U.S. Air Force, Air Force Research Laboratory under Award No. FA9550-10-1-0289.

## References

- [1] A. Hodges, "Alan turing and the turing machine," in *The Universal Turing Machine A Half-Century Survey*. Springer, 1995, pp. 3–14.
- [2] G. McGraw and G. Morrisett, "Attacking malicious code: A report to the infosec research council," *Software, IEEE*, vol. 17, no. 5, pp. 33–41, 2000.
- [3] C.-T. W. Association *et al.*, "Wireless quick facts," 2013.
- [4] T. Abou-Assaleh, N. Cercone, V. Keselj, and R. Sweidan, "Detection of new malicious code using  $n$ -grams signatures." in *PST*, 2004, pp. 193–196.
- [5] M. Christodorescu and S. Jha, "Static analysis of executables to detect malicious patterns," DTIC Document, Tech. Rep., 2006.
- [6] F. Cohen, "Computer viruses: theory and experiments," *Computers & security*, vol. 6, no. 1, pp. 22–35, 1987.
- [7] J. Z. Kolter and M. A. Maloof, "Learning to detect and classify malicious executables in the wild," *The Journal of Machine Learning Research*, vol. 7, pp. 2721–2744, 2006.
- [8] N. Weaver, V. Paxson, S. Staniford, and R. Cunningham, "A taxonomy of computer worms," in *Proceedings of the 2003 ACM workshop on Rapid malware*. ACM, 2003, pp. 11–18.
- [9] C. E. Landwehr, A. R. Bull, J. P. McDermott, and W. S. Choi, "A taxonomy of computer program security flaws," *ACM Computing Surveys (CSUR)*, vol. 26, no. 3, pp. 211–254, 1994.
- [10] B. Le Charlier, A. Mounji, M. Swimmer, and V. T. Center, "Dynamic detection and classification of computer viruses using general behaviour patterns," in *International Virus Bulletin Conference*, 1995, pp. 1–22.
- [11] P. Ferrie, "Attacks on more virtual machine emulators," *Symantec Technology Exchange*, 2007.
- [12] Symantec, "Symantec internet security threat report," 2013.
- [13] T. Atkison, "Aiding prediction algorithms in detecting high-dimensional malicious applications using a randomized projection technique," in *Proceedings of the 48th Annual Southeast Regional Conference*. ACM, 2010, p. 80.
- [14] J. Durand and T. Atkison, "Using randomized projection techniques to aid in detecting high-dimensional malicious applications," in *Proceedings of the 49th Annual Southeast Regional Conference*. ACM, 2011, pp. 166–172.
- [15] O. Henchiri and N. Japkowicz, "A feature selection and evaluation scheme for computer virus detection," in *Data Mining, 2006. ICDM'06. Sixth International Conference on*. IEEE, 2006, pp. 891–895.
- [16] I. Santos, Y. K. Peña, J. Devesa, and P. G. Bringas, "N-grams-based file signatures for malware detection." in *ICEIS (2)*, 2009, pp. 317–320.
- [17] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.
- [18] R. Baeza-Yates, B. Ribeiro-Neto *et al.*, *Modern information retrieval*. ACM press New York, 1999, vol. 463.
- [19] T. Abou-Assaleh, N. Cercone, V. Keselj, and R. Sweidan, "N-gram-based detection of new malicious code," in *Computer Software and Applications Conference, 2004. COMPSAC 2004. Proceedings of the 28th Annual International*, vol. 2. IEEE, 2004, pp. 41–42.
- [20] R. Bellman, *Adaptive control processes: a guided tour*. Princeton university press Princeton, 1961, vol. 4.
- [21] S. Ponomarev, J. Durand, N. Wallace, and T. Atkison, "Evaluation of random projection for malware classification," in *Software Security and Reliability-Companion (SERE-C), 2013 IEEE 7th International Conference on*. IEEE, 2013, pp. 68–73.
- [22] N. Goel, G. Bebis, and A. Nefian, "Face recognition experiments with random projection," in *Defense and Security*. International Society for Optics and Photonics, 2005, pp. 426–437.
- [23] T. Atkison, "Applying randomized projection to aid prediction algorithms in detecting high-dimensional rogue applications," in *Proceedings of the 47th Annual Southeast Regional Conference*. ACM, 2009, p. 23.
- [24] D. Achlioptas, "Database-friendly random projections," in *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. ACM, 2001, pp. 274–281.
- [25] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013.

# Firewall Policy Query Language for Behavior Analysis

Patrick G. Clark\*<sup>§</sup> and Arvin Agah\*

\**Department of Electrical Engineering and Computer Science  
University of Kansas, Lawrence, KS 6045 USA*

<sup>§</sup>*Cooresponding Author: Email: patrick.g.clark@gmail.com*

**Abstract**—Firewalls are one of the most important devices used in network security today. Their primary goal is to provide protections between parties that only wish to communicate over an explicit set of channels, expressed through protocols. These channels are implemented and described in a firewall using a set of rules, collectively referred to as a firewall policy. However, understanding the policy that a particular firewall is enforcing has become increasingly difficult. Many industry forces are converging that cause managing these devices to be much more complex than the premise of rules suggest.

Recently work has been done modeling a firewall policy in a concise and efficient data structure referred to as a Firewall Policy Diagram (FPD). The structure facilitates the canonical representation of a policy as well as human comprehension of the policy. This work builds on top of the data structure to provide a language for asking the data structure questions about the space that is represented in a policy, either the accepted, denied, or remaining traffic. Firewall Policy Query Language (FPQL) is a language loosely modeled after the Structured Query Language often seen related to database systems and relational algebra. It essentially provides a group of set mathematics based operators for deriving knowledge from a very large solution space. This work seeks to provide a simple, yet powerful, query language that is useful for human comprehension of a firewall policy as represented by an FPD.

## I. INTRODUCTION

Firewalls, network devices, and the access control lists that manage traffic provide the protection between networks that only wish to communicate over an explicit set of channels, expressed through the protocols, traveling over the network. The typical placement of a firewall is at the entry point into a network so that all traffic must pass through the firewall to enter the network. The traffic that passes through the firewall is typically based on existing packet-based protocols, and a packet can be thought of as a tuple with a set number of fields [1]. Examples of these fields are the source/destination IP address, port number, and the protocol field. A firewall will inspect each packet that travels through it and decide if it should allow that traffic to pass based on a sequence of rules. This sequence of rules is generally named an *access control list* and is made up of individual rules matched from top to bottom that follow the general form:

$$\langle predicate \rangle \rightarrow \langle decision \rangle$$

On the surface, firewall rule sets are relatively easy to understand. In the context of this research each rule consists of four fields and when the set contains a few rules, an individual firewall administrator can quickly comprehend

the access level and immediately know the appropriate location for granting additional access for a given request. However, because the IP address, protocol, and port solution space can cover a very large number of permutations, an individual's ability to fully understand access diminishes as a rule set grows.

Recent studies show that that the average firewall administration team has to accurately manage about 160,000 rules where 16,000 of those are changing on a monthly basis [2], [3]. Therefore, the ability to accurately and confidently understand the firewall policy and know what changes have occurred are more difficult than ever, and continue to increase in complexity.

### A. Key Contributions

This work presents a query language for comprehension of large network access. The language extends a previous work named Firewall Policy Diagram (FPD) [4], which is a set of data structures and algorithms capable of representing a firewall policy in a concise and canonical form. The primary contribution of this work is to provide an expressive query language that a team of firewall administrators would be able to use to answer important questions about what is contained in a large, and often impossible to understand, set of firewall policies. In addition to describing the tenets of the language, this paper presents the results of experiments run against large policies (up to 20,000 rules). Finally, we present how this language is capable of formal verification of single and multiple policies in a manner similar to FIREMAN [5].

## II. FIREWALL POLICY DIAGRAM

A Firewall Policy Diagram (FPD) is a set of data structures and algorithms used to model a firewall policy into an entity allowing efficient mathematical SET operations. The entity also has the ability to reconstitute the policy into a set of human comprehensible rules [4]. The FPD forms the base of the FPQL processing engine and allows the fast and efficient manipulation of the space. The internal storage mechanism of an FPD uses Reduced Ordered Binary Decision Diagrams (ROBDD or BDD) [6]. These data structures were introduced as an efficient way to capture hierarchical binary data and related works have described their use in firewall policy validation [7], [5], [8]. A full description of the algorithms involved with manipulation and extraction of human comprehensible rules can be found in related work [4].

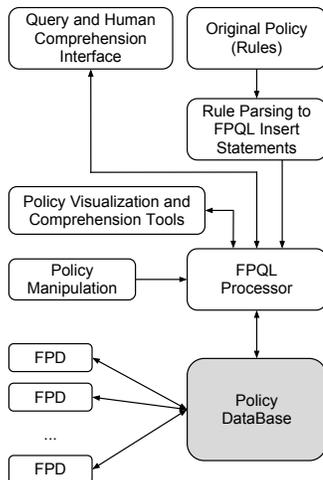


Figure 1: Architecture of the FPQL processor and policy database.

### III. FIREWALL POLICY QUERY LANGUAGE

The intent behind the design of FPQL is to provide a mechanism for firewall administrators and other security professionals to take a known set of complex firewall policies, load them into an FPD database, and query the policies and across policies to uncover valuable information about access.

#### A. Policy Database

Figure 1 illustrates the general architecture of our policy database, which is based on the FPD data structure. The database interaction is managed with an FPQL processor such that the language may be used to create and manipulate firewall policies stored within. The important notion here is that the generation and storage of multiple policies in one system enables comparisons and manipulations between the stored policies. Examples include storing versions of a policy over time to understand how it has changed, or storing policies from multiple vendors to better understand the differences.

#### B. FPQL Grammar

In the subsequent sections of this paper Extended Backus-Naur Form (EBNF) [9] is used to accurately describe the grammar of FPQL. The EBNF grammar of common elements used in the more specific *Insert*, *Query* and *Delete* grammar is described by Table I and Table II.

The EBNF meta-language uses abstractions for syntactic structures. The abstractions in EBNF descriptions, or grammar, are composed of *non-terminal* and *terminal* symbols. The non-terminal symbols are the EBNF descriptions, such as *var*, *id* and *field* such that they are composed of other non-terminal or terminal symbols. Terminal symbols are the individual characters, combination characters (strings), punctuation marks or digits; together called the *lexemes* or *tokens*. Terminals are considered the lowest level of derivation and may be matched with the actual input.

$\langle \textit{alpha} \rangle$	$\rightarrow$	$a-z \mid A-Z$
$\langle \textit{digit} \rangle$	$\rightarrow$	$0-9$
$\langle \textit{id} \rangle$	$\rightarrow$	$( \langle \textit{alpha} \rangle \mid \langle \textit{digit} \rangle \mid \_ )$ $\{ ( \langle \textit{alpha} \rangle \mid \langle \textit{digit} \rangle \mid \_ ) \}$
$\langle \textit{field} \rangle$	$\rightarrow$	$S \mid D \mid \textit{Prot} \mid \textit{Port} \mid \textit{Pol}$
$\langle \textit{var} \rangle$	$\rightarrow$	$\langle \textit{id} \rangle . \langle \textit{field} \rangle$
$\langle \textit{val} \rangle$	$\rightarrow$	$( \langle \textit{digit} \rangle \mid . \mid / \mid - )$ $\{ ( \langle \textit{digit} \rangle \mid . \mid / \mid - ) \}$
$\langle \textit{lparen} \rangle$	$\rightarrow$	$($
$\langle \textit{rparen} \rangle$	$\rightarrow$	$)$
$\langle \textit{comma} \rangle$	$\rightarrow$	$,$

Table I: FPQL Token definitions.

$S$	=	Source IP Space
$D$	=	Destination IP Space
$\textit{Prot}$	=	Protocol
$\textit{Port}$	=	Destination Port
$\textit{Pol}$	=	Entire policy Space
$\textit{alpha}$	=	Identifies $a$ through $z$ or $A$ through $Z$
$\textit{digit}$	=	Identifies $0$ through $9$
$\textit{var}$	=	An FPD name ( $\textit{id}$ ) and $\textit{field}$
$\textit{val}$	=	A field value that will an address or number format, based on the field type
$\textit{id}$	=	A user selected identifier composed of $\textit{alpha}$ , $\textit{digit}$ , or underscores
$\textit{field}$	=	The known fields of an FPD, $S$ , $D$ , $\textit{Prot}$ , $\textit{Port}$ or $\textit{Pol}$
$\textit{lparen}$	=	A left parenthesis “(”
$\textit{rparen}$	=	A right parenthesis “)”
$\textit{comma}$	=	A comma “,”

Table II: FPQL field definitions.

Therefore, a collection of these grammar rules comprises a full EBNF description. The syntax statements are read like a derivation, beginning with the start symbol of the particular grammar, and is processed through the BNF structure. The BNF will use the definitions from Table II.

Most of the common elements are straight forward in the descriptions; however, the *var* definition will benefit from some additional explanation. The *var* element is made of two sub elements, *id* and *field*, separated by a “dot”. Using these sub elements in-conjunction with a “dot” notation allows the language to dereference a policy identifier with the policy element for use in the operation. For example, in the query one might define a policy as  $p1\_accept$  and add known *rules* to the policy in the policy database. In subsequent queries of the policy database, the policy elements are referenced by  $p1\_accept.S$  representing the source IP address space for the accept  $p1$ . For example,  $p1\_accept.S$ .

Another element that requires some additional discussion is *val*. While the grammar allows *val* to be a *digit*, forward slash ( $/$ ), period ( $.$ ) or dash ( $-$ ); the correct *val* identifier format is dependent on the *field* being manipulated. For

$\langle insert \rangle \rightarrow insert\ into\ \langle id \rangle\ \langle insExpr \rangle$
$\langle insExpr \rangle \rightarrow \langle field \rangle == \langle insVal \rangle,$
$\quad \{ \langle comma \rangle \langle field \rangle == \langle insVal \rangle \}$
$\langle insVal \rangle \rightarrow \langle lparen \rangle \langle val \rangle$
$\quad \{ \langle comma \rangle \langle val \rangle \} \langle rparen \rangle$

Table III: FPQL *Insert* statement.

example, for *S* or *D*, then the language would expect an IP address or CIDR form for the correct interpretation. In this work, the validation is processed when the parse tree is traversed for interpretation. It is at this time the FPQL Processor, identified in Figure 1, returns a parse error back to the calling program if invalid identifiers are used in the FPQL statement.

### C. Policy Generation and Manipulation

Before a policy can be queried or visualized, it must first be created and loaded. Using FPQL this can be done with the *Insert* command and keywords. The intent behind the insertion syntax is to linearly process a firewall ruleset and insert each rule, one at a time. The EBNF in Table III describes the grammar of the *Insert* statement.

The following FPQL inserts a rule into the FPD identified by *p1\_accept* such that the rule source is host address *192.168.1.1*, the rule destination is the network address *10.1.1.0/24*, the rule protocol is *TCP* (6), and the rule destination port range is *80* to *90*.

$insert\ into\ p1\_accept\ S == (192.168.1.1),$
$\quad D == (10.1.1.0/24),\ Prot == (6),$
$\quad\quad Port == (80-90)$

This example also demonstrates the concept of how FPQL and the underlying FPD policy database interact to capture the action of the original security rule. In this work the action for a particular rule is constrained to either *accept* or *deny*, therefore to capture the different spaces for a policy, it is strictly a policy naming standard to append the action to the policy name. Not only does this allow for the easy identification of the *space* being manipulated, it also allows other actions to be represented in extended works, such as *encrypt* or *log*, without modifying the grammar.

### D. Queries

Retrieving policy information from a populated policy database is accomplished using the policy query grammar. In addition to identifying what sort of data are wanted from the policy database, the grammar provides operators that can be applied across defined policies. The FPD database can be queried and compared because each operation or grouping that is defined in a FPQL statement results in an FPD. This means that as *sub-expressions* and other nested operations are executed from the language parse tree, an FPD is constructed and manipulated until the tree is traversed from leaf up and reaches the root. Not only does this simplify

$\langle query \rangle \rightarrow (\ count\ \langle field \rangle \mid \langle field \rangle )$
$\quad\quad\quad where\ \langle expr \rangle$
$\langle expr \rangle \rightarrow \langle comparison \rangle$
$\quad\quad\quad \{ (\ \&\ \mid\ or\ )\ \langle comparison \rangle \}$
$\langle comparison \rangle \rightarrow \langle var \rangle$
$\mid \langle lparen \rangle \langle expr \rangle \langle rparen \rangle \mid \langle varStat \rangle$
$\langle varStat \rangle \rightarrow \langle var \rangle \langle op \rangle \langle lparen \rangle$
$\quad \langle val \rangle \{ \langle comma \rangle \langle val \rangle \} \langle rparen \rangle$
$\langle op \rangle \rightarrow :: \mid !:: \mid \sim$
$\quad \mid !\sim \mid == \mid !== \mid \&\ \mid or \mid -$

Table IV: FPQL *Query* statement.

$:: \rightarrow In\ (subset)$
$!:: \rightarrow Not\ In$
$\sim \rightarrow Contains\ (superset)$
$!\sim \rightarrow Not\ Contains$
$== \rightarrow Equals$
$!== \rightarrow Not\ Equals$
$\&\ \rightarrow And$
$or \rightarrow Or$
$- \rightarrow Difference$

Table V: FPQL operators.

reasoning about how the results are constructed, it also allows a common expected result from each operation.

The grammar of a *Query* statement uses the same *id* and *field* definitions as the *Insert* grammar, allowing consistency in use of the language. The EBNF grammar in Table IV describes structure of the *Query* statement.

One small difference from the actual implementation is the use of the *or* operator. In order to make the EBNF easier to read in this work, the word *or* is used, however the actual implementation uses a pipe character. The  $\langle op \rangle$  operators are defined in Table V.

As an example of how the query expression grammar may be used, is the case of an administrator who wants to know the sources allowed to access an important server at *10.2.1.100* over *tcp/80* from a policy loaded into the Policy Database named *policy1*.

$S\ where\ policy1.D == (10.2.1.100)\ \&$
$\quad policy1.Prot == (6)\ \&\ policy1.Port == (80)$

A more complicated way to use a policy database and FPQL interpreter is to track a policy as it changes over time. We suppose that a policy *policy\_september* is loaded into the database and the administrator would like to know how the policy changes over the next month. One way for an administrator to find this information is to go back through the potentially thousands of changes that may have occurred in the month of October (up to 16,000 [3]). However, using FPQL, the same policy from the next month (*policy\_october*) can be loaded into the policy database and the following FPQL will list the differences between the

---

$\langle \text{deleteFromPol} \rangle \rightarrow \text{delete from policy}$   
 $\langle \text{id} \rangle \text{ where } \langle \text{expr} \rangle$

---

Table VI: FPQL *Delete* statement.

two policies, thus providing the administrator with a list of human comprehensible rules representing access changes that are different from the previous month.

---

$\text{Pol where policy\_october.Pol} - \text{policy\_september.Pol}$

---

### E. Delete

The final manipulation language grammar is for deleting portions of *space* from a policy that has been loaded into the policy database. While not useful for direct comprehension of a policy as it relates to the query language, it is necessary for formal verification of a particular policy for anomalies such as those checked by FIREMAN [5]. Table VI describes the EBNF grammar of the *Delete* statement.

If one wants to remove a particular rule from an existing policy, the *Delete* grammar allows selectively removing a portion of the *space* from a policy.

---

$\text{delete from policy } p1 \text{ where } p1.S == (192.168.1.1)$   
 $\& p1.D == (10.1.1.0/24) \& p1.Prot == (6)$   
 $\& p1.Port == (80-90)$

---

### F. Miscellaneous Policy Operations

In addition to the more formal *Insert*, *Query*, and *Delete* language syntax, FPQL also provides the ability for the administrator to check what policies have been loaded into the policy database using the *list policies* command and delete those policies using the *delete policy* command. This work uses these commands as examples of many needed for general maintenance of the policy database in operational situations by firewall administrators. Other commands such as these may be useful for the user to accomplish day to day tasks.

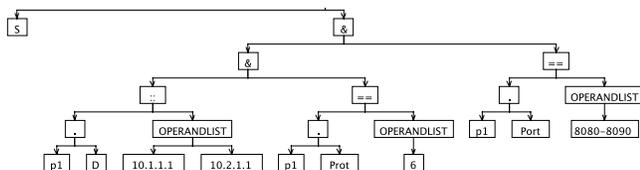


Figure 2: An example AST parsed from a FPQL grammar.

### G. Syntax Tree Parsing

When a FPQL query is parsed, an Abstract Syntax Tree (AST) is produced such that the leaves are processed and transformed as the language is traversed to its root. As each node is visited, an FPD is produced that will then be used as an operand in the next operator. An example is a simple query for discovering the sources which are allowed

access to the destinations 10.1.1.1 or 10.2.1.1, over service tcp/8080-8090:

---

$S \text{ where } p1.D :: (10.1.1.1, 10.2.1.1) \& p1.Prot == (6)$   
 $\& p1.Port == (8080-8090)$

---

Figure 2 shows the abstract syntax tree that is produced as a result of processing the EBNF grammar of the example query. As the tree is traversed from leaf to root, nodes representing operators are acting on operands that are interpreted as portions of an existing policy (as in the case of *p1.D*) or lists (as in the case of '10.1.1.1,10.2.1.1'). These constructs are combined as the tree is pruned into FPD data structures where finally the results are extracted from the final FPD at the root, based on the requested field (*S,D,Prot,Port,Pol*) at the left child and the FPD produced from the right sub-tree.

## IV. FPQL AND POLICY DATABASE PERFORMANCE

The experiments designed for this work seek to evaluate the performance of FPQL when dealing with policies of sizes ranging from 1,000 to 20,000 rules and focuses on processing and executing FPQL statements against the policy database. The rulesets are randomly generated and are generally distributed over the solution space.

Figure 3 charts the performance of FPQL inserting and querying operations on the policies. The insertion operations averaged approximately 500 microseconds with no growth as the number of rules increased. The query operation performance averaged 120 microseconds, again with no growth as the number of rules increased. The constant processing time reflects the constant time performance of the underlying ROBDD data structures in the FPD. However, while the operations and testing appear to be constant, that is a reflection of the queries being run against the system having a constant seven operators. Therefore, the actual complexity is related to the number of operators present in the query and will grow linearly with those.

This work focuses on an extension of a policy comprehension model FPD. Because comprehension is the end goal, it is very important to provide very fast access to ad-hoc queries described by FPQL. The performance numbers presented here achieve this goal by allowing the firewall administration team to very quickly and unerringly gain an understanding of what is allowed through the security policies implemented in their organizations.

## V. FPQL POLICY ANOMALY DETECTION

FIREMAN is a related work that analyzed intra-firewall and inter-firewall anomalies with a framework and algorithm for identifying those inconsistencies [5]. A good firewall configuration is consistent with the administrator's intention and the assertion that FIREMAN makes is that inconsistencies in the ruleset are an indication of mistakes or misconfigurations. FIREMAN defines four primary inconsistencies based off of related work [10], [11] and we show an FPQL algorithm that is able to identify those anomalies.

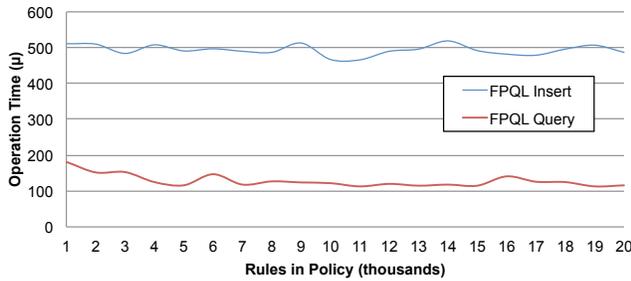


Figure 3: FPQL performance with 1,000 to 20,000 rule policies.

- 1) *Shadowing*: An inconsistency in a firewall policy such that the entire *space* a rule represents and the associated action is contained in one (or a combination of) previous rules with the opposite action.
- 2) *Generalization*: A problem that identifies a case where a portion or subset of the *space* represented by a rule had been previously matched by one or multiple rules with the opposite action.
- 3) *Correlation*: Represents a problem where the *space* of the current rule intersects with one or multiple previous rules with the opposite action.
- 4) *Redundant*: A situation identifying that a previous rule or rules already handled the current rule *space* with the same action. Therefore, this rule would never be matched and is considered redundant.

FIREMAN identifies *shadowing* as an error, but considers *generalization* and *correlation* as potentially not being an error because administrators may be using these more broadly defined networks as a way to keep the policy sizes smaller. *Redundant* rules are considered an error in this situation because it reflects a rule that will never be used. Leaving *redundant* rules in policy makes the policy unnecessarily larger, increasing processing time of the firewall and decreasing the comprehension of the policy.

#### A. FPQL Intra-firewall Anomaly Detection

An algorithm may be used in conjunction with FPQL to identify if a rule is *shadowing*, *generalization*, or *correlation* in an individual firewall policy. The initialization of the policy database begins with three policies represented as FPDs: *accept*, *deny*, and *remain*. The *accept* and *deny* policies are initialized to  $\emptyset$ , with the *remain* policy starting off as  $U$  (meaning, every possible rule field combination). In addition, each rule  $R$  of the test policy  $P$  contains the known FPQL fields  $S$ ,  $D$ ,  $Prot$ , and  $Port$  with two new fields:  $A$  meaning an *Action*, either *accept* or *deny*; and  $R$  meaning a Rule as an FPD  $S+D+Prot+Port$ .

Rule  $R$  fields are dereferenced by a dot (i.e.  $R.S$ ). Each rule  $R$  in policy  $P$  is processed linearly from top to bottom as to model how an actual firewall processes a packet for a matching rule. Notably, not all parts of the algorithm are FPQL. Some of the operations are based on FPD SET capabilities and are predicated on FPQL processing

**Input:** Policy  $P$  to test for inconsistencies

**Output:** Inconsistent rules and the type of anomaly

```

1: procedure TESTPOLICY( $P$ )
2:   FPQL: insert into remain  $S == (0.0.0.0/0)$ ,
3:      $D == (0.0.0.0/0)$ ,  $Prot == (0 - 255)$ ,
4:      $Port == (0 - 65535)$ 
5:   for all Rule  $R \in P$  do
6:     TESTRULE( $R$ )
7:   end for
8: end procedure

```

Figure 4: FPQL Intra-firewall anomaly detection.

returning an FPD. Figures 4 and 5 detail the algorithms for computing intra-firewall anomaly detection.

#### B. FPQL Inter-firewall Anomaly Detection

FIREMAN identifies the reality of many firewalls existing in a network and the potential for data to flow through those multiple policies [5]. A computer network is a graph of nodes (firewalls, routers, switches, hosts, etc.) and edges (traffic flow connections). A well designed network includes multiple paths from one node to another. By modeling the network as a graph, and using DFS or BFS algorithms, a network can be converted into a spanning tree.

Figure 6 shows a spanning tree of a network providing all known directed paths from a *starting* node to a resulting *ending* node. For large and dense graphs, a large number of paths might result in the subsequent spanning tree. These paths may go through routers and other non-filtering devices used primarily for traffic management. However, because we are only concerned with the filtering devices in the network, i.e., firewalls, the size of the resulting spanning tree can be greatly reduced by only including firewall nodes [5]. Figure 6 is a reformulation of the solution presented in [5] to identify the *starting* node once and build the spanning tree out from that location to multiple *ending* nodes. This simplifies the algorithm [5] presented such that the processing at each node is just the inbound solution space  $I$  having been potentially manipulated by its parents. The FPQL reformulation is presented in Figure 7 and traverses a depth first search from the root node *starting* to the leaf nodes *ending*, identifying anomalies as the graph is traversed.

An additional classification used when processing Inter-firewalls for anomalies is the identification of a *raised security level*. This classification primarily identifies those packets that were accepted by a previous firewall, but denied downstream. A situation where this may not be considered an anomaly, but because traffic was allowed through one firewall and denied by the next, should be reviewed by firewall administrators to ensure that this does not indicate unintended access in a previous firewall.

#### C. FPQL Policy Anomaly Detection Performance

The purpose of the intra-firewall and inter-firewall formal verification model presented in this section is to demonstrate

```

1: procedure TESTRULE(R)
2:   accept ← accept.Pol from the policy database
3:   deny ← deny.Pol from the policy database
4:   remain ← remain.Pol from the policy database
5:   if remain = ∅ then
6:     if  $R.R \subseteq \textit{accept}$  then
7:       if  $R.A = \textit{“accept”}$  then R is Redundant
8:       else R is Shadowed
9:       end if
10:    else if  $R.R \subseteq \textit{deny}$  then
11:      if  $R.A = \textit{“deny”}$  then R is Redundant
12:      else R is Shadowed
13:      end if
14:    else R is Correlated
15:    end if
16:  else if  $R.R \cap \textit{remain} = \emptyset$  then
17:    if  $R.R \subseteq \textit{accept}$  then
18:      if  $R.A = \textit{“accept”}$  then R is Redundant
19:      else R is Shadowed
20:      end if
21:    else if  $R.R \subseteq \textit{deny}$  then
22:      if  $R.A = \textit{“deny”}$  then R is Redundant
23:      else R is Shadowed
24:      end if
25:    else R is Correlated
26:    end if
27:  else if  $R.R \cap \textit{remain} \neq \emptyset$  and  $R.R \not\subseteq \textit{remain}$ 
then
28:    if  $R.A = \textit{“accept”}$  then
29:      if  $R.R \cap \textit{deny} \neq \emptyset$  and  $R.R \not\subseteq \textit{deny}$  then
30:        R is Correlated
31:      end if
32:    else if  $R.A = \textit{“deny”}$  then
33:      if  $R.R \cap \textit{accept} \neq \emptyset$  and  $R.R \not\subseteq \textit{accept}$ 
then
34:        R is Correlated
35:      end if
36:    end if
37:  end if
38:  if  $R.A = \textit{“accept”}$  then
39:    FPQL: insert into accept  $S == (R.S)$ ,
40:     $D == (R.D)$ ,  $Prot == (R.Prot)$ ,
41:     $Port == (R.Port)$ 
42:  else
43:    FPQL: insert into deny  $S == (R.S)$ ,
44:     $D == (R.D)$ ,  $Prot == (R.Prot)$ ,
45:     $Port == (R.Port)$ 
46:  end if
47:  FPQL: delete from policy remain where
48:     $S == (R.S) \ \& \ D == (R.D) \ \&$ 
49:     $Prot == (R.Prot) \ \& \ Port == (R.Port)$ 
50: end procedure

```

Figure 5: TestRule function.

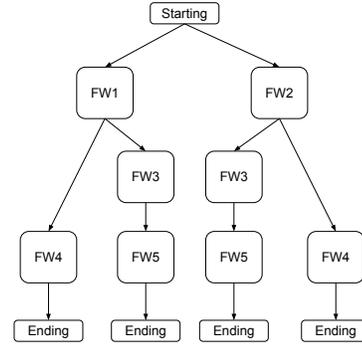


Figure 6: Network spanning tree.

**Input:** Root Node *P* of Spanning Tree

**Output:** Inconsistent rules and the type of anomaly

```

1: procedure TESTINTERFIREWALL(P)
2:   FPQL: insert into I  $S == (0.0.0.0/0)$ ,
3:    $D == (0.0.0.0/0)$ ,  $Prot == (0 - 255)$ ,
4:    $Port == (0 - 65535)$ 
5:   for all Node C ∈ Children(P) do
6:     PROCESSNODE(C, I)
7:   end for
8: end procedure
9: procedure PROCESSNODE(P, I)
10:  if P ∈ ending then return
11:  end if
12:   $I' \leftarrow I \cap U$ 
13:  for all Rule R ∈ P.Policy do
14:    if  $R.R \subseteq I$  then
15:      if  $R.A = \textit{“deny”}$  then R raised security
16:      level
17:    end if
18:    else if  $R.R \subseteq \neg I'$  then
19:      if  $R.A = \textit{“accept”}$  then R is Shadowed
20:      else if  $R.A = \textit{“deny”}$  then R is Redundant
21:      end if
22:    end if
23:    FPQL: delete from policy I' where
24:       $S == (R.S) \ \& \ D == (R.D) \ \&$ 
25:       $Prot == (R.Prot) \ \& \ Port == (R.Port)$ 
26:  end for
27:  for all Node C ∈ Children(P) do
28:    PROCESSNODE(C, I')
29:  end for
30: end procedure

```

Figure 7: FPQL Inter-firewall anomaly detection.

the formal verification capabilities of the FPQL language. This section has shown that FPQL and the Policy Database architecture is capable of formally modeling policies in both an individual and networked environment. In general, the performance of the presented algorithms have a computational complexity similar to the original FIREMAN system. The complexity is bound by the number of rules being verified and will grow in relation to that number.

Therefore the algorithms presented were a reformulation of those presented in other works with the exception of one algorithm. The computation of the network spanning tree for inter-firewall verification included an improvement based on only having to process a single input FPD space *I*.

## VI. RELATED WORK

Over the past decade there have been many research efforts that involved analyzing and understanding firewall policies, both from a single and multiple policy level. Fewer involve allowing ad-hoc querying of a policy and allow comprehension of large policies.

One of the earliest works, and the most closely related to ours, built a query engine on top of a formal verification system called Voss [7]. In a later work, Liu *et al.* (2005) propose a query language called Structured Firewall Query Language that is executed on a data structure called a Firewall Policy Tree [12] and later on a Firewall Decision Diagram [13].

Other associated efforts that do not include a query language have been done by modeling firewall policies; however, most of the models reflect their intended use and not all are capable of the sort of operations described in this work. Much of the research has focused on rule processing and validation of those rules where the goal is to identify hidden, shadowed, and inconsistent rules [14], [15], [10], [11], [16], [5], [17]. In general the focus is on algorithms for finding policy anomalies both from a single policy model to a multi-policy model. A portion of the related research introduced the use of BDDs for the models and became the foundation for some of the algorithms in our work [6], [7], [5].

## VII. CONCLUSIONS

In this paper we presented FPQL (Firewall Policy Query Language), an efficient and powerful language built on top of a policy database and the Firewall Policy Diagram data structure. There are three primary contributions in this work:

- Provided an expressive language for a firewall administrator to understand access through a policy or set of policies.
- Demonstrated that FPQL can run in microseconds of time, even against very large rulesets. This speed encourages comprehension of policies through ad-hoc queries, further supporting the overall goal of understanding network access.
- Demonstrated how FPQL could be used in formal verification of both intra and inter firewall policies.

It is important to note that while this work focuses on four tuples of a firewall rule, there are potentially more tuples to include. In addition, next generation firewalls have begun to expand their application layer filtering to include fields at layers higher in the protocol stack. Both FPQL and the underlying policy database with FPDs are capable of being extended to include fields in future work. The processing

time complexity becomes larger; however, it is still bound by the depth of the ROBDD in the case of queries.

## REFERENCES

- [1] J. F. Kurose and K. W. Ross, *Computer Networking: A Top-Down Approach*, 4th ed. Addison Wesley, 2007.
- [2] A. Wool, "A quantitative study of firewall configuration errors," *Computer*, vol. 37, no. 6, pp. 62–67, 2004.
- [3] M. J. Chapple, J. D'Arcy, and A. Striegel, "An analysis of firewall rulebase (mis)management practices," *ISSA Journal*, pp. 12–18, February 2009.
- [4] P. G. Clark, "Firewall policy diagram: Novel data structures and algorithms for modeling, analysis, and comprehension of network firewalls," Ph.D. dissertation, University of Kansas, 2013.
- [5] L. Yuan, J. Mai, Z. Su, H. Chen, C. Chuah, and P. Mohapatra, "Fireman: A toolkit for firewall modeling and analysis," *IEEE Symposium on Security and Privacy*, pp. 199–213, 2006.
- [6] R. E. Bryant, "Symbolic boolean manipulation with ordered binary-decision diagrams," *ACM Computing Surveys*, vol. 24, pp. 293–318, September 1992. [Online]. Available: <http://doi.acm.org/10.1145/136035.136043>
- [7] S. Hazelhurst, A. Attar, and R. Sinnappan, "Algorithms for improving the dependability of firewall and filter rule lists," in *Proceedings of the 2000 International Conference on Dependable Systems and Networks*, 2000, pp. 576–585.
- [8] K. Ingols, M. Chu, R. Lippmann, S. Webster, and S. Boyer, "Modeling modern network attacks and countermeasures using attack graphs," in *Proceedings of the 2009 Computer Security Applications Conference*, December 2009, pp. 117–126.
- [9] R. W. Sebesta, *Concepts of Programming Languages*, 9th ed. Pearson, 2009.
- [10] E. S. Al-Shaer and H. H. Hamed, "Modeling and management of firewall policies," *IEEE Transactions on Network and Service Management*, vol. 1, no. 1, pp. 2–10, April 2004.
- [11] —, "Discovery of policy anomalies in distributed firewalls," in *Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 4, March 2004, pp. 2605–2616.
- [12] A. X. Liu, M. G. Gouda, H. H. Ma, and A. H. Ngu, "Firewall queries," in *Principles of Distributed Systems*, ser. Lecture Notes in Computer Science, T. Higashino, Ed. Springer Berlin Heidelberg, 2005, vol. 3544, pp. 197–212. [Online]. Available: [http://dx.doi.org/10.1007/11516798\\_15](http://dx.doi.org/10.1007/11516798_15)
- [13] A. X. Liu and M. G. Gouda, "Firewall policy queries," *IEEE Transactions on Parallel and Distributed Systems*, vol. 20, no. 6, pp. 766–777, June 2009.
- [14] Y. Bartal, A. Mayer, K. Nissim, and A. Wool, "Firmato: a novel firewall management toolkit," in *Proceedings of the IEEE Symposium on Security and Privacy*, 1999, pp. 17–31.
- [15] E. S. Al-Shaer and H. H. Hamed, "Design and implementation of firewall policy advisor tools," School of Computer Science, Telecommunications and Information Systems, DePaul University, Chicago, USA, Technical Report, August 2002.
- [16] M. G. Gouda and A. X. Liu, "Firewall design: Consistency, completeness, and compactness," in *Proceedings of the 24th International Conference on Distributed Computing Systems*, 2004, pp. 320–327.
- [17] C. Chao, "A flexible and feasible anomaly diagnosis system for internet firewall rules," in *Proceedings of the 13th Asia-Pacific Network Operations and Management Symposium*, September 2011, pp. 1–8.



**SESSION**

**SPECIAL TRACK ON IOT AND SCADA  
CYBERSECURITY EDUCATION**

**Chair(s)**

**Prof. George Markowsky  
University of Maine - USA**



# MODBUS Covert channel

Carlos Leonardo (cal3678@rit.edu)<sup>1</sup> and Daryl Johnson (dgjics@rit.edu)<sup>2</sup>

<sup>1</sup> and <sup>2</sup> Department of Computing Security, Rochester Institute of Technology, Rochester, NY, United States

**Abstract**—*The security community already has seen some examples of actual attacks against real SCADA installations, like the Stuxnet case in 2010 [3]. MODBUS is one of the most used communication protocols in industrial control systems. Even though the protocol itself is known to lack basic security features; there is not much detail available about real world cases where MODBUS has been used as an attack vector. Covert channels have been mentioned several times as part of the security vulnerabilities of SCADA systems [1], [2] and [4]. This research targets the MODBUS protocol characteristics to introduce a covert channel. This covert channel allows information leakage from one device called covert Master to another one called covert Slave. The covert Master is supposed to be on the internal LAN, while the covert Slave is expected to be on a separate subnet. The Slaves subnet could be based on either Ethernet and TCP/IP or a serial BUS (RS485) using a media converter to reach the LAN.*

**Keywords:** SCADA, MODBUS, Covert Channels

## 1. Introduction

To appreciate the importance of the MODBUS communication protocol and the impact that a MODBUS covert channel could have, it is necessary to briefly describe SCADA Systems. SCADA stands for Supervisory Control and Data Acquisition. These systems usually control critical infrastructure for economy stability, such as power generation and distribution. MODBUS was created in 1979 by Modicon (now Schneider) and has been used since then as one of the main communication protocols for SCADA installations [1]. MODBUS is so important in first place because it is used to control and monitor critical infrastructure that is everywhere.

Being a layer 7 protocol, MODBUS was initially used on RS485 serial networks. However, it is now common to see MODBUS working over Ethernet networks with TCP/IP or in a combination of both. These recent configurations that integrate MODBUS on the LAN networks and internet have also opened the possibility of attacking the SCADA systems in similar ways that information systems are. This covert channel can use the MODBUS infrastructure in place to leak information from the LAN network, which is a valuable ability to have when approaching a target with valuable data. SCADA devices often are physically located outside of the protected building and gated environments. For example, an electrical power meter connected to the MODBUS network would be located on a remote sub-station

or even a residential area. This makes it a candidate for data exfiltration and infiltration.

Although some SCADA firewall solutions exist [5], the common approach is to block non valid or malformed MODBUS traffic as well as *write* transactions. This covert channel uses valid and *read-only* transactions to operate which makes it harder to detect or stop. Also, there are very low chances of having a SCADA system using such type of security implementation since those are still considered a new field compared with the time SCADA systems have been in use. MODBUS has been chosen for this covert channel research because of its importance and common use in SCADA systems as well as the lack of security in its design. SCADA security is still considered, as are covert channels, a novel subject that usually is not taken into account when implementing information security controls.

## 2. Related work

While the MODBUS protocol has been used for decades and SCADA systems are everywhere; it is not common to see MODBUS being used for purposes other than the monitoring and control of SCADA equipment. Some published papers and books have raised the flag indicating that SCADA systems need more attention from the information security community because of their importance [2] and [3].

A good example is the Stuxnet case, mentioned by Kim-Kwang [3] as a recent and high profile attack committed against SCADA systems. Other publications [4] also talk about the SCADA Security issue and how it has changed over the years. Originally, the security community was more worried about physical threats affecting the SCADA systems like sabotage; however, now the concern is also about a new wave of electronic and information based threats affecting them.

Knapp [1] even mentions that the MODBUS protocol lacks some basic security features such as authentication and encryption and says that SCADA Intrusion prevention systems (IPS) could monitor malicious activities using MODBUS signatures [1]. He also proposes that MODBUS sessions could be validated to ensure that MODBUS has not been "hijacked" and used for covert communication [1]. However, there is no evidence of an actual implementation of a MODBUS covert channel.

The SANS Institute (SysAdmin, Audit, Networking, and Security Institute) issued a document called "Using SNORT for intrusion detection in MODBUS TCP/IP communications" describing a method of intrusion detection (IDS)

by developing rules for SNORT [8]. The objective of that document is to provide useful information about how to implement IDS systems for MODBUS by using open source technologies, instead of expensive and limited ones that are commercially available.

avtheir email addresses (unless they really want to.)

### 3. MODBUS Protocol and Data structure

The MODBUS protocol is well defined in the official documents [6] and [7] issued by the MODBUS Organization. This paper only exposes the protocol information that is closely related to the covert channel intended to be implemented.

MODBUS was originally created to work on serial networks like RS485 (figure 1) and that is why some terminology is different from what it is used in modern Ethernet and TCP/IP based networks (figure 2). In a MODBUS network there are always two types of devices: several field Slaves (called Servers in MODBUS TCP/IP) and a single Master (or several Clients as called in MODBUS TCP/IP). Only one Master can exist in a MODBUS RS485 serial network. Slaves generate errors if they receive requests from more than one master. This is obviously different in TCP/IP based networks where several clients can coexist as long as they are synchronized to avoid sending simultaneous requests. Client(s) and servers are usually in different subnets, the client being a computer using TCP/IP with SCADA software and the servers being field devices wired within a serial BUS (RS485) network and using a media converter to reach the LAN (figure 3). No matter what lower layer topology is in use, the protocol data structure changes minimally and it is relatively simple when compared with HTTP or other layer 7 protocols. We use Client for the Master device and Server for each Slave. Three different MODBUS networks layouts are shown below.

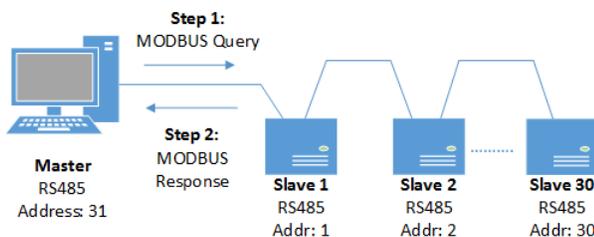


Fig. 1: Legacy MODBUS RTU working on a Serial BUS network (RS485).

One Master → Many Slaves.

The MODBUS servers are the field devices capable of measuring and/or controlling the environment by using inputs and outputs. The inputs can be digital (Discrete Inputs) and analog (Input Registers). Servers measure the

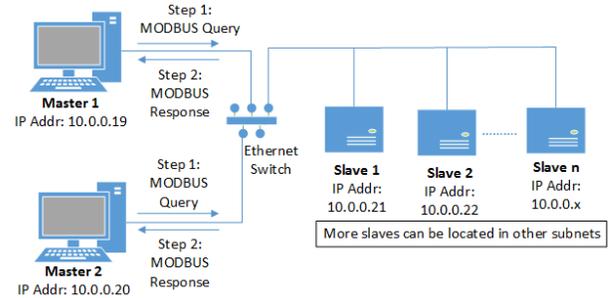


Fig. 2: New MODBUS TCP working on a TCP/IP network (used by this covert channel).

One / many TCP Clients → Many TCP Servers

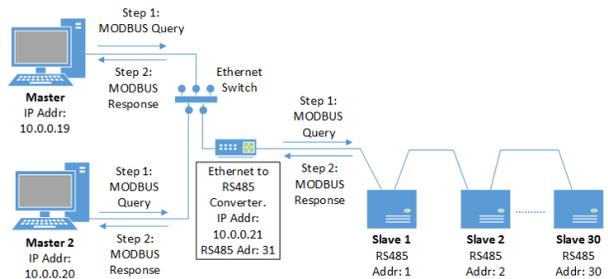


Fig. 3: Mixed MODBUS network (using media converters from RS485 to TCP/IP).

One / many TCP Clients → Converter → Many serial Slaves

environment variables by using these inputs and store their updated values in local memory tables (Table 1). The client can later request a server for the actual value of its inputs. The outputs can be also digital (Coils) and analog (Holding Registers). Coils are used to open or close a digital switch while holding registers are used to vary an analog output value within a range. The client can request a server to change the value of these outputs in the local memory tables (Table 1), thus controlling the environment. Not all servers have all types of inputs, outputs and functions since this is a vendor-specific decision.

Servers organize the data in four primary tables allocated for Discrete Inputs, Coils (digital outputs), Input Registers (analog inputs) and Holding Registers (analog outputs) [6]. Table 1 shows some details about each one of these primary tables [6]. The old version of the MODBUS protocol allows 9,999 data objects in each one of the four primary tables while the new versions of MODBUS allow 65,536 data objects. The proposed covert channel uses object numbers between 0 and 9,999 so it can be implemented in MODBUS networks using both schemas.

The MODBUS client device on the other hand only requests data from the servers and stores it in long term memory for future processing. Servers always wait for a client's request and never initiate a conversation because the

Primary Tables	Object Size	Object access	Comments
Discrete Inputs	Single bit	Read-Only	This data can be provided by an I/O system.
Coils	Single bit	Read-Write	This data can be alterable by an application program.
Input Registers	16-bit word	Read-Only	This data can be provided by an I/O system.
Holding Registers	16-bit word	Read-Write	This data can be alterable by an application program.

Table 1: Primary Tables in a MODBUS Server device.

protocol is based on requests and responses.

The MODBUS Application Data Unit (ADU) contains a simple Protocol Data Unit (PDU) and some additional fields introduced by the network topology in use [6] as illustrated in the figure 4:

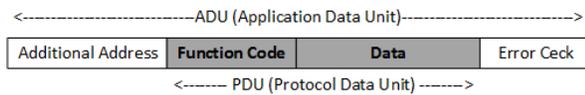


Fig. 4: MODBUS Application Data Unit (ADU)

The messages MODBUS client and servers use to communicate to each other contain a function code of 1 byte long and a data field of variable size. The function code specified in the client request specifies the action (read or write) and the type of object (digital or analog). The data field contains what object(s) will be affected by the function [6]. A complete list of Function codes and their meaning is available on the protocol standard document [6].

#### 4. A Covert channel taking advantage of function codes

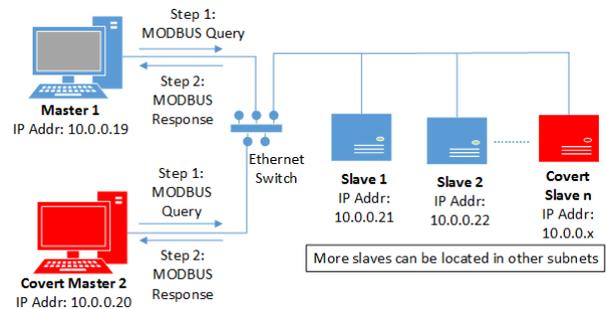
The MODBUS protocol uses two principal elements to establish communication between client and servers that are also used in this covert channel: Tables and Function Codes. The four primary tables (discrete inputs, coils, input registers and holding registers) are the four different sets of variables that can exist in a server device (see table 1). For each one of these types of variables, also called objects, there are up to 9,999 in the older MODBUS versions and 65,535 in newer MODBUS versions.

The second important element in the protocol is the set of function codes available that indicate what action is to be taken by both, the client and the server. The client uses different function codes to specify if it is going to read a discrete input, to write a coil, to read an input register, to write holding register, etc. The servers use function codes to indicate if they are sending a response with the value of a discrete input or a holding register, throwing an error exception code, etc. [6]. There are up to 127 function codes (1 byte value) divided in three groups: Public (1 to 64), User-defined and Reserved. Public function codes are guaranteed

to be unique, publicly documented and widely used by the majority of devices [6].

### 5. Implementation

The covert channel proposed in this paper operates by using the read-only public function codes from 1 to 4 and the object numbers from 0 to 9,999 of MODBUS servers. The covert channel exists between a covert client and a covert server both connected to a TCP/IP network and using MODBUS TCP (figure 5); however, the channel is designed so it can also work with MODBUS RTU in a serial bus network (RS485). Both devices, client and server are separate hosts with a software-based MODBUS implementation. For the purpose of this work, the Python library *pymodbus* will be used [9]. This software library allows the instantiation of MODBUS servers and clients at will by using two different Python scripts. A script called *sender* resides on the client device while another script called *receiver* resides on the server device.

Fig. 5: Covert devices: They can be software based (*pymodbus*) and work in a RS485 or TCP/IP network

To establish the covert channel, a covert client sends a request asking to read a covert server's object (coil, discrete input, holding register or input register). Then, the covert server receives the request, verifying the function code and the object number that is requested. This object number is mapped to a pre-defined ASCII character. This pre-defined "covert" value is the value that the client was intending to send to the server.

To circumvent SCADA firewalls that might be in place, the covert client always uses a read-only function code like 01 to read coils, 02 to read discrete inputs, 03 to read holding registers or 04 to read input registers. The covert client also specifies the number of the object it wants to read (0 to 9,999), which is in fact the ASCII value it wants to transmit.

An example of a normal MODBUS transaction with two steps (Response and Request) is explained below:

1 - The Client sends a Request to the Server asking to read the object 110 (0x6E):

Function Code	Input Register number (8 bits long)
04	110 (0x6E)

2 - The Server receives the Request and sends back a Response with the value of the object number 110. If the value is "30", the Response will look like this:

Function Code	Input Register number	Input Register value (16 bits long)
04	110 (0x6E)	30 (0x001E)

When the covert channel is implemented, the server executes a third step without changing the Request/Response schema. In the example, what the covert channel requires is to have the server interpreting the requested object number 110 (0x6E) as an ASCII character, which is the character "n". The covert server also answers the request with the value of the register 110 to make this look like a legitimate MODBUS request/response message. However, the actual value stored in the object 110 is irrelevant to the covert channel. Every time the client is requesting to read the server's Input Register number 110 (0x6E), in fact, the covert client is *sending* the ASCII character "n" to the covert server (see section 8, alphabet in use).

This behavior is illustrated with the figure 6. The first and second steps represent a normal MODBUS transaction, while the third step on the right represents the covert behavior.

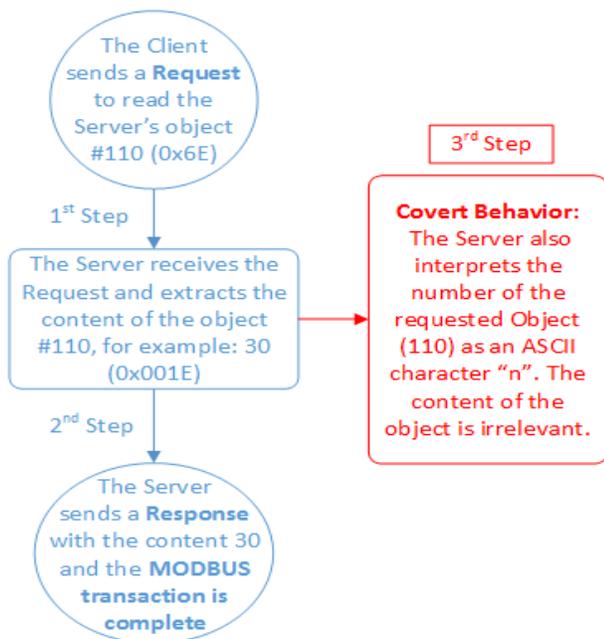


Fig. 6: First and second steps represent the normal behavior of a MODBUS transaction while the third step is the covert channel implemented

## 6. Challenges

The covert channel proposed in this paper is implemented by changing the way the server interprets the MODBUS PDU without changing the protocol structure. The covert channel consists of a series of MODBUS messages that are interpreted by the covert client and server. The messages are valid MODBUS client requests and server responses. Although SCADA systems are usually not taken into account when implementing security controls on the data networks [1]; there are solutions available to secure MODBUS installations.

Such controls represent a challenge for the covert channel and have been taken into account as an obstacle to overcome. As an example, it is important to consider that a SCADA installation could have a Tofino Firewall [5], SNORT IDS [8] or similar solutions implemented. These security solutions can look into the MODBUS PDU and block or alert of dangerous or abnormal messages. However, their common approach is to validate that the MODBUS requests and responses are well structured and valid, blocking the *write* requests as well as the malformed packets and non-MODBUS traffic. This covert channel has taken into account that some security measures might be in place and uses *read-only* and well-formed MODBUS requests that would be considered normal traffic.

## 7. Drawbacks

If an IDS solution is in place and filters all the commands sent to the SCADA network, this covert channel might have problems to send and receive all or some of the characters in the alphabet. However, for the SCADA system to work, whatever solution that is in place must allow the SCADA devices to communicate using the commands, device addresses and register numbers that are valid for that installation. Since the covert channel is aimed to use the same valid commands, addresses and register numbers, the covert channel should work properly.

## 8. Alphabet in use

The actual implementation of this covert channel uses the following alphabet structure:

Object_Number_X	=	ASCII_Character_X
-----------------	---	-------------------

How to use this alphabet is explained in more detail in the section 5 (implementation). This alphabet consists in a pure conversion from the object number to the ASCII character number. When the cover client sends a request, the covert server verifies the object number specified in the request and interprets what ASCII character corresponds to it. For example, the object number 110 (0x6E) maps to the ASCII character "n". This direct mapping keeps the test alphabet simple while supporting all the ASCII characters. However, this is not necessary and it is possible to have a

different alphabet by using a substitution table to obfuscate the characters that are being sent. In that case, the object number 110 (0x6E) can be mapped to any other ASCII character. This model needs at least 256 objects in the covert server to be able to receive all the ASCII characters, including the non-printable and extended ones. The section 10 (future work) explores other schemas for the alphabet.

## 9. Covert Channel Classification

### 9.1 Type

The covert channel described in this paper is considered as a behavioral covert channel. This is because the covert data is not contained as a payload within the MODBUS Request itself, but it is extracted depending on how the Request is interpreted. The covert server extracts the covert data when it interprets the requested object number. The covert server takes the number of the requested object and looks for the ASCII character associated with that number. Depending on which object is requested, the covert server interprets the ASCII character that was sent.

### 9.2 Throughput

The throughput of this covert channel will depend on the characteristics of the targeted installation. Some SCADA installations only work with RS485 serial devices, which reduce the available bandwidth considerably. However, if 256 objects are used, a minimum of one byte can be transmitted per Request/Response transaction. It is normal to have one transaction every one or two minutes in the older installations that only use RS485 serial devices.

To support all the ASCII characters (one byte per transaction), the covert channel needs to be implemented in a MODBUS network where it is normal to have 256 different objects in the servers. In the future work section, a hexadecimal alphabet is considered, which would require only 16 different objects to work while it cuts the throughput in half.

It is important to understand that the receiver Python script, located on the server device and explained on the implementation section, allows the attacker to emulate up to 65,535 objects in the server. That number of objects can be used to create a bigger alphabet, which will allow transferring more data. However, this will increase the chances of being discovered, as explained in the detection section.

### 9.3 Robustness

Even though the majority of SCADA systems are usually not protected with security controls on the network level; there are commercial and open source solutions available to secure MODBUS networks. The MODBUS security solutions available can look into the PDU and block or alert of dangerous or abnormal transactions. This covert channel was designed considering these solutions as an

obstacle to overcome. All the covert messages are legitimate Request/Response transactions that must be accepted by the security controls in place, if any, to allow the MODBUS devices to work, making the covert channel considerably robust.

### 9.4 Detection

To avoid detection, the attacker must know the normal behavior of the targeted network. Especially, it is critical to know how many objects are available in the servers of the network and use a similar number of objects in the covert server, which is emulated with *pymodbus* in the prove of concept. After taking care of this aspect, the critical point is to hide the sender and receiver scripts installed in the compromised devices. At this point, at least one client and one server are required, but in the future work section, another approach is considered to avoid using a compromised server, which will reduce the chances to be discovered.

Even though only 256 objects are required, the MODBUS original standard indicates that the servers could have up to 9,999 objects for each one of the four tables. If a server device with these characteristics is used, it is possible to send up to  $9,999 * 4 = 39996$  different values, but chances are the covert channel is found easily because those object numbers are not usually in use. The same happens if this is implemented in a modern MODBUS network, using a server device with 65,535 different objects per table. The cover client will be able to send up to  $65,535 * 4 = 262140$  different values but the communication can be suspicious.

### 9.5 Prevention

Measures that can be implemented to prevent this covert channel are related to how well the SCADA network configuration is documented and audited over time. If security controls like Tofino firewall or Snort IDS are implemented, they have to be configured to allow not only valid MODBUS objects, but only the object numbers that are used in that particular network. If the MODBUS network uses at least 256 objects in one sever device, then, that is enough to implement the covert channel. If the Hexadecimal alphabet is used as described in the future fork section, only 16 objects are enough. The physical security is extremely important, since the MODBUS network usually include devices installed in remote and unattended locations where contractors have access.

## 10. Future Work

A future implementation of this covert channel would use a slightly different alphabet that links an ASCII character with a combination of a Sever ID and an Object Number, thus, resulting in this structure:

“Sever\_ID\_Y + Object\_No\_X = ASCII\_Char\_Z”

In this new approach, the covert server is listening to all the requests sent by the covert client, including the ones sent to other servers in the network. Then, the covert server interprets the ASCII character that the covert client is sending by combining the server ID and the object number contained in the request. The advantage here is that the alphabet is “distributed”, augmenting the stealth level. The 256 ASCII characters can be interpreted by using several servers IDs and the object of those servers instead of using 256 objects of one single server.

A future alphabet based on 16 characters would reduce the amount of objects needed from 256 (to send all the ASCII Characters individually) to only 16 (to send values from 0x0 to 0xF). The actual ASCII characters would be “assembled” by using two 4-bits values.

Another future implementation of this covert channel would work with serial devices (RS485) only instead of TCP/IP based systems, since an important number of MODBUS installations run over serial BUS networks. This can be done using the same Python library (*pymodbus*) used for this covert channel.

Since a MODBUS server can't initiate a communication (it just responds to requests), the two-way communication could be implemented by having a timing mechanism. For this to work; the covert client has to send a request to the covert server with a special pre-defined character (like 0x05, the Enquiry Character in ASCII). This character is used for asking if new data is available and ready to be sent from the covert server. If the covert server has data to send or not, it will respond with another pre-defined pair of special characters indicating so. The covert client will then keep requiring the next character from the server until it receives the last one indicating there is no more data (0x04 or End of Transmission character for example).

An interesting experiment would be implementing this covert channel as well as an IDS solution based on Snort as described by Díaz [8]. The experiment would consist in determining if it is possible to detect the covert channel with the filters proposed [8]. It would be useful to establish if the covert channel can overcome this security control and how it can be improved to leave a smaller footprint if necessary. The results could demonstrate how deep the SNORT rules have to look into the packets in order to detect the covert channel. The hypothesis is that the IDS rules should not interfere with the valid MODBUS device addresses and object numbers. If the rules interrupt the traffic, trying to block the covert channel, they will also make the SCADA system unusable. This is because the covert channel uses only valid device addresses and registry numbers to operate.

## References

- [1] Industrial Network Security: Securing Critical Infrastructure Networks for Smart Grid, SCADA, and other Industrial Control Systems. Chapter 4 - Industrial Network Protocols. Eric Knapp. Syngress Publishing 2011. ISBN:9781597496452
- [2] Cyber security risk assessment for SCADA and DCS networks. P.A.S. Ralston, J.H. Grahamb, J.L. Hiebb. University of Louisville, JB Speed School of Engineering, Louisville, KY, United States. Department of Computer Engineering and Computer Science, University of Louisville, KY, United States. Available online 10 July 2007.
- [3] The cyber threat landscape: Challenges and future research directions. Kim-Kwang Raymond Choo. School of Computer and Information Science, University of South Australia, Mawson Lakes campus (Room F2-28), Mawson Lakes, SA 5095, Australia.
- [4] The SCADA challenge: securing critical infrastructure. By Steve Gold. Article from Network Security. August 2009.
- [5] Tofino security Appliance home website. <http://www.tofinosecurity.com/>
- [6] MODBUS Application Protocol Specification V1.1b by MODBUS Organization. December 28, 2006. Retrieved from <http://www.Modbus.org>
- [7] MODBUS Messaging On TCP/IP Implementation Guide V1.0a by MODBUS Organization. June 4, 2004. Retrieved from <http://www.modbus.org>
- [8] Using SNORT for intrusion detection in MODBUS TCP/IP communications. By Javier Jiménez Díaz (javier.jimenez@coit.es) and Robert Vandenbrink. SANS Institute InfoSec Reading Room. December 7th, 2011.
- [9] pymodbus Python Library. Retrieved from <https://pymodbus.readthedocs.org/en/latest/index.html>
- [10] Covert Channels in the HTTP Network Protocol: Channel Characterization and Detecting Man-in-the-Middle Attacks. By Erik Brown (erik.t.brown@gmail.com), Bo Yuan (bo.yuan@rit.edu), Daryl Johnson (daryl.johnson@rit.edu), Peter Lutz (peter.lutz@rit.edu). Rochester Institute of Technology, Rochester, NY, USA

# CloudWhip: A Tool for Provisioning Cyber Security Labs in the Amazon Cloud

A. Kevin Amarin, B. Shekar NH, and C. Leena AlAufi

College of Computer and Information Science, Northeastern University, Boston, MA, USA

**Abstract**—*Many traditional techniques of teaching cyber-security lack realistic environments to gain practical experience. In this paper we present CloudWhip, an open source framework to assist educators with the creation of security labs on Amazon Cloud Services. CloudWhip is developed to be accessible to even those people new to IaaS. We have successfully implemented various network security labs over a three year period in the cloud and our results suggest that the application of cloud computing in cybersecurity education not only saves costs, but also relieves the educational institutions of the burden of handling and maintaining complex IT Infrastructure. Cloud also better emulates managed IT service environments which is essential for SCADA security education. Our lab modules have initiated interest among students and spurred other faculty to conduct numerous security projects using Cloud services.*

**Keywords:** Cyber Security Labs, Cloud Computing in Education, Amazon Web Services(AWS), SCADA Security Labs, CloudWhip.

## 1. Introduction

As the number of organizations reporting data breaches in 2013 has increased 30% over 2012[1], the number of attacks continue to rise at a similar rate (about 47k security incidents in 2013[2]). The demand for security professionals continue to increase to handle this threat. According to the International Information Systems Security Certification Consortium (ISC)<sup>2</sup> more than 300,000 additional trained cybersecurity professionals are required in 2014[3] to meet the growing demand. This workforce gap has encouraged various government and private organizations to help fund programs designed to train security professionals in higher education. However to apply these core security concepts in industry students need to “*practice the science and the art of computer security*”[4] and many institutions fall short in crossing this chasm between textbook and practical learning.

To bridge the gap in hands-on training, institutions must invest a significant amount in hardware computing resources and SCADA devices. Even with the resources, faculty are tasked with creating challenging and engaging lab exercises using advanced security tools. Unfortunately, during lab exercises students using these programs can inadvertently attack public network computers that are not part of the target environment. Therefore, these exercises require precautions

and an appropriate level of isolation from the main university network to avoid collateral damage. This can have extreme side effects if SCADA production environment were affected inadvertently. Needing these isolated clusters, requires the university to devote more resources to maintain and firewall these environments.

The solution to isolating these security labs would be using virtual machines on a different campus LAN network as described in [5], [6] but, as noted above, the main drawback of these architectures are that they require additional resources, time and management. Moreover the scalability and flexibility in such a framework is constrained by the available budget from the university. One other alternative is utilizing a service provider for computing resources. In fact, the use of public cloud computing can present a flexible and cost effective solutions to address these concerns. These services can scale to fit any class load and be customized with policies to allow varying degrees of access to the students. Additionally, the infrastructure services are built to provide redundancy, including backup and storage which prevents downtime or data-loss due to equipment failure. Furthermore, online access and remote access requirements are built into the cloud platform as a requirement.

Cloud computing generally consists of either or a combination of these three main service models - Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). It is beyond the scope of this paper to go into each in detail, but IaaS in this context is where the Cloud Service Providers(CSP) can allow the educator to run virtual machines within the service provider's infrastructure. IaaS provides the option to choose the amount of Disk, Security, CPU, and Bandwidth resources you would like to consume. It also provides the ability to configure these resources to create a very specific custom network environment. One such IaaS provider is Amazon Web Services (AWS).

In this paper we propose an open source framework for deploying security lab environments on Amazon's AWS Cloud Services. The goal of this framework is to be able to implement an existing information security lab in a cloud with minimal knowledge of IaaS. This framework allows instructors to include cloud concepts into the lab or abstract them away if it is beyond the scope of the project. If included, students will be able to control all aspects of the computing platform including provisioning, configuration,

security, termination, monitoring and alerting. Our survey results show that the majority of students have never before had access to manage cloud computing resources. After these labs, if given the choice, most will opt to use similar IaaS features in future security projects. Thereby increasing their knowledge of IaaS along with basic security functionality.

What follows is a review of the related work in Section 2. We then discuss the tools and technologies used to create the lab modules in Section 3. Section 4 illustrates three lab modules that we have implemented in our course work. Next, Section 5 presents the results of the survey conducted and impact of the course modules and environment on student interest and the ease of usage for both students and faculty. Finally, Section 6 concludes this paper and presents future work.

## 2. Related Work

### 2.1 In-house Computer Security Labs

In-house computer security labs are those which demand the physical presence of students on campus and hence pose a challenge in current higher education environments. Many in-house security labs such as in [7], illustrate the difficulty in deploying such infrastructure in campus labs. These labs require physical isolation from the main campus network which is time consuming to install and configure and also requires additional resources and maintenance. Another difficulty the authors discuss is maintaining the state of the lab machines throughout the coursework. The execution of lab steps changes the state of the target host and it is not a trivial matter to revert the systems back to their initial state manually.

Another approach illustrated in the NetSecLab[6], consisted of several team machines, victim machines and traffic generator machines. The Traffic generator used a set of scripts to emulate a realistic environment. Such emulators will be restricted to generate traffic based on the pre-configured parameters and thus can only provide pseudo realistic environment. Due to the number of different components and the complexity of the environment, the initial provisioning and maintenance of this lab requires IT to dedicate resources for a significant period of time. If this lab is provisioned in the institution this would require support staff and computing resources. Additionally, scaling a complex support infrastructure with class size will also require scaling the support staff hours.

### 2.2 Virtual Lab Environments

Efforts have been made to isolate and decentralize virtual lab environment such as in [8]. The authors presents a security lab framework, where pre-configured images of virtual systems are distributed to the students and installed on student's personal computers which provides mobility and flexibility for students while maintaining the state of

the system by instructors, as discussed earlier. If the state changes and the system can no longer be utilised for a lab, the initial state can be reverted to through the initial image of the virtual machine created. However, it is hard to conduct labs which require collaboration among students using this system and, this framework is not suitable for a dynamic lab modules. This is because a small change or update in the initial image of the virtual machine by the instructor requires redistribution of the entire image. The uncertainty of a student's personal resources adds a challenge to debugging during lab exercises.

Other virtual lab environments include [5]. They present a distributed virtual laboratory architecture based on the Tele-Lab framework using resources from two different universities with similar course structures. Though these exercises in some way provide realistic implementation for students, there exists scalability issues and collaboration among universities is usually difficult as there is no standard architecture for network and security.

In general virtual environment labs address the issues of mobility, flexibility and maintenance to a certain extent but share the same issue of scaling physical infrastructure as in in-house security labs, adding further cost and time to already overextended in-house staff and infrastructure.

### 2.3 Private Cloud

Private cloud computing model offers the same basic features as a cloud service but this infrastructure is implemented within the university firewalls, which offers better control over user data and an option to move away from proprietary vendor lock in. In [9] authors used Tele-Lab environments with a middleware layer integrating OpenNebula taking advantage of the cloud framework functionality. Xu and et al., in [10] presented their Cloud based lab called V-Lab which provides a contained experimental environment for hands-on experiments. Building on a private cloud allows educators to utilize existing in-house hardware with the abstraction flexibility of virtualization. Though, image management and debugging become easier compared to a VM decentralization method, this requires the necessary computing, memory and network resources owned and operated by the university to meet the lab's scalability requirements.

### 2.4 Cloud vs Dedicated Servers

Several research studies like [11] and [12] suggest that the use of cloud computing by educational institutions benefits students by raising their computing resource accessibility irrespective of location, increases availability and mobility. For faculty, it enables them to create custom images for a specific course and share the same infrastructure for different courses if necessary. For administration, cloud computing standardizes application and processes, lightens the burden of software version control and maintenance, optimizes

resource allocation and brings greater visualization. Importantly, it can be cost effective as it saves money on underutilized computing resources, software licensing, and IT staff time. The setup of lab infrastructure in public clouds can be done in few minutes and there exists no downtime during scaling hardware resources. In-house labs or a private cloud, setting up the infrastructure to a working state demands the Instructor or teaching assistant to have a great knowledge of infrastructure management, which is not simple and can be time consuming. Instead, using our proposed framework (CloudWhip), the Instructor can leverage the infrastructure management to the Cloud Service Providers and spend more time on designing the lab modules.

Through our analysis of above mentioned lab environments we observe, that security lab environments are usually designed in isolated network spaces with limitations related to hardware resources and maintenance during scaling along with access restrictions depending on resource availability. Our approach address these issues, while presenting a solution with effective provisioning, as well as a mobile and scalable infrastructure on a Cloud.

### 3. Tools and Technologies Used

#### 3.1 AWS and AWS Education Grant

Amazon Web Services(AWS)[13], is a collection of IT infrastructure or Cloud Computing services. These services include global computing, storage, database, analytics, application, and deployment to foster organizations scale applications and computing resources on demand at lower IT costs. All the lab modules mentioned in this work were built on AWS services.

*AWS in Education* is a program that assists educators, academic researchers, and students by providing free usage credits to utilize the on-demand infrastructure of the Amazon Web Services to teach advanced courses, tackle research endeavors, and explore new projects. We have received a grant each of the past three years which helped us provide the labs to the students without any cost to the university. We found the grant application process to be fairly simple and it is available online at [14] for any institution.

Once the grant was approved, the Instructor has the option to receive the AWS credits on his account, or provide it directly to each student in the form of a credit code. The credit code would allow students to manage their own usage, however it does require the student to sign up for an AWS account with a credit card. For these security labs we choose the single central account model to avoid any account provisioning issues. Access to instances is authorized through the Instructor's AWS account by creating accounts in the AWS Identity and Access Management(IAM)[15] service.

#### 3.2 AWS CLI and Boto

Amazon AWS Command Line interface (CLI) allows the user to automate and control multiple AWS services via

simple to use tools. Boto is an AWS Software Development(SDK) Kit for Python. It provides Application Programming Interface(API) to many AWS services which eases the process of scripting and automation. The documentation for the AWS CLI and Boto can be found at [16] and [17] respectively. Section 4 illustrates to how we used these tools in our lab environment on Amazon Cloud Services.

#### 3.3 AMIs and EBS

An Amazon Machine Image (AMI) is a template that provides the requisite information (Operating System and applications) to launch an instance. The advantage of creating such a template is that it can be used to launch any number of instances assuring idempotence in the initial state of the virtual machines and also include launch permissions that control the instance, thus easing user and access management in a large deployment. We can also configure an AMI to use an Elastic Block Store (EBS) which allows you to create storage volumes acting like an external block device. Customized AMIs can either be created from scratch or use one of the Amazon provided images as a base to install the required application on top of it. The process of creating your own AMI depends on the root storage of the device - it can either be an Amazon EBS-backed AMI or an Instance store-backed AMI. The steps to create each type can be found at [18][19] and [20][21] respectively.

### 4. Design and Implementation of Lab Modules

We designed three labs modules for our Network Security Course on AWS. In this section we will walkthrough the steps used to design and implement these labs.

#### 4.1 Lab 1: Gaining Access to OS & Application

The first lab was designed to give an hands-on experience with attacking a target computer. [22] defines first three phases of the attack architecture as Reconnaissance, Scanning and Gaining Access to OS & Application. For diverse exposure in operating systems and applications we build 3 customized AMIs for the lab. The configuration for these AMIs are as shown in Table 1 and the entire architecture for the lab environment is as shown in Figure 1. Every student was assigned to a Point Of Delivery(POD) consisting three systems; an attacker system (Kali Linux) and two victim machines (Windows 2008, CentOS). To access the POD, students would use VNC client such as TightVNC[23] to connect to the X Windows GUI of the attacker system. The VNC port on the attacker system was the only item accessible to external users.

All of the PODs were placed under one large subnet (172.16.0.0/20) and an additional subnet (172.16.255.0/24) acted as a Demilitarized Zone(DMZ) Network. The DMZ consisted two instances running a web application and

MySQL-Server, emulating a Multi-Tier Architecture[24]. Kali Linux was chosen as the attacker system because, this distribution is packed with a wealth of pre-configured security tools such as Metasploit, Nmap and other open source penetration testing tools. Also note that in this architecture, only Kali Linux had a public IP assigned to it so students can reach the system remotely and firewall rules were applied to these subnets such that outbound attack traffic from this system was contained within its own subnet. In case of more granular isolation requirement, each POD can be configured to reside on its own subnet as shown in Figure 2. This configuration requires creation and configuration of more subnets.

In the first phase of this lab, students were allowed to conduct reconnaissance on the network and identify the target systems within the subnet assigned to them. The second involved students performing intense network scan using Zenmap[25] to determine the services that were running on the target systems exploring for any vulnerable application using Nessus[26] in Kali Linux. The final phase of the lab was to use the knowledge gained from the first two phases and try to gain access to OS and applications running on these target systems using the tool Armitage[27]. Amazon Windows AMIs are patched with latest Microsoft security updates and older non-patched versions are not available. Due to the up to date security patches, it is difficult to a student to use common Windows OS exploits available in Armitage. In an effort to in-secure the OS, we tried to remove patches from the default Windows AMIs (2003, 2008). This ended up being counter productive as two issues occurred; first the uninstaller crashed on a number of security patches and failed to back out the change, and second the patches that were removed semi-often caused instability in the OS leading to kernel lockups. For this reason, we focused the attacks in the lab on the applications installed vs. the OS itself. We believe focusing on the application also represents the shift to APT style attacks which have increased in the past decade[28] since Blaster Worm[29]. For this lab we installed a vulnerable Oracle MySQL application, and students exploited the application using the *mysql\_payload*[30] module found in Metasploit for UDF payload execution vulnerability.

We included two bonus question for the lab; the first was to brute force *ssh* and gain access to CentOS system, and the second was to exploit a vulnerable e-commerce site in the DMZ and dump all the credit card information stored in a MySQL database. At the end of the lab students were asked to submit a short report on their findings and how they can defend against each phases of the attack architecture covered in this lab.

To implement the lab infrastructure, we first configured the VPC and Subnets using the AWS Console VPC Wizard tool. A AWS security group was created that allowed only the required inbound and outbound traffic to carry out lab

exercises, allowing us to contain the attack traffic within the internal lab environment. Finally we associated the subnet with Internet Gateway in the route table console. This enabled any explicitly allowed network traffic to flow out of the VPC to the general internet. A step-by-step guide to manually set up your VPC and subnets can be found here[31]. This entire provisioning and configuration process is very well documented and the AWS console has number of wizards to walk you through the process. Once the VPC and subnet were configured, we utilized the AWS CLI and developed a script to deploy instances in our subnets according to the architecture shown in Figure 1. The script was rewritten to be much more flexible and formed the basis for the CloudWhip tool.

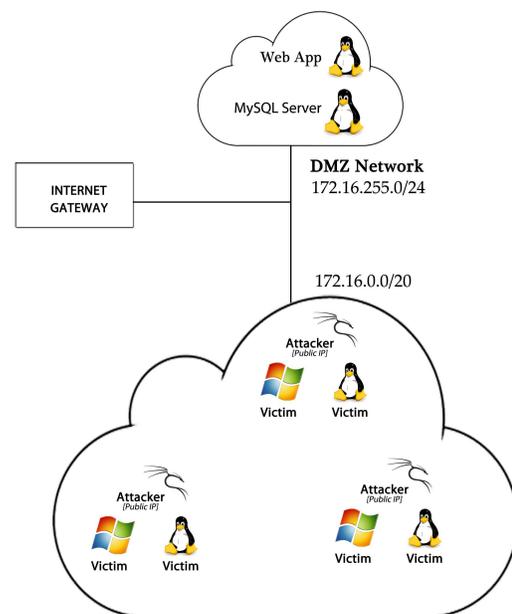


Fig. 1: Lab 1 - Architecture with PODs in Same Subnet

Table 1: Configuration Details of Customizes AMIs.

Operating System	Packages Installed and Additional Configurations
Kali Linux	openSSH, VNC Server, Nmap, Nessus, Metasploit, Armitage
Windows 2k3 R2	mysql (Oracle 5.5.9), Enabled File and Print server roles and removed security updates
CentOS	dovecot, apache web server

## 4.2 Lab 2: AWS Services and Snort IDS

In this lab students were introduced to AWS Cloud services to deploy and run Snort[32], an Intrusion Detection System. Here we utilized the Identity and Access Management(IAM)[15] service on AWS to create multiple users and manage permissions through Role Based Access

Control(RBAC) system. In the first part of the lab, students login to the AWS Management Console using the credentials emailed to them and launch an existing customized AMI, which is a Linux distribution with Snort pre-installed in it. They also create and apply a new security group while initializing the instance. In this case the security group is wide open to all traffic from any source. This was done to allow the snort instance to get an uncensored view of incoming traffic. In most cases 10-15 minutes after an instance is launched it will start receiving incoming unsolicited requests from scanning systems. These requests are a mix of other AWS instances and external compromised hosts and will generate IDS alerts allowing students to experience a realistic attack traffic environment. Also students are able to experiment with the snort sensor signatures at greater depth. This flexibility would not have been possible with a virtual machine running on student's laptop or virtual machine hosted on our college without significant IT configuration. The rest of the lab focused on configuring Snort sensor and creating rules to alert to various scenarios such as a ssh connection to a particular system, alerting when a particular URL is accessed from the internal network and others. Students used BASE, which is one of the GUI for Snort IDS, to manage and visualize the alert notifications. Instructors can also incorporate a SCADA honeypot as explained in [33] which could use snort alerts to capture packets that match any known SCADA attack profiles. Later this packet capture can be used to replay the attack in a SCADA lab environment. Students can then dissect the attacks and discuss the various appropriate defenses. Optionally students could create and test IPS rules to block these specific attack vectors and apply them to the SCADA honeypot.

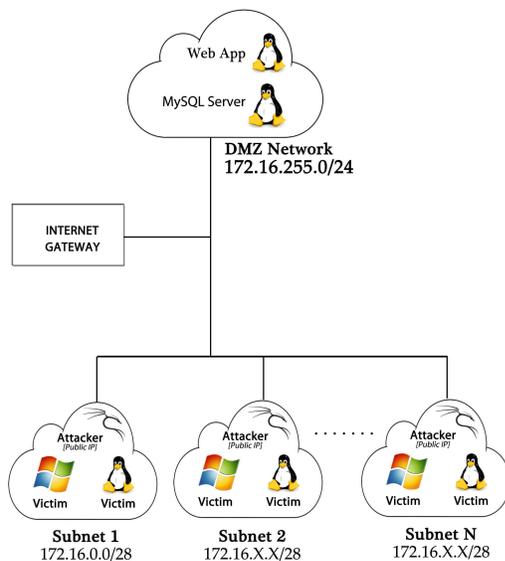


Fig. 2: Lab 1 - Architecture with PODs in its Own Subnet

### 4.3 Lab 3: Online Brute Force Attack

Verizon Data Breach Report[2] shows 76% of network intrusions in 2013 exploited weak or stolen user credentials, that is by far the largest attack vector. In this lab students performed an online brute force attack from their local computer against a web target hosted in AWS cloud. Each student was assigned an Amazon virtual machine running a web server configured with a basic HTTP authentication for “secret” URLs. Each student had previously in the course installed a local virtual machine of Kali Linux on their personal computers. The first phase of the lab was reconnaissance, where students gathered about 200 user email information associated with target web application using an open source tool *theharvester*[34] from their local VM. Later the students used a password dictionary containing 10k most common passwords[35] and the *hydra-gtk*[36] network logon cracker to brute force a user account gathered in the reconnaissance phase. Once they were able to logon as a valid user, the web page provided them instructions for bonus question. The bonus question comprised of an additional secured URL with a different username and a password generated from a larger phpBB dictionary. This dictionary, which is publicly available at [37], contains 184k clear text passwords from users of phpBB.com. This site was compromised and the MD5 hashes were posted to pastebin, and later were brute forced by [38] and others and made available during DefCon17. A similar attack scenario can be crafted as a lab module to gain access to a publicly facing control system with admin privileges in a SCADA environment, as illustrated in [39]

For this lab the Instructor used the community Ubuntu AMI, installed open source Nginx web server and configured HTTP authentication on specified URLs. A set of well known weak user credentials were used for this purpose so that the students will be able to brute force accounts using the common password dictionary. The goal of the lab was to show level of difficulty in online brute force attacks based on password complexity.

## 5. Survey and Results

The above mentioned lab modules were implemented on Amazon Cloud Services and used in our Network Security Practices coursework over a period of three years (6 classes: 3 online and 3 on-campus). At the end of the semester during Spring 2014, to evaluate the effectiveness of our Lab modules and the Cloud environment, we conducted an online survey and the results are as follows.

95% of the students agreed that these lab modules aid in better understanding of concepts taught in the class room and 82% of them noticed that conducting labs on a cloud provided them mobility and flexibility in completing their lab exercises. The survey results suggest that our lab modules encouraged most of the class to use Cloud Services in their

Table 2: Survey Results

New Cloud Service users	47%
Difficulty level using AWS Services	Easy: 79%, Moderate: 16%, Hard: 5%
Had performance or access issues	21%
Prefer Cloud Services over VMs on Localhost or College Servers	82%

future security projects, the main reason being flexibility and scalability. Students also commented that they would use Cloud Services more often if it was free. In fact they can register to AWS Free Tier[40] which allows them to use most of the AWS Services for a year free of cost. Table 2 summarizes our survey results.

## 6. Conclusion and Future Work

In this paper we discussed various drawbacks of some of the traditional cybersecurity teaching methods and how educational institutions, students and faculty can benefit by implementing cyber security labs on a cloud instead. We presented to you three sample lab modules and also demonstrated how to build the environment using AWS-CLI. To further automate the process of deployment, we developed CloudWhip, a wrapper using AWS Boto API, which allows instructors to specify their requirements in a configuration file and deploy the entire lab environment including VPC, Subnets, Instances and Internet Gateway in the Amazon Cloud Services within minutes.

CloudWhip is used to automate the process of deploying and configuring the lab environments. It's goal is to take the time necessary to create a AWS security lab environment from hours to minutes in a simple extensible way. It is under development and is made available at [github.com/NUCyberEd/CloudWhip](https://github.com/NUCyberEd/CloudWhip) under the *MIT License*. The labs discussed above used a very primitive version of CloudWhip. The tool was re-written to support a variety of lab architectures, not only the ones listed above. We would like to extend the CloudWhip project further to cover all the features on Amazon Web Services and provide more granular configuration of lab infrastructures.

We highly encourage course instructors to make use of this wrapper and provide us with reviews and suggestions and share the labs they created using our tool for improvement towards this project.

## 7. Acknowledgments

This work was possible because of the *AWS in Education Grant* program, which funded us to conduct the security labs on Amazon Cloud Services. We would also like to thank the students of IA5150 for their feedback and faculty and staff at College of Computer and Information Science, Northeastern University for their support.

## References

- [1] ITRC, "Identity Theft Resource Center - 2013 Breach List," <http://goo.gl/ZVzu5k>, Tech. Rep., 2013.
- [2] Verizon, "2013 Data Breach Investigations Report," 2013.
- [3] M. Suby, "The 2013 (ISC)2 Global Information Security Workforce Study," <http://goo.gl/dfguAI>, Tech. Rep., 2013.
- [4] M. Bishop, "Education in information," *IEEE Concurrency*, vol. 1, pp. 4–8, October-December 2000.
- [5] C. Willems, T. Klingbeil, L. Radvilavicius, A. Cenys, and C. Meinel, "A distributed virtual laboratory architecture for cybersecurity training," in *Internet Technology and Secured Transactions (ICITST), 2011 International Conference for*, Dec 2011, pp. 408–415.
- [6] C. P. Lee, A. S. Uluagac, G. S. Member, K. D. Fairbanks, J. A. Copeland, and L. Fellow, "The Design of NetSecLab : A Small Competition-Based Network Security Lab," vol. 54, no. 1, pp. 149–155, 2011.
- [7] L. ben Othmane, V. Bhuse, and L. Lilien, "Incorporating lab experience into computer security courses," in *Computer and Information Technology (WCCIT), 2013 World Congress on*, June 2013, pp. 1–4.
- [8] L.-C. Chen and L. Tao, "Teaching web security using portable virtual labs," in *Advanced Learning Technologies (ICALT), 2011 11th IEEE International Conference on*, July 2011, pp. 491–495.
- [9] D. Moritz, C. Willems, M. Goderbauer, P. Moeller, and C. Meinel, "Enhancing a Virtual Security Lab with a Private Cloud Framework," pp. 314–320, August 2013.
- [10] L. Xu, D. Huang, and W.-t. Tsai, "Cloud-Based Virtual Laboratory for Network," pp. 1–6, 2013.
- [11] M. D. B. Chandra, Deka Ganesh, "Cost Benefit Analysis of Cloud Computing in Education," *2012 International Conference on Computing, Communication and Applications (ICCCA)*, pp. 1–6, Feb 2012.
- [12] F. Abidi and V. Singh, "Cloud servers vs. dedicated servers; a survey," in *Innovation and Technology in Education (MITE), 2013 IEEE International Conference in MOOC*, Dec 2013, pp. 1–5.
- [13] AWS, "Amazon Web Services (AWS) - Cloud Computing Services," <http://aws.amazon.com/>, [Online; Accessed 02-April-2014].
- [14] AWS Grants, "AWS in Education (Grants)," <http://aws.amazon.com/grants/>, [Online; Accessed 31-March-2014].
- [15] "AWS Identity and Access Management (IAM) in the Cloud," <http://aws.amazon.com/iam/>, [Online; Accessed 04-April-2014].
- [16] AWS CLI, "AWS Command Line Interface," <http://aws.amazon.com/cli/>, [Online; Accessed 31-March-2014].
- [17] AWS Boto, "AWS SDK for Python," <http://aws.amazon.com/sdkforpython/>, [Online; Accessed 01-April-2014].
- [18] AMI Linux EBS, "Creating an Amazon EBS-Backed Linux AMI - Amazon Elastic Compute Cloud," <http://goo.gl/u83ypX>, [Online; Accessed 31-March-2014].
- [19] AMI Win EBS, "Creating an Amazon EBS-Backed Windows AMI - Amazon Elastic Compute Cloud," <http://goo.gl/lfbwR6>, [Online; Accessed 31-March-2014].
- [20] AMI Linux: Store-backed, "Creating an Instance Store-Backed Linux AMI - Amazon Elastic Compute Cloud," <http://goo.gl/QcMAQS>, [Online; Accessed 31-March-2014].
- [21] AMI Win: Store-backed, "Creating an Instance Store-Backed Windows AMI - Amazon Elastic Compute Cloud," <http://goo.gl/5sRxvZ>, [Online; Accessed 31-March-2014].
- [22] T. L. Edward Skoudis, *Counter Hack Reloaded: A Step-by-Step Guide to Computer Attacks and Effective Defenses*, 2nd ed. Pearson Education, Inc., 2006.
- [23] "TightVNC: VNC-Compatible Free Remote Control / Remote Desktop Software," <http://www.tightvnc.com/>, [Online; Accessed 04-April-2014].
- [24] Oracle, "Application and Networking Architecture," <http://goo.gl/JwS2O2>, 2013, [Online; Accessed 18-April-2014].
- [25] Zenmap, "Zenmap - Official cross-platform Nmap Security Scanner GUI," <http://nmap.org/zenmap/>, [Online; Accessed 04-April-2014].
- [26] Tenable, "Nessus Vulnerability Scanner," <http://goo.gl/sC3OZM>, [Online; Accessed 04-April-2014].
- [27] "Armitage Tutorial: Cyber Attack Management for Metasploit," <http://goo.gl/SS3v0s>, [Online; Accessed 04-April-2014].

- [28] Mandiant, "Mandiant 2013 Threat Report," <http://goo.gl/Ir5JxM>, Tech. Rep., 2013.
- [29] G. Keizer, "Blaster from the past: The worm that zapped XP 10 years ago - Computerworld," <http://goo.gl/ZBa8uH>, [Online; Accessed 17-April-2014].
- [30] Rapid7, "Oracle MySQL for Microsoft Windows Payload Execution," <http://goo.gl/JYEj0l>, [Online; Accessed 15-April-2014].
- [31] AmazonVPC, "Getting Started with Amazon VPC - Amazon Virtual Private Cloud," <http://goo.gl/xzKmHf>, [Online; Accessed 15-April-2014].
- [32] "Snort :: Home Page," <http://www.snort.org/>, [Online; Accessed 04-April-2014].
- [33] K. Wilhoit, "The SCADA That Didn't Cry Wolf," <http://goo.gl/Amw1VG>, Trend Micro Forward-Looking Threat Research Team, Tech. Rep. Part 2, 2013.
- [34] Edge-Security, "theharvester - The Information Gathering Suite," <http://www.edge-security.com/theharvester.php>, [Online; Accessed 17-April-2014].
- [35] Hood3dRobIn, "10k Most Common," <http://goo.gl/2NO0X7>, [Online; Accessed 17-April-2014].
- [36] Van Hauser, "THC-HYDRA - fast and flexible network login hacker," <https://www.thc.org/thc-hydra/>, [Online; Accessed 17-April-2014].
- [37] SkullSecurity, "Passwords - SkullSecurity," <https://wiki.skullsecurity.org/Passwords>, [Online; Accessed 18-April-2014].
- [38] S. A. Matt Weir, "Cracking 400,000 Passwords," <http://goo.gl/0vRjp5>, Tech. Rep., 2009.
- [39] ICS-CERT, "Incident response activity," <http://goo.gl/gRsXtb>, Tech. Rep., April 2014.
- [40] AWS Free, "AWS Free Usage Tier," <http://aws.amazon.com/free/>, [Online; Accessed 18-April-2014].

# SCADA Cybersecurity Education from a Curriculum and Instruction Perspective

R. T. Albert<sup>1</sup>

<sup>1</sup>Professional Management Division, University of Maine at Fort Kent, Fort Kent, ME, USA

**Abstract** – *Concerns over the cybersecurity risks associated with Supervisory Control and Data Acquisition (SCADA) components used to manage significant utility infrastructures have continued to rise. Educational institutions have been called upon to prepare a workforce that is sensitive to such concerns and able to effectively address them. Recognition of the absence of a consistent approach by such institutions to provide such education served as the impetus for this study. The aim of this study was to elucidate the role of SCADA in curriculum standards relating to cybersecurity, provide examples of SCADA instructional approaches, and offer recommendations as to which may work best in a typical public undergraduate university setting. The purpose of this paper is to share approaches and recommendations for addressing SCADA cybersecurity education from a curriculum and instruction perspective.*

**Keywords:** SCADA, Cybersecurity, Education, Information Security Education, Curriculum

## 1 Introduction

Concerns over the cybersecurity risks associated with Supervisory Control and Data Acquisition (SCADA) components used to manage significant utility infrastructures have been widely reported [1-4]. The vulnerability to attack exhibited by SCADA components is due in large part to the lack of authentication and confidentiality controls in the SCADA protocols themselves. Physical security of SCADA remote terminal units (RTUs) further aggravates the situation. Cybersecurity risks continue to mount with each report of the potential for loss of a critical infrastructure.

A recent confidential power-flow analysis by the Federal Energy Regulatory Commission was reported as revealing the United States could suffer a coast-to-coast blackout if just nine of the country's 55,000 electric-transmission substations were knocked out in a coordinated attack. The possibility for a nation-wide blackout lasting for weeks or months has also been reported. The analysis triggered increased interest in tightening physical security through imposition of security standards. Efforts to squelch public disclosure of such reports are increasing in frequency as various agencies [5, 6] and companies [7] fall back to "security through obscurity" in an attempt to sure up weak SCADA cybersecurity. Concomitantly, the need to raise awareness, knowledge and skills of those charged with the cybersecurity of SCADA systems has become increasingly pronounced.

Educational institutions have been called upon to prepare a workforce that is sensitive to the cybersecurity risks associated with SCADA components and able to effectively address them. One of the great impediments to progress in this regard has been the absence of a consistent approach by such institutions to provide such education. Several factors have contributed to the exacerbation of the problem. Absence of a consistent approach by such institutions to provide such education served as the impetus for this study.

The aim of this study is to elucidate the role of SCADA in curriculum standards relating to cybersecurity, provide examples of SCADA instructional approaches, and offer recommendations as to which may work best in a typical public undergraduate university setting.

The aim of this paper is to share approaches and recommendations for addressing SCADA cybersecurity education from a curriculum and instruction perspective in hopes of furthering a more consistent and effective approach to addressing SCADA cybersecurity education.

## 2 The Role of SCADA in Curriculum Standards

The current ambiguous state of SCADA cybersecurity in curriculum standards is due to many factors. For example, differences continue to exist in the opinions and interpretations of cybersecurity as a concept. Another example is the ongoing development of control system security standards by many professional organizations. Some of these organizations have made great strides. In a similar sense, the emergence of Information Security and Assurance as an academic entity and, concomitantly, the identification of a suitable accrediting body, remain in flux. Still other factors have slowed the advancement of SCADA cybersecurity education, not least of which is formation of a consensus on the best placement of the broader cybersecurity curriculum within established academic programs.

Rowe [8] argues that Information Technology programs are "uniquely best-suited to an advanced cybersecurity curriculum" (p. 115) since the core "pillars" serve as prerequisites for cybersecurity. Purdue University's Center for Education and Research in Information Assurance and Security (CERIAS) has made substantial contributions to the broader discussion of the role of security education within computing programs but the wide-spread adoption of their

recommendations relating to the use of a layered approach to cybersecurity, remains to be realized. This is likely due in part to the inherent cross-disciplinary nature of cybersecurity concepts.

The cross-disciplinary nature of SCADA cybersecurity education, falling between curriculum standards, as it were, has limited full recognition and adoption in any single domain. For example, SCADA systems coverage has been addressed in numerous ways in undergraduate mechanical engineering and mechatronics education programs, as cited by Senk [9]. In some instances, SCADA cybersecurity has been identified for inclusion in curriculum to better meet new competencies demanded by the workplace market [10]. In other instances, only partial coverage, such as the “availability” of system functions aspects of SCADA cybersecurity are promoted, according to Papa [11]. More often than not, the focus of such coverage has not been on cybersecurity aspects of SCADA systems. Collectively, such factors can retard the advancement of SCADA cybersecurity education. Nevertheless, progress continues and SCADA is becoming more pronounced in curriculum standards.

Government and societal norms, economic, political, technological, environmental, and audience diversity are among the numerous factors that should be considered when shaping curricula. Each curriculum standard/guideline development entity also takes into account its own particular audience needs (e.g., professional certification requirements).

## 2.1 SCADA and ACM/IEEE Curriculum Standards

The most recent 2013 Association of Computing Machinery (ACM)/Institute of Electronic and Electrical Engineers Computer Society (IEEE) Curriculum Guidelines for Undergraduate Degree Programs in Computer Science (2013) [12] are a reflection of the evolution of the field over many decades. The guiding principles followed in its development are similar to those followed by other curriculum development entities. Included among these principles are:

- “Computer science curricula should be designed to provide students the flexibility to work across many disciplines.” (p. 20)
- “Computer science curricula should be designed to prepare graduates for a variety of professions, attracting the full range of talent to the field.” (p. 21)
- Curricular guidelines “must be relevant to a variety of institutions”(p. 21)
- Curricular guidelines “should provide the greatest flexibility in organizing topics into courses and curricula.” (p. 21)

A three-tiered classification of a “Body of Knowledge Units” was identified (Core Tier 1 – Essential topics, Core Tier 2 – Important topics, Electives). A very similar classification approach has most recently been used by the National Security Agency (NSA)/Department of Homeland Security (DHS) in its establishment of criteria for recognition of Centers of Academic Excellence in Information Assurance/Cyber Defense (CAE-IA/CD) [13]. In addition, three levels are identified for depth of coverage in a topic (Knowledge, Application and Evaluation)

For the first time, Information Assurance and Security (IAS) are recognized as a knowledge area and have been added to the body of knowledge “in recognition of the world’s reliance on information technology and its critical role in computer science education.” (p. 97). IAS is identified as unique among the set of knowledge areas “given the manner in which the topics are pervasive throughout other knowledge areas.” (p. 97). IAS concepts are classified in all three tiers and are widely dispersed. Over 30 hours of coverage are associated with each of Core Tier 1 and Core Tier 2. Cross-core coverage supports the potential value of a modular instructional approach.

No specific reference to SCADA cybersecurity exists in the ACM/IEEE Computer Science (CS) Curriculum Guidelines, however, the concepts essential to developing knowledge and skills in SCADA cybersecurity are present. This suggests that SCADA cybersecurity education might best be addressed within computer science programs through use of an “exemplar” approach in which topics/outcomes are presented/achieved through exploration of their value in the context of SCADA cybersecurity design and operations.

Information Technology (IT) is the newest computing discipline covered by the ACM/IEEE computing curricula recommendations [14]. These recommendations are also evolving and are presented in a separate volume. Nevertheless, information assurance and security is identified as overarching the IT pillars including programming, networking, human-computer interaction, databases, and web systems. One of the guiding principles used in the development of the IT curriculum guidelines is “The curriculum must reflect those aspects that set Information Technology apart from other computing disciplines” (p. 22). By earlier recognition and inclusion of IAS in its curriculum guidelines, it could be argued that IAS is more aligned with IT than CS. This would be a mistake given the ever-evolving nature of these curricular guidelines and the apparent influence this earlier inclusion has had on the more recent CS curriculum guidelines.

## 2.2 Critical Infrastructure and Control Systems Security Curriculum

The model proposed in the Critical Infrastructure and Control Systems Security Curriculum [15] is very comprehensive. The model is presented as a collection of

modules that tend to emphasize policy aspects of the management of critical infrastructure and control systems. Nevertheless, its modularity aids the adoption of components that are well suited to a variety of educational contexts, particularly those at the masters degree level. “The curriculum focuses primarily on the role of control systems in energy, cyber, and other infrastructures [and] provides materials from which instructors can design a specific syllabus to meet the needs and requirements of their particular circumstances.” (p. 1) [15]

## 2.3 SCADA and NSA/DHS CAE-IA/CD Recognition Guidelines

The NSA/DHS have established criteria for recognition of Centers of Academic Excellence in Information Assurance/Cyber Defense (CAE-IA/CD). These criteria have evolved substantially over the past few years and currently include “Industrial Control Systems/SCADA Security” as a “specialty area” that educational institutions may optionally be assessed against in their application for NSA/DHS recognition [13].

On one hand, inclusion of SCADA security among the criteria may be viewed as beneficial to the cause. On the other hand, however, relegating it to the status of an optional specialization area may be viewed as detrimental.

Collectively, these curriculum standards, particularly the inclusion of SCADA as an identified component, aid in directing the evolution of academic program curricula and instructional practices. The evolution of the curriculum standards themselves is driven by industry and societal needs. As these needs become more pronounced, the rate of evolutionary advancement will likely increase. In the interim, identification of demonstrably effective instructional approaches is left to those instructional staff engaged in the scholarship of teaching and learning.

Based on the legacy work of Ernest Boyer's Scholarship Reconsidered [16], the scholarship of teaching and learning has been variously defined as promoting “teaching as a scholarly endeavor and a worthy subject for research, producing a public body of knowledge open to critique and evaluation. Its intent is not only to improve teaching but also to create a community of ‘scholarly teachers’ who add to the body of knowledge about teaching and learning as well as benefiting from the SoTL research of others” [17].

## 3 SCADA Instructional Approaches

SCADA instructional approaches exist in many forms. Variability due to differences in student learning objectives and student characteristics is to be expected. So too is variability due to experimentation with various instructional approaches. Those instructors engaged in the scholarship of teaching and learning may best exemplify such active

experimentation. Variability then is expected, natural and demonstrative of the potential for instructional advancement.

Consistency in the application of instructional approaches demonstrated to be effective is essential to adequate preparation of the cybersecurity work-force. It is also essential to addressing the concerns over cybersecurity risks associated with SCADA components used to manage significant utility infrastructures. Until such time the role of SCADA cybersecurity becomes precisely and prominently defined in curriculum standards, consistency of adoption of demonstrably effective instructional strategies will remain an elusive goal. Sheen has argued the extent to which SCADA system security is currently incorporated into computing disciplines varies from little to none [18]. To date, different SCADA cybersecurity instructional strategies have, nevertheless, shown promise.

Note the following sample of instructional approaches represents a variety of approaches that have been studied. These approaches are not mutually exclusive but often complementary when appropriately adopted and implemented.

### 3.1 Modular Approach

Guillermo [19] reported on a modular approach to incorporation of SCADA cybersecurity education that is intended to “... augment an existing course on critical infrastructure with slightly advanced technological and information security-related materials without overwhelming non-computer science students” (p. 55). The modules and course learning outcomes focus on critical infrastructure and control systems (CICS) security. Four modules are presented for consideration.

- CICS Technology
- Exploration of Prominent CICS Security Standards and Vulnerability Assessment
- CICS Risk Assessment and Mitigation Techniques
- CICS Security Policies

The value and role of supportive laboratories and lab activities are also addressed as is the ongoing work with development of a Critical Infrastructure Security and Assessment Laboratory (CISAL). The introduction of laboratory activities into the curriculum is identified a major challenge [19].

Perhaps the greatest promise to the value of this approach lies in the potential to adopt modules as supplements to other courses pertaining to information security, risk management, and emergency preparedness. Recognition of the multi-disciplinary nature of SCADA cybersecurity is a key to advancement.

### 3.2 Hands-On Toolkit/Laboratory Testbed Approach

Experiential methods of teaching are widely known to be appropriate and effective. Such methods are perhaps most effective with learners who learn best through application of knowledge and skills. No SCADA cybersecurity instructional approach exhibits more variety than those focused on the provision of hands-on laboratories and laboratory exercises. Numerous approaches have been explored and reported, some of which are included here as exemplars.

Guillermo [20] discusses the design and implementation of a Critical Infrastructure Security and Assessment Laboratory (CISAL) as an approach to augment the National SCADA Test Bed (NSTB) at the Idaho National Laboratory. The intent behind establishment of the CISAL facility was to "... simulate research and education in the STEM disciplines by providing a facility that is openly accessible to the academic community." (p. 74). The value of such openly accessible facilities to SCADA cybersecurity education rests in the quantity and quality of the laboratory exercises that are supported and effectively linked to an overarching curriculum. The challenge, according to Guillermo, is in the "continual development" and revision of such laboratory activities and "introduction of novel practices that will leverage the availability of state-of-the-art equipment and system tools" (p. 76).

Later contributions by Guillermo [21] included the specification of SCADA security toolkits as a cost-effective means of equipping educators with the SCADA components essential to supporting hands-on components of SCADA cybersecurity education. Security instructional modules based on these tool-kits are also presented to reinforce the concepts of "wireless communication, information security, control protocols such as Modbus/TCP and DNP3, HMI design and implementation, automation programming and circuit design" (p. 270).

In attempting to "bridge the cultural thinking gap" between control system engineers (responsibly for designing and maintaining critical infrastructure) and information technology professionals (responsible for protecting systems these systems from cyber attacks), Foo [22] suggest a postgraduate curriculum aimed at providing theoretical and practical exercises to raise awareness and preparedness of both groups. A key component of this curriculum is the availability of a number of SCADA system simulators (e.g., Water Reservoir, Smart Meter). Of particular import is the implementation of the simulators from SCADA components commonly used in the field and the utilization of virtual machines to simulate networking components and RTU's. Similarity between the equipment used in the laboratory with that used in the field can favorably influence instructional effectiveness. Simulators can be more affordable and therefore more accessible in educational contexts.

Instructional methods are often limited by available resources. Virtualized environments show promise in bridging gaps (e.g., financial). This author would be remiss without referring to at least one example of an approach that exemplifies this concern directly. Sahin [23], alludes to the financial challenges associated with availing students laboratory experiments based on industrial components. The use of LabVIEW software to virtualize a SCADA environment and instrumentation and the positive effect on student learning is reported. Favorable outcomes suggest that, even with limited resources, SCADA cybersecurity education is attainable.

#### 3.2.1 Web-based Virtual Hands-On Laboratory Testbed Approach

The value and benefits of remotely accessible web-based laboratory resources to increased instructional effectiveness in online, blended, and hybrid delivery modalities have been widely reported, as has recognition of the trend to share such resource between institutions [24, 25]. Such approaches do have limitations (e.g., the requirement for qualified staff members who can effectively configure and maintain a variety of configuration profiles in support of differing research and instructional initiatives) and these should be carefully considered. Much work remains to determine the best approach to making such resources more suitable to widespread adoption.

### 3.3 Case-Study/Group Work Approach

Rowe [8] identifies several advantages associated with the adoption of a case-study instructional approach including availing students opportunities to discuss and develop deeper insight into the "motives, targets, threats, risk, and incident response in the real world." (p. 118). By engaging in critical analysis of the effectiveness of current practices and formulation of recommendations that will improve effectiveness, students are engaged at higher levels of thought and reasoning.

Collectively, the above instructional approaches to SCADA cybersecurity education exemplify the variability that exists among institutions as they continue to explore and seek out those that are most effective. Instructional staff are encouraged to explore these, as well as other novel approaches for effectiveness in their own setting. Thus, the concerns over cybersecurity risks associated with SCADA, may best be addressed through broader engagement in the scholarship of teaching and learning.

## 4 Recommendations

The outcomes of this study illustrate that several instructional approaches have and continue to be explored. Each approach is targeted to addressing the need to educate students about one or more aspects of SCADA cybersecurity. Each emphasizes a particular cognitive domain (e.g.,

comprehension, application, evaluation) and delivery modality (e.g., online, face-to-face).

SCADA cybersecurity education can perhaps best be viewed as a microcosm within the larger cybersecurity education domain. As such, recommendations pertaining to the positioning, design, and implementation of SCADA cybersecurity curriculum and instructional approaches should be informed by adoption of demonstrably effective approaches to cybersecurity education.

One particularly promising approach proposed by Rowe [8] provides an adaptable framework named “Prepare, Defend, Act” (p. 113). When viewed as categories, the three elements can be contextualized, according to the authors, through the following questions:

1. “What cyber-threats are there and how can we prepare for, and minimize potential attacks? (Preparing)
2. How to design and maintain secure systems? (Defending)
3. What should be done in the event of a cyber-attack and how can one place attribution? (Acting)” (p. 117)

The approach is readily adaptable and focusable on SCADA cybersecurity education. In a similar vein, the instructional methods identified by Rowe [8], including “hands-on exposure”, “collaboration”, and “case studies” have also been demonstrated to be effective as discussed above. All approaches should be designed and implemented to support the broadest range of delivery modalities.

The continuing efforts to refine professional and curriculum standards and engage more deeply in the scholarship of teaching and learning will undoubtedly lead to more effective instructional approaches. These in turn, will be more widely adopted with increased consistency of application and overall effectiveness.

Ultimately, each educational institution must take into account its own unique mission, educational programs, and student learning objectives when deciding which SCADA instructional approach(es) to adopt, modify and implement. This process should be informed through the findings reported by those educators engaged in the scholarship of teaching and learning of SCADA cybersecurity. Greater engagement will speed discovery of the most effective SCADA cybersecurity instructional approaches.

## 5 Conclusion

The concerns over risks associated with SCADA cybersecurity are clearly warranted as vulnerabilities continue to be identified. Preparation of a knowledgeable and skilled

cybersecurity workforce is one of the great challenges facing education institutions today. Factors including evolving professional and curriculum standards have contributed to inconsistent application of effective instructional strategies. Instructional approaches continue to be explored with each targeting specific cognitive domains and delivery formats of SCADA cybersecurity. Each institution must take into account its own mission, educational programs, and student learning outcomes when deciding which SCADA instructional approach(es) to utilize. Simultaneously, educators are urged to engage more deeply in the scholarship of teaching and learning of SCADA cybersecurity. Through a collective approach, we may best prepare a future workforce sensitive to such concerns and able to effectively address them.

## 6 References

- [1] President’s Commission on Critical Infrastructure Protection (1997), “Critical Foundations-Protecting America’s Infrastructures”.
- [2] J. Meserve (2007). “Mouse Click Could Plunge City into Darkness, Experts Say”, CNN, Sept. 27, 2007.
- [3] United States Government Accountability Office (2008). Critical Infrastructure Protection DHS Needs to Fully Address Lessons Learned from Its First Cyber Storm Exercise”.
- [4] G. Francia III (2012). “Cyberattacks on SCADA Systems”, Proceedings of the 16<sup>th</sup> Colloquium for Information Systems Security Education, pp. 9-14.
- [5] R. Smith (2014). “U.S. Risks National Blackout From Small-Scale Attack”, Wall Street Journal, March 12, 2014.
- [6] P. Behr (2014). “White House Official Questions FERC’s ‘Next to Impossible’ Grid Assault Scenario”. EnergyWire, April 11, 2014.
- [7] E. Mills (2011). “SCADA Hack Talk Cancelled After U.S., Siemens Request”, CNET News, May 18, 2011.
- [8] D. Rowe, B. Lunt & J. Ekstrom (2011). “The Role of Cyber-Security in Information Technology Education”, Proceedings of the Special Interest Group on Information Technology Education (SIGITE), October 2011, pp. 113-121.
- [9] I. Senk, G. Ostojic, V. Iovanovic, L. Tarian & S. Stankovski (2013). “Experiences in Developing Labs for a Supervisory Control and Data Acquisition Course for Undergraduate Mechanical Education”, Computer Applications in Engineering Education, Wiley Periodicals Inc.

- [10] M. Hentea, & H. Dhillon (2007). "New Competencies for Control Engineers to Meet the Market Demands in Control Systems", Proceedings of the International Conference on Engineering and Education.
- [11] S. Papa, W. Casper \* S. Nair (2011). "Availability Based Risk Analysis for SCADA Embedded Computer Systems", Proceedings of the International Conference on Security and Management (SAM'11), pp. 541-547.
- [12] Joint Task Force on Computing Curricula: Association for Computing Machinery (ACM)/ IEEE Computer Society (2013). "Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science".
- [13] National Security Agency/Department of Homeland Security (2014). "National Centers of Academic Excellence in IA Education: Criteria for Measurement". Retrieved from [http://www.nsa.gov/ia/academic\\_outreach/nat\\_cae/cae\\_ia\\_e\\_program\\_criteria.shtml](http://www.nsa.gov/ia/academic_outreach/nat_cae/cae_ia_e_program_criteria.shtml)
- [14] Association for Computing Machinery (ACM)/ IEEE Computer Society (2008). "Information Technology 2008: Curriculum Guidelines for Undergraduate Degree Programs in Information Technology".
- [15] P. Auerswald, L. M. Branscomb, S. Shirk, M. Kleeman, T. M. Porte & R. N. Ellis (2008). "Critical Infrastructure and Control Systems Security Curriculum", Department of Homeland Security, version 1.0, Washington, DC.
- [16] E. Boyer (1990). "Scholarship Reconsidered: Priorities of the Professoriate", The Carnegie Foundation for the Advancement of Teaching. John Wiley & Sons, New York, NY.
- [17] L. Rosen (2014). "Scholarship of Teaching and Learning (SoTL) Resources", Office of Faculty & Organizational Development; Office of the Provost, Michigan State University (2014).
- [18] J. Sheen, D. Rowe & R. Helps (2012). "Large Scale, Real-Time Systems Security Analysis in Higher Education", Proceedings of the Annual Conference of the American Society for Engineering Education.
- [19] G. Francia III (2011). "Critical Infrastructure Security Curriculum Modules", Proceedings of the 2011 Information Security Curriculum Development Conference, pp. 54-58. ACM, New York, NY, USA.
- [20] G. Francia III, N. Bekhouche & T. Marbut (2011). "Design and Implementation of a Critical Infrastructure Security and Assessment Laboratory", Proceedings of the International Conference on Security and Management (SAM'11), pp. 72-76.
- [21] G. Francia III, N. Bekhouche, T. Marbut & C. Neuman (2012). "Portable SCADA Security Toolkits", International Journal of Information & Network Security (IJINS), Vol. 1, No. 4, October 2012, pp. 265-274.
- [22] E. Foo, M. Branagan & T. Morris (2013). "A Proposed Australian Industrial Control System Security Curriculum", 2013 46<sup>th</sup> Hawaii International Conference on System Sciences, pp. 1754-1762., 2013 46th Hawaii International Conference on System Sciences (HICSS).
- [23] S. Sahin, M. Olmez & Y. Isler (2010). "Microcontroller-Based Experimental Setup and Experiments for SCADA Education", IEEE Transactions on Education, Vol. 53, No. 3, pp. 437-444.
- [24] Z. Aydogmus & O. Aydogmus (2009). "A Web-Based Remote Access Laboratory Using SCADA", IEEE Transactions on Education, Vol. 52, No. 1, pp. 126-132.
- [25] S. Seiler (2013). "Current Trends in Remote and Virtual Lab Engineering. Where are we in 2013?", International Journal of Online Engineering, pp. 12-16.

# Audio Steganography Using Stereo Wav Channels

Douglas C. Farmer, Daryl Johnson  
B. Thomas Golisano College of Computing & Information Sciences  
Rochester Institute of Technology, Rochester, NY  
{dcf2929,daryl.johnson}@rit.edu

## ABSTRACT

This paper presents a new and novel method of audio steganography that allows for the encoding of a textual data within the audio channels of a wav file. This method differs from the more typical forms of audio steganography as it involves the modeling of existing audio channels in order to build a dictionary and then the addition of one or more channels containing encoded data in a form closely related to that of the carrier's wav forms. This method presents a fairly robust, high-bandwidth channel through which to communicate.

**Keywords:** audio steganography, channel, covert channel, covert communications, steganography, WAV, WAVE

## 1. INTRODUCTION

Steganography is the art or practice of encoding some data, be it textual, visual, or of some other form, inside another medium not specifically designed to carry such information. Audio Steganography then is the encoding of such information inside audio data. Such data hiding may take many forms. Common practices include encryption, which attempts to conceal the secret itself using cryptography, generally in the form of some mathematical algorithm to encode the data to a not easily readable form. Steganography however, is different in that the goal is to conceal that the fact that there ever was a communication taking place, by hiding said communication within another expected form. The precipitous rise, and popularity of digital media in recent times provides many convenient new avenues in which to employ these practices.

## 2. RELATED WORK

Steganography is commonly used to transmit data across pre-existing communication channels. Over the years devices have become increasingly interconnected. This Internet of Things (IoT), as it has become known, has been constructed out of numerous protocols, domains, and applications to facilitate the communication of the various devices with each other and the internet as a whole. Steganography then as the

art of hiding data inside other mediums provides a natural avenue through which a person can pursue secure communications over the channels created by the IoT. By encoding data into one of the many communication streams that make up the IoT it is possible to transmit data without it ever being apparent that an aberrant communication ever took place.

There are several popular variants of Audio Steganography. These all address separate issues in dealing with the encoding of data in an auditory stream and have their own respective strengths and drawbacks. Some such encoding schemes include least significant bit encoding, phase encoding, and echo encoding.

### 2.1 Least Significant Bit Encoding

Least Significant Bit Encoding is by far the most common type of audio steganography in use. In this scheme the least significant bit of each audio frame is modified to encode binary information [3]. All data in WAV files is stored in 8-bit bytes arranged in little endian format with the low-order (i.e. least significant) bytes first for multi-byte values. Changing this byte usually doesn't result in noticeable changes to the resulting waveform. This scheme is rather simple however, and unfortunately easily defeated. Techniques such as random bit shifting can interfere with the extraction process, as can normal audio operations like compression and file conversion. Additionally due to the way binary information is represented its statistical shape is easily detectable when embedded in other media.

### 2.2 Phase Encoding

Phase encoding is much more difficult to detect. Phase Encoding involves breaking down audio into chunks, separated by phase groups and then shifting it based on the binary data to be encoded. This technique can still be susceptible to random bit shifts, but is much harder to detect. The main drawback of Phase encoding lies in the low bandwidth, which is a direct result of encoding using the audio phase. Simply put there isn't a great deal of information that can be sent at one time.

### 2.3 Echo Encoding

Echo encoding is another common method for audio steganography. As the name implies this scheme involves the insertion of 'echoes' into an audio signal. The echo is varied along three parameters, initial amplitude, decay rate, and offset/delay. By using a short delay it is possible to hide data using this

technique without noticeable effect on the resulting waveform. Likewise amplitude and decay rate can be set to values below the audible range of the human ear. This makes this encoding scheme incredibly hard to detect. There is however a chance that some mix of echoes can combine to produce to a noticeable effect.

### 3. BACKGROUND

#### 3.1 WAV Background

In order to understand this covert channel it is necessary to provide some further background. This channel deals primarily with the Waveform Audio File Format commonly referred to as WAV due to its file extension. WAV is an audio file format standard for the storing of audio bit-streams. This file format is the general raw, uncompressed format digital audio takes during recording and processing before being compressed into the various lossy and lossless streams developed for general distribution. WAV is an application of the RIFF (Resource Interchange File Format) bit-stream format for storing data chunks and thus closely follows the RIFF file format. WAV files support any number of bit resolutions, sample rates and channels of audio[4] which will be leveraged in the creation of the proposed covert channel. They are a collection of different types of 'chunks' each containing important information. Because RIFF is a tagged format, interpreters of RIFF and WAV files are designed to interpret only tags they understand and ignore the rest. All WAV file interpreters though are expected to understand and read the two required chunks Format and Data. Figure 1 is an illustration of a single WAV file containing the two required chunks.

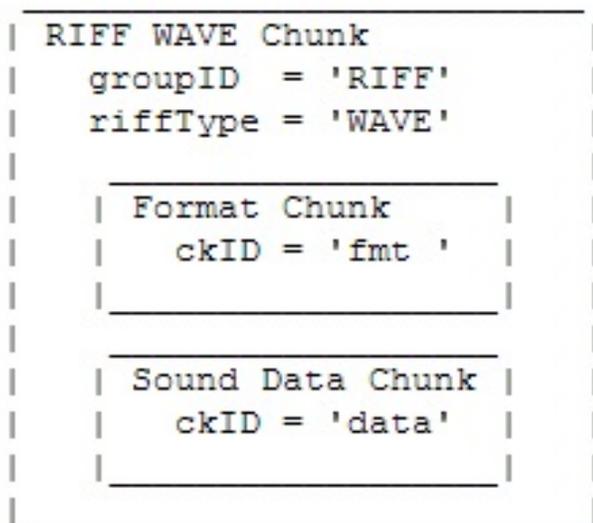


Figure 1: WAV File Require Chunks

The Format chunk describes the characteristics of the waveform data. These include fields such as sample rate, bit resolution, and number of channels. This chunk always has an id of 'fmt'. Analyzing this chunk is incredibly important for the robustness of the channel but will remain largely unmodified.

The Data chunk contains the actual sample frames. Sample frames contain all the waveform data for all channels of the audio. This format chunk while generally the largest chunk in terms of actual bytes, contains only three fields, an ID that is always 'data', a field denoting the number of bytes contained in the waveform data, and the waveform data itself. This data is arranged by sample frame. Every channels first frame followed by every channels second frame and so on and so forth. For a stereo track containing 5 frames, a typical waveform data array might appear as L1R1 L2R2 L3R3 L4R4 L5R5 where L and R denote the Left and Right audio channels respectively, and the number the frame. As mentioned before all data in WAV file is stored as 8-bit bytes in little endian format as pictured in Figure 2. It is also important to note that a properly formatted WAV file should only ever contain one data chunk.

#### 3.2 Audio Channels

Most of the previously outlined encoding techniques for audio steganography involve modifying the pre-existing waveform to add significance to data that wouldn't otherwise carry meaning. This method is different in that it doesn't actually involve the changing of the pre-existing data but in fact adding to it. Audio files may contain many different audio channels. These channels act as a storage device for multi-track recording and playback. Monaural sound or mono refers to a single channel where as stereophonic sound or stereo refers to more than one channel. Most commonly stereo is two channels but there is no real limit on the number of channels that can comprise a stereo sound. 5.1 and 7.1 Surround Sound both refer to stereo audio with 5 and 7 full range channels respectively and .1 to reflect the limited range of the Low Frequency Channel (e.g. bass). This is useful in the creation of a covert channel using audio steganography because of the way these channels are interpreted by audio playback devices.

Mono systems can have multiple speakers but because there is only one audio signal it is simply replicated over each playback device (eg. speakers). Stereo on the other hand may contain level and time/phase information to simulate direction cues or in other cases contain explicitly different audio for each channel. In stereo systems each channel is mapped to a different speaker. The increasing number of channels defines more and more precise sound targeting [5]. One of two things generally happen when a stereo sound with more channels than a system has capabilities to decode is played. Either the excess channels are ignored completely or they are sent to the closest matching speakers. For example on a 5.1 stereo system 7.1 audio would have its extra two channels which correspond to surround back left and surround back right mapped to the closest matching speaker. In this example that would be surround left and surround right. These speakers would then play two channels each, their originally mapped one as well as the additional channels. Alternatively, a given system might choose to ignore these extra channels altogether rather than deal with the overhead of mapping additional channels. Again this behavior is undefined and varies from system to system.

### 4. DESIGN OF CHANNEL

#### 4.1 Encoding Method

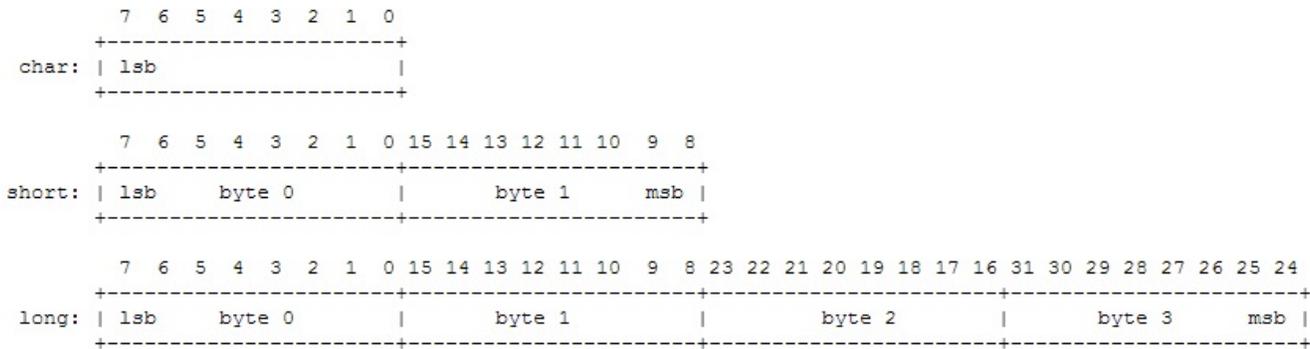


Figure 2: WAV File Byte Organization

The proposed covert channel aims to take advantage of stereo audio processing by adding channels to an original audio file and then encoding data in it. This process has several steps. Before beginning to write data, the file must first be read and parsed for relevant information from the Format chunk. Specifically, the frame rate, number of frames, frame width, and number of channels is required. Because the data chunk is read as a byte string it must be parsed in order to be useful. Thus the frame width and number of channels is needed. The frame width tells how many bytes constitute a discrete sample while the number of frames combined with the number of channels denotes how many samples each frame has. This information is important when writing the file later.

Using the python module numpy the byte string can be converted to an array of integers of the proper size as denoted by the sample width. Even further it can be reshaped to an array of arrays in which each inner array represents a frame in the original audio file and each index in an array is a sample from a channel of audio. Remember that WAV data is represented by channel by frame in the data chunk.

In the simplest form an ASCII message can be obtained and converted to an array of its decimal representation and written directly into a new audio channel. It is important to ensure that the new channel is the same number of frames as the other channels, and so the new channel must be padded. The pad value makes no difference, however using zero's results in no white noise. When writing the file, one denotes the frame rate, new number of channels, and the sample width. The number of frames will be updated after writing is complete.

## 4.2 Decoding Method

Decoding follows the same steps as reading and parsing an audio file for writing. The format chunk is read and parsed. Frame rate is unimportant for decoding but the number of channels, sample width, and number of frames are again key. Numpy is used to covert the binary string to a useful format. Now each sample in the last channel can be iterated over and used in conjunction with the python function 'chr' to convert and append the binary representation to its ASCII equivalent.

## 5. RESULTS

The proposed channel was implemented in two Python3.3 programs titled `encode_wav.py` and `decode_wav.py`. As previously mentioned numpy was used extensively as a more convenient way to work with formatted byte data. Additionally the built-in python Wave and Struct modules were used to read and pack WAV files for writing respectively. The last module was the Matplotlib module Pylab, which was used to generate graphical representations of the sample distribution. Sample widths for which there is a corresponding integer type (8, 16, 32, 64 bits) are trivial to implement, however, only 16-bit audio has been tested. 24-bit WAV files are quite common and can be written with use of the Struct module, but as of now they are beyond the scope of this implementation.

Figure 3 and Figure 4 are graphical representations of two test WAV file, the former mono, and the latter stereo. These plots were normalized from -1 to 1. Note the full spectrum spread. Both the mono channel and the various stereo channels in Figures 3 and 4 are well distributed.

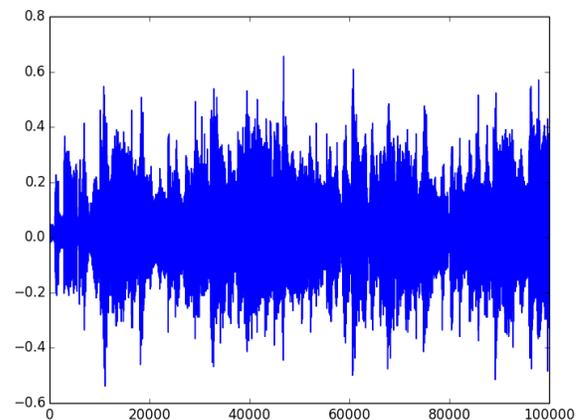


Figure 3: Mono Channel Non-Encoded Woman.wav Plot

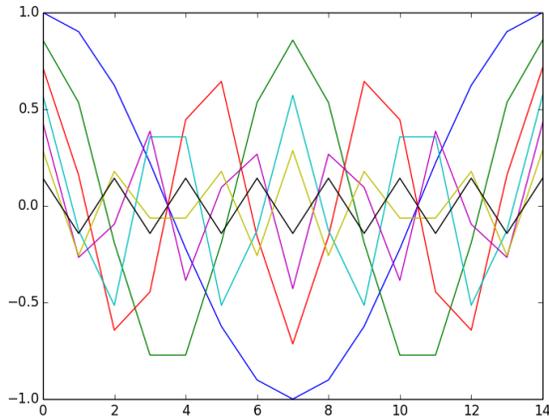


Figure 4: 7 Channel Non-Encoded Plot

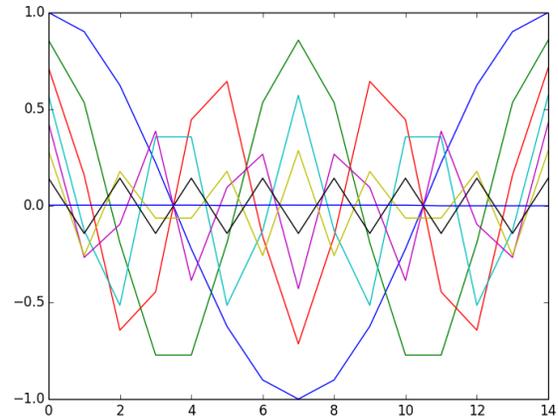


Figure 6: 8 Channel Encoded Plot

Several individuals were asked to listen to test audio files both before and after the encoding of data. When adding channels to mono (1 channel) audio there was a discernable change in the audio but no real degradation. Adding another channel causes the original channel to be played as a front left audio channel while next to nothing was played on the new channel containing the encoded data. Stereo tracks faired much better as there was no discernible difference in the audio.

Figure 5 and Figure 6 are both stereo channels that have had a message encoded in them. The waveforms are unchanged but if one notes the almost horizontal line around 0 one can see where a message has been encoded. Because the data has been encoded using 0 to 255 in a field that ranges to 32,767 there is barely any appreciable slope.

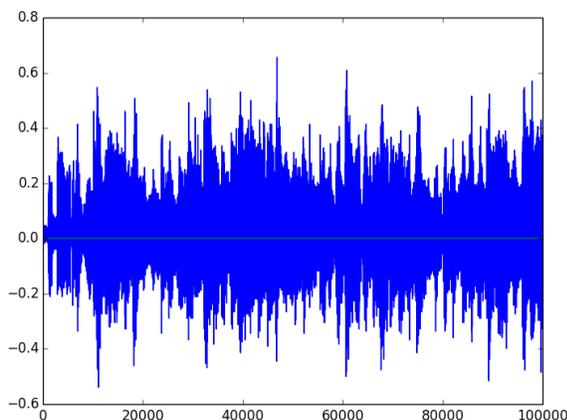


Figure 5: 2 Channel Woman.wav Plot

Normalizing these values based on the existing WAV data greatly increase the covertness of this channel. Each sample frame contains a sample from each channel at that moment in time and thus this implementation is able to build a list of average values for each frame in the WAV file. This list is then used to perform a ROT13 using the corresponding value in the list as the amount to rotate. ROT13 is an example of a Caesar cipher or simple substitution cipher. Applying a ROT13 to a piece of text requires analyzing a piece of text for character and replacing each one with the character 13 spaces farther along in the alphabet.

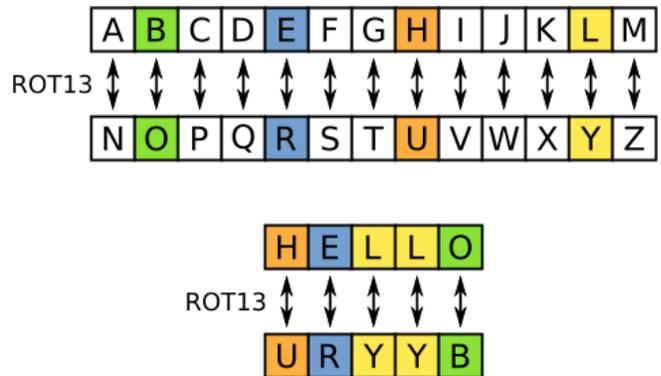


Figure 7: Example of ROT13 on Text

Here, instead of character, each decimal ASCII values is replaced with the number, an average number of places way, as based on the list of averages for that frame. Applying this cipher in reverse decrypts the message.

A covert channel can be characterized by several criterion. A channels throughput describes the amount of data that can be transferred over it within a given amount of time. Robustness refers to a channels ability to withstand errors during execution as well as to withstand change without alterations to the initial configuration. Finally prevention and detection characterize the ease with which a channel can be detected and prevented. [1]

## 5.1 Throughput

In a 9 second clip of audio, a 16-bit mono WAV file contained 100,000 frames. Upon encoding each frame using the 16-bit representation contains a single signed short with value ranging from -32,768 to 32,767. Even using only the extended ASCII range from 0 to 255 and a single frame per character the channel was able to encode 100,000 characters in 2.044 seconds. This implies a theoretical limit of 48,923 characters per second. This calculation doesn't account for the fact that the encoded message was almost entirely comprised of padding zero's, however longer messages would not take significant time to convert to decimal representations and would also decrease the amount of time spent padding. Normalizing the data first would also increase the length of time it took to encode, with each additional channel causing the entire process to take that much longer. At each sample frame, all the samples from every channel have to be averaged to find the amount the bit to be encoded should be shifted. Decoding this message took 1.663 seconds.

## 5.2 Robustness

This channel as presently constituted is robust to an acceptable degree. Most regular file operations will not affect the payload, nor will altering the files meta data. Here the data is hidden in a way that is completely uncommon for audio steganography by adding an entire channel. This also means that the data is interspersed throughout the entire waveform if at regular intervals. There has been little work done to obfuscate or encrypt the data before encoding. It remains to be seen whether compression or conversion to other formats will alter or destroy the encoded data.

## 5.3 Prevention and Detection

This channel is hard to prevent because it is incredibly hard to detect. Most detection schemes rely upon analysis of the WAV data looking for the statistical signature of binary data. This channel won't be detected that way because each ASCII value is being represented as a 16-bit integer rather than directly as bits or event bytes. The statistical shape of the data won't give it away as text. Additionally other prevention techniques such as random bit shifting add noticeable noise to an audio stream and are thus rarely if ever used. While "a signal-to-noise ratio (SNR) above 20dB guarantees for a reasonable audio quality"[2] there is future work that would effectively eliminate this as a viable prevention method. Finally, there is the mean normalization. Because we are making the channel an approximation of the others, there is no audible difference. Plotting the channels will also show the data looks similar to the rest as well at each frame.

## 6. FUTURE WORK

This channel has numerous areas where future work could expand on and improve this channel. The most obvious area of work is to make this channel work for more than just ASCII data. The ability to encode binary data would greatly increase the utility of this channel. Additionally, some work into obfuscating or encrypting the data before encoding would make it even less likely to be detected. There is currently only a 16-bit implementation and making this work for additional bit widths increases the variety of carrier files. Like most of the other encoding schemes mentioned in related work this channel is vulnerable to random bit shifting. While

this isn't guaranteed to destroy a message there is a great method available to mitigate this attack that there simply wasn't enough time to implement. Using mean normalizing the current implementation is able to make the encoded data look that much more like the waveforms of the rest of the audio but also results in audio that plays close enough to the original to blend in. Adding to this, one could also map each ASCII value to a range of values based on the sample frame that way even random bit shifting wouldn't alter the message[2]. Lastly, currently this implementation is limited to a single channel for encoding of information. It would be possible using some sort of formatting or delimiting character to use multiple channels to encode data. This greatly expands the amount of data that can be transmitted at one time.

## 7. CONCLUSION

Steganography is by no means a new field and while its use in audio may be a somewhat recent event there are still several well-researched and established schemes in existence for encoding data. This paper presents a new, novel approach to encoding data within WAV audio files using stereo audio channels. The method proposed is by definition Steganography, but used in conjunction with other channels it presents a very unique and robust means of securely transmitting data.

One such example would be using this method in a plug-in for a VOIP client, wherein one could transfer covert data packets, while holding what would appear to be just a regular conversation. Message boards, and services such as Sound Cloud, and Band Camp would also provide conditions favorable to the use of this method. These services would be in many ways even better as they allow for anonymous posting with nothing more than a user name linked to an email account. Those factors combined would allow one to use the services as blind drops for data, with both sender and receiver of data rendered virtually anonymous.

This method has been shown through both experimentation and research to effectively transmit data unbeknownst to end users. The data is virtually undetectable without doing a thorough cryptographic analysis of the resulting raw audio. This does indeed have weaknesses as do all covert channels but there should be great optimism that proposed future work could go along way to fixing those weaknesses.

## 8. REFERENCES

- [1] E. Brown, B. Yuan, D. Johnson, and P. Lutz. "Covert Channels in the HTTP Network Protocol: Channel Characterization and Detecting Man-in-the-Middle Attacks". 5th International Conference on Information-Warfare & Security, (Ohio, NY, USA, 2010).
- [2] M. Nutzinger, "Real-time Attacks on Audio Steganography". Journal of Information Hiding and Multimedia Signal Processing. Vol. 3, no. 1, January 2012. (Taiwan)
- [3] Gunjan Nehru, Puja Dhar, "A Detailed look of Audio Steganography Techniques using LSB and Genetic Algorithm Approach". IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 2, January 2012. (Republic of Mauritius)
- [4] L. Silvestro and G. Baribault, "Waveform Audio File

Format MIME Sub-type Registration.” [Online].

Available:

<http://tools.ietf.org/html/draft-ema-vpim-wav-00>

- [5] “Multiple Channel Audio Data and WAVE Files.” [Online]. Available: <http://msdn.microsoft.com/en-us/library/windows/hardware/gg463006.aspx>.

- [6] “WebP Container Specification – WebP – Google Developers.” [Online]. Available: [https://developers.google.com/speed/webp/docs/riff\\_container?csw=1](https://developers.google.com/speed/webp/docs/riff_container?csw=1)

# Potential Vulnerabilities of the NextGen Air Traffic Control System

C. Giannatto<sup>1</sup> and G. Markowsky<sup>1</sup>

<sup>1</sup>School of Computing & Information Science, University of Maine, Orono, Maine, USA

**Abstract**—*The FAA is well on its way to replacing the current air traffic control surveillance system with a new system known as Automatic Dependent Surveillance-Broadcast (ADS-B). As with many projects, the focus is on performance and getting the system operational, with security having secondary importance. This paper describes some of the vulnerabilities of the current proposed implementation of ADS-B and offers some suggestions on mitigating these vulnerabilities.*

**Keywords:** ADS-B, ATC, Air Traffic Control, Vulnerabilities, Safety, Air Travel

## 1. Introduction

By the late 1930s commercial air travel was starting to become a popular mode of transportation and the volume of air traffic increased dramatically. As it became more difficult to keep track of the increasing number of aircraft in operation, the airlines developed a system of radio stations to help monitor their en route air traffic. These initial radio stations were located in Chicago, Newark and Cleveland and were the precursor to our current air traffic control system. The Bureau of Air Commerce acquired the radio stations in 1936 and in so doing formed what is considered the First Generation of ATC [1, p. 4].

This First Generation ATC system consisted of no automation and very little radar coverage. The fledgling ATC system relied on manual methods of tracking aircraft using progress strips for each flight. By the late 1950s the volume of aircraft in operation had increased to the point that manual tracking was no longer feasible. In 1959 the Second Generation ATC system was introduced, which automated many of the flight monitoring tasks through the use of computers for processing air traffic data and ground based radar to help track individual aircraft. Two years later, another major improvement to the ATC system was made when the FAA incorporated ground based equipment to interrogate a transponder located on the aircraft, allowing each air traffic radar target to be uniquely identified [1, pp. 4-5].

In the late 1960s, air traffic was again taxing the capabilities of the National Airspace System (NAS). By the early 1970s, advances in computer technology made it possible for Upgraded Third Generation development (UG3d) of the ATC system. UG3d provided a substantial improvement in both the terminal and en route air traffic control structures [1, p. 5]. Through the increased automation of controller tasks and the ability to receive timely flight tracking information,

UG3d enabled air traffic controllers to safely accommodate and monitor the increasing volume of air traffic.

## 2. Air Traffic Control Today

With the exception of the Global Positioning System (GPS) technologies introduced in the late 1990s, the current NAS infrastructure has undergone few changes since the improvements incorporated into UG3d. Currently the NAS consists of a vast number of facilities including 750 ATC installations, over 18,000 airports and more than 4,500 air navigation stations [2, pp. 1-9, 1-10]. The 750 ATC facilities are comprised of 21 Air Route Traffic Control Centers (ARTCCs), 197 Terminal Radar Approach Control (TRACON) facilities and more than 450 airport control towers [2, pp.1-9,1-10]. We note that there is a newer edition of the Instrument Procedures Handbook [3], but it does not contain much of this background information about the air traffic control system.

ARTCCs are responsible for controlling en route traffic within designated control sectors, with the majority of the en route traffic traveling along designated airways at and above 18,000 feet. TRACON facilities control aircraft within a 30 nautical mile radius of the larger airports within the ATC system, while airport control towers are responsible for controlling aircraft within a 5 nautical mile radius of the airport. [2, pp.1-9, 1-10].

Current NAS aircraft position tracking (surveillance) techniques fall into three basic categories; Procedural ATC, Primary Surveillance Radar (PSR) and Secondary Surveillance Radar (SSR) [2, pp. 3-6, 3-8, 3-17]. Procedural ATC is known as a dependent surveillance technique, which means it depends on input from individual aircraft. With Procedural ATC, pilots are required to periodically report their position using radio communications, and it is predominately used for oceanic and remote area flight operations where there is little or no radar coverage. PSR is an independent and non-cooperative surveillance radar system typically used by TRACON facilities and in busy terminal areas which does not depend on any input from the aircraft. SSR is a partially-independent and cooperative surveillance radar typically used for en route tracking by ARTCCs, and determines aircraft position through a combination of radar target return and aircraft transponder reply when interrogated by a ground station [1, pp. 8-9].

Many of the current ATC facilities have been in service for more than 50 years. These installations, and in particular the

ground-based SSR and PSR radar systems, are very costly to operate and maintain. Increased air traffic, aging equipment and a desire to leverage technological advancements necessitate a comprehensive overhaul to the NAS. In its current form, the air transportation system performs adequately but it is once again approaching its capacity limits. Without a makeover, the expected growth in air traffic will likely create costly flight delays and increased flight safety hazards [1, pp. 6-7].

In response to these concerns, the FAA has begun the overhaul of the current air traffic control system and started working on the Next Generation Air Transportation System (NextGen). The primary goal of NextGen is to significantly increase the safety and capacity of air transportation operations. The upgrade requires a fundamental conversion of the entire NAS, including incorporation of satellite-based technologies for surveillance operations and the shutdown of many legacy ground-based systems currently in use [1, pp. 6-7]. A key component of NextGen is a position reporting and tracking technology called Automatic Dependent Surveillance - Broadcast.

### 3. ADS-B

Automatic Dependent Surveillance - Broadcast (ADS-B) is a core feature of NextGen. ADS-B is a satellite-based surveillance technology that also uses aircraft avionics and ground-based systems to provide information on aircraft location to pilots and air traffic controllers [4, pp. 1-5]. The ADS-B system is "automatic" in that it requires no pilot or controller intervention. It is "dependent surveillance" because the aircraft provides input to the air traffic control system based on information derived from the aircraft's GPS receiver, which will allow for much greater position accuracy than the current radar-based system. [1, pp. 9-10].

ADS-B has the potential to improve safety through enhanced pilot and controller situational awareness, better inflight collision and runway incursion avoidance, and the ability to implement accurate ATC surveillance in remote areas with no current radar coverage. Better position monitoring accuracy should allow the air traffic control system to handle a higher volume of aircraft through condensed aircraft separation standards, more direct traffic routings and optimized departures and approach procedures. Another potential benefit of the NextGen ADS-B infrastructure is reducing overall air traffic control system maintenance and operating costs, since the new system is comprised of simple UHF radio stations that are significantly cheaper to install and maintain than the aging surveillance radar ground stations [1, p. 8].

The FAA's NextGen implementation plan includes a network of approximately 800 ADS-B ground stations, placed 150 to 200 miles apart [5, pp. 72-73]. These stations will receive signals on two designated UHF frequencies; 1090 MHz and 978 MHz. Commercial and military aviation traffic

flying in the high-altitude airways structure (at and above 18,000 feet) will utilize the 1090 MHz ADS-B frequency, while general aviation aircraft flying at lower altitudes will use 978 MHz. To facilitate interoperability between aircraft using different frequencies, the system incorporates a support component called Automatic Dependent Surveillance-Rebroadcast (ADS-R). ADS-R receives the traffic information broadcasts on the 1090MHz or 978 MHz links and rebroadcasts the information to aircraft on the opposite data link frequency [5, pp. 72-73].

There are many prospective benefits of NextGen and ADS-B, but there are also a myriad of impending problems that need to be addressed. The ADS-B portion of NextGen is scheduled to be fully deployed by January 1, 2020 and the FAA faces many potential problems as the components of NextGen are put into operation. ADS-B presents challenges to aviation on a number of levels including questions about the true costs of implementing the system and concerns over vulnerabilities to exploitation that do not exist in the current air traffic control system. In the following sections we will examine some of the implementation concerns and potential security risks that this new system imposes.

### 4. Scenarios and Concerns

In this section we will begin by summarizing some of the scenarios that have already been published and analyze them.

#### 4.1 Challenges to Implementing NextGen

The successful deployment of ADS-B throughout the NAS faces significant risks and challenges [4, p. 2]. NextGen is facing cost projection overruns, resistance from aircraft owners and operators, miscalculations in the true benefits that can be realized, and security vulnerabilities. In short, there are areas of the project that have not yet thoroughly vetted.

One of the greatest risks to the successful implementation of ADS-B is aircraft operator reluctance to purchase and install new avionics for their aircraft. This situation is compounded by the FAA's inability to accurately define requirements for the system's more advanced capabilities. Operators have raised justifiable concerns about changing requirements and uncertain equipage costs and benefits. The FAA has yet to fully define requirements for modifying its existing automation systems, and the specifics of how ADS-B information will be integrated into the existing system. Until the FAA effectively addresses these uncertainties, progress with ADS-B will be limited and concern over cost increases, delays, and performance shortfalls will remain [4, p. 2].

NextGen and ADS-B could also be facing significant cost overruns due to peculiarities in the program's contract specification. Specifically, the FAA decided on a service-based contract for the program instead of the traditional

method of owning and operating the system. FAA officials acknowledge that the analysis used to justify the service-based approach and cost savings was flawed but asserted that over the long term, the cost-benefit equation changes in favor of a contractor owning and operating the system. In spite of this claim, the FAA has not updated its cost and benefit analysis to support the service-based approach. This puts the program at risk of realizing minimal return on the FAA's investment and possible delays in achieving NextGen goals [4, pp. 2-3].

Another implementation issue the FAA needs to address is that of ADS-B frequency saturation in congested airspace. Since the ADS-B signal is broadcast, a large number of broadcasting aircraft could overwhelm the system in dense traffic areas. The frequency congestion problem is complex and solutions to address frequency congestion may require changes to the ADS-B baseline or equipment, which would increase the program cost. Currently, the FAA is examining potential solutions and exploring the specific changes needed for ADS-B air and ground components and existing systems [4, pp. 11-12].

## 4.2 Scenarios from Darryl H. Phillips

The scenarios in this section come from a website put together by Darryl H. Phillips [6] in 2000. We will just summarize some of the scenarios here. For more technical details please see [6].

### 4.2.1 Scenario One: Terrorism

This scenario envisions a lone terrorist coming to the United States and obtaining a light aircraft such as a Cessna 172, Beechcraft Bonanza or Piper Arrow equipped with an ADS-B collision avoidance system. He expects that the ADS-B systems installed on all airliners and some business and pleasure aircraft to automatically report the precise position and identity of the aircraft they are on twice per second. The terrorist waits for a day of poor visibility, and then he flies at low altitude and slow speed above a busy highway toward a major airport. He knows that Air Traffic Control radar will not see him because he has disabled his transponder output, thereby assuring that there will be no secondary returns nor any ADS-B transmissions. His aircraft is smaller than the tractor-trailers on the highway below, and the ATC primary radar has been programmed to eliminate highway clutter from the display. He will not be seen.

The terrorist decides to target a large aircraft on its final approach when it is most vulnerable. Using the ADS-B readout to spot his target, he flies up the glideslope and directly toward the target. Even though he cannot see the airliner because of the bad weather, the ADS-B display lets him know its position within a few meters and he is able to crash his plane into the airliner killing hundreds of people.

In our opinion the success of this scenario depends upon whether or not the airport has an operating Primary

Surveillance Radar system. If an airport has a functioning Primary Surveillance Radar system, they would be aware of the rogue aircraft and its position, even if the rogue aircraft's transponder was turned off. The PSR is sensitive enough to detect the attacker, and ATC would divert any aircraft away from the attacker's flight path. If the full implementation of the FAA's NextGen goes into effect and the Primary and Secondary Surveillance Radar Systems are dismantled, then this is indeed a very credible threat.

### 4.2.2 Scenario Two: Extortion

This scenario envisions a sociopath building a large model airplane and using it to crash into an airliner as a way of extorting money from the airline. Again, the idea is to use the ADS-B information as the basis of a guidance system. Aside from the cost of the model airplane, the cost for the electronics is relatively low. A GPS receiver might cost \$100, a wireless LAN card under \$30, and a 1090 MHz ADS-B receiver can be built from a DBS satellite receiver for under \$200. With some software, one can create a guided missile for a relatively low cost.

The credibility of this scenario is similar to that of the previous scenario, and is unlikely to happen at airports with an operational PSR. In order for the model aircraft to pose a substantial danger to an aircraft approaching the airport, it would need be large enough so that it would likely show up as a target to ATC. At airports without an operable PSR, this scenario is certainly plausible.

### 4.2.3 Scenario Three: Revenge

This scenario envisions a disgruntled employee who feels cheated by the company for which he has produced many worthwhile inventions. He is about to retire and will be receiving only 60% of his current salary, while various executives receive millions of dollars based on the products that he has created. He decides to revenge himself against the company by making the executive jets that belong to the company crash. Since the various signals are not encrypted he is able to learn all of the ID numbers of the company planes. This disgruntled employee then waits for bad weather and spoofs the signals that guide the plane's landing so that it comes down a half-mile short of the runway and crashes. Several months later he causes another company plane to crash.

This scenario is plausible, but is a bit less likely to pose a credible aviation threat. In this scenario, the type of aircraft being targeted would be required to have a radio altimeter (which reads feet above ground level) and would likely have some sort of ground-proximity warning system. Both of these devices would warn the pilots of an impending contact with terrain. In addition, most professional flight crews complete multiple cross-checks on final approach, and would notice a discrepancy between the published approach parameters and what the flight instruments were reading. It

is possible that a complacent flight crew could be caught off guard by such as scheme, but the scenario is less likely to be successful than the previous two scenarios.

#### 4.2.4 Scenario Four: Data Mining

In this scenario, a company is created that tracks corporate planes looking for planes from different companies that go to the same or nearby locations at about the same time. The idea is to get tips about which companies might be negotiating with other companies. This information might have great business value.

This scenario is quite plausible and should be an area of concern. Unless the information is somehow protected, the granularity of the data provided into the air traffic control system by ADS-B will allow for data mining opportunities for all types of information gathering purposes. Individuals, news organizations, paparazzi and foreign intelligence agency will have easy access to data that is either unavailable or that is far more time consuming to aggregate from the current air traffic control system.

### 4.3 Information Gathering Scenario

Due to National Security concerns, certain military and government flights need to be operated "off the radar" and in secrecy. The movement of tactical aircraft such as bombers and fighters, logistics movements of troops and supplies, and flights involving government officials should not be information disseminated to the general public. However, this information can easily be pieced together via data collected from ADS-B transmissions, creating an Operational Security risk to our military and elected representatives. In this scenario, a nation state utilizes the granular information provided by ADS-B transmissions to gather intelligence information on military aviation activities.

An enemy nation state has been concerned over a possible bomber strike on one of their chemical weapons production facilities. They have placed intelligence operatives in the United States, and have been watching for signs of bomber aircraft movement. The enemy operatives know that tactical aircraft such as bombers will be flying with ADS-B Out disabled, but they are aware that the air refueling aircraft used to refuel those bombers inflight fly with ADS-B Out enabled. The intelligence operatives have been monitoring flight activity at two KC-135 tanker bases in New England. The operatives are monitoring the ground control and tower frequencies at these bases, and are logging call signs, takeoff and land times, and gathering flight information on departing and arriving tanker aircraft. The operatives note that two tanker aircraft depart simultaneously, one from each base. Utilizing laptop computers and easily developed software, the agents begin gathering data from the 1090 MHz ADS-B Out transmissions. From this data, they are able to tell what airspeed each aircraft was flying at when it left

the ground. Based on publicly available data on the KC-135 and the current weather conditions, they are able to determine the approximate takeoff weight of each aircraft, and consequently, how much fuel it is carrying.

Utilizing data provided to the air traffic control system by ADS-B, the agents are able to monitor the flight paths of each aircraft and observe that the two aircraft rendezvous inflight and continue north as a formation. The operatives watch the aircraft as they proceed up over northern Canada, and turn toward the east. The aircraft continue on this eastern track for several minutes, and then turn back toward the south. The agents follow the flight of each aircraft as they split up and return to their respective bases. When the aircraft arrive at their bases, the agents are able to obtain the landing speed of each aircraft from the ADS-B Out information. Using the information they have gathered, they are able to calculate approximately how much fuel was offloaded during the flight. They are also able to discern, based on readily available information regarding the observed airspeed of the tankers on their easterly heading over northern Canada, that the tankers were refueling a flight of B-2 Stealth Bombers. After analyzing this information, they pass on the likely impending bomber strike, and give their government several hours lead time to relocate sensitive equipment at the chemical weapons facility.

This scenario is very similar to Scenario Four in the previous section and should be an area of concern. The ability of individuals and state-sponsored groups to gather information directly from the ADS-B transmission and indirectly from the information provided by ADS-B into the air traffic control system poses a substantial Operation Security risk to our military and elected officials traveling aboard government aircraft. The obvious solution is for sensitive flights such as these to operate with ADS-B Out disabled. This, however, poses significant challenges for tracking these aircraft in the FAA's non-radar NextGen air traffic control system.

## 5. Hackers + Airplanes = No Good Can Come of This

The title of this section comes from a talk delivered at DefCon 20 by Brad "RenderMan" Haines [7]. In case anyone thinks that hackers have not noticed the vulnerabilities we have been discussing, this is a false hope. DefCon features talks on hacking a wide variety of systems and it is always worth looking at the materials available at [8]. In addition, many of the talks at DefCon feature videos available on YouTube.

We briefly mention and describe two free, open-source projects that have the potential to be widely used by the hacking community to build systems that can interact with NextGen. The first project is the GNU Radio project. The GNU Radio project provides both a C++ API and a Python

API. Nick Foster has provided a demonstration project showing how GNU Radio can be used to track aircraft. His code can be downloaded from github [9]. The following quotation from the GNU Radio website [10] provides some information about this project.

GNU Radio is a free & open-source software development toolkit that provides signal processing blocks to implement software radios. It can be used with readily-available low-cost external RF hardware to create software-defined radios, or without hardware in a simulation-like environment. It is widely used in hobbyist, academic and commercial environments to support both wireless communications research and real-world radio systems.

The second project is a sophisticated, open-source flight simulator designed for research, pilot training, and entertainment purposes called FlightGear [11]. The software has realistic flight dynamics for a variety of military, commercial and general aviation aircraft. The project has a worldwide airport and scenery database, and provides an excellent platform for demonstrating the target injection vulnerabilities in the ADS-B system.

The following quotation comes from [12] and addresses several of the scenarios discussed in this paper and elsewhere.

The FAA said that the ADS-B system is secure and that fake ADS-B targets will be filtered from controllers' displays. "An FAA ADS-B security action plan identified and mitigated risks and monitors the progress of corrective action," an FAA spokeswoman told AIN.

A spokeswoman for key ADS-B contractor ITT Exelis explained, "The system has received the FAA information security certification and accreditation. The accreditation recognizes that the system has substantial information security features built-in, including features to protect against spoofing attacks. [This] is provided through multiple means of independent validation that a target is where it is reported to be."

The FAA has not provided any details on the testing or accreditation process, and is basically saying trust us!

## 6. GPS Vulnerabilities

The NextGen system will rely heavily on GPS. Because of this it is worthwhile to examine the vulnerabilities of systems based on GPS. Of special interest is a briefing given by James V. Carroll in 2001 [13]. Unfortunately, the vulnerabilities discussed in this presentation continue to be vulnerabilities. Because of space constraints we will just mention some of the highlights from Carroll's presentation.

Carroll makes three very important observations in [13]:

- "GPS users are vulnerable to signal loss or degradation."
- "Awareness and planning can mitigate the worst vulnerabilities."
- "*The vulnerabilities will not be fully eliminated.*"

GPS is vulnerable because it uses a very weak signal on a single civilian frequency. In addition, because of military applications, there is a GPS disruption industry. It is relatively easy to build or buy a GPS disruption device. GPS disruption can take a variety of forms including jamming and spoofing. Jamming prevents a GPS receiver from receiving a valid GPS signal and can lead to unwanted behavior. Spoofing can mislead pilots and control systems and lead to disaster. Carroll stresses the need for back ups to the standard GPS systems. We do not have space to examine the subject of backups for the GPS system, but one system to consider is the Nationwide Differential GPS (NDGPS) Service operated by the Coast Guard [14].

## 7. Normal Accidents

The sociologist Charles Perrow has written a book called *Normal Accidents* [15]. The book introduces the concept of a *system accident* which he also calls a *normal accident*. In a nutshell, the idea is the following. Usually, when an accident happens we want to blame some individual or group of individuals. Yet, Perrow argues that in many cases the true culprit is the system that has been created. In particular, the system is set up so that serious accidents are inevitable, hence the term normal accident, and it is just a matter of dumb luck who the individual is who is left holding the bag when the accident occurs.

A famous example of this is the 1999 loss of the Mars Climate Orbiter satellite because of a communication error. A review of the incident [16] showed that the error resulted because there were two software development teams. One of the teams used the metric system and the other used the imperial system of measurements. It is easy to say that the error might be the fault of some individual, but it is clear to most people that it is just a matter of time before such an error would have occurred.

According to Perrow, system accidents happen because various aspects of the system make them more likely to occur. One factor that contributes to system accidents is the complexity of the system. Perrow's book [15] is full of fascinating accounts of accidents that happened because of unexpected interactions of factors. Of course, sometimes the accidents happen because of expected interactions as in the Mars Climate Orbiter disaster mentioned earlier.

Of special relevance to us are the accidents described in Chapter 6, "Marine Accidents," of [15] which are called noncollision-course collisions. These are collisions in which two ships were originally on courses that would have caused them to pass by each other safely, but because of actions on the part of one or both crews, the ships ended up colliding.

Some of these collisions occurred on the open ocean and where it would have been difficult for the ships to collide even if they had been trying to ram each other. This type of accident is relevant to our study because in maritime systems there is generally no central authority controlling the ships' movements and each ship has a radar device that shows the other ship's location, and yet they still collide. It is imperative that aircraft not be involved in noncollision-course collisions.

A recent example [17] illustrates how complexity can cause disruptions. Fortunately, no one was hurt in this series of incidents. In particular, officials at Newark airport noticed that the GPS based system called "Smartpath" would experience interference on a regular basis. They traced it to a driver for an engineering company who often drove by Newark Liberty Airport, but who was using an illegal \$100 GPS jammer to confuse the tracking device placed in the company vehicle by the company. The low cost and easy availability of GPS jammers suggests that the FAA needs to be prepared for GPS malfunctions in the future. In particular, as advised by Carroll [13] the FAA has to have a reliable and robust backup system available in case GPS signals get jammed or spoofed.

## 8. Alternatives

There is a pressing need to develop alternatives and solutions to address the ADS-B frequency saturation capacity and signal security shortcomings. Workable solutions need to employ existing technologies in order to minimize the cost impact as well as decrease the implementation timeline. These solutions need to include methods of encrypting the ADS-B signal within established secure channels and modifying the ADS-B protocol to allow for the signal to be transmitted via HF band and satellite communication channels.

Possible solutions for increasing the security of ADS-B for military and government aircraft involve feeding the ADS-B data streams into an encrypted channel such as Mode 5 Level 2. Mode 5 is a new IFF (Identification Friend or Foe) technology designed to replace the aging and easily compromised Mode 4 IFF system. Mode 5 Level 2 allows for an encrypted communication that sends a unique aircraft PIN along with an aircraft position report. Sending the ADS-B data stream through the Mode 5 Level 2 encryption channel would allow for secure control of military and sensitive government aircraft flights in non-radar environments [1, pp. 35-36].

A solution for the 1090 MHz frequency saturation problem is to develop alternative transmission channels for the ADS-B signal, utilizing communications channels other than UHF. Possible solutions utilizing present technology include HF and satellite communications channels. HF frequencies are notoriously prone to atmospheric disturbances and interferences, but recent work in the area of Wide-Band

HF may hold some promise for increasing the reliability and throughput of HF transmissions. The Wide-Band HF protocol uses a spectral analysis and band compression to create a long-distance communications channel that could be employed to pass flight data to ground stations. Another solution for passing ADS-B position information over long distances is via satellite communications. Both of these communication channels could utilize existing technologies to achieve reliable and precise position information over very long distance for aircraft operating in non-radar environments.

## 9. Drones

A recent near mid-air collision incident [18] involving a US Airways regional jet and a suspected drone has set off a new round of discussion regarding drones operating in the same airspace as conventional aircraft. This discussion is pertinent to the FAA's NextGen and ADS-B. In order for unmanned aircraft to safely fly in close proximity to conventional air traffic, there needs to be a method of precisely monitoring these drones at all times.

In the case of the US Airways flight, it is unclear as to whether the "drone" was a model aircraft being flown too close to the Tallahassee Regional Airport, or if it was indeed a military or commercial drone. The FAA's definition of what qualifies as a model aircraft and what is a "drone" (a UAV) is somewhat ambiguous further complicating the situation. The FAA has strict limits on operating any model, UAV or aircraft within 5 nautical miles of an airport. In the case of the Tallahassee Regional Airport, whoever was flying the aircraft was in clear violation of the Federal Aviation Regulations, and could be facing some rather steep fines if caught.

Obviously the Department of Defense has a significant interest in UAVs for both reconnaissance and tactical employment. Over the past 5 years, interest in UAVs for commercial use has also grown significantly. The availability of cheap and reliable components for propulsion and control has made UAVs a very attractive technology to explore for a host of surveillance and logistics applications. All of this interest leads to the question of how to integrate these devices into the existing and already crowded air traffic control system.

Aside from the airspace integration concerns, this exploding interest in UAVs creates huge potential security risks. For example, the loss of a U.S. stealth drone in 2011 appears to be a case of hacking [19]. The pilots of the UAV lost control of the drone, and it appears that the Iranians were able to "steal" a multi-million dollar piece of hardware along with the underlying technology. The thought of a sky full of UAVs that could be potentially "hacked" is a very sobering one. The damage that someone could inflict with a several thousand pound UAV capable of flying at several hundred miles per hour is a pretty scary scenario to contemplate.

Currently most UAVs are restricted to operating along specially designated corridors and within special use airspace. A major concern for a pilot is when a UAV or drone goes into “homing pigeon” mode. The majority of UAVs are pre-programmed to return home if they lose contact with their controlling ground station. In most cases, the controlling agency notifies ATC of the problem, the drone flies along a pre-determined flight corridor and ATC vectors aircraft away from this flight path. The problem occurs when, whether due to a latency in communications or some other failure in the system, the controllers are unaware that they are no longer controlling the UAV. The potential exists for the UAV to start its return home without anyone being aware it is doing so.

In a radar environment, ATC will eventually notice the UAV off its planned routing, but until then it is a potentially hazardous situation. This potential hazard will be compounded as we move toward integrating UAVs into the broader ATC system. It gets even worse when one considers the impact of the FAA's NextGen in creating a non-radar environment over the contiguous 48 states. We hope that drones delivering packages for Amazon remain more science fiction than science fact. Needless to say, it is clear that hackers are keenly interested in creating their own drones and they are not the most safety conscious when piloting them. [20] describes an illegal flight carried out by hackers who wanted to build their own UAV. Fortunately, no one was injured when the drone crashed.

## 10. Conclusions and Recommendations

It is clear to us that much more thought needs to go into the security of the FAA NextGen system and that it should not be rushed into production. Based on the analysis in this paper we recommend the following.

- 1) A thorough risk assessment needs to be carried out to make sure all scenarios have been considered.
- 2) There should be no rush to dismantle the ground based radar systems, since these might be very useful in a wide variety of situations.
- 3) Much thought should be invested in making sure that there are all sorts of alternatives and back-up systems available. Nothing must compromise the safety of the system.
- 4) All communication should be encrypted. Sending critical information as cleartext should be avoided at all costs.
- 5) The issue of frequency saturation of the 1090 MHz band in congested airspace needs to be addressed. Some alternatives worth considering are opening up additional frequencies in the UHF band, utilize VF and HF frequencies, use satellites and also spread spectrum techniques.
- 6) The FAA wants to tie a unique identifier to a transponder for its lifetime and broadcast that identifier so that everyone can spot it. This opens up the possibility of

all sorts of monitoring and information gathering. We believe that the system should generate a unique ID for each transponder for the duration of each flight.

- 7) It is essential that jammers be made much more difficult to obtain even though they are easy to build. In any case, whatever system the FAA adopts must be able to cope effectively with missing, jammed, degraded or spoofed GPS signals.

## References

- [1] Donald L. McCallie, *Exploring Potential ADS-B Vulnerabilities in the FAA's Nextgen Air Transportation System*, Air Force Institute of Technology, Graduate Research Project, June 2011, <http://www.hsdl.org/?abstract&did=697737>.
- [2] *Instrument Procedures Handbook*, FAA-H-8261-1A, 2007.
- [3] *Instrument Procedures Handbook*, FAA-H-8083-16, 2014, [http://www.faa.gov/regulations\\_policies/handbooks\\_manuals/aviation/instrument\\_procedures\\_handbook/media/FAA-H-8083-16.pdf](http://www.faa.gov/regulations_policies/handbooks_manuals/aviation/instrument_procedures_handbook/media/FAA-H-8083-16.pdf).
- [4] *FAA Faces Significant Risks in Implementing the Automatic Dependent Surveillance - Broadcast Program and Realizing Benefits*, FAA Report AV-2011-002, October 12, 2010, <http://www.oig.dot.gov/library-item/5415>.
- [5] Federal Aviation Administration, “FAA's NextGen Implementation Plan,” Washington, DC. [http://www.faa.gov/nextgen/implementation/media/NextGen\\_Implementation\\_Plan\\_2013.pdf](http://www.faa.gov/nextgen/implementation/media/NextGen_Implementation_Plan_2013.pdf)
- [6] Darryl H. Phillips, “Will ADS-B Increase Safety and Security for Aviation?”, 2000, <http://www.airport-corp.com/adsb2.htm>.
- [7] Brad “RenderMan” Haines, “Hackers + Airplanes: No Good Can Come of This,” Presentation at DefCon 20, <http://www.youtube.com/watch?v=CXv1j3GbgLk>.
- [8] *Defcon Conference*, <https://www.defcon.org/>
- [9] Nick Foster, “Tracking Aircraft with GNU Radio,” 9/14/2011, <http://gnuradio.org/redmine/attachments/download/246/06-foster-adsb.pdf>. Code may be downloaded from <https://github.com/bistromath>.
- [10] GNURadio Project, <http://gnuradio.org>
- [11] Flightgear Flight Simulator, <http://www.flightgear.org/>.
- [12] Matt Thurber, “Hackers, FAA Disagree Over ADS-B Vulnerability,” *AInonline*, August 21, 2012, <http://www.ainonline.com/aviation-news/ainalerts/2012-08-21/hackers-faa-disagree-over-ads-b-vulnerability>.
- [13] James V. Carroll, “Vulnerability Assessment of the Transportation Infrastructure Relying on GPS,” DOT/OST Outreach Meeting, October 5, 2001, <http://www.navcen.uscg.gov/ppt/Volpe%20Slides.ppt>.
- [14] “NDGPS General Information,” United States Coast Guard, <http://www.navcen.uscg.gov/?pageName=dgpsMain>.
- [15] Charles Perrow, *Normal Accidents*, Princeton University Press, Princeton, NJ, 1999.
- [16] CCN, “NASA's metric confusion caused Mars orbiter loss,” September 30, 1999, <http://www.cnn.com/TECH/space/9909/30/mars.metric/>.
- [17] CBS News, “N.J. Man In A Jam, After Illegal GPS Device Interferes With Newark Liberty Operations,” <http://newyork.cbslocal.com/2013/08/09/n-j-man-in-a-jam-after-illegal-gps-device-interferes-with-newark-liberty-operations/>.
- [18] CNN, “FAA official: Drone, jetliner nearly collided over Florida,” May 11, 2014, <http://www.cnn.com/2014/05/09/travel/unmanned-drone-danger/>
- [19] John Walcott, “Iran Shows Off Downed Spy Drone on TV as U.S. Assesses Loss of Technology,” *Bloomberg News*, December 9, 2011, <http://www.bloomberg.com/news/2011-12-09/iran-shows-off-downed-spy-drone-as-u-s-assesses-technology-loss.html>.
- [20] Michael Weigand, Brad Haines, Mike Kershaw, “Build your own UAV 2.0,” DEFCON 18, Las Vegas, NV, July 30-August 1, 2010, <https://www.youtube.com/watch?v=Dim2-rEO9j4>

**SESSION**  
**CRYPTOGRAPHIC TECHNOLOGIES I**

**Chair(s)**

**Dr. Michael Grimaila**  
**Air Force Institute of Technology - USA**



# Modeling Continuous Time Optical Pulses in a Quantum Key Distribution Discrete Event Simulation

Logan O. Mailloux<sup>1</sup>, Michael R. Grimaila<sup>1</sup>, Douglas D. Hodson<sup>1</sup>, L. Elaine Dazzio-Cornn<sup>1</sup>, and Colin McLaughlin<sup>2</sup>

<sup>1</sup>Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio 45433, United States

<sup>2</sup>Naval Research Laboratory, Washington, DC 20375, United States

<sup>1</sup>{Logan.Mailloux, Michael.Grimaila, Douglas.Hodson, Laura.DazzioCornn}@afit.edu

<sup>2</sup>Colin.Mclaughlin@nrl.navy.mil

**Abstract**—Quantum Key Distribution (QKD) is an innovative technology that leverages the fundamental laws of quantum mechanics to securely distribute shared secret cryptographic keys. QKD technology, when paired with the one-time pad encryption algorithm, provides the opportunity for “unconditionally secure” communications between two parties. However, QKD is a nascent technology with non-ideal system implementations where design trade-offs in system architectures are not well understood due to the complexities of physical and system-level interactions. In this paper, we discuss modeling Continuous Time (CT) optical pulses in a hybrid Discrete Event Simulation (DES) framework built to enable performance analysis and characterization of current and future QKD system implementations. Modeling considerations, design decisions, and trade-offs for system-level modeling of QKD systems, and specifically the efficient modeling of CT optical pulses, is described. Finally, the strengths and weaknesses of modeling CT optical pulses in a DES framework are explored.

**Keywords**—Quantum Key Distribution; Model and Simulation; Discrete Event Simulation; Simulation Framework

## I. INTRODUCTION

The beginnings of Quantum Key Distribution (QKD) can be traced back to Stephen Wiesner who developed the idea of securely encoding quantum information in conjugate basis sets during the late 1960s [1]. As a student at Columbia University, he described two applications for quantum coding: (1) a method for the creation of fraud-proof banking notes (i.e., quantum money) and (2) a method for the transmission of multiple messages in such a way that reading one of the messages destroys the others (i.e., quantum multiplexing). In 1984, Charles Bennett and Gilles Brassard operationalized this concept when they proposed the first QKD protocol (i.e., BB84) to securely distribute shared encryption key between two parties in the presence of an eavesdropper [2].

The unique nature of QKD necessitates that any interference on the key distribution channel leaves detectable fingerprints through increased error rates. Thus, QKD provides a secure key distribution technology that when paired with the One-Time Pad (OTP) encryption algorithm offers

“unconditionally secure” communications regardless of an eavesdropper’s computational power [3, 4, 5].

A QKD usage scenario is depicted in Fig. 1, illustrating the customarily named sender “Alice” and receiver “Bob” in context with bulk encryptors. Alice and Bob are connected via a quantum channel and a classical channel in order to generate a shared secret key,  $K$ , used to encrypt network data. In the scenario, the plaintext message,  $m$ , is transformed into the ciphertext,  $E_K(m)$ , transmitted over the Internet, and decrypted at the distant end where  $m = D_K(E_K(m))$  using the secret key  $K$ .

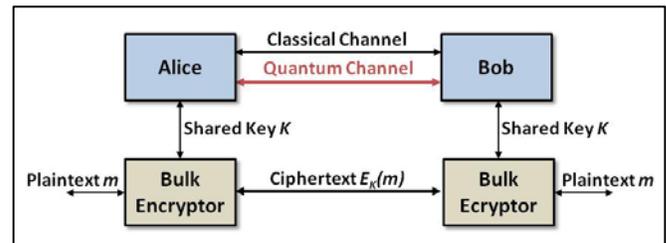


Figure 1. QKD Context Diagram. Note: additional control signals and possible cryptographic parameters are not shown for simplicity.

Over the past decade there has been increasing interest in advancing QKD technology as a practical solution to enable unconditionally secure communications. However, QKD is a nascent technology and non-idealities exist within system implementations [6, 7, 8]. Furthermore, trade-offs in architectural design and implementation choices are not well understood due to the complexities of physical and system-level interactions [9]. To address these needs, the Authors have constructed a hybrid Discrete Event Simulation (DES) framework which enables the efficient modeling and analysis of current and future QKD implementations.

A concise background of QKD technology is provided in Section II to foster an understanding of modeling Continuous Time (CT) optical pulses. Section III introduces the hybrid QKD DES framework used to model, simulate, and analyze QKD systems. Section IV describes modeling considerations and design decisions for CT optical pulses. Section V explores the strengths and weaknesses of modeling CT optical pulses in a DES framework. Finally, conclusions and future work are proposed in Section VI.

## II. BACKGROUND

QKD systems have been demonstrated in optical fiber and terrestrial free-space implementations, while initial commercial offerings are available from ID Quantique, MagiQ and others

This work was supported by the Laboratory for Telecommunication Sciences [grant number 5743400-304-6448]

[10, 11, 12]. Consider for example, the ID Quantique Cerberus system, it takes advantage of QKD generated shared key to frequently rekey conventional encryption algorithms (i.e., rekeys the Advanced Encryption Standard (AES) once a minute) [13]. This QKD implementation reduces the exposure of secure communications with respect to both time and availability of information to an eavesdropper. Conventional encryption algorithms such as AES are used in most encryption applications where the security is based on the computational difficulty of breaking the encryption algorithm measured in long periods of time (e.g., millions of years). However, entities desiring even higher levels of security may implement the OTP encryption algorithm to achieve “unconditionally secure” communications between two parties where the strength of the encryption is not based on time or computational power involved in solving difficult mathematical equations.

#### A. The One-Time Pad Encryption Algorithm

As discovered by Vernam [3] and proven by Shannon [4], the OTP encryption algorithm can achieve unconditionally secure communications between two parties if three strict requirements are followed:

1. The key is random (not pseudo-random)
2. The key is not repeated (used one-time only)
3. The key is at least as long as the data to be encrypted

The main limitation with OTP unconditionally secure communication is distributing enough secret key to encrypt arbitrarily large amounts of data. However, QKD systems have the potential to meet these rigorous requirements as the “Heisenberg uncertainty principle” [1] and Woottter’s “no cloning theorem” [14] are leveraged to ensure two parties can generate enough secret shared key to meet the desire application.

#### B. The BB84 Protocol

In the BB84 protocol, quantum bits (qubits) are single photons polarized into one of four states selected from two conjugate basis sets as described by Wiesner’s quantum multiplexing [2]. As an example, consider Fig. 2 which illustrates how qubits can be polarization encoded in the rectilinear and diagonal basis sets [15]. The rectilinear basis is shown in green comprised of horizontal (i.e.,  $0^\circ$ ) and vertical (i.e.,  $90^\circ$ ) polarizations, while the diagonal basis is shown in blue composed of diagonal (i.e.,  $45^\circ$ ) and anti-diagonal (i.e.,  $135^\circ$  or similarly  $-45^\circ$ ) polarizations. The complementary conjugate basis sets are necessary to ensure security of the quantum channel.

Alice prepares qubits by randomly encoding a bit (0 or 1) and basis (rectilinear or diagonal), while Bob measures the qubit by randomly selecting a basis (rectilinear or diagonal) to measure the arriving polarized photon. If Bob’s randomly selected basis matches Alice’s basis, Bob will measure the encoded bit with a high degree of accuracy. If Bob’s basis does not match Alice’s basis, a random result will be obtained with an equal likelihood (i.e., 50%). In this manner, the BB84 protocol provides a means for passing secret keys between two parties in such a way that an eavesdropper generates detectable errors during the key generation process through increased error rates.

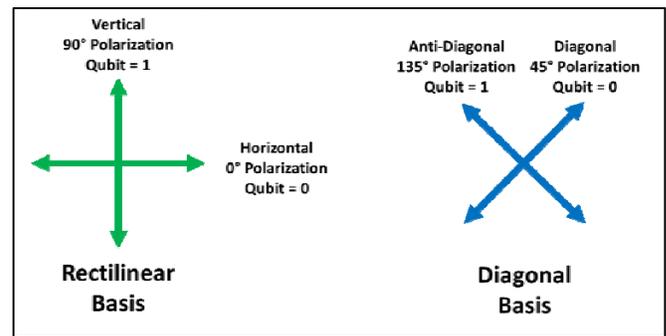


Figure 2. Rectilinear and Diagonal Basis Sets [15].

However, the BB84 protocol assumes several idealities including: (1) Alice emits perfect single photons; (2) the optical link between Alice and Bob is noisy but lossless; (3) Bob has single photon detectors with perfect efficiency; and (4) the basis alignment between Alice and Bob is perfect. If these conditions are met, QKD provides secure key exchange, as shown in QKD security proofs [5, 6, 7]. However, many of these assumptions are not valid when building real-world systems. For example, the BB84 protocol relies on the transmission of single photons, yet reliable on-demand single photon generation is not currently practical. Typically, QKD systems generate coherent optical pulses (with millions of photons) and attenuate optical pulses down to weak coherent pulses with a Mean Photon Number (MPN) of 0.1. A MPN of 0.1 implies a sub-quantum level optical pulse where only 1 in 10 pulses will contain a photon. While this significantly reduces the efficiency of the protocol, it is required to limit and bound the knowledge gained by an eavesdropper for secret key generation [5, 7].

An effective way to study such assumptions and implementation limitations is needed to adequately understand QKD system performance.

### III. A HYBRID QKD SIMULATION FRAMEWORK

In order to study QKD implementations, the Authors desired to build a modular framework for the efficient modeling and simulation of QKD systems. More specifically, the Authors sought to build a flexible framework through modularization and parameterization where each device (or component) stands alone and can be individually conceptualized, implemented, and verified with the desire to enable performance analysis and system characterization of current and future QKD architectures. Designing and building the desired framework required understanding various modeling considerations and making numerous design decisions.

#### A. Model & Simulation Frameworks

In general, Model & Simulation (M&S) is used to analyze complex behaviors, processes, or interactions. A number of industries use M&S to study behaviors which are cost prohibitive to examine or unfeasible to reproduce otherwise (e.g., comparison of proposed systems, deep-space missions, the effects of natural disasters on a metropolitan area, etc.). The two main types of simulation paradigms are discrete (i.e., DES) — where state variables change only at specific points in time and continuous (i.e., CT) — where state variables change

constantly over time [16]. A general outline of pros and cons associated with the simulation paradigms is shown in Table 1.

Table 1: Simulation Paradigm Comparison

<b>Continuous Time</b>	<b>Discrete Event</b>
Pros:	Pros:
<ul style="list-style-type: none"> <li>• Higher fidelity</li> <li>• Most accurate representation</li> </ul>	<ul style="list-style-type: none"> <li>• Computationally efficient</li> <li>• Captures temporal dependencies</li> </ul>
Cons:	Cons:
<ul style="list-style-type: none"> <li>• Can be computationally resource intensive</li> <li>• Copious data output</li> </ul>	<ul style="list-style-type: none"> <li>• Less accurate for continuous time signals</li> <li>• Lower fidelity</li> </ul>

Discrete simulation paradigms are commonly used to model manufacturing and production capabilities, logistics systems, and service queues; these process-oriented systems readily lend themselves to DES. Alternatively, continuous simulation paradigms are generally used to more precisely model detailed behaviors such as power consumption or signal strength.

Choosing an appropriate simulation framework is a critical design decision, one that enables and limits the entirety of the M&S solution. Competing simulation paradigms should be thoroughly examined by both the sponsor and developer to determine how to meet the intended purpose(s).

#### B. The QKD hybrid DES Framework

In some cases, the previously specified M&S paradigms can be combined to form a hybrid M&S solution. The Author's QKD simulation framework (qkdX) is a hybrid M&S solution, where DES is used to schedule and execute the majority of transactions and continuous behaviors are modeled where appropriate (i.e., when studying CT optical pulses). In this way, the qkdX simultaneously supports parameterized physical devices, process-oriented controllers, and quantum phenomenon of weak coherent optical pulses. The framework can be used to study or compare competing architectures, assessments of new QKD protocols, or the study of environmental effects.

The QKD framework is built upon the open source modeling platform OMNeT++ [17], where a unique device library is employed. This feature allows QKD systems to be modeled in a drag and drop fashion selecting from inventoried electrical, optical, and electro-optical devices. Further, the framework enables increasing levels of detail (i.e., fidelity) to be modeled through an object-oriented design. For example, interactions (or transformations) between optical pulses and optical devices can be modeled to a level of detail suitable for performance analysis with or without environmental effects.

This approach accommodates unknown future requirements across a broad spectrum of QKD system implementation possibilities, quantum communication protocols, and plausible electrical, optical, and electro-optical devices in both fiber and free-space applications. Furthermore, the object-oriented qkdX allows modelers to more quickly adapt new devices, protocols, and architectures for experimentation.

#### C. Ensuring Valid Model Representations

Flexibility to model QKD systems accurately (i.e., valid for a specific purpose) and efficiently (i.e., without significant rework) is an expressed objective of the QKD framework. Formally, the aggregation of various modeling components into valid system representations is described as model composability; it is a concept that enables modelers to assemble trustworthy system configurations from a variety of modeled elements to satisfy user requirements in a timely manner [18]. Composability requires that individual model elements be developed such that implementation details (e.g., parameter passing mechanisms, external data access, timing assumptions, etc.) are appropriately accounted for in all potential system configurations. To achieve model composability, detailed system elements should be verified through analytical means based on design specifications and legitimate behaviors.

Model composability is also achieved through conceptual model definition, to ensure simulated models will never enter an unexpected state regardless of what combination of elements are assembled [19]. Conceptual model definition and composability are of particular interest to our project as unique QKD systems are modeled to analyze performance of architectural implementations while optical pulses are passed from one device to another in a structured, assured manner.

The described framework expressly enables the modeling of valid QKD system representations for performance analysis and characterization. Specifically, the qkdX allows for quantum phenomenon and system-level interactions between hardware, software, and protocols to be modeled and studied at user specified levels of detail.

#### IV. MODELING CONSIDERATIONS AND DESIGN DECISIONS FOR CONTINUOUS TIME OPTICAL PULSES

Given infinite time and resources a modeler could conceivably describe every behavior of interest. However this is obviously never the case nor would such a complex effort provide the clarity desired to understand the QKD design and implementation trade space. In a resource constrained environment, model developers are forced to make design decisions and trade-offs. This forces sponsors and developers to more fully understand the system of interest, essential system behaviors, and intended purpose(s). These considerations should provide additional clarity and understanding in the study of difficult problems and complex phenomenon, resulting in more explicitly designed simulation capabilities and value-added simulation results.

Modeling CT optical pulses is one of the most critical design decisions within the QKD framework. QKD systems will generate hundreds of millions of optical pulses during secure key distribution, each of which will propagate through multiple optical devices potentially requiring complex mathematical transformations. As a consequence, an efficient mathematical representation must be chosen to model weak coherent optical pulses at the appropriate level of abstraction necessary to account for the desired resolution and accuracy.

While each pulse is most accurately represented as a CT optical pulse, it is computationally infeasible to model a

complete QKD system using CT simulation. Our approach is to model optical pulses as abstract, parameterized objects (i.e., pulses within temporal packets), where the optical pulse can be manipulated through individual parameters when performing simple transforms and fully reconstructed when performing complex transforms.

#### A. Modeling Light as Weak Coherent Optical Pulses

Light is electromagnetic radiation that can be viewed in two complementary ways: as an electromagnetic wave or as a stream of particles called photons [20]. We adopt the wave nature of light when calculating propagation through the QKD framework. This convention (wave nature) allows for standard calculations of optical effects such as propagation, dispersion, attenuation, or interference. Eq. (1) allows time-dependent propagation of the optical pulse's electric field to be handled through individual components (e.g., fiber, wave plates, beam splitters, etc.) in the OMNeT++ environment. Defining the  $z$ -direction to be along the axis of travel through the optical fiber, the electric field vector of an optical wave then lies in the  $x$ - $y$  plane, where:

$$\vec{E}(t) = \begin{bmatrix} \vec{E}_x(t) \\ \vec{E}_y(t) \end{bmatrix} = E_0 e^{i\omega_0 t} e^{i\theta} \begin{bmatrix} \cos\alpha \\ (\sin\alpha)e^{i\phi} \end{bmatrix} \quad (1)$$

In this continuous-wave representation, the orientation of the electric field is given by  $\alpha$ . The relative phase, or ellipticity, between the  $x$  and  $y$  components of the electric field is given by  $\phi$ . The amplitude, relative phase, and angular frequency of the field are given by  $E_0$ ,  $\theta$ , and  $\omega_0$ , respectively. The angular frequency,  $\omega_0$ , is determined from the optical frequency  $f$  by the relation  $\omega_0 = 2\pi f$ . The  $2 \times 1$  column vector on the right side of in Eq. (1) is the *Jones* vector representation of the electric field [21]. Upon passing through a linear optical device, the polarization state of the emerging light can be determined by multiplication of the *Jones* vector with an appropriate *Jones* matrix. Note: *Jones* calculation applies only to light which is fully polarized, as is generally the case for polarization based QKD systems. Randomly polarized, partially-polarized, and incoherent light cannot be represented in this manner.

Modeling pulsed optical sources requires the inclusion of a time-dependent power envelope,  $G(t)$ , where the power envelope is proportional to the square of the electric field envelope. Thus, the model for the electric field of a pulsed optical source will take the form,

$$\vec{E}(t) = \begin{bmatrix} \vec{E}_x(t) \\ \vec{E}_y(t) \end{bmatrix} = \sqrt{G(t)} E_0 e^{i\omega_0 t} e^{i\theta} \begin{bmatrix} \cos\alpha \\ (\sin\alpha)e^{i\phi} \end{bmatrix} \quad (2)$$

This representation allows classical operations (or transformations) to be efficiently conducted on the optical pulse model. For example, integrating Eq. (2) over the duration of the pulse will yield the total energy contained in the pulse. The photon count of the pulse can then be determined by dividing the total pulse energy by the energy per photon at the given frequency,  $\omega_0$ .

Within the QKD framework we have chosen to use Poisson Distribution statistics to model the presence of photons in a weak coherent optical pulse. The distribution in Eq. (3) gives

the probability of the presence of “ $k$ ” photons in a single pulse given a MPN of  $\lambda$ .

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (3)$$

A stream of weak coherent optical pulses with, for example, a MPN=0.1 ( $\lambda=0.1$ ) will yield no photons per pulse with a probability of 90.48%, one photon per pulse with a probability of 9.05%, and two or more photons with an approximate probability of 0.47%.

#### B. QKD Laser Sources

An example of a commercial sub-nanosecond laser source is the ID Quantique (IDQ) id300 [22]. The id300 laser is capable of generating short pulses at a wavelength of 1310 nm or 1550 nm. The laser source, based on Fabry-Perot (FP) or on distributed-feedback (DFB) laser diodes, is externally triggered to produce sub-nanosecond laser pulses with a pulse rate ranging from 0 to 500 MHz. Fig. 3 shows the time profile of the optical pulse for the id300 laser source when triggered by a 1 MHz clock source [22]. Note that the peak of the pulse is specified as 0 dBm (i.e., 1 mW) and the nominal pulse duration is specified as 0.3 ns or 300 ps Full Width at Half Maximum (FWHM).

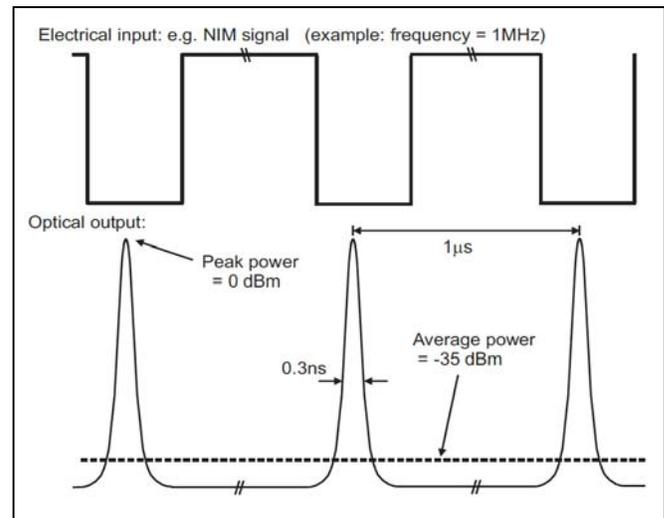


Figure 3. id300 Laser Pulse Specification [22].

Fig. 4 shows Lydersen *et al.*'s measured time profile of the IDQ id300 laser source when using a 12.5 GHz sampling oscilloscope and a 45 GHz optical probe [23]. While the peak of the measured waveform is close to that specified in the id300 data sheet, the pulse shape is quite different from that depicted in Fig. 3.

#### C. Efficient Modeling of Continuous Time Optical Pulses

While the time profile of each pulse could be represented by a set of (time, value) pairs collected in these laboratory measurements, the Authors choose to represent the coherent optical pulse as a CT optical pulse. We believe this representation is more conducive for modeling photon detection and performing interference calculations necessary to understand the impact of quantum interactions on system performance.

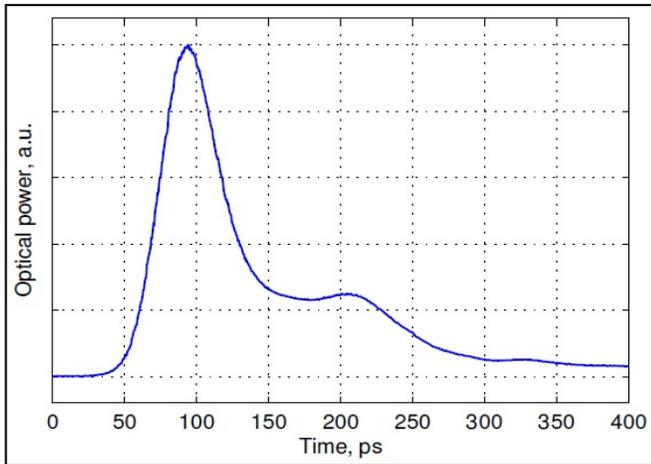


Figure 4. Measured id300 Laser Pulse with Peak at 0 dBm (1 mW) [23].

As a consequence, we approximated the measured pulse's time profile using a sum of Gaussian functions [24]. From Lydersen *et al.*'s experimental measurements, a three Gaussian curve approximation was made as shown in Fig. 5 and detailed in Table 2. Comparing the measured and approximated optical power waveforms, the adjusted R square fit value  $R^2$  is 0.9962, which constitutes a reasonably good fit [25].

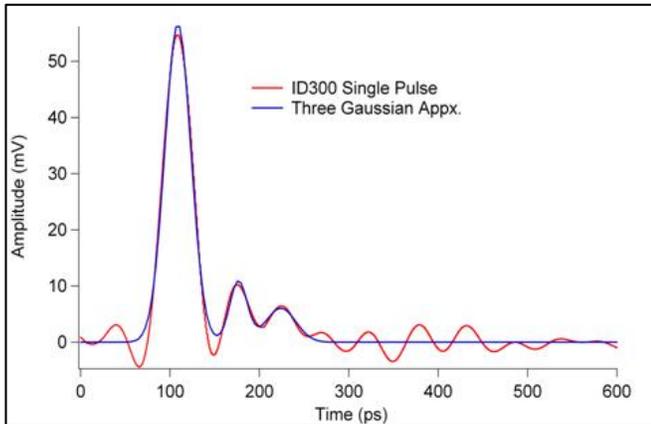


Figure 5. Gaussian Approximation of id300 Laser Pulse.

Approximating the measured shape with Gaussian waveforms allows the QKD framework to take advantage of optimized Gaussian integration techniques and reduces the computational burden placed on the simulation platform.

Table 2. Parameters to Approximate Continuous Time Optical Pulse

Gaussian Curve	Gaussian Amplitude	$\mu$ [ps]	$\sigma$ [ps]
1	45.5	95.48	19.82
2	38.1	181.12	58.14
3	4.76	352.32	49.02

#### D. Object-Oriented Modeling of Optical Pulses

The qkdX employs an object-oriented design to efficiently model weak coherent optical pulses. Several related classes define the CT optical pulse as shown in Fig. 6. The abstract Pulse class serves as the principle interface for a wide variety of pulse definitions. This design allows end users, analysts, or modelers to define pulse representations at an adequate level of

detail (i.e., fidelity) suitable to the desired experimental purpose(s).

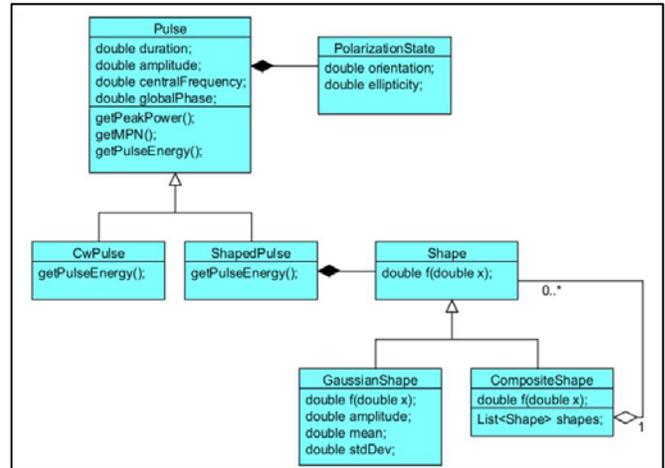


Figure 6. Optical Pulse Class Definition.

Each pulse has a specified *duration* that serves as a windowing function to identify the interval of time which the pulse covers. The duration is represented in the simulator using a double precision variable with a value greater than or equal to zero. The duration is a required parameter to facilitate the efficient processing of optical pulses at the detectors. As described in Eq. (2), each pulse has an electric field amplitude  $E_0$  named *amplitude*, an angular frequency  $\omega_0$  named *centralFrequency*, a global phase  $\theta$  named *globalPhase*, an angle of the vector with respect to the  $x$  axis  $\alpha$  named *Orientation*, and a relative phase  $\phi$  named *ellipticity*. Additionally, the pulse class contains functional calls associated with calculating the pulse energy *getPulseEnergy()*, peak power *getPeakPower()*, and MPN *getMPN()*.

The typical use of the Pulse class will be to represent weak coherent pulses, however it is also desirable to provide a means to represent long duration, Continuous Wave (CW) light. For this reason, we employ inheritance to capture the unique characteristics of each type of coherent pulse represented as *CWPulse* or *ShapedPulse*. In the case of weak coherent optical pulses, we define the pulse shape using a composite pattern. This enables optical pulse shapes to be modeled using multiple Gaussians curves to approximate the optical pulse shape as shown in Fig. 5. Each of the Gaussian curves has an associated *amplitude*, *mean*, and *standardDeviation*. If *pulseType*=CW, none of the Gaussian shape parameters described are used.

The pulse class design simplifies and centralizes the modeling effects due to transmission through optical fiber and transformations associated with optical devices as represented within a modeled QKD architecture. For example, consider a pulse passing through a simple optical attenuator as shown in Fig. 7. Only the e-field magnitude  $E_0$  of the optical pulse is reduced by the attenuator, while the pulse shape remains unchanged. Therefore, the object's amplitude attribute is efficiently updated without using unnecessary computational resources to perform integration.

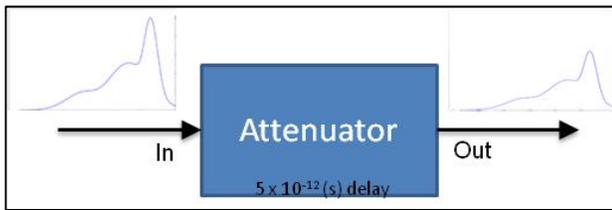


Figure 7. Optical Pulse Passing Through a simple Attenuator.

Centralizing these calculations within the pulse class allows for simplified modeling of optical components throughout the entire QKD framework. Additionally, because all calculations are centralized, efficient integration algorithms can be more easily localized and applied.

## V. STRENGTHS AND WEAKNESSES OF MODELING CONTINUOUS TIME OPTICAL PULSES IN A QKD DES

In this section we explore the strengths and weaknesses of modeling CT, weak coherent optical pulses in a QKD DES framework. In DES future events are scheduled and stored, where the next scheduled event is 'jumped to' saving computational resources between scheduled events. An equivalent continuous model would require constant processing regardless of when the next transaction occurs, causing undue processing burden and potentially generating significant amounts of superfluous data. For systems that have considerable dead time between transactions, the DES modeling construct is generally advantageous. However, this efficiency begins to fall away with large numbers of scheduled events and increasingly smaller dead times, which is the case for QKD systems. For example, during operation potentially millions of events are scheduled with delays on the order of nano ( $10^{-9}$ ) and pico ( $10^{-12}$ ) seconds placing a significant burden on scheduling resources.

### A. Strengths of DES for Modeling QKD Systems

While attempting to discretely model continuous quantum phenomenon may seem counter-intuitive, an investigation of quantum simulation literature demonstrates that continuous simulations of QKD systems are generally more complex than necessary and even impractical to model and/or simulate for system-level behaviors [26]. For example, the wave-particle duality of light in a continuous simulation would require a complete characterization of the optical path as each pulse is created, adding significant computational overhead and negating the desired temporal interactions of a system-level simulation [26].

In the QKD DES framework, these complexity shortcomings are efficiently overcome by probabilistically modeling the frequency of a photon striking the detector after propagating through the system as described in Eq. (3). Despite the heavy scheduling burden, the DES paradigm allows for system-level interactions and CT optical pulses to be modeled and simulated more efficiently than continuous simulations.

Additionally, scheduled events in DES highlight critical dependencies in the subject system. Identifying these system-level interactions and temporal relationships can lead to additional clarity when studying complex systems. This feature of DES is particularly useful when attempting to gain further

understanding of QKD systems or conducting performance analysis and characterization of competing architectures.

### B. Weaknesses of DES for Modeling QKD Systems

Lower fidelity in DES is a limiting issue that needs to be fully understood because it can lead to inaccurate results and incorrect system behaviors. Consider for example, the optical pulse shown in Fig. 4, it has a duration of approximately 400 ps, while the optical attenuator shown in Fig. 7 has a scheduled delay of 5 ps (a significantly smaller time elapse). In the QKD DES framework, the pulse is received, processed, and scheduled for the next event according to the attenuator's 5 ps delay. There is little consideration for the 400 ps pulse propagation time through the attenuator, which can result in erroneous representations of CT coherent optical pulses within the qkdX.

Figs. 8(a-c) demonstrate the problematic nature of low fidelity DES in the case where conflicting events occur within the optical pulse's 400 ps duration. Each of these scenarios can result in invalid outputs because of scheduling constraints inherent in DES. In Fig. 8(a) a conflicting environmental message indicates an "overheat" of the attenuator such that the device's performance and output pulse should be severely degraded. In Fig. 8(b) a continuous light source is shone into the attenuator input which should overpower the weak optical pulse output. In Fig. 8(c) multiple overlapping pulses should cause interference on the output pulse.

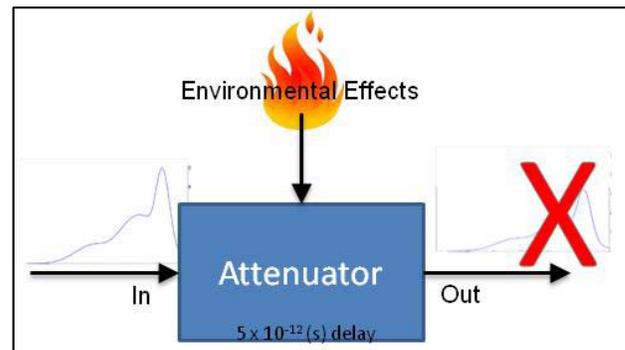


Fig. 8(a): Environmental Overheating.

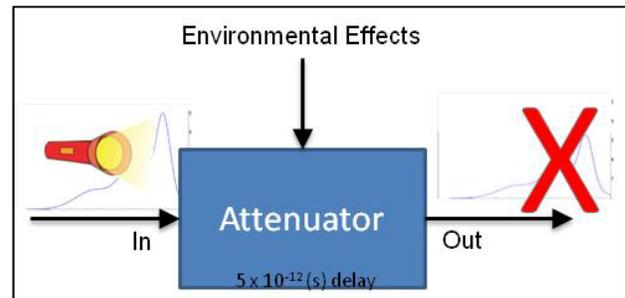


Fig. 8(b): Overpowering Light.

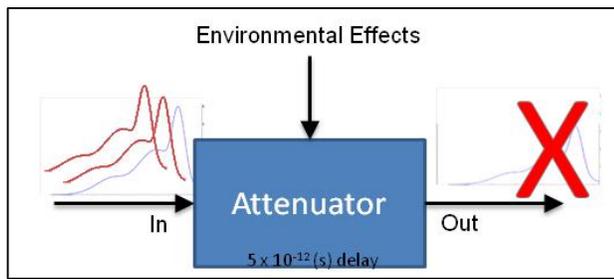


Fig. 8(c). Overlapping Pulse Interference.

While these three examples demonstrate limitations in the lower fidelity QKD DES framework, this weakness needs to be tempered with the fact that only a minimal number of pulses would be incorrectly modeled amongst millions of correctly modeled weak optical pulses.

## VI. CONCLUSIONS

This research is part of an ongoing effort to model, simulate, and analyze QKD system architectures. This paper discusses considerations, design decisions, and trade-offs for modeling CT optical pulses in a hybrid QKD DES framework. A discussion of modeling physical and system-level interactions is facilitated through a QKD simulation framework, and specifically in the modeling of CT weak coherent optical pulses. Strengths and weaknesses of modeling CT optical pulses in a DES framework are explored. Future work includes modeling alternative QKD encoding schemes (i.e., phase-based and entanglement) and free-space implementations.

## 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.

## REFERENCES

- [1] S. Wiesner, "Conjugate coding," *ACM Sigact News*, vol. 15, no. 1, pp. 78-88, 1983.
- [2] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," in *In Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, 1984.
- [3] G. S. Vernam, "Cipher printing telegraph systems for secret wire and radio telegraphic communications," *American Institute of Electrical Engineers, Transactions of the*, vol. 45, pp. 295-301, 1926.
- [4] C. E. Shannon, "Communication Theory of Secrecy Systems," *Bell System Technical Journal*, vol. 28, pp. 656-715, 1949.
- [5] R. Renner, N. Gisin and B. Kraus, "An information-theoretic security proof for QKD protocols," arXiv:quant-ph/0502064, 2005.
- [6] D. Gottesman, H.-K. Lo, N. Lutkenhaus and J. Preskill, "Security of quantum key distribution with imperfect devices," in *In Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*, 2004.
- [7] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus and M. Peev, "The security of practical quantum key distribution," *Reviews of modern physics*, vol. 81, no. 3, p. 1301, 2009.
- [8] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar and V. Makarov, "Hacking commercial quantum cryptography systems by tailored bright illumination," *Nature photonics*, vol. 4, no. 10, pp. 686-689, 2010.
- [9] J. D. Morris, M. R. Grimaila, D. D. Hodson and D. Jacques, "A Survey of Quantum Key Distribution (QKD) Technologies," in *Emerging Trends in ICT Security*, Elsevier, 2013, pp. 141-152.
- [10] ID Quantique, "ID Quantique Main Page," 2013. [Online]. Available: <http://www.idquantique.com/>. [Accessed 1 Nov 2013].
- [11] MagiQ, "MagiQ Main Page," 2014. [Online]. Available: <http://www.magiqtech.com/MagiQ/Products.html>. [Accessed 19 Feb 2014].
- [12] N. Gisin, G. Ribordy, W. Tittel and H. Zbinden, "Quantum cryptography," *Reviews of modern physics*, vol. 74, no. 1, pp. 145-195, 2002.
- [13] ID Quantique, "Cerberis Quantum key Distribution (QKD) Server," 08 Mar 2014. [Online]. Available: <http://www.idquantique.com/network-encryption/products/cerberis-quantum-key-distribution.html>.
- [14] W. K. Wootters and W. H. Zurek, "A single quantum cannot be cloned," *Nature*, vol. 299, no. 5886, pp. 802-803, 1982.
- [15] M. R. Grimaila, J. D. Morris and D. D. and Hodson, "Quantum Key Distribution: A Revolutionary Security Technology," *The Information System Security Association Journal*, vol. 10, no. 6, pp. 20-27, 2012.
- [16] J. Banks, J. S. Carson, B. L. Nelson and D. M. Nicol, *Discrete Event System Simulation*, 5, Ed., Prentice Hall, 2010.
- [17] OMNeT++, "OMNeT++ Main Page," 2013. [Online]. Available: <http://www.omnetpp.org/>. [Accessed 08 11 2013].
- [18] M. D. Petty and E. W. Weisel, "A compossibility lexicon," in *Proceedings of the Spring 2003 Simulation Interoperability Workshop*, 2003.
- [19] B. P. Zeigler, H. Praehofer and T. G. Kim, *Theory of modeling and simulation*, New York: John Wiley, 1976.
- [20] D. C. Giancoli, *Physics for Scientists and Engineers*, Prentice Hall, 1989.
- [21] R. C. Jones, *Journal of the Optical Society of America*, vol. 31, pp. 488-493, 1941.
- [22] ID Quantique, "id300 Series Sub-Nanosecond Pulsed Laser Source Datasheet," 2012. [Online]. Available: <http://www.idquantique.com/images/stories/PDF/id300-laser-source/id300-specs.pdf>. [Accessed 05 Mar 2014].
- [23] L. Lydersen, N. Jain, C. Wittmann, Ø. Marøy, J. Skaar, C. Marquardt, V. Makarov and G. Leuchs, "Superlinear threshold detectors in quantum cryptography," *Phys. Rev.*, vol. A 84, no. 032320, 2011.
- [24] A. Goshtasby and W. D. O'Neill, "Curve Fitting by a Sum of Gaussians," *CVGIP: Graphical Models and Image Processing*, vol. 56, no. 4, pp. 281-288, 1994.
- [25] A. C. Cameron and F. A. Windmeijer, "An R-squared measure of goodness of fit for some common nonlinear regression models," *Journal of Econometrics*, vol. 77, no. 2, p. 1790-2, 1997.
- [26] N. T. Sorensen, *Quantum Channel Modeling for Discrete Event Simulation of Quantum Key Distribution (Master's Thesis)*, Air Force Institute of Technology, 2012.

# A Survey on Certificateless Encryption Techniques

Fahad T. Bin Muhaya

Management Information Systems Department, College of Business Administration  
Prince Muqrin Chair for IT Security (PMC), King Saud University, Riyadh, Kingdom of Saudi Arabia  
E-mail addresses: fmuhaya@ksu.edu.sa

**Abstract**— Certificateless public key encryption is an advanced version of identity-based and public key encryption techniques. It eliminates the inherent key escrow and other key management issues in both techniques, e.g., it does not need any public key infrastructure or digital certificates. This paper presents a survey of the certificateless public key encryption schemes that have been proposed to improve efficiency and security. We examine the construction, infrastructure, and security models to compare published schemes and their security levels. Analyses and comparisons show that many of the different schemes are based on public key infrastructures and/or are identity based. Analyses and comparison results also show that some schemes have correct generally inclusive security concepts, while others are comparatively insecure and inefficient.

## I. INTRODUCTION

The revolutionary public key or asymmetric system was proposed by Rivest, Shamir, and Adleman (called the RSA algorithm) in 1978 [1] and released in 2000. It uses two distinct separate keys, public and private, with a mathematical relationship to each other. A public key is used to encrypt the data and is widely distributed, while a private key is used to decrypt the data and is kept secret. The security is based on the assumption that no one can get one key from the other. The public and private keys are generated with attention and care to make them secure. The concept of using public and private keys ensures that a sender can securely send a message to the receiver without exchanging a shared secret key. This is possible because of using the public-private keys, instead of a single secret key. However, these systems have some drawbacks, like the need for the sender to have the correct public key to encrypt data for the receiver. Thus, a public key infrastructure (PKI) is required to manage and distribute public keys as shown in Figure 1. In such systems, a public key is bound to the respective unique user ID. Trusted third party tools are used to bind the users to unique public keys through an appropriate registration process for the users. The use of an additional third party application makes the public key cryptography expensive and inefficient. To overcome the problems and issues of PKI management systems, many ID-based cryptography systems have been proposed. In such systems, a connection is created between the user and a public key using a unique digital ID that is a combination of text characters. In this way, an ID-based system removes the PKI third party application by creating a link between the public key and digital ID for the user. However, this creates the problem of generating and managing the private keys for users. Again, the system needs a third party application for private key management to generate them secretly and distribute them to users.

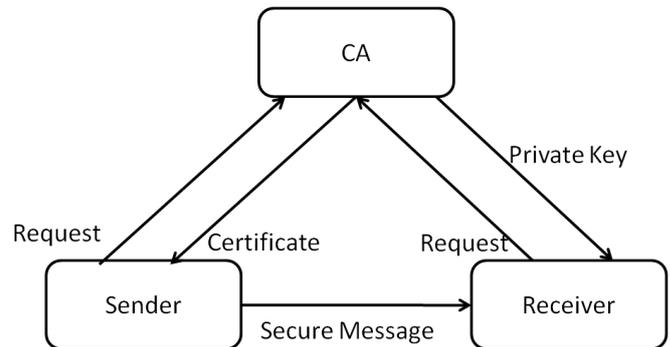


Figure 1. PKI infrastructure

To overcome the PKI and ID-based cryptography key management issues, a new certificateless public key cryptography scheme was proposed by Al-Riyami [2]. It did not need public key distribution. Certificateless public key cryptography merges the key management concepts of both the ID-based and PKI key management techniques. In this technique, to encrypt data, a user gets a public key using a secret value and partial public key. Similarly, to decrypt the data, the user gets a private key using a secret value and partial private key. The Key Generation Center (KGC) generates these partial public and private keys for the user. Thus, in a certificateless cryptography scheme, there is no need for any third party key management application like the PKI and ID-based techniques. In this case, a user can find a public key from the public directory. One time validation is needed, after which the user can change and update the public key according to his needs. The whole process is depicted in Figure 2.

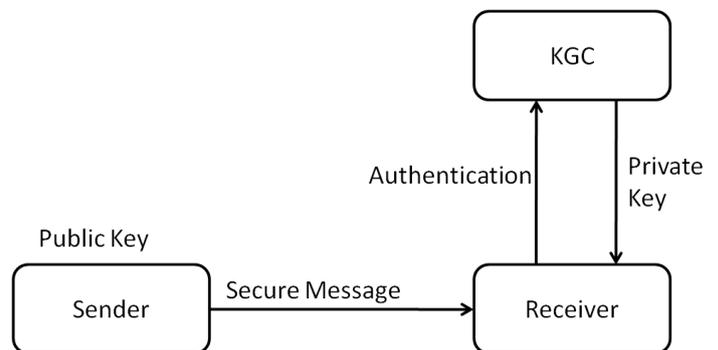


Figure 2. Identity based encryption

This paper presents a survey of some of the known certificateless public key encryption schemes that have been published since the first proposal by Al-Riyami and Paterson. We survey the existing schemes to better determine their reality and applicability.

II. OVERVIEW OF CERTIFICATELESS PUBLIC KEY ENCRYPTION SCHEMES

The objective of a certificateless public key encryption scheme is to provide a system where users can securely share information using publicly available information. This scheme eliminates the need for the certificate used in PKI and ID-based encryption schemes. The basic infrastructure of the scheme is based on three entities: the sender, receiver, and KGC as shown in Figure 3. In the following subsection, brief descriptions of some well-known schemes are provided for a basic understanding of certificateless public key encryption schemes.

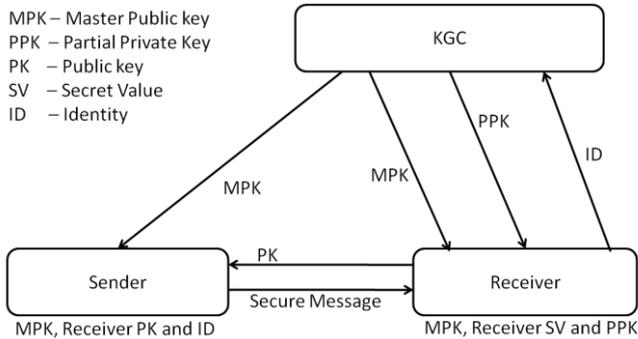


Figure 3. Al-Riyami and Paterson encryption scheme

A. Al-Riyami and Paterson Certificateless public key encryption scheme

In 2003, the first formulation, basic notations, and terminologies of a certificateless public key encryption scheme were introduced by Al-Riyami and Paterson, which was constructed using elliptic curve pairings. The proposed Al-Riyami and Paterson certificateless public key encryption scheme formulation is given below:

- **Initialization setup:** The initialization and startup process is done at the KGC system. It takes the security parameter as input and generates a master private key and public key for the user.
- **Partial private key extraction:** The extraction process is also executed by the KGC system, but only once for each user. It takes the master public and private keys, with the user ID, as inputs and generates a partial private key.
- **Secret value:** The secret value setup process is executed at the user end. It takes the master public key and user ID as inputs and generates a secret value for the user.
- **Private Key:** The one time private key generation process takes the master public key and partial private key, along with the secret value of the user, as inputs to generate the private key for the user. This process is executed by the user.
- **Public Key:** The one time public key generation process takes the master public key and secret value of the user as inputs to generate a public key for the user, after which the public key is distributed to encrypt the data. This process is executed by the user.

- **Encryption:** This process uses the master public key, user ID, and public key to encrypt the message into unreadable form.
- **Decryption:** This process uses the master public key and private key to decrypt the encrypted text back to its original form.

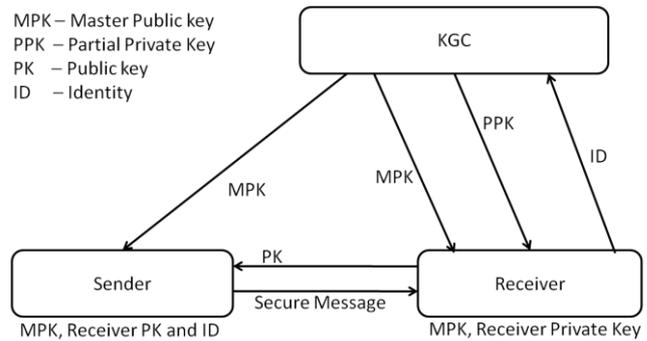


Figure 4. Baek, Safavi-Naini and Susilo encryption scheme

B. Baek, Safavi-Naini and Susilo Certificateless public key encryption scheme

In 2005, Baek, Safavi-Naini, and Susilo proposed a new and modified formulation of the scheme by Al-Riyami and Paterson with an improved architecture that did not use elliptic curve pairings [3]. This system can be seen in Figure 4. In their scheme, the receiver needs a partial private key before publishing the public key. The proposed Baek, Safavi-Naini, and Susilo certificateless public key encryption scheme formulation is given below:

- **Initialization setup:** The initialization setup process is done at the KGC system. It takes the security parameter as input and generates a master private key and public key for the users. This is the same as the Al-Riyami and Paterson setup process.
- **Partial private key extraction:** This process is also the same as the Al-Riyami and Paterson extraction process and is executed by the KGC system once for each user. It takes the master public and private keys, along with the user ID, as inputs and generates a partial private key.
- **Key Generation:** This process is different from that of Al-Riyami and Paterson. It takes the partial private key and identity of the user as inputs and generates a pair of public and private keys for users.
- **Encryption:** This process is also the same as the Al-Riyami and Paterson extraction process. It uses the master public key, user ID, and public key to encrypt the message into an unreadable form.
- **Decryption:** This process is different from the Al-Riyami and Paterson decryption process. It uses the master public key and receiver private key to decrypt the encrypted text back to its original form. It does not require the partial private key.

### III. COMPARISONS AND ANALYSIS OF CERTIFICATELESS PUBLIC KEY ENCRYPTION SCHEMES

This section provides comparisons and an analysis survey of the existing literature about certificateless public key encryption schemes. The results of the comparative analysis survey of the certificateless public key encryption schemes are given in Table 1.

In 2003, the first certificateless public key encryption scheme was introduced by Al-Riyami and Paterson [2]-[4]. In their paper, Al-Riyami and Paterson presented the concept of concrete certificateless public key cryptography. They proposed certificateless public key encryption, signature, and key exchange schemes that were constructed using elliptic curve pairings. The scheme was proven using a fully adaptive adversarial model. In 2005, Al-Riyami and Paterson again proposed a new certificateless public key encryption scheme [5] and proved that its security and efficiency were much better than those of the original scheme. They applied techniques similar to those of Fujisaki and Okamoto [6] to prove the security of the scheme. In 2006, Libert and Quisquater [7] proved that the generic compositions of Al-Riyami and Paterson were insecure against chosen cipher text attacks in a relaxed security model. They proposed a better method to achieve a generic construction and fixed the problem by applying the certificateless Fujisaki and Okamoto transform method. In the same year, Zhang and Feng [8] proposed some fine tuning for the certificateless public key encryption scheme of Al-Riyami and Paterson and proved that the old version was insecure. Zhang and Feng also proposed a solution for the problem, but it resembled the scheme of Cheng and Comley [9] without having proven security.

In these schemes, random oracle models or weak security models were used to prove the efficiency and security. They lacked security proofs that used strong security models without random oracle models for the proposed certificateless public key encryption schemes. The first certificateless public key encryption scheme to have demonstrably secure in strong security models was proposed by Dent, Libert, and Paterson [10]. They included two generic constructions of certificateless encryption schemes and proved the security in strong security models without using random oracles models. They also included concrete constructions of Waters identity-based [11] and hierarchical identity based encryption schemes [12]. After

some time, researchers started thinking that they could merge public key and identity-based encryption techniques for better security, reliability, and efficiency in certificateless public key encryption schemes. The idea was to use a public key for encryption (as used in public key encryption) and a private key for decryption, with a combination of two private keys, one from public key encryption and the other from identity-based encryption. The concept of using both public key and identity-based encryption in certificateless public key encryption was first implemented by Al-Riyami and Paterson in [5] and by Yum and Lee in [13]-[14]. They proposed different schemes by applying different combinations of public key and identity-based encryption techniques. In the first proposal, they applied both techniques sequentially. First, the original message is encrypted using public key encryption, after which the output is again encrypted using identity-based encryption. In the second proposal, they again applied both techniques sequentially, but this time the original message was encrypted using identity-based encryption first. In the third proposal, both techniques were applied in parallel by dividing the original message into two messages.

In 2006, Libert et al. [15] and Galindo et al. [16] proved that these schemes were insecure. Libert et al. proved that the first proposal for generic compositions was insecure against chosen cipher text attacks in a relaxed security model. The second proposal was insecure in a strong security model. While, in the third proposal, the message could be decrypted using two oracle queries. Galindo et al. proved that the second proposal was insecure on a weak security model. In 2008, Dent [17] proved that these schemes have more serious issues, in which an attacker can re-encrypt the message using a public key and submit it again in a new encrypted form. In [18], Bentahar et al. proposed a different method that was similar to that of Cramer et al. [19] by extending the key encapsulation mechanisms (KEM) concept for ID-based and certificateless encryption using data encapsulation mechanisms (DEM), which results in more secure encryption schemes for weak type models. In this scheme, KEM generates symmetric key  $K$  and performs a key encapsulation operation, after which DEM encrypts a message using symmetric key  $K$ . In addition, they proposed generic constructions of ID-based and certificateless encryption based on a KEM and DEM combination, which proved secure on a random oracle model.

Table 1. Comparisons and Analysis of Certificateless Public Key Encryption Schemes

Proposed	Construction	Security Analysis	Broken By
Al-Riyami [4]	Generic Construction	For one scheme weak type models are used but for other schemes no proof is given	Libert et al. [15] , Galindo et al. [16] and Dent et al. [10]
Yum-Lee [13]	Generic Construction	Weak type model is used.	Libert et al. [15] and Galindo et al. [16]
Al-Riyami and Paterson [2]	Concrete Construction	Random Oracle Model is used	
Al-Riyami and Paterson [5]	Concrete Construction	Random Oracle Model is used	Zhang et al. [8] and Libert et al. [15]
Baek-Safavi-Naini-Susilo [3]	Concrete Construction	Random Oracle Model is used	--
Bentahar et al. [18]	Generic Construction	Random Oracle Model is used	--
Cheng-Comley [9]	Concrete Construction	Random Oracle Model is used	--
Dent-Libert-Paterson [10]	Generic and Concrete Construction	Strong Type models are used	--
Libert-Quisquater [7]	Generic and Concrete Construction	Random Oracle Model is used	--
Huang-Wong [20]	Generic Construction	Weak Type models are used	--
Yum-Lee [14]	Generic Construction	Weak Type models are used	Galindo et al. [16]

In 2007, Huang and Wong [20] proposed a generic certificateless KEM scheme based on public key encryption ID-based KEM and message authentication code, in the standard model. They proved the security of the scheme against malicious-but-passive KGC attacks without using a random oracle model. Secondly, they proposed a certificateless tag-based KEM scheme based on the concept proposed in Abe et al. [21]. They also showed the construction of a hybrid certificateless encryption scheme by applying Abe et al.'s transformation to a certificateless tag-based KEM and one time DEM. The schemes were less efficient but comparable to the certificateless KEM scheme of Bentahar et al., in which only random oracle models were used to prove the security. In 2007, Dent et al. [22] proposed certificateless encryption schemes in the standard model and proved the secure against strong adversaries without using random oracle models. The proposed scheme was based on a combination of certificateless encryption schemes, public key encryption schemes, and the extended version of Naor et al. [23] and Sahai [24] for non-interactive zero-knowledge proofs. In [25], Hwang et al. showed that the schemes of Liu et al. [26] and Dent et al. were insecure and required random oracle models to prove their security. Then, they proposed an improved and secure scheme against a malicious KGC attack in the standard model.

#### CONCLUSION

This paper presented a survey of certificateless public key encryption schemes, which are advanced version of identity-based and public key encryption techniques, used to overcome and eliminate key management issues. The paper investigated the issues and problems related to the construction, infrastructure, and security models for certificateless public key encryption schemes. Comparative analyses showed that some of the proposed schemes were based on public key infrastructures, identity-based techniques, or both techniques. Some schemes have a correct generally inclusive concept of security, while others are comparatively insecure and inefficient. In the early stages of development for certificateless public key encryption schemes, mostly random oracle or weak security models were used to prove the security and efficiency. In the development of new schemes, strong security models are used and tested without using random oracle models. However, many questions remain unanswered that need to be answered relate to security, models, efficiency, and implication. Thus, we conclude that more research and development is needed in this area.

#### AKNOWLEDGEMENTS

I would like to thank the anonymous reviewers for their valuable comments. This work is supported by NPST Program 09-INF851-02, by King Saud University.

#### REFERENCES

- [1] Rivest, R.L., Shamir, A., Adleman, L.: A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM* 21, 120–126 (1978)
- [2] Sattam S. Al-Riyami and Kenneth G. Paterson. Certificateless public key cryptography. In C. S. Laih, editor, *Asiacrypt 2003*, volume 2894 of *Lecture Notes in Computer Science*, pages 452–473. Springer Berlin / Heidelberg, 2003.
- [3] J. Baek, R. Safavi-Naini, and W. Susilo. Certificateless public key encryption without pairing. In J. Zhou and J. Lopez, editors, *Proceedings of the 8th International Conference on Information Security (ISC 2005)*, volume 3650 of *Lecture Notes in Computer Science*, pages 134–148. Springer-Verlag, 2005.
- [4] Al-Riyami, S.: Cryptographic schemes based on elliptic curve pairings. Ph.D. thesis, Royal Holloway, University of London (2004). Available from <http://www.isg.rhul.ac.uk/~kp/sattthesis.pdf>
- [5] Al-Riyami, S.S., Paterson, K.G.: CBE from CL-PKE: A generic construction and efficient schemes. In: S. Vaudenay (ed.) *Public Key Cryptography – PKC 2005*, *Lecture Notes in Computer Science*, vol. 3386, pp. 398–415. Springer-Verlag (2005)
- [6] Fujisaki, E., Okamoto, T.: How to enhance the security of publickey encryption at minimal cost. In: H. Imai, Y. Zheng (eds.) *Public Key Cryptography*, *Lecture Notes in Computer Science*, vol. 1560, pp. 53–68. Springer-Verlag (1999)
- [7] Libert, B., Quisquater, J.J.: On constructing certificateless cryptosystems from identity based encryption. In: M. Yung, Y. Dodis, A. Kiayias, T. Malkin (eds.) *Public Key Cryptography – PKC 2006*, *Lecture Notes in Computer Science*, vol. 3958, pp. 474–490. Springer-Verlag (2006)
- [8] Zhang, Z., Wong, D.S., Xu, J., Feng, D.: Certificateless public-key signature: Security model and efficient construction. In: J. Zhou, M. Yung, F. Bao (eds.) *Applied Cryptography and Network Security*, *Lecture Notes in Computer Science*, vol. 3989, pp. 293–308. Springer-Verlag (2006)
- [9] Cheng, Z., Comley, R.: Efficient certificateless public key encryption (2005). Available from <http://eprint.iacr.org/2005/012/>
- [10] Dent, A.W., Libert, B., Paterson, K.G., “Certificateless Encryption Schemes Strongly Secure in the Standard Model”, Springer Berlin / Heidelberg *Lecture Notes in Computer Science*, pp. 344-359, vol. 4939, 10.1007/978-3-540-78440-1\_20, 2008
- [11] Waters, B.: Efficient identity-based encryption without random oracles. In: R. Cramer (ed.) *Advances in Cryptology – EUROCRYPT 2005*, *Lecture Notes in Computer Science*, vol. 3494, pp. 114–127. Springer-Verlag (2005)
- [12] Boyen, X., Mei, Q., Waters, B.: Direct chosen ciphertext security from identity-based techniques. In: *Proc. of the 12th ACM Conference on Computer and Communications Security*, pp. 320–329 (2005)
- [13] Yum, D.H., Lee, P.J.: Generic construction of certificateless encryption. In: A.L. et al. (ed.) *Computational Science and Its Applications ICCSA 2004: Part I*, *Lecture Notes in Computer Science*, vol. 3043, pp. 802–811. Springer-Verlag (2004)
- [14] Yum, D.H., Lee, P.J.: Identity-based cryptography in public key management. In: S.K. Katsikas, S. Gritzalis, J. Lopez (eds.) *Public Key Infrastructure: First European PKI Workshop (EuroPKI 2004)*, *Lecture Notes in Computer Science*, vol. 3093, pp. 71–84. Springer-Verlag (2004)
- [15] Libert, B., Quisquater, J.J.: On constructing certificateless cryptosystems from identity based encryption. In: M. Yung, Y. Dodis, A. Kiayias, T. Malkin (eds.) *Public Key Cryptography – PKC 2006*, *Lecture Notes in Computer Science*, vol. 3958, pp. 474–490. Springer-Verlag (2006)
- [16] Galindo, D., Morillo, P., Rafols, C.: Breaking Yum and Lee generic constructions of certificate-less and certificate-based encryption schemes. In: A.S. Atzeni, A. Liyo (eds.) *Public Key Infrastructure: Third European PKI Workshop (EuroPKI 2006)*, *Lecture Notes in Computer Science*, vol. 4043, pp. 81–91. Springer-Verlag (2006)

- [17] Alexander W. Dent, "A survey of certificateless encryption schemes and security models". *Int. J. Inf. Secur.* 7, 5, 349-377. DOI=10.1007/s10207-008-0055-0, Sep 2008.
- [18] Bentahar, K., Farshim, P., Malone-Lee, J., Smart, N.P.: Generic constructions of identity-based and certificateless KEMs (2005). Available from <http://eprint.iacr.org/2005/058>
- [19] Cramer, R., Shoup, V.: Design and analysis of practical publickey encryption schemes secure against adaptive chosen ciphertext attack. *SIAM Journal on Computing* 33(1), 167–226 (2004)
- [20] Huang, Q., Wong, D.S.: Generic certificateless key encapsulation mechanism. In: J. Pieprzyk, H. Ghodosi, E. Dawson (eds.) *Information Security and Privacy (ACISP 2007)*, Lecture Notes in Computer Science, vol. 4586, pp. 215–299. Springer-Verlag (2007)
- [21] Abe, M., Gennaro, R., Karosawa, K., Shoup, V.: Tag-KEM/DEM: A new framework for hybrid encryption. In: R. Cramer (ed.) *Advance in Cryptology – Eurocrypt 2005*, Lecture Notes in Computer Science, vol. 3494, pp. 128–146. Springer-Verlag (2005)
- [22] Dent, A.W., Libert, B., Paterson, K.G.: Certificateless encryption schemes strongly secure in the standard model (2007). Unpublished Manuscript
- [23] M. Naor and M. Yung. Public-key cryptosystems provably secure against chosen ciphertext attacks. In *Proc. 22nd Symposium on the Theory of Computing, STOC 1990*, pages 427{437. ACM, 1990.
- [24] A. Sahai. Non-malleable non-interactive zero knowledge and adaptive chosen-ciphertext security. In *40th Annual Symposium on Foundations of Computer Science, FOCS '99*, pages 543{553. IEEE Computer Society, 1999.
- [25] Y.H. Hwang, J.K. Liu, S.S.M. Chow, Certificateless public key encryption secure against malicious KGC attacks in the standard model, *Journal of Universal Computer Science* 14 (3) (2008) 463–480.
- [26] J.K. Liu, M.H. Au, W. Susilo, Self-generated-certificate public key cryptography and certificateless signature/encryption scheme in the standard model, in: R. Deng, P. Samarati (Eds.), *Proceedings of the Second ACM symposium on Information, Computer and Communications Security (ASIACCS'07)*, ACM, New York, 2007, pp. 273–283. Also *Cryptology ePrint Archive*, Report 2006/373, <<http://eprint.iacr.org/2006/373>>.

# Binary versus Multi-Valued Logic Realization of the AES S-Box

M. Abd-El-Barr<sup>1</sup>, and A. Al-Farhan<sup>1</sup>

<sup>1</sup>Information Science Department, Kuwait University, Kuwait

**Abstract** - The Sub-byte (S-Box) is a crucial block in the Advanced Encryption Standard (AES) cryptosystem. There has been growing interest in reporting improvement in the hardware realization of the S-Box. In this paper we consider the hardware realization of the S-Box using both binary and multi-valued logic (MVL). We also provide a comparison between the binary and the MVL realization and also with existing S-Box realization.

**Keywords:** Advanced Encryption Standard (AES), S-Box, Binary Hardware realization, Multiple-Valued Logic (MVL), Chip Area, Delay Performance

## 1 Introduction

The Advanced Encryption Standard (AES) is a symmetric-key block cipher algorithm used to encrypt/decrypt data worldwide. Data to be encrypted by the AES is divided into equally sized blocks each is called a *state*. The algorithm performs a series of mathematical operations on each state based on the Substitution-Permutation Network principle to produce the cipher text. Each of the repeated encryption round includes four operations: *sub-byte*, *shift row*, *mix column*, and *add round key*. Among the four operations, the sub-byte operation is performed using what is called the substitution box (S-Box). The S-Box performs substitutes each individual byte of a given state by a different byte as a means for achieving what is called confusion [1].

There exist a number of techniques for realization of S-Box. Those are divided into hardware, software, and combined hardware/software techniques, see the classification presented in [2]. A number of research efforts were devoted to the optimization of the byte substitution both in time, hardware complexity, and power consumption [3]-[12]. In this paper, we consider the hardware realization of the S-Box using both binary and multiple-valued logic (MVL).

The paper is organized as follows. Section 2 provides some background material for the topics covered in the paper. In Section 3, we present the related work. In Section 4, we present the proposed binary and MV realization of the S-Box. In Section 5, we present a performance comparison between the existing S-Box realization and the proposed realization in terms of both chip area measured using gate count and circuit delay. Section 6 concludes the paper.

## 2 Background Material

In this section, we provide some background material.

**Definition 1:** An n-variable r-valued function,  $f(X)$ , is defined as a mapping  $f: R^n \rightarrow R$ , where  $R = \{0, 1, \dots, r-1\}$  is a set of r logic values,  $r \geq 2$  &  $X = \{x_1, x_2, \dots, x_n\}$  is a set of r-valued n-variables.

Examples one-variable and two-variable 4-valued functions  $f_1(x)$  and  $f_2(x, y)$  are shown in Fig. 1, while Fig. 2 shows the definition of a number of unary and two-variable r-valued functions.

X	0	1	2	3
$f_1(x)$	3	0	1	2

(a) One-variable

x	0	0	0	0	1	1	1	1	2	2	2	2	3	3	3	3
y	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
$f_2(x, y)$	0	2	2	3	2	2	3	2	3	3	1	1	1	2	3	0

(b) Two-variable

Fig. 1. Examples of 4-valued functions.

Name of Operator	Definition
Cycle (Cyclic)	$x \rightarrow^k = (x + k) \bmod r$
Successor	$x \rightarrow = (x + 1) \bmod r$
Predecessor	$x \leftarrow = (x - 1) \bmod r$
Negation	$\bar{x} = (r - 1) - x$
Window Literal	$x^a = \begin{cases} (r-1) & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$
Threshold Literal	$x^a = \begin{cases} (r-1) & \text{if } x \geq a \\ 0 & \text{otherwise} \end{cases}$

(a) Unary r-valued functions

Name of Operator	Definition
Min	$x \bullet y = \begin{cases} x & \text{if } x < y; \\ 0 & \text{otherwise} \end{cases}$
Max	$x + y = \begin{cases} x & \text{if } x > y; \\ 0 & \text{otherwise} \end{cases}$
Truncated sum	$x \oplus y = \min((r - 1), (x + y))$

(b) Two-variable r-valued functions

Fig. 2. Definition of unary and two-variable r-valued functions.

Fig. 3 shows an illustration of the S-Box architecture. The substitution values can be pre-computed and stored in one Look Up table (LUT). The substitution of a byte {xy} that belongs to  $GF(2^8)$  in hexadecimal notation can be obtained using Table I. Consider, for example, that we need to find the byte substitution for the byte {c6}. Referring to row (c) and column (6) in Table 1, the substitution value is simply {b4}.

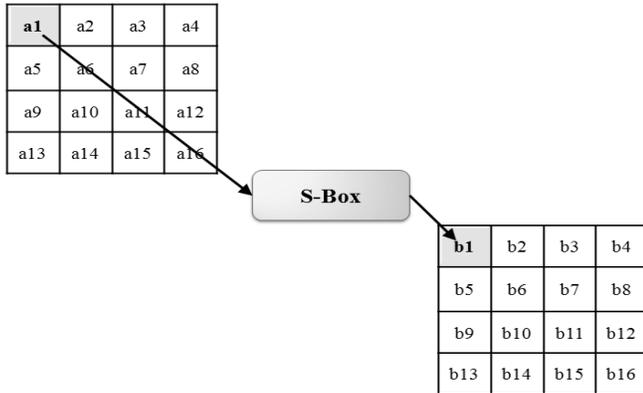


Fig. 3. The S-Box architecture.

Table I. S-BOX LOOKUP TABLE (LUT)

Y X	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
0	63	7c	77	7b	f2	6b	6f	c5	30	01	67	2b	fe	d7	ab	76
1	ca	82	c9	7d	fa	59	47	f0	ad	d4	a2	af	9c	a4	72	c0
2	b7	fd	93	26	36	3f	f7	cc	34	a5	e5	f1	71	d8	31	15
3	04	c7	23	c3	18	96	05	9a	07	12	80	e2	eb	27	b2	75
4	09	83	2c	1a	1b	6e	5a	a0	52	3b	d6	b3	29	e3	2f	84
5	53	d1	00	ed	20	fc	b1	5b	6a	cb	Be	39	4a	4c	58	cf
6	d0	ef	aa	fb	43	4d	33	85	45	f9	02	7f	50	3c	9f	a8
7	51	a3	40	8f	92	9d	38	f5	bc	b6	Da	21	10	ff	f3	d2
8	cd	0c	13	ec	5f	97	44	17	c4	a7	7e	3d	64	5d	19	73
9	60	81	4f	dc	22	2a	90	88	46	ee	b8	14	de	5e	0b	db
a	e0	32	3a	0a	49	06	24	5c	c2	d3	Ac	62	91	95	e4	79
b	e7	c8	37	6d	8d	d5	4e	a9	6c	56	f4	ea	65	7a	ae	08
c	ba	78	25	2e	1c	a6	b4	c6	e8	dd	74	1f	4b	bd	8b	8a
d	70	3e	b5	66	48	03	f6	0e	61	35	57	b9	86	c1	1d	9e
e	e1	f8	98	11	69	d9	8e	94	9b	1e	87	e9	ce	55	28	df
f	8c	a1	89	0d	bf	e6	42	68	41	99	2d	0f	b0	54	bb	16

### 3 Related Work

There has been a number of hardware realizations of the S-Box reported in the literature, see for example [13]-[16]. The realization in [14] is based on the use of a multi-stage PPRM (Positive Polarity Reed-Muller (AND-XOR)) realization of a composite field S-Box. It has been shown that the use 3-stage PPRM can lead to a typical 50% saving and about 1/3 power saving as compared to a realization using the SOP (Sum-of-Product) form.

The hardware realization reported in [13] is based on the idea considering the S-Box shown in Table 1 as consisting of eight output functions  $Z_0, Z_1, \dots, Z_7$  each function takes the form  $Z_i = f(x_0, x_1, x_2, x_3, y_0, y_1, y_2, y_3)$  where the  $x_i$ 's are used to identify a given row in the S-Box and the  $y_i$ 's are used to identify a given column in the S-Box. The technique analyzes each truth table of a target function searching for fragments, which correspond to selected types of sub-functions which are easily minimized in the current-mode gate algebra. For example a minimized form of the function  $Z_0$  has been shown to require 49 product terms using the Quine-McCluskey method.

### 4 Proposed S-Box Realizations

Byte Substitution is one of the essential repeated operations in AES. It is performed for each bytes of an AES 16 byte state. The S-Box lookup table accepts eight binary bits as input and generates eight binary bits as output. In pursuing the hardware realization of the S-Box we use the model shown in Fig. 4.

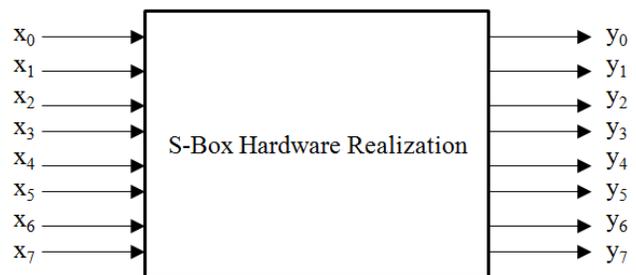


Fig. 4. S-Box Hardware Realization Model.

If we extract the least significant bits of all S-Box bytes shown in Fig. 5, then we end up with table Z0 shown in Fig. 6. The same idea applies to the rest of the bits resulting in eight sheets containing single bit cells.

X	01100011	01111100	01110111	01111011	...	10101011	01110110
	11001010	10000010	11001001	01111101	...	01110010	11000000
	10110111	11111101	10010011	00100110	...	00110001	00010101
	00000100	11000111	00100011	11000011	...	10110010	01110101
	00001001	10000011	00101100	00011010	...	00101111	10000100
	01010011	11010001	00000000	11101101	...	01011000	11001111
	11010000	11101111	10101010	11111011	...	10011111	10101000
	01010001	10100011	01000000	10001111	...	11110011	11010010
	11001101	00001100	00010011	11101100	...	00011001	01110011
	01100000	10000001	01001111	11011100	...	00001011	11011011
	11100000	00110010	00111010	00001010	...	11100100	01111001
	11100111	11001000	00110111	01101101	...	10101110	00001000
	10111010	01111000	00100101	00101110	...	10001011	10001010
	01110000	00111110	10110101	01100110	...	00011101	10011110
	11100001	11111000	10011000	00010001	...	00101000	11011111
	10001100	10100001	10001001	00001101	...	10111011	00010110
	Y						

Fig. 5. A portion of the S-Box represented in binary.

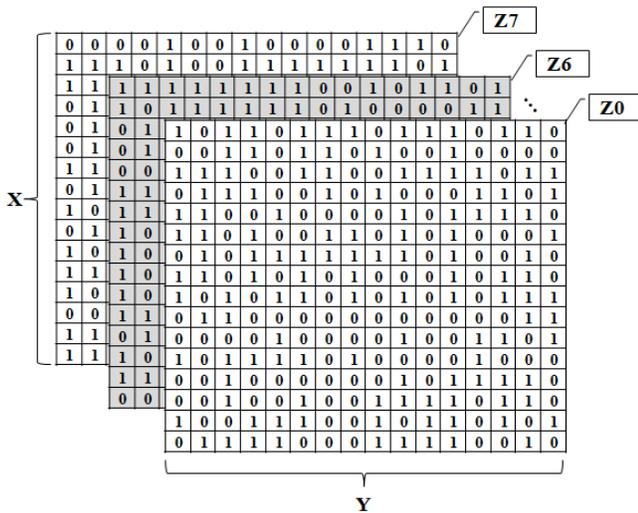


Fig. 6. S-Box sliced into eight boxes having single bit cells.

Here, we investigate incorporating multiple-valued data representations for S-Box input and/or output. Four valued logic is considered in this paper for simplicity as it maps easily to binary format. This will yield a completely different function minimization options which are expected to outperform the purely binary minimizations. Table II shows the binary and multiple-valued representations combinations for S-Box input and output. The four possible cases are discussed next.

Table II. S-BOX INPUT AND OUTPUT COMBINATIONS.

Input \ Output	Binary	Multiple-Valued
Binary	Case A (B-B)	Case B (M-B)
Multiple-Valued	Case C (B-M)	Case D (M-M)

### 4.1 Case A: Binary Input Binary Output (B-B)

is the basic and straight forward case in which all the possible 256 S-Box cells are indexed using eight binary input bits namely  $x_3, x_2, x_1, x_0, y_3, y_2, y_1, y_0$ . Rather than storing the output values as bytes, the S-Box values are generated as outputs of eight Boolean functions namely  $Z_7, Z_6, Z_5, Z_4, Z_3, Z_2, Z_1$ , and  $Z_0$ . Given that  $X$  and  $Y$  are four bit binary numbers. A general expression of the output function  $Z_i$  can be expressed as follows:

$$Z_i = f(x_3, x_2, x_1, x_0, y_3, y_2, y_1, y_0) \quad (1)$$

Where  $i \in \{0, 1, 2, 3, 4, 5, 6, 7\}$ ,  
 $X = x_3x_2x_1x_0$ ,  
 $Y = y_3y_2y_1y_0$ .

### 4.2 Case B: Multiple-Valued Input Binary Output (M-B)

Unlike the input in the previous case, Case B uses four variable four valued input. Equation 2 provides the formal specification of the binary output function  $Z_i$ .

$$Z_i = f(A_1, A_0, B_1, B_0) \quad (2)$$

Where  $i \in \{0, 1, 2, 3, 4, 5, 6, 7\}$ ,  
 $X = A_1A_0$ ,  
 $Y = B_1B_0$ .

### 4.3 Case C: Binary Input Multiple-Valued Output (B-M)

Similar to Case A, the input here is an eight bit binary number, while the output is a four variable of four-valued numbers. The output functions  $M_3, M_2, M_1$ , and  $M_0$  are obtained by pairing  $Z$  values discussed earlier. For instance, we can pair the  $Z$  functions as:  $M_3 = Z_7Z_6$ ,  $M_2 = Z_5Z_4$ ,  $M_1 = Z_3Z_2$ , and  $M_0 = Z_1Z_0$ . Other pairings are also possible by permuting the  $Z$  functions. This gives a total of 8 choose 2 combinations, where  $\binom{8}{2} = 8!/2!(8-2)! = 28$  combinations.

Equation 3 has the formal specification for this case:

$$M_j = f(x_3, x_2, x_1, x_0, y_3, y_2, y_1, y_0) \quad (3)$$

Where  $j \in \{0, 1, 2, 3\}$ ,  
 $X = x_3x_2x_1x_0$ ,  
 $Y = y_3y_2y_1y_0$ .

### 4.4 Case D: Multiple-Valued Input Multiple-Valued Output (M-M)

The totally multiple-valued logic case has the input of Case B and the paired output of Case C. Equation 4 has the formal specification.

$$M_j = f(A_1, A_0, B_1, B_0) \quad (4)$$

Where  $j \in \{0, 1, 2, 3\}$ ,  
 $X = A_1A_0$ ,  
 $Y = B_1B_0$ .

## 5 Minimization Results and Comparisons

In this section we compare our findings with the work done in [13]. The results reported in this paper are limited to the first two forms, i.e. B-B and B-M. To have a fair comparison, we compare our purely binary Case A to the binary minimization presented in [13]. Then we compare the current mode realization proposed in [13] to our Case B (B-M) realization.

Espresso [17] is a binary logic minimization program developed by Robert Brayton at the University of California, Berkeley. Later in 1986, Rudell developed a variant version of Espresso specialized for Multiple-Valued Logic synthesis

and named it Espresso-MV [18]. The later program is the one used in this paper for near optimal minimization. When used for binary minimization for Z0 of Case A the following minimization containing 44 product terms was obtained:

$$Z_0 = x_3x_1x_0y_3y_2y_1y_0' + x_3x_2x_1x_0y_3'y_1'y_0' + x_2x_1x_0y_3'y_2'y_0 + x_3x_2x_1y_3y_2'y_0' + x_3x_1'y_3y_2y_1y_0' + x_3x_0y_3'y_2'y_1y_0' + x_3x_1x_0'y_3'y_1'y_0' + x_3x_2x_1x_0y_3'y_2y_0' + x_3x_2x_1x_0'y_3'y_1'y_0' + x_2x_1x_0y_3y_2y_1y_0' + x_1x_0y_3'y_2'y_1y_0' + x_1x_0'y_3y_2y_1y_0' + x_3x_2x_1x_0'y_3'y_1'y_0' + x_2x_1x_0'y_3y_2y_0' + x_3x_2x_0'y_3'y_2'y_0' + x_3x_2x_1x_0'y_2'y_1'y_0' + x_3x_1x_0y_3y_2y_1'y_0' + x_3x_2x_1y_3y_2y_1y_0' + x_3x_2x_1x_0'y_3y_2'y_1' + x_2x_0'y_3y_2'y_1'y_0' + x_3x_2x_1'y_2'y_1y_0' + x_3x_2x_1'y_2'y_1'y_0' + x_1x_0'y_3y_2'y_0' + x_3x_2x_0'y_3y_2'y_1' + x_2x_1x_0'y_3y_2y_0' + x_3x_2x_1y_3'y_0' + x_3x_1x_0'y_3y_2y_0' + x_2x_1y_3y_2y_1'y_0' + x_3x_1x_0'y_2y_1y_0' + x_3x_1y_3'y_2'y_1'y_0' + x_2x_1x_0'y_3'y_2y_1' + x_1x_0'y_3y_1'y_0' + x_3x_1y_3'y_2y_1'y_0' + x_3x_2x_0y_3y_2'y' + x_3x_2'y_3'y_1y_0' + x_3x_2x_1x_0y_2y_1'y_0' + x_3x_2x_1'y_3y_2y_1' + x_3x_2x_1x_0y_3'y_2'y_1y_0' + x_3x_2x_0y_3y_2'y_1'y_0' + x_3x_1'y_3'y_2'y_1y_0' + x_3x_2x_0y_3'y_2'y_1' + x_3x_2x_1'y_3'y_2y_1'y_0' + x_3x_2x_1x_0y_1y_0' + x_2x_1y_2'y_1y_0' \quad (5)$$

Table III compares the binary minimization given in [13] using Quine-McCluskey Algorithm to our binary Case A using Espresso-MV.

Table III. COMBINATIONAL CIRCUIT PARAMETERS COMPARISON OF Z0 FUNCTION.

Gate Type	Number of Gates		Number of Transistors per Gate	Number of Inputs		Whole Transistor Count	
	[13]	Case A		[13]	Case A	[13]	Case A
NAND8	1	1	18	8	8	18	18
NAND7	22	13	16	154	91	352	208
NAND6	22	24	14	132	144	308	336
NAND5	4	6	12	20	30	48	72
NOR49	1	0	100	49	0	100	0
NOR44	0	1	90	0	44	0	90
<b>Total</b>	<b>50</b>	<b>45</b>		<b>363</b>	<b>317</b>	<b>826</b>	<b>724</b>

The table shows that our findings have significantly less number of total transistors needed for Z0 minimization. The work in [13] unfortunately did not report results for the rest of Z values to compare with. However, we included these values here for our Case A shown in Table IV to provide the total gate count and transistor count required by the whole S-Box. The minimization expressions for the multiple-valued input binary output Case B were obtained using Espresso-MV. As an example, Parts of the minimized of Z0 was found to require 36 product terms. These are shown in equation (6). It should be noted that equation (6) is written using the MVL operators shown in Fig. 2.

$$Z_0 = 1 \bullet A_1^1 \bullet A_0^3 \bullet B_1^3 \bullet B_0^2 + 1 \bullet A_1^1 \bullet A_0^2 \bullet B_1^2 \bullet B_0^1 + 1 \bullet A_1^1 \bullet A_0^1 \bullet B_1^3 \bullet B_0^3 + 1 \bullet A_1^1 \bullet A_0^2 \bullet B_1^1 \bullet B_0^3 + 1 \bullet A_1^1 \bullet A_0^2 \bullet B_1^2 \bullet B_0^0 + 1 \bullet A_1^2 \bullet A_0^1 \bullet B_1^3 \bullet B_0^3 + 1 \bullet A_1^0 \bullet A_0^3 \bullet B_1^1 \bullet B_0^2 + 1 \bullet A_1^3 \bullet A_0^3 \bullet B_1^2 \bullet B_0^0 + \dots \quad (6)$$

Table IV. COMBINATIONAL CIRCUIT PARAMETERS COMPARISON OF Z FUNCTIONS FOR CASE A.

Gate Type	# of Trans. Per Gate	Number of Gates							Whole Transistor Count								
		Z0	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z0	Z1	Z2	Z3	Z4	Z5	Z6	Z7
NAND8	18	1	1	1	2	0	0	1	0	18	18	18	36	0	0	18	0
NAND7	16	13	14	9	9	12	18	15	17	208	224	144	144	192	288	240	272
NAND6	14	24	33	28	37	26	28	28	28	336	462	392	518	364	392	392	392
NAND5	12	6	2	7	1	7	6	4	4	72	24	84	12	84	72	48	48
NOR52	106	0	0	0	0	0	1	0	0	0	0	0	0	0	106	0	0
NOR50	102	0	1	0	0	0	0	0	0	0	102	0	0	0	0	0	0
NOR49	100	0	0	0	1	0	0	0	1	0	0	0	100	0	0	0	100
NOR48	98	0	0	0	0	0	0	1	0	0	0	0	0	0	0	98	0
NOR45	92	0	0	1	0	1	0	0	0	0	0	92	0	92	0	0	0
NOR44	90	1	0	0	0	0	0	0	0	90	0	0	0	0	0	0	0
<b>Zi Total</b>		45	51	46	50	46	53	49	50	724	830	730	810	732	858	796	812
<b>S-Box Total</b>		<b>390</b>							<b>6292</b>								

Table V compares the current mode minimization given in [13] to our Case B minimized using Espresso-MV. We observe that the total number of product terms is equal for our Case B and [13] solution when it comes to Z0 value. The total number of product terms required for a whole S-Box is found to be 297 terms, which is notably less than the number of terms in Case A.

Table V. PRODUCT TERMS COUNT COMPARISON.

S-Box Realization	Case B							Current Mode [13]	
Zi	Z0	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z0
Number of Product Terms	36	40	35	38	38	38	37	35	36
<b>S-Box Total Product Terms</b>	<b>297</b>							<b>NA</b>	

The minimization expressions for the binary input multiple-valued output Case C were obtained using Espresso-MV for the paired function Z1Z0. Parts of the 113 minimized terms are shown in equation (7):

$$Z_1Z_0 = 1 \bullet x_2x_1x_0'y_3'y_2y_1'y_0' + 1 \bullet x_2x_1x_0'y_3y_2'y_1y_0' + 1 \bullet x_2x_1x_0'y_3y_2y_1'y_0' + 2 \bullet x_3x_2x_1'y_3'y_2y_1'y_0' + 2 \bullet x_3x_2x_1y_3y_2'y_1y_0' + 3 \bullet x_3x_2x_0'y_3'y_2'y_1'y_0' + 1 \bullet x_3x_1x_0y_3'y_2'y_1y_0' + 3 \bullet x_2x_1x_0y_3'y_2'y_1y_0' + 1 \bullet x_3x_2x_1x_0y_3'y_2'y_0' + 2 \bullet x_3x_2x_0y_3'y_2y_1'y_0' + \dots \quad (7)$$

The number of product terms for one selected combination out of 28 available for Case C is shown in Table VI. The total number of product terms required for a whole S-Box is found to be 448 terms, which is relatively larger than the previous realizations.

Table VI. PRODUCT TERMS COUNT COMPARISON.

S-Box Realization	Case C			
	Z1 Z0	Z3 Z2	Z5 Z4	Z7 Z6
Zi Pair				
Number of Product Terms	108	117	110	113
S-Box Total Product Terms	<b>448</b>			

The remaining Case D needs more research and requires the use of other tools to obtain near optimal minimization.

## 6 Conclusions

In this paper new S-Box realization forms were proposed and compared to existing realizations. The results found here shows that Espresso heuristic minimization yields less number of transistors required for the particular functions that generate AES S-Box compared to Quine-McCluskey approach. Our result for Case A has a transistor count 14% less than the technique used in [13]. When blending multiple-valued logic with binary logic, the multiple-valued input binary output (Case B) resulted in better minimization compared to binary input binary output (Case A) and binary input multiple valued output (Case C). The totally MVL input/output case can be implemented using the recent findings by Abd-El-Barr et al. [19]. This is currently explored by the authors and would be implemented as part of future work.

## 7 Acknowledgements

The authors would like to acknowledge the financial support of Kuwait University under Grant WI 04/10 and the expected travel financial support from Kuwait Foundation for the Advancement of Science (KFAS).

## 8 References

- [1] Daemen, J., and Rijmen, V., AES proposal: Rijndael, First Advanced Encryption Standard (AES), 1998. Available: <http://www.nist.gov/aes>
- [2] Abd-El-Barr, M., and Al-Farhan, A., "A Highly Parallel Area Efficient S-Box Architecture for AES Byte-Substitution", Proceeding the 2<sup>nd</sup> International Conference on Security Science and Technology (ICSST 2014) to be held March 20-21, 2014, Manila Philippines. Number of pages is 5.
- [3] D. Chen, G. Shou, Y. Hu, and Z. Guo, "Efficient architecture and implementations of AES," in *Third International Conference on Advanced Computer Theory and Engineering (ICACTE)*, 2010, vol. 6, pp. 295-298.
- [4] A. Satoh, S. Morioka, K. Takano, and S. Munetoh, "A compact Rijndael hardware architecture with S-box optimization," in *Advances in Cryptology – ASIACRYPT*, 2001, LNCS Vol. 2248, pp. 239–254.
- [5] H. Li, "A parallel S-box architecture for AES byte substitution," in *International Conference on Communications, Circuits and Systems (ICCCAS)*, 2004, vol. 1, pp. 1-3.
- [6] D. Canright, "A very compact S-Box for AES," in *Cryptographic Hardware and Embedded Systems – CHES 2005*, Springer Verlag, vol. 3659 of Lecture Notes in Computer Science, pp. 441–455, 2005.
- [7] S. Tillich, M. Feldhofer, and J. Großschädl, "Area, delay, and power characteristics of standard-cell implementations of the AES S-box," in *Embedded Computer Systems: Architectures, Modeling, and Simulation*, Springer Verlag, vol. 4017 of Lecture Notes in Computer Science, pp. 457–466, 2006.
- [8] S. Nikova, V. Rijmen, and M. Schläffer, "Using Normal Bases for Compact Hardware Implementations of the AES S-Box," in *Security and Cryptography for Networks*, Springer Verlag, vol. 5229 of Lecture Notes in Computer Science, pp. 236-245, 2008.
- [9] V. Rijmen, "Efficient Implementation of the Rijndael S-box," [online]. Available: <http://www.esat.kuleuven.ac.be/rijmen/rijndael/sbox.pdf>
- [10] J. Wolkerstorfer, E. Oswald, and M. Lamberger, "An ASIC implementation of the AES SBoxes," in *Topics in Cryptology—CT-RSA 2002*, Springer Verlag, pp. 67-78, 2002.
- [11] G. Bertoni, M. Macchetti, L. Negri, and P. Frangneto, "Power-efficient ASIC Synthesis of Cryptographic Sboxes," in *Proceedings of the 14th ACM Great Lakes symposium on VLSI (GLSVLSI 2004)*, pp. 277-281, ACM Press, 2004.
- [12] V. Tomashau, "Pipeline AES S-box Implementation Starting with Substitution Table," LC Engineers Inc., USA, [Online]. Available: <http://www.design-reuse.com/articles/30375/pipeline-aes-s-box-implementation-starting-with-substitution-table.html>
- [13] O. Maslennikow, M. Rajewska, R. Berezowski, Hardware Realization of the AES Algorithm S-Block Functions in the Current-Mode Gate Technology, Proc. of the 9th Int. Experience of Designing and Application of CAD Systems in Microelectronics, CADSM, 2007, IEEE Catalog Number 07EX1594, pp. 211-217.
- [14] Morioka, S., and Satoh, A., "An Optimized S-Box Circuit Architecture for Low Power Design", Proceedings 4<sup>th</sup> International Workshop on Cryptographic hardware and Embedded Systems (CHES 02), pp. 172-186.
- [15] Moradi, A., Poschmann, A., Ling, S., Paar, C., and Wang, H., "Pushing the limits: A very compact and threshold implementation of AES", Lecture Notes in Computer Science, 6632, pp. 69.88.
- [16] Rais, M. and Qasim, S., "Efficient Hardware Realization of Advanced Encryption Standard Algorithm using Virtex-5 FPGA", International Journal of Computer Science and Network Security (IJCSNS), vol. 9, no. 9, September 2009, pp. 59-63.
- [17] University of California, Berkeley, "Espresso Source Code and Documentation," [online]. Available: <http://embedded.eecs.berkeley.edu/pubs/downloads/espresso/index.htm>
- [18] Rudell, R. L., "Multiple-Valued Logic Minimization for PLA Synthesis", Memorandum No. UCB/ERL M86-65 (Berkeley), 1986.
- [19] Abd-El-Barr, M., Khan, E. A., "Improved direct cover heuristic algorithms for synthesis of multiple-valued logic functions", International Journal of Electronics, vol. 101, no. 2, 2014, pp. 271-286.

# Privacy-Preserving Protocol for Reduced Cancer Risk on Daily Physical Activity

Hiroaki Kikuchi  
School of Interdisciplinary  
Mathematical Sciences,  
Meiji University  
4-21-1 Nakano, Nakano Ku,  
Tokyo, 164-8525  
kikn@meiji.ac.jp

Shigeta Ikuji  
Cyber Communications Inc.  
2-14-1, Higashi-shimbashi,  
Minato-ku, Tokyo, Japan  
s.ikuji@cci.co.jp

Manami Inoue  
National Cancer Center,  
5-1-1 Tsukiji, Chuo-ku,  
Tokyo 104-0045, Japan  
mminoue@gan2.res.ncc.go.jp

**Abstract**—In a field of medical information, an integration of multiple data sets will lead more accurate and effective results in comparison to the research using one data set. Medical information integration has a risk of disclosure of confidential information and hence datasets don't have the common identification information in order to reduce possible risk factor to determine the identity of an individual.

In this paper, we will examine the condition of privacy-preserving data mining for the actual cohort, studied in the National Cancer Center. We study the necessary condition for data sets without common identification information to be integrated and generating more accurate and effective results.

**Keywords**-privacy, privacy-preserving data mining, epidemiology, hypothesis testing

## I. INTRODUCTION

### A. Background

In a study of the risk factors for several kinds of cancer, a long-term large-scale investigations are performed. For examples, Cardis et al. at International Agency for Research on Cancer published the epidemiological study on the risk of cancer after low doses of ionizing radiation in [6]. They carried out the 15-countries collaborative study of cancer risk among nearly 600,000 radiation workers in the nuclear industry. The main result is that the excess relative risk for cancers other than leukemia was 0.97 per Sv, 95% confidence interval 0.14 to 1.97. They concluded that an excess risk of cancer exists, even at the low doses and dose rate typically received by nuclear workers.

In the epidemic study of cancer, the exposure of cancer is confidential and critical private information. With the era of big-data, we are continuously monitored by the ubiquitous sensors, the smart-phones, and the portable devices. By integrating multiple datasets, we perform epidemiological study more accurately. To do so, we have the following issues;

- privacy issues for patients in a confidential dataset. No cancer patient wants to be exposed even though it contributes for progress of medical study.
- inconsistent identities in multiple datasets. An institute identifies individuals with proprietary identifiers. In most cases, it is hard to assume a global identities. Hence,

to join multiple datasets with inconsistent identifiers, we need to find alternatives of identities.

A set of personal attributes, e.g., name, telephone numbers, can be used to identify individuals. However, it is not trivial to find the optimal combination of attributes because we have no guarantee of uniqueness of personal attributes. Hence, we need to study some model to configure the appropriate set of attributes.

### B. Our contributions

In this paper, we propose some secure schemes for estimation of relative risk of cancers in privacy-preserving way. Our schemes use cryptographic protocols, e.g., *Private Set Intersections (PSI)*, to carry out the epidemiological processing including set intersection for mortality rate, and evaluation of test statistics for hypothesis testing. The confidentiality of data can be preserved even after the intersection of two subsets is computed.

Our contributions of the paper are

- a privacy-preserving protocol for hypothesis testing using the set of personal attributes as a quasi-identifier,
- an experimental results of the proposed protocol to estimate a risk of cancer in terms of quantity daily physical activities.

## II. PRELIMINARY

1) *Relative Risk*: In a cohort study, given two groups of individuals with and without exposure, we examine the risk factors for a disease. The *relative risk* (RR) is the chance that a member of a group receiving some exposure will develop a disease relative to the chance that a member of an unexposed group will develop the same disease[14].

For example, consider a  $2 \times 2$  table of observed frequencies with a sample of size  $N$ , which is known as a *contingency table*, as shown in Table I. In the table,  $m_1$  individuals smoke and  $m_2$  do not. Among the smokers,  $a$  suffered from cancer at the time of the investigation, whereas  $c$  did not. The RR of smoking is defined as the probability of cancer in the exposed

TABLE I  
CONTINGENCY TABLE FOR A CASE-CONTROL STUDY

	Smoking	Nonsmoking	Total
Cancer	$a$	$b$	$n_1$
Noncancer	$c$	$d$	$n_2$
total	$m_1$	$m_2$	$N$

(smoking) group divided by the probability in the unexposed group, as follows.

$$\begin{aligned} RR &= \frac{Pr(\text{cancer}|\text{smoking})}{Pr(\text{cancer}|\text{non-smoking})} \\ &= \frac{a}{n_1} / \frac{c}{n_2} = \frac{a(c+d)}{(a+b)c} \approx \frac{ad}{bc} \end{aligned} \quad (1)$$

A RR greater than 1.0 implies that there is an increased risk of disease among those in the exposed group.

To examine the confidence of the RR, we test the null hypothesis:

$H_0$ : The proportion of people who suffer from cancer in the population who smoke daily is equal to the proportion of people who suffer from cancer among those who do not smoke,

against the alternative hypothesis:

$H_A$ : The proportion of people who suffer from cancer is not identical in the two populations.

Under the null hypothesis  $H_0$ , the expected counts are computed for each cell of the contingency table by multiplying two independent probabilities,  $Pr(\text{cancer}) = n_1/N$  and  $Pr(\text{smoking}) = m_1/N$ , as follows:

$$E_1 = \frac{n_1}{N} \frac{m_1}{N} N = \frac{n_1 m_1}{N}.$$

The *chi-squared test* is used to compare the observed frequencies in each category of the contingency table,  $O_i$ , with the expected frequencies,  $E_i$ . To perform the test for the counts in a  $2 \times 2$  contingency table, we compute

$$\begin{aligned} \chi^2 &= \sum_{i=1}^{2 \times 2} \frac{(|O_i - E_i| - 1/2)^2}{E_i}, \\ &= \frac{N(|ad - bc| \pm N/2)^2}{n_1 n_2 m_1 m_2}, \end{aligned} \quad (2)$$

where there are  $2 \times 2$  cells in the table, which are called the *degrees of freedom*. The probability distribution of  $\chi^2$  is approximated by a *chi-squared distribution* with  $(2-1)(2-1)$  degrees of freedom. Given a chi-squared distribution with one degree of freedom, the outcome of  $\chi_1^2 = 3.84$  cuts off the upper 5% of the tail of the distribution. Alternatively, if employ  $\chi$  with a normal distribution  $N(0, 1)$ , we can test whether

$$\chi = \frac{\sqrt{N-1}\{(ad - bc) \pm N/2\}}{\sqrt{n_1 n_2 m_1 m_2}}, \quad (3)$$

is less than  $Z(0.05/2) = 1.960$  with 95 % confidence.

In general, the following three are well-known schemes for identifying the size of intersection  $|X_A \cap X_B|$  of two private datasets, i.e., *private set intersection*.

1) AES03 (Commutative One-way Function) [9].

Agrawal, et. al. used a commutative Pohlig-Hellman cipher, which is performed only for active (i.e. not missing) elements and therefore is more appropriate for sparse datasets. Algorithm1 shows the procedure.

2) SSP (Secure Scalar Product)[11]

Scalar product of two vectors is performed securely using a additive homomorphic public-key algorithm.

3) FNP04 (Oblivious Polynomial Evaluation) [12]

The scheme presented in [12] uses oblivious polynomial evaluation that suffers from the linear relation between computational cost and the order of the polynomial.

---

### Algorithm 1 Secure Intersection Protocol

---

Input: Alice has subset  $X = \{x_1, \dots, x_{n_A}\}$ , Bob has subset  $Y = \{y_1, \dots, y_{n_B}\}$ .

Output: Intersection  $|X \cap Y|$ .

---

Let  $p$  and  $q$  be prime numbers such that  $p = 2q + 1$  and  $p > \max(n_A, n_B)$ . Let  $Z_p$  be a multiplicative group with order  $q$  and  $Z_q$  is a set of integer less than  $q$ . Let  $H$  be a secure hash function that maps into range  $Z_p$ .

- 1) Alice chooses random  $u \in Z_q$  and send to Bob  $H(x_1)^u, \dots, H(x_{n_A})^u \pmod p$  in random order.
  - 2) Bob chooses random  $v \in Z_q$  and send to Alice  $H(y_1)^v, \dots, H(y_{n_B})^v \pmod p$  and  $(H(x_1)^u)^v, \dots, (H(x_{n_A})^u)^v \pmod p$  as well.
  - 3) Alice computes  $(H(y_1)^v)^u, \dots, (H(y_{n_B})^v)^u$  and selects pairs  $(x_j, y_i)$  such that  $H(y_i)^{vu} = H(x_j)^{uv} \pmod p$ , whose number is the size of intersection  $= |X \cap Y|$ .
- 

### III. PRIVACY-PRESERVING HYPOTHESIS TESTING

#### A. Objective

In the example of risk of radiation, party  $A$  can be an agency that maintains a list of all workers exposed to dose of radiation. In many countries, there is a regulation specifying the limit of total annual dose of radiation and workers in nuclear-power station are supposed to declare the recode of dose of radiation. In Japan, working under more that 50 mSv is prohibited.

Party  $B$  is a hospital for cancer and keeps dataset of cancer patients. Both parties should keep the confidentiality of their dataset,  $X_A$  and  $X_B$ , but are interested in determining the correlation between the risk of cancer and the dose of radiation.

To clarify the correlation, we need to compare the death rates, or *mortality rate*, for both datasets. The morality rate needs to be adjusted for differences of distribution of ages in two datasets. Let  $X_{A,y}$  be a subset of  $X_A$  whose age is in between  $y$  and  $y+10$ , and the  $X_A$  can be partitioned as  $X_A = X_{A,30} \cup X_{A,40} \cup \dots \cup X_{A,80}$ . Then, the expected numbers of subjects to death can be known as *standardized mortality rate*.

#### B. Hypothesis Testing

The *Poisson distribution* is used to model discrete events that occur infrequently in time, e.g., cancer, and death. Let  $X$

be a random variable that represents the number of occurrences of some events over a given interval. Let  $\lambda$  be a constant that denotes the average number of occurrence of the event in an interval. If the probability that  $X$  assumes the value  $k$  is

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad (4)$$

then  $X$  is said to have a Poisson distribution with parameter  $\lambda$ .

Suppose we observe  $X = O$  deaths when  $E$  expected number of deaths is given. We consider a *Standardized Mortality Ratio* (SMR) defined by

$$SMR = \frac{O}{E} = \frac{\sum d_j}{\sum q_j n_j}, \quad (5)$$

where  $d_j$  is the observed number of deaths at the  $j$ -th age interval in the interested condition to be tested.

We wish to determine whether the SMR is close to 1 or not. Namely, if the SMR in workers in nuclear-power station is equal to that of ordinary SMR, the risk of radiation is not significant. Hence, we test null hypothesis

$$H_0 : \lambda = E$$

against the alternative hypothesis

$$H_1 : \lambda \neq E.$$

If we conduct a one-sided test,

$$\begin{aligned} p &= P(O|E) + P(O+1|E) + \dots, \\ &= 1 - \sum_{j=0}^{O-1} \frac{E^j}{j!} e^{-E} \end{aligned}$$

gives  $p$ -value of the test. Employing approximation of Poisson distribution when  $E \geq 5$ , the test statistic

$$Z = \frac{O - E \pm 0.5}{\sqrt{E}} \quad (6)$$

has an normal distribution  $N(0,1)$  with mean 0 and the standard deviation 1. Note that 0.5 is the constant. If we conduct a two-sided test, the test statistic satisfying

$$Z = \frac{|O - E| - 0.5}{\sqrt{E}} > Z(\alpha/2) \quad (7)$$

would reject the null hypothesis at the  $\alpha$  level of significance.

#### IV. PRIVACY-PRESERVING FOR ESTIMATION OF REDUCED CANCER RISK

##### A. Problem Definition

Consider Alice is a national cancer center that maintains comprehensive personal attributes for patients of gastric cancer, lung, colon, and so on. In our study, we focus on colon cancer because the risk of colon cancer has been known as significantly correlated to daily physical activity. Let  $X$  be the set of colon cancer own by Alice.

Bob is a party that has a dataset of exposure of interest on personal physical activity. The examples include a sport club that stores frequencies of exercises of members, a public health

center that periodically investigate citizen, or a commercial health company that monitors daily physical activity quantities from vital devices. The daily total physical activity level is a quantity, called a *metabolic equivalents* (METs), based on questionnaire about the hours/day in heavy physical works, the hours/day in walking, the hours/day in sitting, and the days/week in leisure-time sports or exercise [2]. With the METs score, Bob classifies the people into four ( $q = 4$ ) orthogonal classes; Lowest (L), Second (S), Third (T), and Highest (H), specified by four subsets of  $U$ ,  $Y_1, Y_2, Y_3$  and  $Y_4$ , respectively.

1) *Number of People with exact same name*: An institute identifies individuals with proprietary identifiers. Hence, to join multiple datasets with inconsistent identifiers, we need to find alternatives of identities. As for *quasi-identifier*, we study a set of personal attributes that are significant enough to identify unique person. For example, a name is known as almost unique attribute, but with some exceptions. Some peoples have the exactly same surname and given name. For instance, Figure 2 shows the population of a set of people in which  $x$  individuals have the exact same surname and given name in some datasets, JPHC<sup>1</sup>, Univac[1], and<sup>2</sup> Note both vertical and horizontal axis are plotted in log scale. In JPHC, there are about 100 thousand people with unique name ( $x = 1$ ), which becomes about 2 thousand when two people having the same name ( $x = 2$ ).

Figure 1 shows the number of people that  $x$  people have the identical name written in Chinese character (Kanji), and Japanese character (Hiragana). Based on the observation of the distribution of people with the exact same names, we find a mathematical model of *Zipf's law*, which states that the number of people,  $f(x)$ , with the  $x$ -th order is proportional to  $1/x$ . With fitting to the dataset, we have the population of  $x$  individuals with same name written in Hiragana as

$$f(x) = \frac{a}{x^s} = \frac{110000}{x^{3.87}}$$

where  $s$  is a constant characterized by given dataset.

A generalized Zipf's model allows to estimate the number of people with the same attribute. The total population  $D$  is given by

$$D = a \sum_{k=1} \frac{1}{k^{3.87}}. \quad (8)$$

By solving Equation 8, we have the constant  $a$ . For instance, the number of people with not unique name is given from total population in Japan 120 million as  $a = \frac{D}{(\sum_{k=1} \frac{1}{k^{3.87}})} \simeq \frac{D}{1.1} \simeq 109e^6$ . Consequently, we estimate that 109 million people have the same name written in Hiragana. Hence, the name in Hiragana is not significant to identify individuals.

2) *Combination of Attribute to Unique Identifier*: With our Zipf's model, we quantify the entropy of personal attribute  $S$ ,

<sup>1</sup>the Japan Public Health Center-based Prospective Study (JPHC)

<sup>2</sup>NTT telephone book 2001.

TABLE II  
ENTROPY OF PERSONAL ATTRIBUTES

	entropy [bit]	ambiguity	Max # of Duplicated IDs	note
name in Chinese char. (Kanji)	27	high	24	same name can be written in several ways.
name in Japanese char. (Kana)	N/A	low	30	several representations can be unified.
sex	1	none	61,020	male or female (1 bit)
birthday and year	15	none	86	365 days (15 bit)
mailing address	26	low	56	almost unique but several representations in font.
city (ku, machi, mura)	14	high	12,131	not very unique for historical reason.
prefecture (states)	6	none	22,336	unique.
c/a	2	high	n/a	ocasionally specified. not complete attribute.

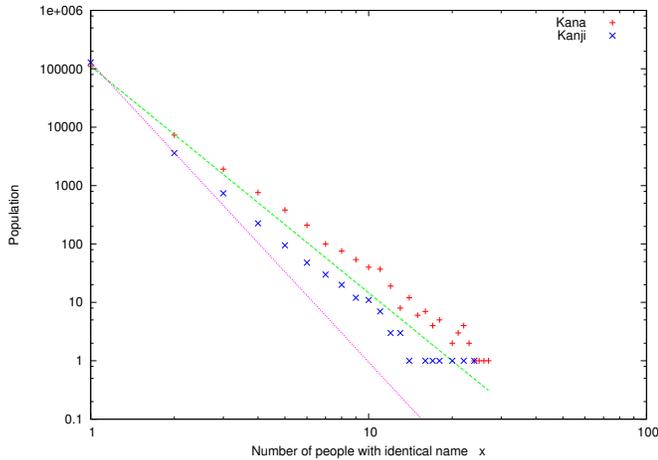


Fig. 1. Distribution of Population for numbers of people written with identical name (in Chinese character, labeled as “Kanji”, in Japanese character, labeled as “Kana”)

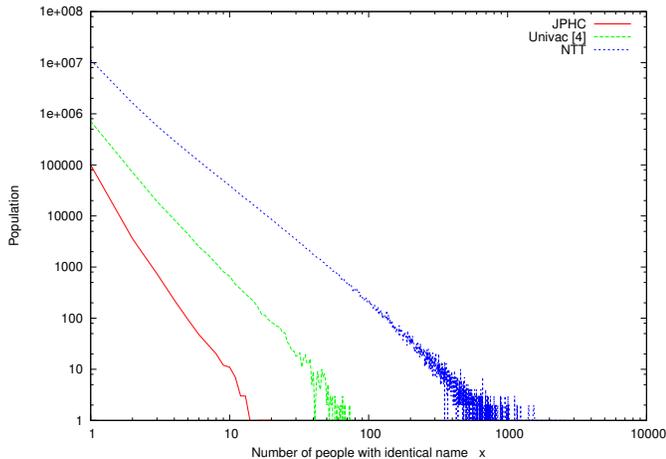


Fig. 2. Distributions of population for numbers of people written with identical name in dataset “JPHC”, Univac [4] and NTT

defined by

$$H(S) = \sum_k P(k) \log(P(k)) \text{ [bit/symbol]}$$

where  $P(k)$  is a probability of symbol  $k$ , i.e., value of attribute in  $S$ . Accordingly, the entropies of name in Kanji and in

Hiragana in JPHC dataset of 140,420 records are 14.63 and 13.71 bit/symbols, respectively.

In the similar way, we examine the JPHC dataset of 111,458 recodes<sup>3</sup> Table III shows the number of duplicated (more than two) recodes for some combinations of personal attributes. For example, option *A* uses the combination of name written in Hiragana and sex as quasi-identifier but there are 30,180 recodes can not be uniquely resolved because more than two recodes match the exactly same name and the same sex. The most common name is shared by 30 individuals.

According to Table III, we find options *D* (name, sex, birthday, address) and *E* (name, birthday, address) uniquely identify all individuals in JPHC dataset.

3) *Proposed Scheme*: Identities used by Alice are not consistent with that used by Bob. Instead of proprietary identities, we use a combination of significant personal attributes, e.g., names in Hiragana and birthday, as *quasi-identifiers*, which can be computed using secure hash function, e.g., SHA256, as

$$i = \text{Hash}(\text{name} \parallel \text{birthday} \parallel \text{address})$$

where  $\parallel$  is a symbol of concatenation. A person who belongs to both datasets  $A$  and  $B$  is uniquely identified by the quasi-identifier defined over  $U$ , the range of secure hash function.

We propose a cryptographic protocol between Alice with  $X$  and Bob with  $Y_1, \dots, Y_q$  for privacy-preserving for relative risk estimation without revealing identities to the other party in Algorithm 2. It uses Algorithm 1 as subprotocol.

## V. EVALUATION

### A. Experiment with JPHC Dataset

We have implemented the proposed protocol and apply to JPHC Dataset with 99,127 individuals. Table V shows the experimental results. For menfs third METs class ( $T$ ), the test statistic is  $\chi^2_T = 6.54$ . For chi-square distribution of 1 degree of freedom, we see the probability  $p < 0.025$  and hence we reject  $H_0$ . Therefore, daily physical activities in  $T$ , and  $H$  for menfs reduces the relative risk of cancer with significant level of confidence.

However, test statistics in womenfs are not significant in our experiment. Some possible reasons why the METs scores in women are not significant include the distribution of ages was

<sup>3</sup>it excludes the missing records in some attribute.

TABLE III  
ENTROPIES OF SOME COMBINATIONS OF PERSONAL ATTRIBUTES

option	set of attributes	entropy [bit]	Max # of duplicated recodes	# of unresolved records
A	name in Kana, sex	14	30	30,180
B	name in Kana, sex, birthday	30	2	16
C	name in Kana, sex, birthday, state	36	2	12
D	name in Kana, sex, birthday, address	56	0	0
E	name in Kana, birthday, address	55	0	0
F	name in Kana, address	40	2	16
G	sex, birthday, address	42	2	10

TABLE IV  
CONTINGENCY TABLE AND RELATIVE RISKS WITH TEST STATISTIC

	$ X \cap Y_p $	$ Y_p - (X \cap Y_p) $	$Y_p$	$RR$	$N_{(p)}$	$\chi^2_{(p)}$
$Y_1$	$c$	$d$	$c + d$	1.0	-	Reference
$Y_2$	$a_2$	$b_2$	$a_2 + b_2$	$\frac{a_2}{a_2 + b_2} / \frac{c}{c + d}$	$a_2 + b_2 + c + d$	$\frac{N_{(2)}(a_2 d - b_2 c)^2}{(a_2 + b_2)(c + d)(a_2 + c)(b_2 + d)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$Y_q$	$a_q$	$b_q$	$a_q + b_q$	$\frac{a_q}{a_q + b_q} / \frac{c}{c + d}$	$a_q + b_q + c + d$	$\frac{N_{(q)}(a_q d - b_q c)^2}{(a_q + b_q)(c + d)(a_q + c)(b_q + d)}$

**Algorithm 2** Privacy-Preserving Relative Risk Estimation for  $q \times 2$ -contingency table

Input: Alice has target subset  $X$  of a set of all identities  $U$ . Bob has  $q$  attribute subsets  $Y_1, Y_2, \dots, Y_q$  of  $U$ , where  $Y_1, \dots, Y_q$  are partition of  $U$ , i.e.,  $Y_1 \cup Y_2 \cup \dots \cup Y_q = U$  and  $Y_i \cap Y_j = \varnothing$  for all  $i \neq j$ .

Output: relative risks  $RR_1, \dots, RR_q$  of  $q$  attributes for target attribute  $X$

Step 1. Alice and Bob use Algorithm 1 to compute  $c = |X \cap Y_1|$  and

$$a_i = |X \cap Y_i|$$

for  $i = 2, \dots, q$ .

Step 2. Given  $c$  and  $a_i$ , Alice computes  $d = |Y_1| - c$ ,

$$b_i = |Y_i| - a_i$$

for  $i = 2, \dots, q$ .

Step 3. Alice computes relative risks  $RR_2, \dots, RR_q$  and the corresponding  $\chi^2_2, \dots, \chi^2_q$  according to Table IV.

skewed, or the other exposure factors such as smoking habit causes some effect.

For the reference, we show the existing results in [2] in Figure 4 as well as our results in Figure 3. We observe that both results look similar behavior correlation between cancer risk and the daily physical activities. Note that risk in [2] is evaluated with hazard ratio, or odds ratio, while our results are in relative risk which is more easy to compute in privacy-preserving way. It is know that when dealing with a rare disease, the relative risk can be approximated by the odds ratio.

**B. Performance**

With JPHC dataset of 140,000 individuals, we evaluate the proposed algorithm in our trial implementation. Table VI

TABLE V  
RELATIVE RISK OF COLON CANCER WITH RESPECT TO DAILY TOTAL PHYSICAL ACTIVITY LEVEL

	$X$	$ Y_i - (X \cap Y_i) $	$ Y_p $	$RR$	$\chi^2_{(i)}$
Men $n = 46,236$					
	(178)	(41,108)	(41,286)		
$L(16,374)$	79	13915	13994	1.00	Reference
$S(9,594)$	36	8229	8265	0.77	1.68
$T(9,085)$	25	7865	7890	0.56	6.54
$H(11,184)$	32	9830	9862	0.57	7.20
Women $n = 52,891$					
	(130)	(46,330)	(46,460)		
$L(17,404)$	40	14347	14387	1.00	Reference
$S(13,795)$	32	11703	11735	0.98	0.01
$T(11,865)$	32	10283	10315	1.12	0.21
$H(9,827)$	19	8473	8492	0.80	0.61

shows the experimental environment for measurement of performance. Figure 5 and 6 shows processing time with respect to size of datasets, 10k, 35k, 70k, 140k recodes. We iterate our test 10 times and takes an average value for processing time.

The dominant source of performance is the overhead of modular exponentiation necessary for each of 140k elements. Note that the process can be distributed with multiple machines to improve the performance.

TABLE VI  
EXPERIMENTAL ENVIRONMENTS

modulus size $ p $	2048 bit
order of $G$	160 bit
domain of $u \in \mathbb{Z}_p$	160 bit
Application impl.	Scala
SHA-1	Java sphlib
modulo	Java Big integer
Data Structure	Java HashSet Collection
OS	Ubuntu 12.10 amd64
CPU	Intel Celeron Processor G1610
Memory	4 GB (DDR3 SDRAM PC3-10600)
network speed	46 Mbps (measured values average)

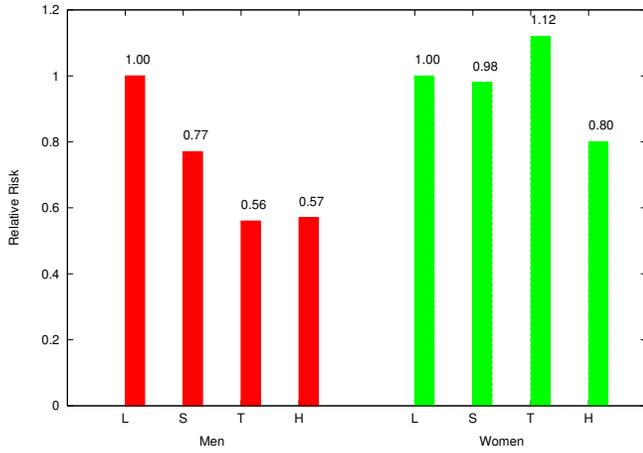


Fig. 3. Relative risk of colon cancer for METs estimated in the proposed algorithm

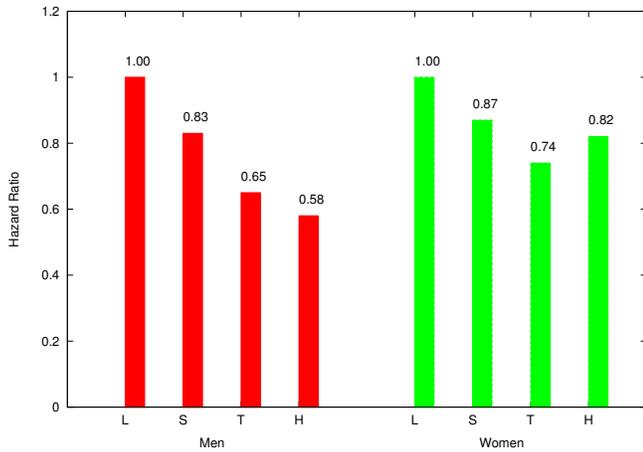


Fig. 4. Relative risk of colon cancer for METs estimated in [2]

C. Security

We assume that the parties are *honest-but-curious*, which is known as *semi-honest* model, with parties that own private datasets following protocols properly but trying to learn additional information about the datasets from received messages.

In [9], assuming the random oracle model and no hash collisions, and in semi-honest model, there is no polynomial-time algorithm that can distinguish between a random value and  $H(x)^u$  given  $x$ . This means that Algorithm 1 preserves the privacy of input subsets  $X$  and  $Y$ .

The threat of malicious party to figure out the particular individual depends on the chance to identify random numbers in the algorithm. The probability to pick the correct random number is

$$S = \frac{1}{|u|} = 2^{-160},$$

which is almost infeasible. Hence, the proposed scheme is secure against the malicious party to guess the individual.

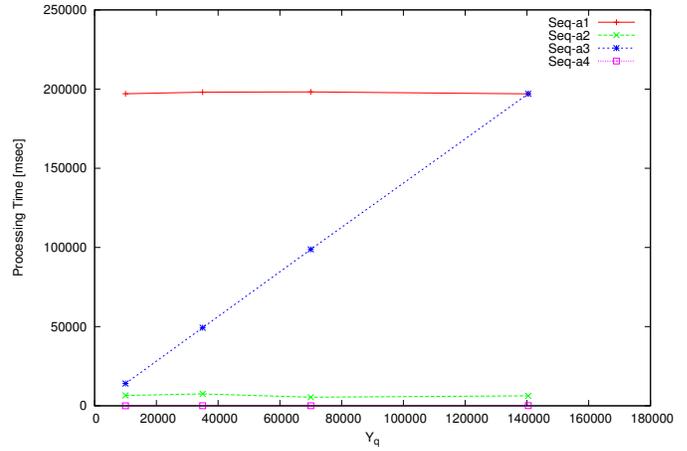


Fig. 5. Processing time for dataset size in Alice

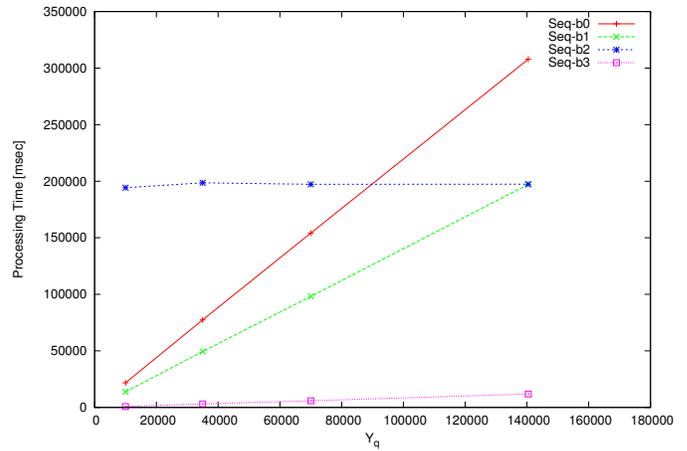


Fig. 6. Processing time for dataset size in Bob

D. Comparison to Existing Scheme

Our scheme is designed for privacy-preserving hypothesis testing for the first time. Hence, there is no direct comparison with the related works. Instead, we show the relationship between possible building blocks and the proposed two schemes in Table VII.

VI. CONCLUSIONS

We have proposed a privacy-preserving hypothesis testing for epidemic studies. The proposed schemes allow independent parties with confidential dataset to perform computing correction between any interested attributes. Our experiment

TABLE VII  
COMPARISON OF BUILDING BLOCKS

	AES03[9]	SSR[10]	FNP04[12]
intersection	available	no (size only)	available
input form	set	vector	set
complexity	$O(n)$	$O(N)$	$O(n^2)$
performance	360 elements/s	10 dim/s	-

shows the daily physical activities reduce a risk of cancer for some experiment in significant level of confidence.

#### ACKNOWLEDGEMENT

We thank Mr. Koyanagi, Mr. Taguchi, Mr. Kato, Mr. Ohkubo for help of our experiment and useful suggestions.

#### REFERENCES

- [1] Y. Tanaka, "Frequency of people with same first and last names", IPSJ SIG Technical Report on Natural Language (NL), Vol. 1977-NL-010, pp. 1-7, 1977.
- [2] Inoue et al. Daily Total Physical Activity Level and Total Cancer Risk in Men and Women: Results from a Large-scale Population-based Cohort Study in Japan. *Am J Epidemiol*, 168, pp. 391-403, 2008.
- [3] Rakesh Agrawal, Alexandre Evfimievski, and Ramakrishnan Srikant, "Information sharing across private databases", in proc. of ACM SIGMOD International Conference on Management of Data, 2003.
- [4] Report of a Joint WHO/FAO Expert Consultation, "Diet, nutrition and the prevention of chronic diseases", WHO technical report series, 916, pp. 100, 2003.
- [5] Hiroaki Kikuchi, Tomoki Sato and Jun Sakuma, "Privacy-Preserving Protocol for Epidemiology in Effect of Radiation", Proceedings of the 2013 Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS '13), pp. 831-836, IEEE, 2013.
- [6] E Cardis, M Vrijheid, M Blettner, E Gilbert, et al., "Risk of cancer after low doses of ionizing radiation: retrospective cohort study in 15 countries.", *BMJ Online First*, pp. 1-6, 2005.
- [7] Radiation Effects Association, Annual Report on radiation epidemiological study for workers in nuclear-power station, 2010. (written in Japanese)
- [8] Ministry of Health, Labor and Welfare, the Vital Statistics of Japan. (available from <http://www.mhlw.go.jp/english/database/index.html>)
- [9] Rakesh Agrawal, Alexandre Evfimievski, and Ramakrishnan Srikant, "Information sharing across private databases", in proc. of ACM SIGMOD International Conference on Management of Data, 2003.
- [10] Vaidya, J. & C. Clifton, Privacy preserving association rule mining in vertically partitioned data, in The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, SIGKDD, ACM Press, Edmonton, Canada, pp. 639-644, 2002.
- [11] Bart Goethals, Sven Laur, Helger Lipmaa, and Taneli Mielikainen, On Secure Scalar Product Computation for Privacy-Preserving Data Mining, In Choonsik Park and Seongtaek Chee, editors, The 7th Annual International Conference in Information Security and Cryptology (ICISC 2004), volume 3506, pp. 104-120, December 2.3, 2004.
- [12] M. J. Freedman, K. Nissim, and B. Pinkas, "Efficient private matching and set intersection", *EUROCRYPT 2004*, LNCS 3027, pp. 1-19, Springer-Verlag, 2004.
- [13] Dahlia Malkhi, Noam Nisan, Benny Pinkas, and Yaron Sella, "Fairplay – A Secure Two-Party Computation System", Usenix Security Symposium, 2004.
- [14] Marcello Pagano and Kimberlee Gauvreau, "Principles of biostatistics – 2nd ed.", *Brooks/Cole*, 2000.

**SESSION**  
**CRYPTOGRAPHIC TECHNOLOGIES II**

**Chair(s)**

**Dr. Sushil Kumar**  
**Univ. of Delhi - India**



# Formal Verification of Improved Numeric Comparison Protocol for Secure Simple Pairing in Bluetooth Using ProVerif

Kenichi Arai and Toshinobu Kaneko

Department of Electrical Engineering, Faculty of Science and Technology, Tokyo University of Science  
2641 Yamazaki, Noda, Chiba, 278-8510 Japan

**Abstract**—Recently, research has been conducted on automatic verification of cryptographic security protocols with the formal method. An automatic verifier is very useful because the risk of human error in such complicated protocols can be reduced. In this paper, we introduce our formalization of an improved Numeric Comparison protocol for Secure Simple Pairing in Bluetooth proposed by Yeh et al. and verify its security using ProVerif as an automatic cryptographic protocol verifier. As a result, we show that this improved protocol is subject to attacks. Moreover, we propose countermeasures against these attacks on this improved protocol. Our proposal provides this improved protocol with a higher level of security.

**Keywords:** Formal Verification, Security, ProVerif, Bluetooth, Secure Simple Pairing, Improved Numeric Comparison Protocol

## 1. Introduction

Generally, cryptographic protocols hold some security properties, but it is difficult for a non-security specialist to verify the security of cryptographic protocols because of their complexity. Recently, research has been conducted on automatic verification of security with the formal method. ProVerif[1], [2] is a known successful automatic verifier for cryptographic protocols defined in the formal model (the so-called Dolev–Yao model[3]). It is based on a representation of the protocol by Horn clauses[4] and can verify the security properties of secrecy and authentication. Therefore, many cryptographic protocols have been verified by ProVerif[5], [6], and it has succeeded in determining their security weaknesses. The main objective of using an automatic verifier is to reduce the risk of human error in such complicated protocols.

On the other hand, Bluetooth[7], [8], which is built into many devices, is a wireless communication standard connecting digital devices. The mutual authentication procedure between Bluetooth devices is called “pairing.” The pairing protocol is only able to select the protocol that uses a personal identification number (*PIN*) in Bluetooth Core Specification Version 2.0 + EDR[7] and earlier versions. However, an attacker can obtain the *PIN* with relative ease because many of these Bluetooth devices use a 4-digit *PIN* or a fixed *PIN* of commonly known values.

Secure Simple Pairing (SSP) is a protocol that improves the security weakness of the pairing protocol that uses a *PIN*. SSP is a new pairing protocol specified in Bluetooth Core Specification Version 2.1 + EDR[8]. It uses four models: “Numeric Comparison,” “Just Works,” “Out Of Band,” and “Passkey Entry.” However, potential attacks against SSP have been identified in recent years. Chang and Shmatikov proposed an attack against the Numeric Comparison protocol using ProVerif[5]. Lindell, Phan, and Mingard proposed an attack against the Passkey Entry protocol[9], [10]. Moreover, Nomura and Matsuo proposed a more practical attack on this protocol[11]. Yeh et al. pointed out a security weakness in the Numeric Comparison protocol different from Chang and Shmatikov, and proposed an improved version[12].

In this paper, we introduce our formalization of the improved Numeric Comparison protocol proposed by Yeh et al. and verify its security using ProVerif. This paper also discusses countermeasures against attacks on this improved protocol.

The remainder of this paper is organized as follows. In Section 2, we briefly introduce ProVerif, and in Section 3, we introduce Secure Simple Pairing and the improved Numeric Comparison protocol proposed by Yeh et al. In Sections 4 and 5, we introduce our formalization of the improved Numeric Comparison protocol. In Section 6, we show verification results of executing our formalization of the improved Numeric Comparison protocol on ProVerif. In Section 7, we present attacks against the improved Numeric Comparison protocol derived using ProVerif, and in Section 8, we discuss countermeasures against these attacks. We conclude the study in Section 9.

## 2. ProVerif

ProVerif is an automatic cryptographic protocol verifier in the formal model (the Dolev–Yao model) and enables the verification of the security of cryptographic protocols under the assumption that the cryptographic primitives are idealized. Since the attacker has complete control of the communication channels, it may read, modify, delete, and inject messages.

In ProVerif, cryptographic protocols are described using the syntax (grammar) of Blanchet’s process calculus, based on applied  $\pi$ -calculus[13]. The syntax used in this paper is shown as follows:

$M, N ::=$	terms
$a, b, c, k, m, n, s$	names
$x, y, z$	variables
$(M_1, \dots, M_k)$	tuple
$h(M_1, \dots, M_k)$	constructor/destructor
$M = N$	application
$M <> N$	term equality
$\text{not}(M)$	term inequality
	negation
$P, Q ::=$	processes
$0$	null process
$P Q$	parallel composition
$!P$	replication
$\text{new } n : t; P$	name restriction
$\text{in}(M, x : t); P$	message input
$\text{out}(M, N); P$	message output
$\text{if } M \text{ then } P \text{ else } Q$	conditional
$\text{let } x = M \text{ in } P \text{ else } Q$	term evaluation
$R(M_1, \dots, M_n)$	macro usage
$\text{event } e(M_1, \dots, M_n); P$	events

Please refer to [2] for more details of this syntax. The cryptographic protocol described using this syntax is automatically translated into a set of Horn clauses by ProVerif. It is also possible to describe the cryptographic protocol using a set of Horn clauses from the start.

A clause is a Horn clause if it contains at most one positive literal and is defined as  $F_1 \wedge \dots \wedge F_n \Rightarrow F$  ( $\equiv \neg F_1 \vee \dots \vee \neg F_n \vee F$ ), where  $n \geq 0$  and  $F$  is the only positive literal.  $F$  is also called a fact. A Horn clause  $F_1 \wedge \dots \wedge F_n \Rightarrow F$  means that, if all facts  $F_1, \dots, F_n$  are true, then  $F$  is also true. A Horn clause with no hypothesis  $\Rightarrow F$  is simply written as  $F$ . Here, a fact  $F = p(M_1, \dots, M_n)$  expresses a property of the messages  $M_1, \dots, M_n$ .  $p$  denotes predicates, and several predicates can be used. The term  $M$  also represents messages that are exchanged between the protocol's participants. The main predicate used by the Horn clause representation of protocols is `attacker`: the fact `attacker( $M$ )` means "the attacker may have the message  $M$ ." Actions of the adversary and the protocol participants can be modeled because of this predicate.

A set of Horn clauses obtained by automatic translation is called an initial clause. This is composed of the attacker's computational abilities, its initial knowledge, and the cryptographic protocol itself. ProVerif executes a resolution algorithm using initial clauses and verifies whether a fact in contradiction to the desired security property can be derived. When it can, there is an attack against the desired security property. In this case, ProVerif displays an explanation of the actions that the attacker has to perform to break the desired security property. Conversely, when the fact in contradiction to the desired security property cannot be derived, there is no attack. Please refer to [4], [14] for details of the resolution algorithm.

ProVerif can verify the security properties of secrecy[14]

and authentication[15]. The verification of secrecy is the most basic capability in ProVerif. To test secrecy of the term  $M$ , ProVerif attempts to verify that the state in which the term  $M$  is known to the adversary is unreachable. Authentication means "if Alice thinks she is talking to Bob, then she really is talking to Bob." Authentication can be defined using correspondence assertions. These are used to capture relationships between events that can be expressed in the form "if an event  $e$  has been executed, then event  $e'$  was previously executed."

### 3. Secure Simple Pairing

In this section, we briefly review SSP[8], and review the improved Numeric Comparison protocol proposed by Yeh et al.[12].

SSP is a new pairing protocol specified in Bluetooth Core Specification Version 2.1 + EDR and has two security goals: protection against passive eavesdropping and protection against man-in-the-middle attacks. It also aims to exceed the maximum security level provided by the use of a *PIN* with the pairing algorithm used in Bluetooth Core Specification Version 2.0 + EDR and earlier versions.

There are five phases of SSP. Phases 1,3,4, and 5 are the same for all protocols, whereas Phase 2 is different depending on the protocol used.

Phase 1 (Public Key Exchange) exchanges public keys using the Elliptic Curve Diffie-Hellman (ECDH) protocol, and a shared key between both devices is generated. Devices  $A$  and  $B$  first generate their own ECDH private-public key pair  $(sk_A, pk_A)$  and  $(sk_B, pk_B)$ , respectively, and then each sends its own public key to the other device. Devices  $A$  and  $B$  then each compute a shared key *DHKey* using the other device's public key and its own private key.

Phase 2 (Authentication Stage 1) exchanges authentication parameters used by Phases 3 and 4 and confirm these parameters. Phase 2 has three different protocols: Numeric Comparison, Out-of-Band, and Passkey Entry. Note that the Just Works model uses the Numeric Comparison protocol. These protocols are chosen based on the I/O capabilities of both devices. The Numeric Comparison protocol is designed for scenarios where both devices are capable of displaying a 6-digit number and of having the user enter "yes" or "no." The user is shown a 6-digit number on both displays, and then asked whether the numbers are the same on both devices. If "yes" is entered on both devices, the pairing is successful. An example of this protocol is the cell phone/PC scenario.

Phase 3 (Authentication Stage 2) confirms that both devices have successfully completed the exchange.

Phases 4 (Link Key Calculation) and 5 (LMP Authentication and Encryption) compute a link key and an encryption key, respectively. The link key is used to maintain the pairing. The final phase is the same as the final steps in legacy pairing.

Please refer to [8] for more details of the five phases of SSP.

### 3.1 Improved Numeric Comparison Protocol

The ECDH protocol (Phase 1 of standard SSP) has a security weakness against man-in-the-middle attacks because the senders of the public keys ( $pk_A, pk_B$ ) are not authenticated.

A man-in-the-middle attack occurs when a user wants to connect devices  $A$  and  $B$  but instead of directly connecting them, they unknowingly connect to an attacker device  $M$  that masquerades as the intended device.

To prevent this attack, a visual number confirmation is designed in the Numeric Comparison protocol (Phase 2 of standard SSP). However, Nokia Research Center conducted a usability experiment and pointed out the possibility that user error occurs when conducting visual number confirmation[16]. SSP has remained vulnerable to man-in-the-middle attacks because of this user error. Therefore, Yeh et al. proposed an improved Numeric Comparison protocol[12]. This improved protocol is composed of three phases and uses a  $PIN$  instead of confirming the displayed numbers. We review this improved protocol as follows (Figure 1):

#### Phase 1: Public Key Exchange and Authentication.

1. The user inputs a  $PIN$  on both devices  $A$  (the initiating device) and  $B$  (the responding device). Devices  $A$  and  $B$  then each generate their own ECDH private-public key pair ( $sk_A, pk_A$ ) and ( $sk_B, pk_B$ ), respectively.
2. Device  $A$  XORs  $pk_A$  with the  $PIN$  and sends  $A, IOcapA$  and its XOR value to device  $B$ . Here,  $IOcapA$  and  $A$  are the I/O capability of  $A$  and Bluetooth address of  $A$ , respectively.
3. Device  $B$  XORs the received ( $pk_A \oplus PIN$ ) with the  $PIN$  entered by the user to obtain  $pk_A$  and computes a shared key  $DHKey$  to  $pk_A$  and its own private key.  $DHKey$  is computed as a function P192 of these values. Device  $B$  then computes a commitment value  $C_B$  to  $DHKey, IOcapB, IOcapA, B,$  and  $A$ .  $C_B$  is computed as a function  $f1$  of these values. Device  $B$  XORs its own public key with the  $PIN$ , and then sends  $B, IOcapB,$  its XOR value, and  $C_B$  to device  $A$ . Here,  $IOcapB$  and  $B$  are the I/O capability of  $B$  and Bluetooth address of  $B$ , respectively.
4. Device  $A$  XORs the received ( $pk_B \oplus PIN$ ) with the  $PIN$  entered by the user to obtain  $pk_B$  and computes  $DHKey$  to  $pk_B$  and its own private key. Device  $A$  then computes  $C_B$  and compares its  $C_B$  with the received  $C_B$ . If this check fails, the protocol is aborted. Device  $A$  then computes a commitment value  $C_A$  to  $DHKey, IOcapA, IOcapB, A,$  and  $B$ .  $C_A$  is computed as the function  $f1$  of these values. Device  $A$  then sends  $C_A$  to device  $B$ .

5. Device  $B$  computes  $C_A$  and compares its  $C_A$  with the received  $C_A$ . If this check fails, the protocol is aborted.

#### Phase 2: Link Key Calculation.

Devices  $A$  and  $B$  compute a link key  $LK$  to the previously shared key ( $DHKey$ ) and the publicly exchanged data (constant string “ $btlk$ ,”  $A$ , and  $B$ ). This link key  $LK$  is computed as a hash function  $f2$  of these values.

#### Phase 3: LMP Authentication and Encryption.

After the link key is computed by Phase 2, devices  $A$  and  $B$  compute an encryption key  $K_C$  to the link key ( $LK$ ), the random number  $EN\_RAND$ , and ciphering offset number  $COF$ . This encryption key  $K_C$  is computed as a hash function  $E3$  of these values.

See [8] for details of the function P192 and hash functions  $f1, f2, E3$ . Note that the ECDH private-public key pair needs to be generated only once per device and may be computed in advance of pairing. Moreover, devices  $A$  and  $B$  may, at any time, choose to discard the ECDH private-public key pair and generate a new one, although it is not required to do so. These assumptions are the same as those of the standard SSP protocol.

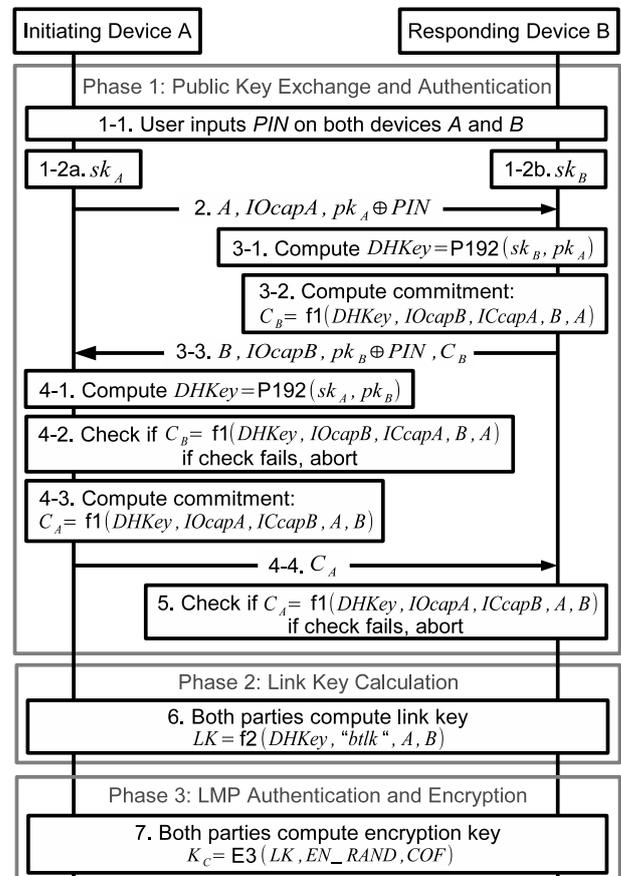


Figure 1: Improved Numeric Comparison protocol

## 4. Formalization of Cryptographic Primitives

Many cryptographic primitives can be modeled in ProVerif. It is necessary to model function P192, exclusive OR (XOR), hash functions, and symmetric encryption for the formalization of the improved Numeric Comparison protocol. Symmetric encryption and hash functions have already been modeled in ProVerif[2]. In this section, we formalize function P192 and XOR.

### 4.1 Function P192

Function P192 is defined as follows: given a scalar  $a$  and a point  $P$  on curve  $E$ , the value  $P192(a,P)$  is computed as the  $x$ -coordinate of the  $a$ -th multiple  $aP$  of point  $P$ . Therefore, function P192 means scalar multiplication on elliptic curves.

We formalize function P192 as follows:

```

1 type G1.
2 type scalar.
3 const P:G1 [data].
3 fun P192(scalar,G1):G1.
4 equation forall a:scalar, b:scalar;
   P192(a,P192(b,P)) = P192(b,P192(a,P)).

```

This formalization is based on the model of the Diffie–Hellman key agreement[2] that has already been formalized and models the ECDH key agreement. This key agreement relies on scalar multiplication in a cyclic group  $\mathbb{G}_1$  of prime order  $q$ ; let  $P$  be a point of  $\mathbb{G}_1$ . Alice selects a random scalar  $a$  and sends  $aP$  to Bob. Similarly, Bob selects a random scalar  $b$  and sends  $bP$  to Alice. Alice and Bob then compute  $a(bP)$  and  $b(aP)$ , respectively. These two keys are equal since  $a(bP) = b(aP)$  and cannot be obtained by an adversary who has  $aP$  and  $bP$  but neither  $a$  nor  $b$ . Here, the elements of  $\mathbb{G}_1$  have type G1, the scalars have type scalar, and  $P$  is point P. P192 also models scalar multiplication  $P192(a,P) = aP$ . The equation at Line 4 means that  $a(bP) = b(aP)$ .

### 4.2 Exclusive OR

We formalize XOR as follows:

```

1 fun xor(G1,G1):G1.
2 equation forall x:G1,y:G1; xor(xor(x,y),y) = x.
3 equation forall x:G1; xor(x,xor(x,x)) = x.
4 equation forall x:G1; xor(xor(x,x),x) = x.
5 equation forall x:G1,y:G1; xor(y,xor(x,x)) = y.

```

Here, xor models  $xor(a,b) = a \oplus b$ . Line 2 means that  $((x \oplus y) \oplus y) = x$ . Lines 3,4, and 5 refer to the idempotent and associative properties. Note that ProVerif cannot handle the commutative property (that means  $(x \oplus y) = (y \oplus x)$ ) together with the property of Line 2.

## 5. Improved Numeric Comparison Protocol

### 5.1 Declarations

The declarations specify a public channel  $c$  and cryptographic primitives (constructors/destructors). We formalize declarations as follows (for brevity, we omit cryptographic primitive declarations formalized in Section 4):

```

1 free c:channel.
2 free PIN:G1[private].
3 type tag.
4 const COF,EN_RAND,btlk:tag.
5 (* Shared key encryption *)
6 fun enc(bitstring, G1): bitstring.
7 reduc forall x:bitstring,y:G1;
   dec(enc(x,y),y) = x.
8 (* Hash functions *)
9 type nonce.
10 type key.
11 fun f1(G1,tag,tag,tag,tag):nonce.
12 fun f2(G1,tag,tag,tag):key.
13 fun E3(key,tag,tag):key.

```

Here, we assume that a  $PIN$  of the same value is always input to devices  $A$  and  $B$ . That is, the  $PIN$  always uses the same value. Please refer to [2] for details of function/type declarations.

### 5.2 Security Properties

Authentication can be defined using correspondence assertions[2]. The syntax to query a basic (non-injective) correspondence assertion is **query**  $x_1:t_1, \dots, x_n:t_n$ ; **event** $(e(M_1, \dots, M_j)) \implies$  **event** $(e'(N_1, \dots, N_k))$ . The query is satisfied if for each occurrence of the event  $e(M_1, \dots, M_j)$ , there is a previous execution of the event  $e'(N_1, \dots, N_k)$ . When the query is not satisfied, the cryptographic protocol of the verification target is subject to an “impersonation attack.” The definition of the basic (non-injective) correspondence assertion is also insufficient to capture authentication in cases where a one-to-one relationship between the number of protocol runs performed by each participant is desired. Injective correspondence assertions capture the one-to-one relationship and are denoted as **query**  $x_1:t_1, \dots, x_n:t_n$ ; **inj-event** $(e(M_1, \dots, M_j)) \implies$  **inj-event** $(e'(N_1, \dots, N_k))$ . This correspondence asserts that for each occurrence of the event  $e(M_1, \dots, M_j)$ , there is a distinct earlier occurrence of the event  $e'(N_1, \dots, N_k)$ . When this query is not satisfied, the cryptographic protocol of the verification target is subject to a “replay attack.”

The main objective of SSP is mutual authentication of devices  $A$  and  $B$ . Accordingly, when device  $A$  reaches the end of the protocol with the belief that it has done so with device  $B$ , then device  $B$  has indeed engaged in a session with device  $A$ . The opposite is also true for device  $B$ . We declare four events as follows.

**event** beginAkey(G1,G1), which is used by device  $B$  to record the belief that the initiator whose public key and

shared key are supplied as a parameter has commenced a run of the protocol with device *B*. **event** endAkey(G1,G1), which denotes that device *A* believes it has successfully completed the protocol with device *B*. This event is executed only when device *A* believes it runs the protocol with device *B*. Device *A* supplies its public key and shared key *DHKey* as the parameter. **event** beginBkey(G1,G1), which denotes device *A*'s intention to initiate the protocol with an interlocutor whose device public key and shared key are supplied as a parameter. **event** endBkey(G1,G1), records device *B*'s belief that it has completed the protocol with device *A*. Device *B* supplies its public key and shared key *DHKey* as the parameter.

If device *A* believes it has completed the protocol with device *B*, and hence executes the event endAkey, then there should have been an earlier occurrence of the event beginAkey, indicating that device *B* started a session with device *A*. Moreover, the relationship should be injective. A similar property should hold for device *B*.

In addition, we test whether the shared key *DHKey* is secret at the end of the protocol. The reason for testing the secrecy of *DHKey* is because the link key and encryption key are computed using *DHKey*. *DHKey* is a name created by variables such as *DHKeyA* and *DHKeyB*, while the standard secrecy queries of ProVerif deal with the secrecy of private free names. To solve this problem, the following general technique is used in ProVerif: instead of directly testing the secrecy of the shared keys, ProVerif uses them as session keys to encrypt some free name and test the secrecy of that free name. For example, in the process for device *A*, we describe enc(secretA,*DHKeyA*) at the end of the protocol and test the secrecy of secretA. SecretA is secret if and only if *DHKeyA* (that is, the shared key *DHKey* that device *A* has) is secret. We proceed symmetrically for device *B* using secretB.

The ProVerif code to verify the properties of secrecy and authentication can be described as follows:

```

14 (* Secrecy queries *)
15 free secretA, secretB:bitstring[private].
16 query attacker(secretA); attacker(secretB).
17 (* Authentication queries *)
18 event endAkey(G1,G1).
19 event beginAkey(G1,G1).
20 event endBkey(G1,G1).
21 event beginBkey(G1,G1).
22 query x:G1,y:G1; inj-event(endAkey(x,y)) ==>
    inj-event(beginAkey(x,y)).
23 query x:G1,y:G1; inj-event(endBkey(x,y)) ==>
    inj-event(beginBkey(x,y)).
24 (* Secrecy assumptions *)
25 not attacker(new skA).
26 not attacker(new skB).
27
28 (* Device A *)
29 let processA(skA:scalar, pkA:G1,
    A:tag, B:tag, IOcapA:tag, IOcapB:tag) =
30   out(c, (A, IOcapA, xor(pkA, PIN)));
31   in(c, (X:tag, IOcapB':tag, m1:G1, CB1:nonce));
32   let pkX=xor(m1, PIN) in

```

```

33   let DHKeyA=P192(skA, pkX) in
34   event beginBkey(pkX, DHKeyA);
35   let CB1'=f1(DHKeyA, IOcapB', IOcapA, X, A) in
36   if CB1=CB1' then
37     let CA1=f1(DHKeyA, IOcapA, IOcapB', A, X) in
38     out(c, CA1);
39     let LKA=f2(DHKeyA, btlk, A, X) in
40     let KCA=E3(LKA, EN_RAND, COF) in
41     event endAkey(pkA, DHKeyA);
42     out(c, enc(secretA, DHKeyA)).
43
44 (* Device B *)
45 let processB(skB:scalar, pkB:G1,
    A:tag, B:tag, IOcapA:tag, IOcapB:tag)=
46   in(c, (Y:tag, IOcapA':tag, m0:G1));
47   let pkY=xor(m0, PIN) in
48   let DHKeyB=P192(skB, pkY) in
49   event beginAkey(pkY, DHKeyB);
50   let CB1=f1(DHKeyB, IOcapB, IOcapA', B, Y) in
51   out(c, (B, IOcapB, xor(pkB, PIN), CB1));
52   in(c, CA1:nonce);
53   let CA1'=f1(DHKeyB, IOcapA', IOcapB, Y, B) in
54   if CA1=CA1' then
55     let LKB=f2(DHKeyB, btlk, Y, B) in
56     let KCB=E3(LKB, EN_RAND, COF) in
57     event endBkey(pkB, DHKeyB);
58     out(c, enc(secretB, DHKeyB)).
59
60 (* Main *)
61 process
62   new skA:scalar; let pkA = P192(skA, P) in
63   new skB:scalar; let pkB = P192(skB, P) in
64   new IOcapA:tag; out(c, IOcapA);
65   new A:tag; out(c, A);
66   new IOcapB:tag; out(c, IOcapB);
67   new B:tag; out(c, B);
68   ((!processA(skA, pkA, A, B, IOcapA, IOcapB)) |
69   (!processB(skB, pkB, A, B, IOcapA, IOcapB)))

```

Queries for secrecy and authentication are specified in Lines 15–16 and Lines 18–23, respectively. Lines 25–26 refer to security assumptions and inform ProVerif that the attacker cannot have the ECDH private key  $sk_A$  and  $sk_B$ . Process macros for devices *A* and *B* are specified in Lines 29–42 and Lines 45–58, respectively. The main process is also specified in Lines 61–69. This process begins by constructing the ECDH private–public key pair  $(sk_A, pk_A)$  and  $(sk_B, pk_B)$  for devices *A* and *B*, respectively. IOcapA, IOcapB, A, and B are then output on the public communication channel *c*, ensuring they are available to the adversary. An unbounded number of instances of processA and processB are then instantiated with the relevant parameters.

## 6. Verification Results

Verification results of executing our formalization of the improved Numeric Comparison protocol on ProVerif are shown in Table 1.

This means that the non-injective authentication of device *B* to *A*, secrecy for device *A* (secrecy of secretA), and secrecy for device *B* (secrecy of secretB) hold; whereas the injective authentications of device *A* to *B* and of device *B* to *A*, and the non-injective authentication of device *A* to *B* are violated.

Property		Result
Secrecy for Device A		True
Secrecy for Device B		True
Injective Authentication	A to B	False
	B to A	False
Non-Injective Authentication	A to B	False
	B to A	True

Table 1: Security properties

The non-injective authentication of device  $A$  to  $B$  is “false,” meaning that device  $B$  may end the protocol thinking it has been talking to device  $A$  when device  $A$  has never run the protocol with device  $B$ . This means an impersonation attack. When the injective authentication of device  $A$  to  $B$  (device  $B$  to  $A$ ) is false, it means that replay attacks are possible for the attacker. If secrecy for device  $A$  (device  $B$ ) is “true,” it means that the attacker cannot obtain the shared key  $DHKey$ . If the non-injective authentication of device  $B$  to  $A$  is true, it means that impersonation attacks are impossible for the attacker.

## 7. Derived Attacks

In this section, we review attacks against the improved Numeric Comparison protocol derived using ProVerif.

### 7.1 Replay Attacks

When the  $PIN$  always uses the same value, devices  $A$  and  $B$  are subject to replay attacks. We explain this attack derived using ProVerif as follows.

Device  $A$  sends  $(A, IOcapA, pk_A \oplus PIN)$  and  $C_A (= f1(DHKey, IOcapA, IOcapB, A, B))$  to device  $B$  in Phase 1, but these values are always the same. Therefore, an attacker can eavesdrop on communication between both devices during a certain session and obtain these values. The attacker then sends these values to device  $B$ , causing a replay attack. Device  $A$  is similarly compromised.

### 7.2 Impersonation Attacks

Device  $B$  is subject to impersonation attacks. We explain this attack derived using ProVerif as follows.

Device  $A$  sends  $(A, IOcapA, pk_A \oplus PIN)$  to device  $B$  in Phase 1. An attacker device  $M$  intercepts these values, modifies them to  $(B, IOcapB, m_M)$ , and sends  $(B, IOcapB, m_M)$  to device  $B$ . Here,  $m_M$  is a random number that device  $M$  generated. Device  $B$  then computes  $DHKey' = P192(sk_B, xor(m_M, PIN))$  and  $C'_B = f1(DHKey', IOcapB, IOcapB, B, B)$  using these modified values and sends  $(B, IOcapB, pk_B \oplus PIN, C'_B)$  to device  $A$ . Device  $M$  eavesdrops on the communication and obtains these values. Device  $A$  then computes  $DHKey$  and  $C_B$  and compares its  $C_B$  with the received  $C'_B$ . This check fails because its  $C_B$  is not equal to the received  $C'_B$ , and device  $A$  aborts the protocol. Therefore, device  $A$  is not subject to impersonation attacks.

Meanwhile, device  $M$  sends  $C'_B$  instead of sending  $C_A$  to device  $B$ . Device  $B$  then computes  $C'_A = f1(DHKey', IOcapB, IOcapB, B, B)$  and compares its  $C'_A$  with the received  $C'_B$ . However, this check succeeds because its  $C'_A$  is equal to the received  $C'_B$ . Therefore, device  $B$  is subject to impersonation attacks because there is no check (comparison) after Phase 2.

## 8. Countermeasures against Attacks

In this section, we propose countermeasures against the attacks mentioned in Section 7.

### 8.1 Countermeasure against Replay Attacks

In Phase 1, device  $A$  sends  $(A, IOcapA, pk_A \oplus PIN)$  and  $C_A$  to device  $B$ , and device  $B$  sends  $(B, IOcapB, pk_B \oplus PIN, C_B)$  to device  $A$ . However, these values are always the same; because  $(pk_A \oplus PIN)$  and  $(pk_B \oplus PIN)$  values are always the same,  $DHKey$  value is always the same. That is,  $C_A$  and  $C_B$  values are also always the same. Therefore, devices  $A$  and  $B$  are subject to replay attacks.

Since the values sent by devices  $A$  and  $B$  are always the same, we change them to a different value, changing the computational method of obtaining  $DHKey$ . We explain this countermeasure as follows.

**[Phase 1-2]:** Device  $A$  first selects a random number  $N_A$ . Device  $A$  then concatenates its own public key with  $N_A$ , XORs its concatenation value  $(pk_A || N_A)$  with the  $PIN$ , and sends  $(A, IOcapA, (pk_A || N_A) \oplus PIN)$  to device  $B$ .

**[Phase 1-3]:** Device  $B$  first selects a random number  $N_B$ . Device  $B$  XORs the received  $(pk_A || N_A) \oplus PIN$  with the  $PIN$  entered by the user to obtain  $pk_A, N_A$  and computes a shared key  $DHKey$  to  $N_A$ , its own random number, its own private key, and  $pk_A$ .  $DHKey$  is computed using a hash function  $f$  and function P192 as follows:

$$DHKey = f(N_A, N_B, P192(sk_B, pk_A)).$$

Here, we define hash function  $f$  using hash functions already defined in SSP. Device  $B$  also computes a commitment value  $C_B = f1(DHKey, IOcapB, IOcapA, B, A)$ . Device  $B$  then concatenates its own public key with its own random number, XORs its concatenation value  $(pk_B || N_B)$  with the  $PIN$ , and sends  $(B, IOcapB, (pk_B || N_B) \oplus PIN, C_B)$  to device  $A$ .

**[Phase 1-4]:** Device  $A$  XORs the received  $(pk_B || N_B) \oplus PIN$  with the  $PIN$  entered by the user to obtain  $pk_B, N_B$  and computes  $DHKey (= f(N_A, N_B, P192(sk_A, pk_B)))$ . Device  $A$  then computes  $C_B$  and compares its  $C_B$  with the received  $C_B$ . Device  $A$  then computes a commitment value  $C_A = f1(DHKey, IOcapA, IOcapB, A, B)$  and sends  $C_A$  to device  $B$ .

In this countermeasure, we add random numbers to the values sent by devices  $A$  and  $B$ , and can change these values

to a unique value. Moreover,  $DHKey$  can be changed to a unique value using hash function  $f$  and random numbers  $(N_A, N_B)$ .

Note that we have formalized with the assumption that the ECDH private-public key pair is generated only once per device.  $DHKey$  can be changed to a unique value by generating the ECDH private-public key pair for each pairing. Therefore, devices  $A$  and  $B$  are not subject to replay attacks by generating the ECDH private-public key pair for each pairing.

## 8.2 Countermeasure against Impersonation Attacks

In Phase 1-3, in receiving  $B$  and  $IOcapB$  sent from the attacker, device  $B$  is subject to impersonation attacks. Therefore, device  $B$  checks whether the Bluetooth address that it has received is its own ( $B$ ), and similarly checks whether the I/O capability that it received is its own ( $IOcapB$ ). We add the following procedures to Phase 1-3.

**[Phase 1-3]:** Device  $B$  compares the received Bluetooth address with its own. If it is not the same, the protocol is continued, otherwise the protocol is aborted (A1). Device  $B$  then compares the received I/O capability with its own. Again, the protocol is continued if the received I/O capability is not equal to its own, otherwise the protocol is aborted (A2).

```

46  in(c, (Y:tag, IOcapA':tag, m0:G1));
A1  if Y <> B then
A2  if IOcapA' <> IOcapB then
47  let pkY=xor(m0, PIN) in

```

## 8.3 Verification Results after Countermeasures

Verification results of executing our formalization of the proposed countermeasures on ProVerif are shown in Table 2.

Property		Result
Secrecy for Device $A$		True
Secrecy for Device $B$		True
Injective Authentication	$A$ to $B$	True
	$B$ to $A$	True
Non-Injective Authentication	$A$ to $B$	True
	$B$ to $A$	True

Table 2: Security properties after countermeasures

This means that all properties of secrecy and authentication are held. That is, we have succeeded in making replay and impersonation attacks against the improved Numeric Comparison protocol impossible.

## 9. Conclusion

In this paper, we introduced our formalization of the improved Numeric Comparison protocol for Secure Simple Pairing in Bluetooth proposed by Yeh et al. and verified its

security using ProVerif. We also formalized cryptographic primitives needed to formalize this improved protocol. As a result, we succeeded in deriving replay attacks and impersonation attacks against this improved protocol. We also proposed countermeasures against these attacks on the improved protocol, making them impossible. In future, we would like to verify the security of many cryptographic protocols using ProVerif.

## References

- [1] B.Blanchet(Project leader), "ProVerif: Cryptographic protocol verifier in the formal model," Available at <http://prosecco.gforge.inria.fr/personal/bblanche/proverif/>.
- [2] B.Blanchet, B.Smyth, and V.Cheval, "ProVerif 1.88: Automatic Cryptographic Protocol Verifier, User Manual and Tutorial," Available at <http://prosecco.gforge.inria.fr/personal/bblanche/proverif/manual.pdf>.
- [3] D.Dolev and A.Yao, "On the Security of Public Key Protocols," IEEE Transactions on Information Theory, Vol.29(2), pp.198–208, 1983. doi: 10.1109/TIT.1983.1056650.
- [4] B.Blanchet, "Using Horn Clauses for Analyzing Security Protocols," Formal Models and Techniques for Analyzing Security Protocols, Cryptology and Information Security Series, Vol.5, pp.86–111, 2011. doi: 10.3233/978-1-60750-714-7-86.
- [5] R.Chang and V.Shamatkov, "Formal Analysis of Authentication in Bluetooth Device Pairing," In Proc. of LICS/CALP Workshop on Foundations of Computer Security and Automated Reasoning for Security Protocol Analysis, 2007. Available at [http://www.cs.utexas.edu/~shmat/shmat\\_fcs07.pdf](http://www.cs.utexas.edu/~shmat/shmat_fcs07.pdf).
- [6] M.Christofi and A.Goujet, "Formal Verification of the mERA-Based eServices with Trusted Third Party Protocol," Information Security and Privacy Research, IFIP Advances in Information and Communication Technology, Vol.376, 2012, pp.299–314. doi: 10.1007/978-3-642-30436-1\_25.
- [7] Bluetooth SIG, "Bluetooth 2.0 + EDR Core Specification," 2004. Available at [https://www.bluetooth.org/docman/handlers/DownloadDoc.ashx?doc\\_id=40560](https://www.bluetooth.org/docman/handlers/DownloadDoc.ashx?doc_id=40560).
- [8] Bluetooth SIG, "Bluetooth 2.1 + EDR Core Specification," 2007. Available at [https://www.bluetooth.org/docman/handlers/downloaddoc.ashx?doc\\_id=241363](https://www.bluetooth.org/docman/handlers/downloaddoc.ashx?doc_id=241363).
- [9] A.Lindell, "Attacks on the Pairing Protocol of Bluetooth v2.1," In Blackhat USA, 2008.
- [10] R.Phan and P.Mingard, "Analyzing the Secure Simple Pairing in Bluetooth v4.0," Wireless Personal Communications, Vol.64(4), pp.719–737, 2012. doi: 10.1007/s11277-010-0215-1.
- [11] D.Nomura and K.Matsuo, "A Man-in-the-Middle Attack against Secure Simple Pairing in Bluetooth," IPSJ (Information Processing Society of Japan) Journal, Vol.53(9), pp.2225–2233,2012.(in Japanese)
- [12] T.Yeh, J.Peng, S.Wang, and J.Hsu, "Securing Bluetooth Communications," International Journal of Network Security, Vol.14(4), PP.229–235,2012.
- [13] M.Abadi and C.Fournet, "Mobile Values, New names, and Secure Communication," In Proc. of the 28th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, pp.104–115, 2001. doi: 10.1145/360204.360213.
- [14] B.Blanchet, "An Efficient Cryptographic Protocol Verifier Based on Prolog Rules," In 14th IEEE Computer Security Foundations Workshop, pp.82–96, 2001. doi: 10.1109/CSFW.2001.930138.
- [15] B.Blanchet, "From Secrecy to Authenticity in Security Protocols," Static Analysis, Lecture Notes in Computer Science Vol.2477, pp.342–359, 2002. doi: 10.1007/3-540-45789-5\_25.
- [16] E.Uzun, K.Karvonen, and N.Asokan, "Usability Analysis of Secure Pairing Methods," Financial Cryptography and Data Security, Lecture Notes in Computer Science Vol.4886, pp.307–324, 2007. doi: 10.1007/978-3-540-77366-5\_29.

# Simple method to find primitive polynomials of degree $n$ over $GF(2)$ where $2^n - 1$ is a Mersenne prime

Jiantao Wang<sup>1</sup>, Dong Zheng<sup>2</sup>, and Qiang Li<sup>2</sup>

<sup>1</sup>School of Information Security Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>2</sup>National Engineering Laboratory for Wireless Security, Xi'an University of Posts and Telecommunications, Xi'an 710121, Shaanxi Province, China

**Abstract**—The paper describes the group structure of cyclotomic cosets modula  $2^n - 1$ , the group is cyclic when  $2^n - 1$  is a prime. The integers modula  $2^n - 1$  can be regarded as the exponents of a primitive element  $\alpha \in GF(p^n)$ . The traces of  $\alpha^i$  show the same structure as the cyclic group of the cyclotomic cosets modula  $2^n - 1$ . The coefficients of the minimal polynomial of a specific  $\alpha^j$  consist of the sum of the traces of different  $\alpha^i$ , which follow the cyclic group structure. We demonstrate that all the primitive polynomials can be calculated fast through the permutation of the traces of  $\alpha^i$ .

**Keywords:** Finite Fields, Cyclotomic Cosets, Primitive Polynomial

## 1. Introduction

The theory of finite fields has played important roles in code design and cryptography[1], [2]. The irreducible polynomials of degree  $n$  over  $GF(p)$ , where  $p > 0$  is a prime, is of special interest[3], [4]. Many algorithms require the calculations of different irreducible polynomials of a fixed degree  $n$ .

There has been various methods for constructing irreducible polynomials of the same degree  $n$ [1], [2], [4] from a given primitive polynomial. And one direct way is to use the relations between the coefficients and the roots of the irreducible polynomials[1], [2], [5]. For a defining element  $\alpha$  of a finite field  $GF(p^n)$ , the coefficients of the minimal polynomials of different  $\alpha^k$  are the sum of different  $\alpha^t$ . This means that one specific power  $\alpha^t$  appears in different positions in the coefficients of minimal polynomials of different elements. In this paper, we show that the reason is the group structure of cyclotomic cosets. For a Mersenne prime, which is defined to be the primes of the form  $2^n - 1$ , the group structure of the cyclotomic cosets reduces the computing work to simple group permutations. The group structure can also explain why some former classical algorithms[2], [6] using the cubic root and permutation succeeded.

The paper are organized as follows. In Section 2 some preliminary results are given. Section 3 introduces our main

theory. Experiment results are given in Section 4. Section 5 concludes our work.

## 2. Newton Formula and Cyclotomic Cosets

We first give some preliminaries that are useful for our theory. In a finite field  $F = GF(p^n)$ , where  $p$  is a prime and  $n > 0$  is an integer, the trace function of an element  $\alpha \in F$  is defined as:

$$tr(\alpha) = \alpha + \alpha^p + \alpha^{p^2} + \cdots + \alpha^{p^{n-1}} \in GF(p) \quad (1)$$

Assume  $f(x)$  to be an irreducible polynomial over  $GF(p)$  of degree  $n$  whose roots are  $x_1 = \alpha, x_2 = \alpha^p, \cdots, x_n = \alpha^{p^{n-1}}$ . The elementary symmetric polynomials  $\sigma_1, \sigma_2, \cdots, \sigma_n$  are the coefficients of  $f(x)$ :

$$\begin{aligned} f(x) &= (x - x_1)(x - x_2) \cdots (x - x_n) \\ &= x^n - \sigma_1 x^{n-1} + \sigma_2 x^{n-2} \\ &\quad - \cdots + (-1)^{n-1} \sigma_{n-1} x + (-1)^n \sigma_n \end{aligned} \quad (2)$$

Another kind of symmetric polynomial is defined as:

$$\begin{aligned} s_k &= s_k(x_1, x_2, \dots, x_n) \\ &= x_1^k + x_2^k + \cdots + x_n^k \\ &= \sum_{1 \leq t \leq n} (\alpha^{p^{t-1}})^k, \end{aligned} \quad (3)$$

where  $k \geq 1$  is an integer.

The Newton Formula is [7, p.12]:

$$\begin{aligned} s_k - s_{k-1} \sigma_1 + \cdots + (-1)^i s_{k-i} \sigma_i \\ + \cdots + (-1)^{k-1} s_1 \sigma_{k-i} \\ + (-1)^k k \sigma_k = 0, \sigma_j = 0 \quad \text{for } j > n. \end{aligned} \quad (4)$$

The trace of  $\alpha^k$  equals to the symmetric polynomial  $s_k$  induced by the roots of  $f(x)$ . If  $f(x)$  is primitive, then  $\alpha^k$  can denote all the elements in the finite fields, and we can use the Newton Formula to compute the trace of any element of the finite field via linear iteration.

In the expansion of  $f(x)$ ,  $\sigma_i$  is the sum of all powers of  $\alpha$  having exponents which, when written as  $p$ -ary  $n$ -tuples,

have  $i$  ones and  $n - i$  zeroes. The exponents of  $\alpha$  in one trace function also have the same ones when written as  $p$ -ary  $n$ -tuples. So the coefficients  $\sigma_i$  could be decomposed into the sum of traces of some specific elements.

Cyclotomic cosets[2, p.42] are a classification of the non-zero residues modula  $p^n - 1$ . Each coset contains the numbers that are congruent to each other modula  $p^n - 1$  by multiplying a power of  $p$ , e.g.  $\{1, 2, 4\}, \{3, 6, 5\}$  are two cyclotomic cosets modula  $2^3 - 1 = 7$ . Every coset equals to a set of the exponents of the powers appeared in the trace function of a finite field element  $\beta = \alpha^k, k$  is an integer. All  $\alpha^i$  where  $i$  runs through a cyclotomic coset have the same minimal polynomial in the finite field[2].

### 3. Group Structure and Minimal Polynomials

We present our main theory in this section. According to number theory, the residues modula  $q = p^n - 1$  forms an Abelian group with respect to multiplication. The group is cyclic for  $p = 2$  and  $q = 2^n - 1$  prime, and we denote the cyclic group as  $G$ . And all the cyclotomic cosets have the same length  $n$  as all the irreducible polynomials are primitive polynomials for  $q$  prime.

Considering cyclotomic coset modula  $q, H = \{1, 2, 2^2, \dots, 2^{n-1}\}$ . It is also a cyclic subgroup of  $G$ . We have the following relations between  $G$  and  $H$ .

*Proposition 1:*  $H$  is a normal subgroup of  $G$ . The quotient group  $G/H$  is a cyclic group. Multiplying any  $1 \leq k < 2^n - 1$  to all the elements in  $G/H$  means a permutation of the quotient group.

*Proof:* Because  $G$  is commutative, the subgroups of  $G$  are all normal, so does  $H$ . Both  $G$  and  $H$  are cyclic,  $G/H$  is also cyclic by group theory.

Let  $H_1 \in G/H$ , then  $H_1 = k_1H$ , where  $1 \leq k_1 < 2^n - 1$ , then  $k \cdot H_1 = k \cdot k_1H = kH \cdot k_1H = k_2H$  where  $1 \leq k, k_2 < 2^n - 1$ . Consider  $kH$  as a group member of  $G/H$ , hence the theorem follows. ■

*Example 1:*  $p = 2, n = 5, q = 31, G = \{1, 2, \dots, 30\}, H = \{1, 2, 4, 8, 16\}$ .

Then  $G/H$  is an cyclic group of order 6.

$$\begin{aligned} H_1 &= H = \{1, 2, 4, 8, 16\}, \\ H_2 &= 3H = \{3, 6, 12, 24, 17\}, \\ H_3 &= 5H = \{5, 10, 20, 9, 18\}, \\ H_4 &= 7H = \{7, 14, 28, 25, 19\}, \\ H_5 &= 11H = \{11, 22, 13, 26, 21\}, \\ H_6 &= 15H = \{15, 30, 29, 27, 23\}, \\ G/H &= \{H_1, H_2, \dots, H_6\}. \end{aligned}$$

$H_2$  is a generator of  $G/H$ .  $H_2^2 = 9H = 5H = H_3, H_2^3 = 15H = H_6, H_2^4 = 45H = 14H = H_4, H_2^5 = 11H = H_5, H_2^6 = 2H = H_1$ .

We want to find all the primitive polynomials from a given primitive polynomial of degree  $n$  over  $GF(2)$ . The coefficients of the primitive polynomials consists of the traces of  $\alpha^k$  where  $k$  belongs to cyclotomic cosets leaders for a Mersenne prime  $q$ .

From finite field theory,  $q$  is the smallest integer such that  $\alpha^q = 1$ . So any  $\beta \in GF(q)$  can be written in the form  $\alpha^k$  and has the same order  $q$  as  $\alpha$ . This means the minimal polynomials of all  $\alpha^k$  where  $k$  belongs to different cyclotomic cosets, are all the primitive polynomials of degree  $n$ .

From the general structure of minimal polynomials discussed in the former section, the exponents of powers of  $\alpha$  contained in the coefficients of  $f(x)$  cover all the cyclotomic cosets. Every exponent needs to be multiplied by  $k$  to compute the minimal polynomial of a specific element  $\alpha^k$ . The numbers in the same cosets appear as a whole in the same coefficient of a primitive polynomial, as proved in Proposition 1. The coefficients of the minimal polynomial of  $\alpha^k$  are the sum of permuted elements of the quotient group defined in Proposition 1.

For example, the minimal polynomial of a primitive element  $\alpha \in GF(2^5)$  has the following form.

$$\begin{aligned} f_\alpha(x) &= x^5 + tr(\alpha)x^4 + (tr(\alpha^3) + tr(\alpha^5))x^3 \\ &\quad + (tr(\alpha^7) + tr(\alpha^{11}))x^2 + tr(\alpha^{15})x + 1 \end{aligned} \quad (5)$$

The minimal polynomial of any element  $\beta = \alpha^k \in GF(q)$  is a primitive polynomial of the same form shown in (5). The trace function has the exponent property  $tr((\alpha^k)^t) = tr(\alpha^{kt})$ , so the minimal polynomial of  $\beta$  can be represented by  $\alpha$ . Continued from (5), let  $k = 3$ , then:

$$\begin{aligned} f_\beta(x) &= x^5 + tr(\beta)x^4 + (tr(\beta^3) + tr(\beta^5))x^3 \\ &\quad + (tr(\beta^7) + tr(\beta^{11}))x^2 + tr(\beta^{15})x + 1 \\ &= x^5 + tr(\alpha^3)x^4 + (tr(\alpha^9) + tr(\alpha^{15}))x^3 \\ &\quad + (tr(\alpha^{21}) + tr(\alpha^{33}))x^2 + tr(\alpha^{45})x + 1 \quad (6) \\ &= x^5 + tr(\alpha^3)x^4 + (tr(\alpha^5) + tr(\alpha^{15}))x^3 \\ &\quad + (tr(\alpha^{11}) + tr(\alpha))x^2 + tr(\alpha^7)x + 1 \end{aligned}$$

The last step in the deduction is due to  $tr(\alpha^{31}) = 1$  and the trace is the same for the exponents of powers of  $\alpha$  in the same cyclotomic coset. Comparing (5) with (6), the coefficients of the minimal polynomials of  $\alpha$  and  $\beta = \alpha^3$  are permutations of the traces  $tr(\alpha), tr(\alpha^3), tr(\alpha^5), tr(\alpha^7), tr(\alpha^{11}), tr(\alpha^{15})$ .

A generator of the cyclic cyclotomic cosets group is needed to get all the primitive polynomials of degree  $n$ . Multiplying the generator gives a permutation chain among all the cyclic group elements. Then all the primitive polynomials can be calculated by iteration.

*Example 2 (continued from Example 1):*  $p = 2, n = 5, q = 31$ . Then  $f_1(x) = x^5 + x^3 + 1$  is a primitive

polynomial over GF(2) with a root  $\alpha$ . So  $\sigma_1 = \sigma_3 = \sigma_4 = 0, \sigma_2 = \sigma_5 = 1, s_1 = \sigma_1 = 0$ .

By Newton Formula,  $s_2 - \sigma_1 s_1 + 2\sigma_2 = 0, s_2 = 0$ , then  $s_3 = tr(\alpha^3) = 0, s_4 = 0, s_5 = tr(\alpha^5) = 1$ . For  $k > 5$ , we have  $s_k = s_{k-2} + s_{k-5}$ . We get  $s_7 = tr(\alpha^7) = 1, s_{11} = tr(\alpha^{11}) = 1, s_{15} = tr(\alpha^{15}) = 0$ .

The basic structure of the minimal polynomial of degree 5 over GF(2) is shown in (5). We use the form  $(\gamma_1 \gamma_2 \dots \gamma_m)$  to show a permutation  $\sigma$  over some elements  $\{\gamma_1, \gamma_2, \gamma_3, \dots, \gamma_m\}$  of a group where  $\sigma(\gamma_i) = \gamma_{i+1}$ , for  $1 \leq i \leq m - 1, \sigma(\gamma_m) = \gamma_1$ .

Example 1 shows that  $H_2 = 3H$  is a generator of the cyclic group  $G/H$ . The cyclic relation can be written in a permutation form.

$$\begin{aligned} & (H \ 3H \ 5H \ 15H \ 7H \ 11H) \\ & = (1 \ 3 \ 5 \ 15 \ 7 \ 11) \\ & = (tr(\alpha) \ tr(\alpha^3) \ tr(\alpha^5) \ tr(\alpha^{15}) \ tr(\alpha^7) \ tr(\alpha^{11})) \end{aligned}$$

This leads to the conversion from (5) to (6). So the other primitive polynomials of degree 5 over GF(2) can be computed by the permutation sequently.

$$\begin{aligned} f_{\alpha^5}(x) &= x^5 + tr(\alpha^5)x^4 + (tr(\alpha^{15}) + tr(\alpha^7))x^3 \\ &\quad + (tr(\alpha) + tr(\alpha^3))x^2 + tr(\alpha^{11})x + 1 \\ f_{\alpha^{15}}(x) &= x^5 + tr(\alpha^{15})x^4 + (tr(\alpha^7) + tr(\alpha^{11}))x^3 \\ &\quad + (tr(\alpha^3) + tr(\alpha^5))x^2 + tr(\alpha)x + 1 \\ f_{\alpha^7}(x) &= x^5 + tr(\alpha^7)x^4 + (tr(\alpha^{11}) + tr(\alpha))x^3 \\ &\quad + (tr(\alpha^5) + tr(\alpha^{15}))x^2 + tr(\alpha^3)x + 1 \\ f_{\alpha^{11}}(x) &= x^5 + tr(\alpha^{11})x^4 + (tr(\alpha) + tr(\alpha^3))x^3 \\ &\quad + (tr(\alpha^{15}) + tr(\alpha^7))x^2 + tr(\alpha^5)x + 1 \end{aligned}$$

Combining with the traces computed by Newton Formula, it follows:

$$\begin{aligned} f_{\alpha}(x) &= x^5 + x^3 + 1, \\ f_{\alpha^3}(x) &= x^5 + x^3 + x^2 + x + 1, \\ f_{\alpha^5}(x) &= x^5 + x^4 + x^3 + x + 1, \\ f_{\alpha^{15}}(x) &= x^5 + x^2 + 1, \\ f_{\alpha^7}(x) &= x^5 + x^4 + x^3 + x^2 + 1, \\ f_{\alpha^{11}}(x) &= x^5 + x^4 + x^2 + x + 1. \end{aligned}$$

### 4. Experiments

We give some numerical experiments to show the efficiency of our algorithm.

The next Mersenne prime after 31 is  $2^7 - 1 = 127$ , and we know its primitive root is 3. Starting from a given primitive polynomial of degree 7,  $f(x) = x^7 + x + 1$ , the next table shows the cosets  $3^k \text{ mod } 127$  and their binary representative,

Table 1: The Coset Leaders And Their  $s_k$

$k$	$3^k \text{ mod } 127$	binary form	$s_k$
1	3	1100000	0
2	9	1001000	0
3	27	1101100	1
4	81	1000101	1
5	116	0010111	0
6	94	0111101	1
7	28	0011100	1
8	84	0010101	1
9	125	1011111	1
10	121	1001111	1
11	109	1011011	0
12	73	1001001	1
13	92	0011101	0
14	22	0110100	0
15	66	0100001	0
16	71	1110001	0
17	86	0110101	1

from low digits to high digits, and their  $s_k$  computed by Newton Formula.

The number of ones in the binary form in the table shows the position the coset belongs in  $f(x)$ . The order of Table 1 shows the permutation structure. For a fixed permutation, one coset is replaced by a coset whose place has a fixed distance from the former one in Table 1. We compute all the other primitive polynomial according to the sum of the permuted  $s_k$  in Table 2. The polynomial is written in a short form, where 10101011 stands for  $x^7 + x^5 + x^3 + x + 1$ .

We have also tested the Mersenne prime  $2^{13} - 1, 2^{17} - 1, 2^{19} - 1$ , the result is too long for our paper, but we get all the polynomials in this simple way.

Table 2: The Primitive Polynomials Of Degree 7 Over GF(2)

$k$	minimal polynomial of $\alpha^k$
3	10101011
9	10111001
27	11110111
81	11100101
116	10010001
94	11110001
28	11111101
84	11001011
125	11000001
121	11010101
109	10011101
73	11101111
92	10100111
22	10001001
66	10001111
71	10111111
86	11010011

### 5. Conclusion

This paper associates the computation of minimal polynomial with the group structure of cyclotomic cosets modula  $2^n - 1$  for  $2^n - 1$  is a Mersenne prime. From the examples and experiments, we see the computation of the primitive polynomials is simple and efficient due to the cyclic group

structure of the cyclotomic cosets, and the usual knowledge of the tables of the sums and products in the finite field is not required. This cyclic group structure is also the reason for the “rational algorithm” in [2, p.48] and for the valid assignments yielded by permutation in [6].

## Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant No 61272037, Key Program of Natural Science Foundation of Shaanxi Province(Grant No.2013JZ020) and “New Generation of Broadband Wireless Communications Network” Ministry of Industry and Information Technology major projects (Project No.2013ZX03002004).

## References

- [1] R. Lidl and H. Niederreiter, *Introduction to finite fields and their applications*, Cambridge University Press, 1986.
- [2] S. W. Golomb and G. Gong, *Signal design for good correlation*, Cambridge University Press, 2005.
- [3] J. A. Gordon, “Very simple method to find the minimum polynomial of an arbitrary nonzero element of a finite field,” *Electronic Letters*, vol. 12, pp. 663–664, 1976.
- [4] V. C. da Rocha Jr. and G. Markarian, “Simple Method to Find Trace of Arbitrary Element of a Finite Field,” *Electronic Letters*, vol. 2, No. 7, Mar, 2006.
- [5] O. Ahmadi and A. Menezes, “On the number of trace-one elements in polynomial bases for  $GF(2^m)$ ,” *Designs, Codes and Cryptography*, vol. 37, pp. 493–507, 2005.
- [6] S. W. Golomb, “Obtaining specified irreducible polynomials over finite fields,” *SIAM J. ALG. DISC. MATH.* vol. 1, No. 4, December, 1980.
- [7] H. M. Edwards, *Galois Theory*. Springer, 1997.

# Digital Identities and Accountable Agreements in Web Applications

J. Kannisto<sup>1</sup>, J. Harju<sup>1</sup>, and T. Takahashi<sup>2</sup>

<sup>1</sup>Department of Pervasive Computing, Tampere University of Technology, Tampere, Finland

<sup>2</sup>National Institute of Information and Communications Technology, Tokyo, Japan

**Abstract**—*The current security model of web applications is centered around trust to the web service provider. The consequence of this is that the digital identities that people use are not under their own control. Therefore, the accountability of actions is questionable as objective evidence cannot be gathered. The technology – public key encryption – that enables the use of digital identities has been available for long, but it has not been utilized by the end-users in the web domain. We explore the obstacles for end-users' public key based identities in the web, and present ways to overcome them.*

**Keywords:** identity management, public key, javascript

## 1. Introduction

Digital identities are required to do business online. These digital identities come in many shapes and forms. For example, website accounts, email addresses, credit cards and public keys, can all be considered to be digital identities. On the basic level identification is the authentication of identity [1]. Therefore, digital identities need properties whose ownership can be authenticated by computers, either through a federated authentication or directly. For example, identity associated with a password can be authenticated by those who have means to verify the password and trust that it is sufficient proof of authenticity. Usually this requires, that the verifier ends up in the possession of the password. Compared to passwords, public key based identities are superior in this regard as the authentication done by a signature can be validated without sensitive information, only thing needed is knowledge of the correct public key for a particular identity.

With public key based identities, a major problem has traditionally been that the mapping between a public key and a meaningful identity context requires trust. Some proposals have been made to reduce the need for this linking, and basically establish systems where the public key itself is a meaningful identifier. The assumption is that any public key used frequently enough will eventually become a pseudonymous digital identity, even when it may not be the intention[2]. Such opportunistic identity management assumes that public key identities can be mapped into a meaningful identity context later, if they are taken into use first.

Web is available on a multitude of very heterogeneous devices, which are used by practically all user segments. Applications and protocols are delivered using HTML and

JavaScript, and the user does not have to do any cumbersome installation steps to use the applications. Instead of only an information network, the web has become one of the most important application platforms. Increasingly, even native applications use web compatible interfaces to communicate with each other. However, web applications have lagged behind desktop applications in their usage of cryptography, and in particular public key identities. Multitude of issues contribute to this: private key storage, security issues of JavaScript, lack of libraries, and consequently also lack of services requiring, for example, public key authentication. We present a case for public key based authentication as a web application.

The presented use case is to build a contract on the service's accepted security level (Security Service Level Agreement) [3]. For the Service Provider it is necessary to get confirmation that the user accepts the security level of the service, and is willing to submit data to the service. This confirmation needs to be accountable to an identity that is associated with the owner of the data, so that the confirmation can free the service provider from liability. Specifically, a security service that the user uses, may not want to carry the liability itself, so the authentication has to be made by the user in a way that is accountable for all the parties.

### 1.1 Contribution

The paper reviews current authentication mechanisms that are used in the web, and different proposals that have been made to improve it. In the later part, the paper presents a set of boundary conditions for distributed secure end-user web applications, and solutions that are able to overcome some these restraints. Private key storage is identified as one of the major issues hindering the adoption of public key systems [4]. This issue is looked in detail in Section 3.3, and several solutions to this problem are compared.

Part of the presented case is a REST based service model for private key operations, such as signing of data. Such service enables the use of the same public key identity in different applications and devices, in a simple unified manner. To verify the practicality of the scheme, we have implemented a JavaScript prototype that implements the necessary cryptographic operations. We also investigate suitable models for client identity management for public key web applications.

## 2. Related Work

The security on the Web mostly relies on the Transport Layer Security (TLS) [5] protocol. TLS guarantees the confidentiality and integrity of the sessions. Optionally, TLS offers mutual authentication for the client and the server. However, as TLS is a tunneling protocol, only the endpoints of the tunnel can be assured about the authenticity of the messages. In addition, as TLS acts below the applications it does not integrate into them particularly well. For instance, authentication errors show browser generated error messages to the users, which can not be customized by the end service. In addition, there is no fallback option built-in to the TLS client authentication, which means that either custom solutions are needed, or the same site cannot function for SSL and non-SSL clients. Also, federated authentication, based on the Certificate Authority (CA) model that is used to authenticate TLS web server certificates, may not function well outside of closely connected organizations. As a consequence, TLS client authentication is not widely used in web applications.

TLS-OBC (TLS Origin-Bound Certificates), and PhoneAuth [6], [7] present a model for client certificate based authentication for the web. These certificates are short-lived and only valid for a single service. Authentication wise, TLS-OBC mainly replaces authentication cookies, which may already be relatively secure when TLS is used. One problem with PhoneAuth is that it uses Bluetooth and requires extensions or special software to both ends, this makes it unsuitable for short term use in, for example, many Internet cafés, and on devices that do not have Bluetooth.

Grey [8] is a general authentication system that is built on the paradigm of using smart phone as general authentication device. It has the ambitious goal of eventually extending into physical space, with even door locks functioning with it. Fongen [9] presents a federated identity management system for Android based smartphones. The system uses shortlived identity statements from an identity provider (IdP). This scheme is oriented towards authentication for temporary purposes rather than establishing stable public key identities.

UbiKiMa [10], a system for ubiquitous authentication using a smartphone, utilizes a smartphone as a key management device for password and public key authentication. UbiKiMa firstly offers password management capabilities, but extends it to cover also public key based authentication. UbiKiMa is similar to the TLS-OBC, in that it intends to register different public keys for each relaying party. UbiKiMa is relying heavily on JavaScript[10], and presents an application layer authentication protocol, instead of TLS authentication like in TLS-OBC.

Hardware Security Modules (HSMs), Trusted Platform Modules (TPMs) and Secure Elements (SEs) tend to be underutilized by applications, and especially web applications. In addition, software emulation of SE is emerging also in mobile payments because of the politics associated

with the control over the SE[11]. Web Cryptography API [12] is developed in the W3C. This API enables JavaScript applications to use platform's security functions, which tend to be faster than the ones written in JavaScript. In addition, this API should standardize the use of HSMs by JavaScript applications. It is difficult to do private key operations securely in JavaScript [13]. For example, until recently there was a lack of a secure random number generator [14]. Currently there exists also browser specific extensions for seeding the random number generator from the browser's entropy pool.

Because strong public key based mechanisms have not been able to extend to web, substitutes that require considerably more trust have been developed. For instance, credit cards have been the liability mechanism for consumer purchases, and can be considered to be a form of online identity. However, the magnet stripe equivalent authentication in which credit cards are used in the web, does not provide strong authentication and puts the users under an increased risk of fraudulent transactions [15]. Password based payment confirmation protocols are available, but they are mainly used to limit the merchant's liability [16].

Stronger authentication measures are required for use cases where the risk is not something that could be mitigated by a simple credit limit. For example, authentication based on the credentials used for online banking has been utilized in Finland to provide strong authentication for contracts and public online services[17]. In this system the client is directed to authenticate with their bank, and the client's social security number is transmitted to the end service as a proof of the authentication. While functional, this system transmits more personal information about the user than necessary and the service provider does not get strong guarantees of the authentication. Moreover, the required mutual trust limits the uses of such authentication to a limited number of contractually agreed parties.

In general, connecting physical world identity to a digital one has been a challenge. Recent cryptocurrency proposal Bitcoin [18], builds a pseudonymous digital identity[2] that is not connected with any real life identity by default. This identity connects to real life identities only through purchases and currency exchanges. Indeed, one basic principle of Bitcoin is to avoid the trust issues of some earlier cryptocurrencies by establishing a currency that is intended to have an independent value. Moreover, Bitcoin wallet identities are typically used with software based wallets, which may have been a contributing factor for the prevalence of the system.

## 3. Public Key Web Application Security

### 3.1 Identity Management Model

In Public Key Infrastructure (PKI) there is a Certification Authority (CA) that guarantees that a public key is bound to

certain attributes or identity in a context it holds authority over. The advantage of the CA system is that its signature can be included with the certificate in question. However, this advantage vanishes when trust agility is required, in essence, the CA may not actually be an authority from the recipient's perspective.

Moreover, the value of a digital signature depends on the identity it is connected to. The party verifying the identity takes up liability on the fact that its testimony about the identity is correct. In such case the trust exists between the certifying party and the party accepting the assertion. However, federated identity management models often assume that the relying service trusts the home organization of the users. Yet, in Internet context there may not exist a trustworthy home organization for many user-service pairings. Therefore, the requirement of a certifying authority acceptable for all service providers is impractical. Most commonly, this leads to the user generating multiple separate identities. One way to avoid this is to build the trust chain from the verifying end. Indeed, public key and certificate authenticity can be guaranteed through also other trust management systems than certificate hierarchies. For example, PGP's Web of Trust [19], DNS Authentication of Named Entities (DANE) [20], and Certificate Transparency [21] are systems that are able to work from the verifier towards the public key and not vice versa.

It would make sense for end-user identities to be validated in separate contexts with separate directories. However, if a single public key is used to identify users in different contexts, the users' privacy could be endangered. Yet, cross-linking of separate public keys is also a problem that is hard to avoid. Nevertheless, to avoid misuse it might be necessary to require authorizations for querying public key directories that hold end-user public keys. For example, hash tree based directories support such authorizations quite naturally.

### 3.2 Operating Environment

In order to assess the needed properties for a public key identity web application we use these assumptions about the users and platforms:

- Global adversaries are more powerful than local but lack physical access
- Private key leak is worse than signature for wrong data or compromise of decrypted data.
- Internet access is ubiquitous, but not necessarily high bandwidth or low latency [10]
- Peripherals, such as, smart card readers, and bluetooth interfaces are not ubiquitous [10] and may not have interfaces available to web applications [22]
- JavaScript is ubiquitous [10], although it may have security issues on some platforms[13]
- Users will sign almost anything presented to them and will not verify data actively

- Users are not able to memorize strings with strong entropy
- Comparison based authentication measures are not effective for typical end users. For example, they do not expect attacks to happen [23], [24].

Traditionally, the security model for web applications has been different from the distributed desktop application model. With web applications the application and data usually belong to the same domain, and as such the security of traditional web applications is based on server side security mechanisms that the client blindly trusts. While both types of applications are commonly downloaded from the Internet there are differences in the trust management. The difference in web applications and desktop applications is that the desktop application is able to "tack" itself to the host, saving the configuration and the integrity of the application on the local machine. Also, the data and services which are used with these applications may not have any relation with the application provider, which leads to less centralized trust. In contrast, the web application may be delivered fresh on every access to the target site.

Cloud based applications fit somewhere in the middle as they save most of their state in the cloud, and are able to start where they left off in a completely different host. However, this requires that the user is able to authenticate to the cloud service so that the cloud service can deliver the user's configuration (Figure 1). In addition, security sensitive data cannot be stored in the cloud unencrypted, particularly if the cloud service provider is not a trusted party. Therefore, the private key should be held in a separate application, which has its state saved on the user side. Moreover, there is a difference between providing information used for security bootstrapping and the other party being in full control over the credentials.

### 3.3 End-User Key-Management

Managing access to private keys is critical for the safety of public key systems [4]. However, multiuser devices and multiple devices by the same user make this task considerably harder. For instance, a user who has multiple mail clients and wants to use S/MIME mail with encryption has to copy the same private key to all of his/her devices. Otherwise, the encrypted messages open only on the device that has the private key corresponding to the public key known to the message's sender. Yet, multiple devices that share a common private key makes it difficult to handle situations with lost keys. Indeed, tracking a shared private key across different devices is difficult, and may lead to keys being forgotten on decommissioned devices [4]. Key management is less problematic, when only data signing is used, as it is possible to use certificate chains with cross-validation, or get a signed certificate for multiple keys from authorities in PKI.

In general, a public key system with end-user managed private keys is restricted either in its key entropy or avail-

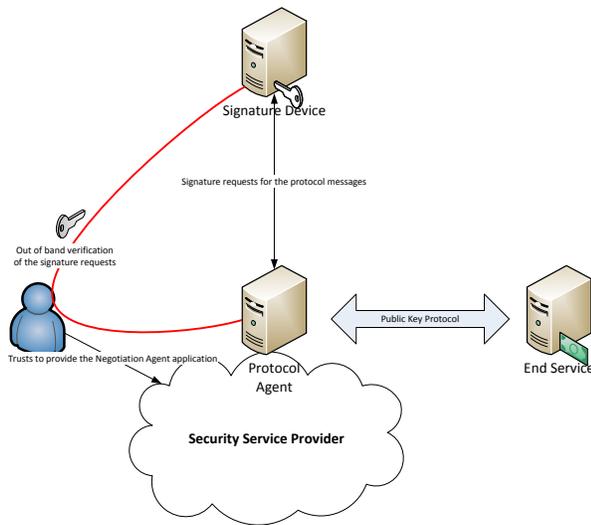


Fig. 1

PUBLIC KEY WEB APPLICATION DEPLOYMENT ILLUSTRATION

ability. This is caused by the limited capacity and accuracy of human memory, which we must compensate for. For instance, the private key can be stored on a portable device, stored on local computer, or derived from a lower entropy password with key derivation functions that require a lot of computation. In all cases the intention is to transform the user's advantage to a massive physical resource cost for the attacker, without losing the practical ability of the user having the key when needed.

### 3.3.1 Password Based Private Keys

As password authentication is prevalent, a question arises, whether it would be possible to derive a private key from a password? For example, ECDSA private keys can be chosen at random, which allows one to use a Key Derivation Function's (KDF) output as the private key. The used KDF should be such that key guessing channel available for an attacker can be reduced. For example, some schemes have used a server to function as part of the KDF [25].

Nevertheless, passwords are a problematic source for public keys since they tend not to contain much entropy under a password cracking entropy model [26]. These low entropy passwords are preferred by users as they are easier to remember. However, a balance between usability and security could be reached with a scheme that would require more computational work for low entropy passwords both from the user and the attacker. Indeed, the amount of iterations for the KDF can be defined to be  $2^{C-L_{PW}}$  in which  $C$  is a constant that determines the safety margin and  $L_{PW}$  is the approximate entropy of the password string.

Entropy is always relative to some model, and it cannot be objectively measured. For instance, character estimating Shannon entropy for strings like "password", is relatively high, i.e. one needs around 24 bits to encode it optimally under a natural language model. Yet, in terms of password entropy, that string could be encoded in a few bits as it is among the top three passwords used. Therefore, only the entropy of system assigned passphrases or passwords can be estimated with simple models [27]. The aim of the KDF is that the resulting private key requires a known effort to brute force. Naturally, a user specific but public component such as email address should be used as a salt in the key generation.

Nevertheless, the problem with such password based public keys is that the difference in computing power between a powerful adversary and the most constrained legitimate user terminal is not small. Particularly if the problem is parallelizable, as dictionary attacks with a computational bottleneck are, the difference can be huge. For example, mobile processors and high end graphics processors are measured to have a performance ratio of around  $10^3$  [28], which means that an adversary with resources to buy a 1000 graphic cards would be able to try out  $10^6$  passwords in the time that it takes for a mobile user to regenerate a private key (the key validation time is negligible). Also, as the adversary is not under a usability related time constraints it could, for example, use a year instead of 3 seconds (factor of  $10^6$ ). From this we can estimate attacker to be able to make guessing attacks against  $\log_2(10^{12}) \approx 40$  bit passwords<sup>1</sup>.

Indeed, the attacker capability can be estimated to be at least 40 bits. This is more than what could be expected from an unrestricted user choice, but not completely unattainable [27]. Yet, that is only the lower threshold, and to have a reasonable margin more entropy should be required for the private keys. Therefore, for most users password based public keys cannot actually be public, and should be constrained for single service like the TLS-OBC certificates [7]. It seems that public key based passwords' advantage, compared to traditional password authentication is that the authenticating server does not end up in the possession of the password, without active dictionary attack.

### 3.3.2 Smartcards

From a global viewpoint smartcards as digital identity devices have failed. On most markets consumers have not opted for them, even though they are very secure. Suggested reasons for their failure have been the lack of card readers [22], and the existence of insecure partial solutions, such as magnet stripe credit cards and password based authentication. Also, while there are many smart card related standards, there does not seem to exist accepted standards outside

<sup>1</sup>Adjusting the scrypt parameters to require more memory could change these calculations

certain vertical markets [22].

Smartcard adoption is on the rise in the payment processing industry, as they are a lot more secure than the magnetic stripes. Yet, smartcards are unable to complete transactions independently, which means that they have to rely on trusted infrastructure, such as card readers. These trusted components could easily capture PINs or display a different transaction from the one the user is actually authorizing. These risks are mitigated by the payment industry by contracts and auditing processes that the merchants need to follow. Over all, the smart card security seems to favor control over the whole infrastructure.

### 3.3.3 Mobile Phone as a Security Device

Better entropy source than human memorizable passwords is required when the services that use the identity are not confined to a single provider, expected public key lifetime is longer, or the identity holds significant value and cannot be effectively revoked or restored. Still, public key based identities need to be portable, which requires using a smart card, or other device, such as a smart phone. Smart phones have an advantage over smart cards in that they are able to communicate over a multitude of mediums, and do not need to rely on external input and output devices for communication with the user either.

As smartphones present a tempting target for private key storage, the threats need to be evaluated. At least three major threats can be identified for key storage on modern smartphones:

**Mobile Malware:** Applications tend to be sandboxed, and accessing keys used by other applications requires a root exploit. Mobile malware has mostly consisted of SMS sending trojans [29], although more sophisticated malware is bound to appear if there is more financial incentive for it.

**Theft:** Physical access by an attacker is a relevant threat, keys can be left unprotected, at least to volatile memory[30], and keys protected with a password can be cracked<sup>2</sup>. While the resources that most probable attackers have are not very good, larger collectives of attackers may reasonably break key storage mechanisms, which do not rely on hardware security.

**Broken or Lost:** Mobile devices may get destroyed or lost, which causes the owner to lose access to their keys.

The risk of losing or breaking the security device, is not unique to mobile phones but the probability is amplified by the conditions imposed by mobility. This threat could be mitigated by allowing the user to export the keys for backup.

<sup>2</sup>if the corresponding public key is known, or valid private keys can be identified like in RSA

Yet, if the keys are not used for encryption, key continuity can be established also with signed statements.

While it is possible to protect a private key with a password when it is not used, malware can access the private key in memory if it is able to gain full system access or utilize a side channel. Also, the password protections can be broken by a determined attacker, provided that there is access to the encrypted password, due to malware or theft of the device. HSMs are able to alleviate the risks that can be caused by malware and theft. However, as the control of these elements can be locked in applications, their use for emerging services may not be very easy. In addition, while the malware cannot get the private keys from a HSM, it may still trick the user into accepting fraudulent transactions, which may be a relevant threat in some situations.

### 3.4 Interface to Private Key Operations

For the end users the concept of identity is independent of devices and software used to access services in which the identity context exists. Indeed, the user may use multiple devices and platforms, and each one of those devices should be able to access the same identity. Using the same private key distributed to all these devices is impractical and unsafe. By concentrating all the identity related functions to a single device, only the messages need to be transmitted. In the Figure 1 the signing functionality was separated to a separate Signature Device (SD).

As the SD may reside in a restricted network, it might not have a remotely accessible interface. Instead it will communicate through a Mirror Service (MS) (Figure 2) that enables the user applications to make signature requests and for the SD to respond to them in a RESTful manner.

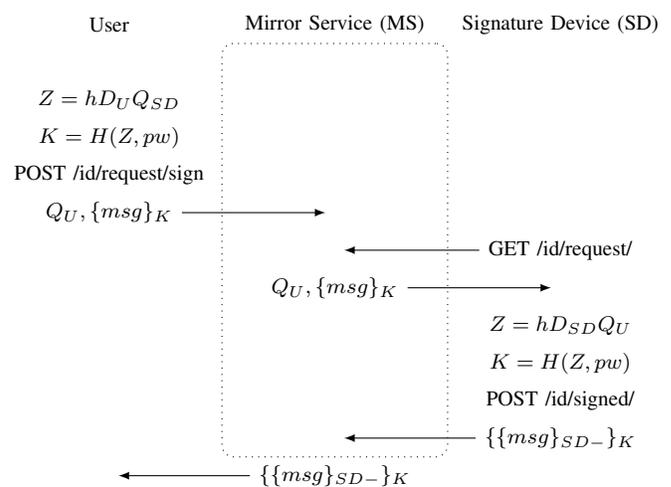


Fig. 2

MESSAGE SEQUENCE FOR REMOTE SIGNING

The MS in Figure 2 might not be a trusted component. By modifying messages the MS could try to get the user to sign something other than is intended. Therefore, using Transport Layer Security (TLS) to the MS is not enough, and also application layer (eg. data object) end-to-end security is needed (Figure 2). Nevertheless, TLS should be employed for confidentiality and privacy purposes. However, the addition of application layer encryption and out-of-band verification makes the signature protocol a lot more complicated than it would need to be otherwise.

The passphrase (*pw* in Figure 2) that is transmitted out-of-band by the user is a relatively short system assigned pronounceable password. Strong symmetric key would of course be preferable, and could be used without public key cryptography. For example, UbiKiMa [10] uses a 128-bit symmetric secret, and communicates the symmetric secret with QR-code, which is not error-prone like manual entry. However, QR-code readers require interfaces, which may not be available on all platforms. Short single case pronounceable passwords are a user friendly [27] alternative. Yet, such passwords are able to provide only approximately 20–30 bits of entropy in usable lengths, which is not enough to deter a resourceful attacker even with moderate key strengthening. In order to prevent the attacker from validating passphrase guesses, a public key based key derivation function should be used. Figure 2 presents Elliptic Curve Diffie-Hellman (ECDH) with static and ephemeral key pair. This is defined in standards as ECDH-ES [31].

Because of the ECDH-ES key derivation, the required minimum length for the confirmation password is considerably smaller than what would be required if the authenticity and confidentiality of the message would rely on solely on the passphrase. If an attacker changes the encrypted message and the ephemeral key, and tries to use the user to validate passphrase guesses, the attacker is able to validate only one passphrase at a time, and will break the protocol with incorrect guesses. The protocol requires that the static key that the user is using is correct, otherwise a man in the middle can make passphrase guesses. This is a potential weak point, as the user may not be able to verify it or is being misled by the Security Service that maps user friendly names with public keys.

Again, this does not prevent modification by a man in the middle who is able to fool the user to use a wrong static ECDH static key. To counter this attack, the Signature Device could ask the user to verify the ephemeral Diffie-Hellman public key. The static key could be verified as well. However, users have been shown to be unable to verify public keys [24], [23], which means that these measures cannot be relied on.

### 3.5 Example of a Public Key Protocol

To evaluate the model of web applications with accountable identities, we present an example of Security Service

Level Negotiation [3]. In this public key protocol, the user and the Service Provider (SP) agree on the service's security functions as well as the protection targets, i.e. user assets that the SP needs to access in delivering the service.

The SP initiates exchange with the user by exposing a protocol handler. This handler should be registered in the user's browser, and can be handled by a browser extension, JavaScript bookmarklet or a trusted secure website, which either delivers a standalone clientside application or partially relies on server side code. The selection of the protocol handler can be seen as the user's main trust decision, as it points to the selected trust anchors, and acts as a negotiation agent. Indeed, the negotiation agent is responsible for presenting the user the information that they can understand, and is assumed to act in the user's best interest. In addition, the negotiation agent also presents the mappings from user friendly names to public keys. The negotiation agent can use external resources for security functions, but it should be able to authenticate the returned information.

In the example below (Listing 1) the Service Provider presents the protection targets it wants to agree in the URI as a Base64 encoded JSON array. The identifier of the Service Provider needs to be included as well. The user starts the negotiation process by following the link.

```
web+ssla-negotiation://service.example.com?\
idSP=BL4GQ4G6B6o-XSKGdDGrbWXCHSDBT11-osLRX7Ycv\
BDbD0cVmqz1X5eqJDyhZiw4LD7uuJB5bIGTKrIPETEmJe8\
&targets=WyIxLjQuMSIsIjeuNC4yIiwiMi4yLjIuMS4yI10
```

Listing 1  
NEGOTIATION PROTOCOL URI

After receiving the service identifier and the protection targets, the negotiation agent creates an SSLA that satisfies user requirements. For this the agent uses a translation service, a Knowledge Base [3]. The translation service should be contacted over a secure channel (TLS or secured data objects). This requires that the negotiation agent holds the trust relationship by saving information about the public key of the Knowledge Base.

The user trusts the translation service in that the machine readable representation corresponds to the user's security profile. The natural language form of the agreement can be presented to the user, if it is required by the user's security profile. Indeed, the natural language form of the agreement contents should be presented by trusted entities, as it could be possible to do carefully worded agreement that are under more serious inspection actually void of any liability. Furthermore, our assumption is that the user will not verify what they sign, unless the verification step is made compulsory.

When the initial agreement template has been made, the negotiation agent starts a negotiation with the service

provider using signed messages. If the negotiation agent does not hold the user's private key, the user may use an external signature device to sign the protocol messages, as described in Figure 2. To complete the transaction, the user has to input the symmetric secret that authenticates the request, and possibly a short passphrase that unlocks the private key.

## 4. Conclusions

This paper has evaluated the concept of digital identities in web applications. Secure storage of private keys and multi-device usage of the same digital identity were identified as obstacles for end-user public key identities. Also, the special nature of web applications requires that the application code and trust anchors are delivered to the user on the time of the application use. Important issue is, how can the user trust the underlying application in that its actions match the user's mental model of the applications function.

The technical challenges can be solved by utilizing different security services, and by making the user's public key identity to be device independent and portable. However, challenges remain that are not purely technical but instead revolve around liability issues and trust. For example, which parties would authenticate end-user public keys, and how it can be done without sacrificing the user's right to privacy.

We presented an example of public key based agreement over service's security properties as such agreement needs to consider many different trust issues, and has to be separated from the trust domain of the service provider. Yet, we see benefits for many other application domains, particularly those which need to show accountable evidence of actions, such as applications in the billing domain.

## References

- [1] L. J. Camp, "Digital identity," *Technology and Society Magazine, IEEE*, vol. 23, no. 3, pp. 34–41, 2004.
- [2] E. Androulaki, G. O. Karame, M. Roeschlin, T. Scherer, and S. Capkun, "Evaluating user privacy in Bitcoin," in *Financial Cryptography and Data Security*. Springer, 2013, pp. 34–51.
- [3] T. Takahashi, J. Kannisto, J. Harju, S. Heikkinen, B. Silverajan, M. Helenius, and S. Matsuo, "Tailored security: Building non-repudiable security service level agreements," *IEEE Vehicular Technology Magazine*, 2013.
- [4] P. Gutmann, "Lessons learned in implementing and deploying crypto software," in *Usenix Security Symposium*, 2002, pp. 315–325.
- [5] T. Dierks and E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.2," RFC 5246 (Proposed Standard), Internet Engineering Task Force, Aug. 2008.
- [6] A. Czeskis, M. Dietz, T. Kohno, D. Wallach, and D. Balfanz, "Strengthening user authentication through opportunistic cryptographic identity assertions," in *Proceedings of the 2012 ACM conference on Computer and communications security*. ACM, 2012, pp. 404–414.
- [7] M. Dietz, A. Czeskis, D. Balfanz, and D. S. Wallach, "Origin-bound certificates: A fresh approach to strong client authentication for the web," in *USENIX Security*, 2012.
- [8] L. Bauer, S. Garriss, J. McCune, M. Reiter, J. Rouse, and P. Rutenbar, "Device-enabled authorization in the Grey system," in *Information Security*, ser. Lecture Notes in Computer Science, J. Zhou, J. Lopez, R. Deng, and F. Bao, Eds. Springer Berlin Heidelberg, 2005, vol. 3650, pp. 431–445.
- [9] A. Fongen, "Federated identity management for android," in *SECUREWARE 2011, The Fifth International Conference on Emerging Security Information, Systems and Technologies*, 2011, pp. 77–82.
- [10] M. Everts, J.-H. Hoepman, and J. Siljee, "Ubikima: Ubiquitous authentication using a smartphone, migrating from passwords to strong cryptography," in *Proceedings of the 2013 ACM workshop on Digital identity management*. ACM, 2013, pp. 19–24.
- [11] M. Roland, "Software card emulation in NFC-enabled mobile phones: great advantage or security nightmare," in *Fourth International Workshop on Security and Privacy in Spontaneous Interaction and Mobile Phone Use*, 2012, pp. 1–6.
- [12] R. Sleeve and M. Watson, "Web cryptography API," W3C, 2014, W3C Editor's Draft 7 March 2014. [Online]. Available: <http://dvcs.w3.org/hg/webcrypto-api/raw-file/tip/spec/Overview.html>
- [13] "JavaScript cryptography considered harmful," Matasano Security. [Online]. Available: <http://matasano.com/articles/javascript-cryptography/>
- [14] E. Stark, M. Hamburg, and D. Boneh, "Symmetric cryptography in JavaScript," in *Computer Security Applications Conference, 2009. ACSAC'09. Annual*. IEEE, 2009, pp. 373–381.
- [15] H. K. Lu and A. Ali, "Prevent online identity theft—using network smart cards for secure online transactions," in *Information Security*. Springer, 2004, pp. 342–353.
- [16] S. J. Murdoch and R. Anderson, "Verified by Visa and Mastercard SecureCode: or, how not to design authentication," in *Financial Cryptography and Data Security*. Springer, 2010, pp. 336–342.
- [17] "Tupas identifiaton service." Federation of Finnish Financial Services, Apr. 2011. [Online]. Available: <http://www.fkl.fi/en/themes/e-services/tupas/Pages/default.aspx>
- [18] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2009. [Online]. Available: <http://www.bitcoin.org/bitcoin.pdf>
- [19] P. R. Zimmermann, *The official PGP user's guide*. MIT press, 1995.
- [20] P. Hoffman and J. Schlyter, "The DNS-Based Authentication of Named Entities (DANE) Transport Layer Security (TLS) Protocol: TLSA," RFC 6698 (Proposed Standard), Internet Engineering Task Force, Aug. 2012.
- [21] B. Laurie, A. Langley, and E. Kasper, "Certificate Transparency," RFC 6962 (Experimental), Internet Engineering Task Force, Jun. 2013. [Online]. Available: <http://www.ietf.org/rfc/rfc6962.txt>
- [22] S. Petri, "An introduction to smart cards," Secure Service Provider TM, 2002.
- [23] H.-C. Hsiao, Y.-H. Lin, A. Studer, C. Studer, K.-H. Wang, H. Kikuchi, A. Perrig, H.-M. Sun, and B.-Y. Yang, "A study of user-friendly hash comparison schemes," in *Computer Security Applications Conference, 2009. ACSAC'09. Annual*. IEEE, 2009, pp. 105–114.
- [24] P. Gutmann, "Do users verify ssh keys?" *USENIX; login*, Aug. 2011.
- [25] W. Ford and B. S. Kaliski Jr, "Server-assisted generation of a strong secret from a password," in *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2000. (WET ICE 2000). Proceedings. IEEE 9th International Workshops on*. IEEE, 2000, pp. 176–180.
- [26] S. Schechter, C. Herley, and M. Mitzenmacher, "Popularity is everything: A new approach to protecting passwords from statistical-guessing attacks," in *Proceedings of the 5th USENIX conference on Hot topics in security*. USENIX Association, 2010, pp. 1–8.
- [27] R. Shay, P. G. Kelley, S. Komanduri, M. L. Mazurek, B. Ur, T. Vidas, L. Bauer, N. Christin, and L. F. Cranor, "Correct horse battery staple: Exploring the usability of system-assigned passphrases," in *Proceedings of the Eighth Symposium on Usable Privacy and Security*. ACM, 2012, p. 7.
- [28] "Mining hardware comparison," Mar. 2014. [Online]. Available: [https://litecoin.info/Mining\\_hardware\\_comparison](https://litecoin.info/Mining_hardware_comparison)
- [29] Apville A & Strazzere T, "Reducing the window of opportunity for Android malware gotta catch 'em all," *Journal in Computer Virology*, pp. 1–11, 2012.
- [30] T. Müller and M. Spreitzenbarth, "Frost," in *Applied Cryptography and Network Security*. Springer, 2013, pp. 373–388.
- [31] 800-56A, National Institute for Standards and Technology Std., 2007.

**SESSION**  
**SECURITY MANAGEMENT I**

**Chair(s)**

**Prof. Nizar Al Holou**  
**Univ. of Detroit Mercy - USA**  
**Dr. Mehmet Sahinoglu**  
**Auburn Univ. Montgomery - USA**



## Metrics to Assess and Manage Software Application Security Risk

M. Sahinoglu, S. Stockton, S. Morton, P. Vasudev, M. Eryilmaz

Auburn University at Montgomery (AUM) and ATILIM University, Ankara  
[msahinog@aum.edu](mailto:msahinog@aum.edu), [stephen.stockton@gunter.af.mil](mailto:stephen.stockton@gunter.af.mil), [smorton1@aum.edu](mailto:smorton1@aum.edu);  
[pvasudev@aum.edu](mailto:pvasudev@aum.edu), [meltem\\_eryilmaz@atilim.edu.tr](mailto:meltem_eryilmaz@atilim.edu.tr)

**Abstract:** Software application security risk is of critical importance to modern enterprises and organizations. The very existence of these entities often depends on the successful and secure operation of mission critical applications such as the outer space explorations and high assurance scenarios regarding the surgery table for one striking example. Software application security risk is concerned primarily with how security personnel, facility managers, network personnel, management, and other interested parties rate their experience with the various aspects of software security risk, and overall management and maintenance, including issues such as continuity or availability of service, security design and configuration to name a few. To address this need, the principal author has built the fundamental (probability and game-theory related) computational aspects and an associated automated software tool for quantitative risk management. This tool, the Risk-o-Meter (RoM) provides measurable risk, advice for cost and risk mitigation on vulnerabilities and threats associated with the implementation and operation of software applications.

### I. INTRODUCTION

The identification and management of risk is a critical part of operating an IT system. While there are many approaches to identifying and managing risk, many managers focus only on the security of the software and often neglect the other aspects of system operations. Additionally, once all risks are identified, formulating a cost-optimized solution to mitigate undesirable risks to a tolerable level is often an ad-hoc process.

In this research, we adopt a model of software application risk that quantifies the user's experience with six crucial aspects of the software application security environment. However we will add an original concept of quantification to the existing model through a designed algorithm by the principal author to calculate the software application security risk index [1]. To accomplish this task, numerical and/or cognitive data was collected to supply the input parameters to calculate the quantitative security risk index for software applications. This paper will not only present a quantitative model but also provide a remedial cost-optimized game-theoretic analysis about how to bring

an undesirable risk down to a locally-determined “tolerable level”.

## II. METHODOLOGY

This applied research paper implements a methodology on how to reduce the risk associated with the Software Application Security Risk associated with implementing and operating an Information Technology (IT) system. A software-centered composite security approach is proposed to aid security personnel, facility managers, and network personnel within an organization. In order to control the risk associated with the implementation of an IT system, concepts such as continuity of operations, security design and configuration of the software, the U.S. Department of Defense’s Enclave Boundary Defense construct, a system’s identification and authentication process, the physical environment where the system will be operated, and the personnel associated with operating the system (to name a few), should be considered. The primary author’s innovation, i.e. RoM (Risk-O-Meter), an automated software tool based on game theory, will provide IT managers a measurable assessment of the current security posture of their implemented IT system by detailing associated cost and risk mitigation suggestions for countering identified vulnerabilities and threats associated with the IT system. The RoM will greatly facilitate the assessment and enhancement of the current security of an implemented IT system. Additionally, the Risk of Operation or Unavailability metric out of 100% will be assessed to provide

a remedial cost-optimized game-theoretic analysis to bring an undesirable risk down to a user determined “tolerable level” [2]. While the Risk-O-Meter can be applied to virtually any organization subject to systemic threats/vulnerabilities to their business operations, this particular implementation focuses on six key areas critical in ensuring Software Application security.

- **Continuity:** Regardless of the quality of a given software application, continuous and uninterrupted operations are critical for mission critical applications. In order to achieve such availability, alternate site planning, backup and restoration processes, exercises and drills, enclave boundary defense, and disaster planning have to be considered. Each of these areas must be addressed to ensure optimal user availability of the application.
- **Security Design and Configuration:** This area focuses on secure-by-design and subsequent configuration of the software itself as well as the platform you choose to run the application on. This entails industry best practices for acquisition standards and configuration specifications, as well as compliance testing, operational best security practices, and proper implementation of non-repudiation.

- **Enclave Boundary Defense:** Unless the software application operates in a stand-alone environment, the security defenses of the enclave it runs within are crucial to the overall security health of an application. This is the first line of defense and determines what barriers are in place to prevent unauthorized access to the platform running the application and to ensure authorized access is securely administered. This key area focuses on overall boundary defense, connection rules, remote access, encryption, confidentiality, and proper network and system auditing.
- **Identification and Authentication:** While boundary defenses provide an excellent defense for software systems, one key to a secure system is to ensure that individuals who are allowed access to the application can be identified and authenticated. This key area focuses on account control, use of individual accounts, key management, token-certificate standards, and group authentication.
- **Physical Environment:** While much attention is given to keeping unauthorized/unwanted individuals from gaining access to systems via electronic means, the facilities that house these platforms must also be protected to prevent system compromise.

Additionally, the facilities must also be evaluated for their ability to protect the individuals responsible for maintaining the application and ensuring the application remains online and available to users. This key area focuses on the physical protection provided to an application and covers access to the facilities, protection against data interception, emergency lighting and power, use of screen locks, and storage of data/hardware.

- **Personnel:** Personnel, in various forms, are often the biggest threat to software applications. While there are many opportunities for individuals to impact the security of an application while it is being developed, tested and fielded, this area focuses on the personnel that may impact the operational security of an application. This key area focuses on personnel access to information, maintenance personnel, IA training, rules of behavior, and background checks.

While it is critical to ensure that applications are secure-by-design, the daily challenge is to ensure that a given software application remains operationally secure on a daily basis. This research focuses on the areas vital to secure application operations and provides field managers with an analysis they can use to more efficiently secure their operational environments, including cloud [3].

### III. VULNERABILITY AND THREAT ASSESSMENT STRUCTURE

As previously noted, six vulnerabilities are assessed: Continuity, Security Design and Configuration, Enclave Boundary Defense, Identification and Authentication, Physical and Environmental, and Personnel. Within each vulnerability category, questions pertain to specific threats and countermeasures. For example, within the Continuity vulnerability, users are asked questions regarding Alternate Sites, Backup and Restoration, and Exercises and Drills threats and countermeasures. Within the Enclave Boundary Defense

vulnerability, users are asked questions regarding Boundary Defense, Connection Rules, and Remote Access threats and countermeasures. Within the Physical and Environmental vulnerability, users are asked questions regarding Access to Facilities, Data Interception, and Emergency Lighting threats and countermeasures. See Figure 1 below for the Software Application Security Risk diagram detailing vulnerabilities and threats. The user's responses are then used as input for the RoM to generate a quantitative software application security risk index using a game-theoretical mathematical algorithm.

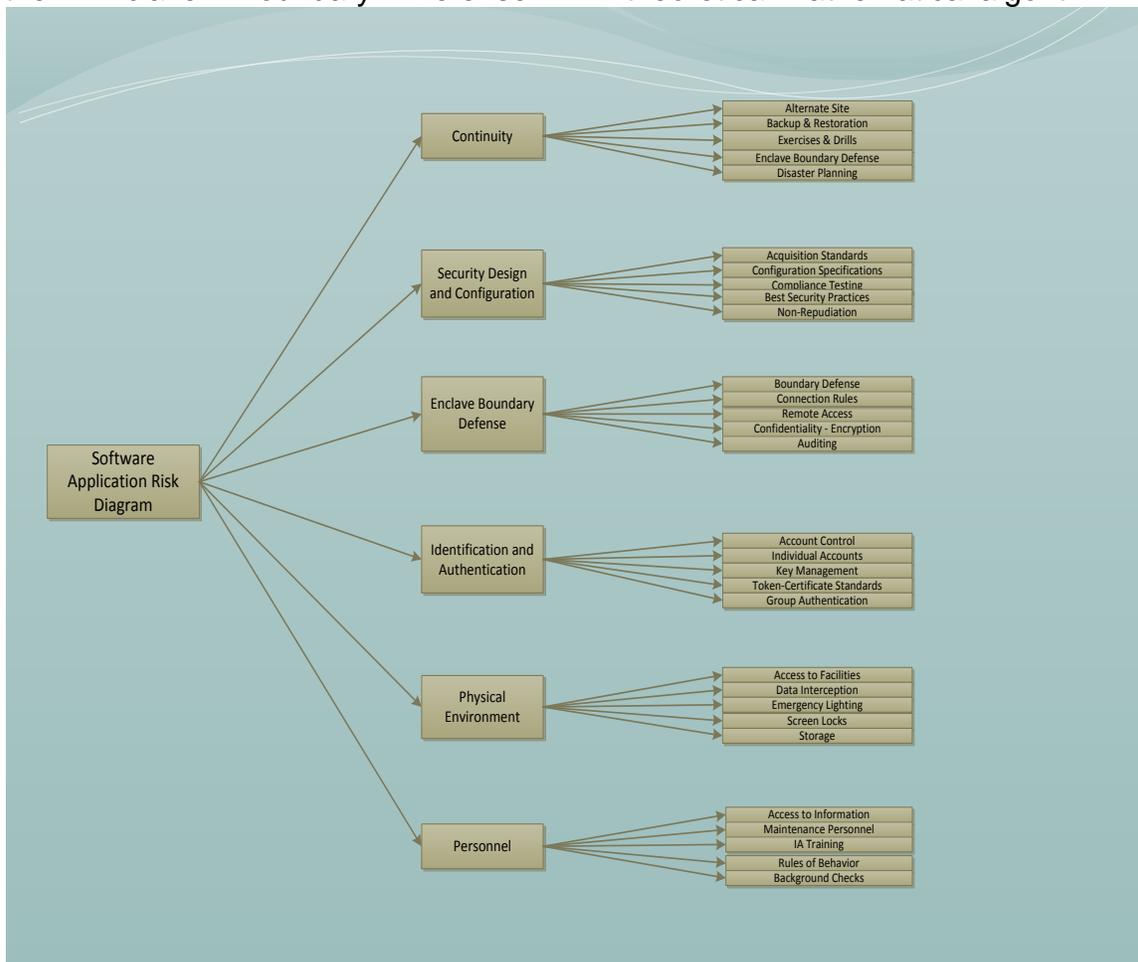


Figure 1: Software Application Security Risk Tree Diagram

#### IV. SAMPLE ASSESSMENT QUESTIONS

Questions are designed to elicit the user's response regarding the perceived risk to software application security from particular threats, and the countermeasures the users may employ to counteract those threats. For example, in the Identification and Authentication vulnerability, questions regarding Account Control include both threat and countermeasure questions. Threat questions would include:

- Is a comprehensive account management process unimplemented?
- Are controls lacking to ensure that only authorized users can gain access to workstations, applications, and networks?
- During the creation of a new account for a system user, does the registration process leave the required information uncollected?

While countermeasure questions would include:

- Are default accounts removed/disabled during installation of servers and workstations?
- Are accounts disabled and user IDs and passwords removed within 48 hours of notification that a user no longer requires or is authorized system access?

- Do system administrators immediately disable any account through which unauthorized user activity has been detected?

#### V. RISK CALCULATION AND MITIGATION

Essentially, the users are responding yes or no to these questions. These responses are used to calculate residual risk. Using a game-theoretical mathematical approach, the calculated risk index is used to generate an optimization or lowering of risk to desired levels [1]. Further, mitigation advice will be generated to guide security personnel, facility managers, network personnel, management, and other interested parties. Or more specifically, in what areas can the risk be reduced to optimized or desired levels (such as from 37% to 27% in the screenshot representing the median response from the study participants). See Figure 2 below for the screenshot of the Median Software Application Security Risk Meter Results Table displaying threat, countermeasure, and residual risk indices, optimization options, as well as risk mitigation advice. For this study, a random sample of 31 respondents was taken. Their residual risk results are tabulated and presented in Appendix A at the end of this paper. Respondents' experience in software application security risk included both corporate and governmental.

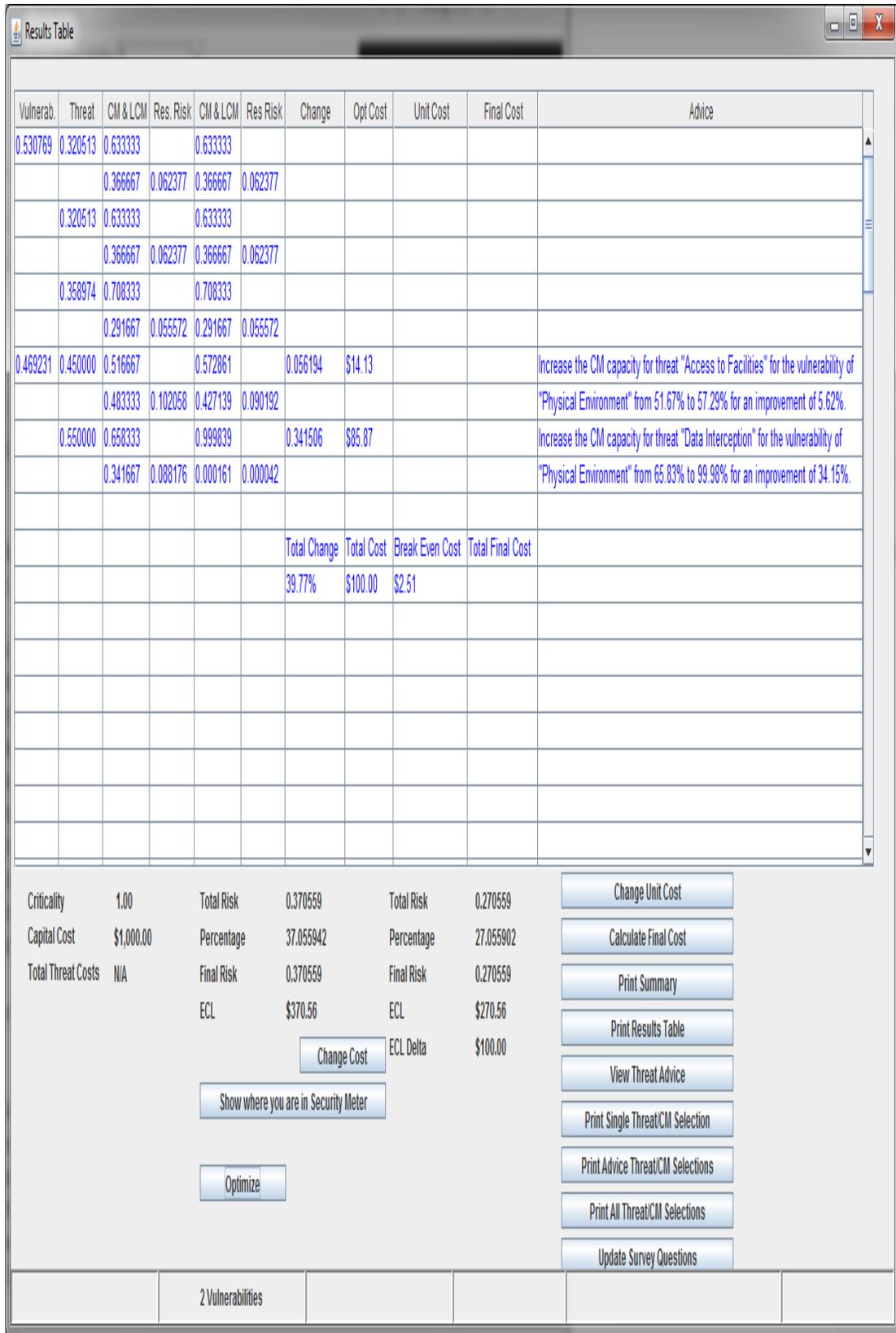


Figure 2: Median Software Application Security Risk Meter Results Table

## VI. CONCLUSION

The Software Application Security Risk Meter breaks new ground in that it provides a quantitative assessment of risk to the user as well as recommendations for mitigating that risk. As such, it will be a highly useful tool for security personnel, facility managers, network personnel, management, and other interested parties seeking to minimize and mitigate software application security risk in an objective, quantitatively based manner. Future work will involve the incorporation of new questions so as to better refine user responses and subsequent calculation of risk and mitigation recommendations. Minimization and mitigation of software application security risk will greatly benefit not only the organizations deploying the applications, but society at large through the minimization of security breaches leading to monetary loss and ID theft. The Software Application Security Risk Meter tool and its future refinement provide the means to do so as there are and always will be software application security risks in cyberspace [4,5].

## VII. REFERENCES

- [1] M. Sahinoglu, Trustworthy Computing, John Wiley, 2007.
- [2] M. Sahinoglu, "An Input-Output Measurable Design for the Security Meter Model to Quantify and Manage Software Security Risk", IEEE Transactions on Instrumentation and Measurement, Vol. 57, No. 6, pp. 1251-1260, June 2008.
- [3] K. Hashizume, D. G. Rosdao, E. Fernandez-Medina, E. B. Fernandez, "An Analysis of Security Issues for Cloud Computing", Journal of Internet Services and Applications, , 2013; doi:10.1186/1869-0238-4-5 <http://www.ijsajournal.com/content/4/1/5> (Accessed 12/18/2013)
- [4] <http://www.infoq.com/news/2010/03/Top-10-Security-Risks> (Accessed 12/18/2013)
- [5] M. Sahinoglu, L. Cueva-Parra , D. Ang , "Game-theoretic computing in risk analysis", WIREs Comput. Stat 2012, doi: 10.1002/wics, 1205, 2012. <http://authorservices.wiley.com/bauthor/onlineLibraryTPS.asp?DOI=10.1002/wics.1205&ArticleID=961931>

### Appendix A: Respondent Residual Risk Results Table

SURVEY TAKER	RESIDUAL RISK %	RANKED OVERALL (OUT OF 31)	REMARKS
Company A1	43.66	8 <sup>th</sup>	1 <sup>st</sup> out of 11 within Company A
Company A2	15.48	31 <sup>st</sup>	11 <sup>th</sup> out of 11 within Company A
Company A3	27.42	28 <sup>th</sup>	8 <sup>th</sup> out of 11 within Company A
Company A4	39.47	15 <sup>th</sup> ~ <i>OVERALL AVERAGE</i>	4 <sup>th</sup> out of 11 within Company A
Company A5	35.45	20 <sup>th</sup>	5 <sup>th</sup> out of 11 within Company A
Company A6	24.27	30 <sup>th</sup>	10 <sup>th</sup> out of 11 within Company A
Company A7	24.96	29 <sup>th</sup>	9 <sup>th</sup> out of 11 within Company A
Company A8	43.27	9 <sup>th</sup>	2 <sup>nd</sup> out of 11 within Company A
Company A9	31.90	26 <sup>th</sup>	7 <sup>th</sup> out of 11 within Company A
Company A10	35.18	21 <sup>st</sup>	6 <sup>th</sup> out of 11 within Company A ( <i>Group Median for Company A</i> )
Company A11	42.34	11 <sup>th</sup>	3 <sup>rd</sup> out of 11 within Company A
Company B1	48.59	6 <sup>th</sup>	5 <sup>th</sup> out of 11 within Company B
Company B2	51.36	2 <sup>nd</sup>	2 <sup>nd</sup> out of 11 within Company B
Company B3	50.93	5 <sup>th</sup>	4 <sup>th</sup> out of 11 within Company B
Company B4	45.59	7 <sup>th</sup>	6 <sup>th</sup> out of 11 within Company B ( <i>Group Median for Company B</i> )
Company B5	53.52	1 <sup>st</sup>	1 <sup>st</sup> out of 11 within Company B
Company B6	41.75	13 <sup>th</sup>	8 <sup>th</sup> out of 11 within Company B
Company B7	42.92	10 <sup>th</sup>	7 <sup>th</sup> out of 11 within Company B
Company B8	34.04	24 <sup>th</sup>	10 <sup>th</sup> out of 11 within Company B
Company B9	51.16	3 <sup>rd</sup>	3 <sup>rd</sup> out of 11 within Company B
Company B10	36.71	17 <sup>th</sup>	9 <sup>th</sup> out of 11 within Company B
Company B11	29.77	27 <sup>th</sup>	11 <sup>th</sup> out of 11 within Company B
Company C1	34.67	23 <sup>rd</sup>	8 <sup>th</sup> out of 9 within Company C
Company C2	40.20	14 <sup>th</sup>	3 <sup>rd</sup> out of 9 within Company C
Company C3	34.85	22 <sup>nd</sup>	7 <sup>th</sup> out of 9 within Company C
Company C4	51.14	4 <sup>th</sup>	1 <sup>st</sup> out of 9 within Company C
Company C5	36.44	18 <sup>th</sup>	5 <sup>th</sup> out of 9 within Company C ( <i>Group Median for Company C</i> )
Company C6	37.06	16 <sup>th</sup> = <i>OVERALL MEDIAN</i>	4 <sup>th</sup> out of 9 within Company C
Company C7	33.98	25 <sup>th</sup>	9 <sup>th</sup> out of 9 within Company C
Company C8	35.93	19 <sup>th</sup>	6 <sup>th</sup> out of 9 within Company C
Company C9	42.04	12 <sup>th</sup>	2 <sup>nd</sup> out of 9 within Company C

**Table 1. Companies (A, B, C) Survey Results for the Risk-O-Meter study, ranked within and overall, where Median: 37.06% (C6) and Average: 38.58% (A4: 39.47% is the result that comes the closest).**

# An Interoperability Framework for Security Policy Languages

Amir Aryanpour, Edmond Parkash, *University of Bedfordshire, UK*  
 Andrew Harmel-Law, Scott Davies, *Technology Service, Capgemini UK*

**Abstract**— In today's economy, whilst corporations are looking to control costs and still driving productivity, the cost of acquiring and maintaining a company's software is under the spotlight. IT departments are under pressure to deliver more services, in shorter amounts of time, and with ever decreasing budgets; hence, IT departments are willing to invest in and choose technologies that provide more business value at a lower cost. Taking these facts into account, and bearing in mind that a number of security policy languages available and the majority of scenarios covered by them are similar, this paper proposes a framework that understands security domains and provides users with an abstract security policy language, which can be translated into the desired policy language as per framework's configuration.

Using such a framework would allow multi-dimensional organizations to use an abstract policy language to orchestrate all security scenarios from a single point, which could then be propagated across the environment. In addition, using such a framework would help security administrators to learn and to use a single, common abstract language to describe and model their environment(s). That in turn would help IT departments to control their security related costs.

**Keywords:** *Security Policy Language, Domain specific Language, Management of Secure Domains, Scala, Software System Design.*

## I. INTRODUCTION

The notation of protecting networked resources came to life at the very same moment computer networking was introduced. Similar to programming languages which facilitate programmers to orchestrate a series of actions to achieve a goal, many access control model and security policy languages have been proposed in order to address the abovementioned concern. These security policy languages, which have undergone a revolution during the last decade, usually come with different specifications, advantages and disadvantages aim to tackle different business requirements.

As an information technology expert, how many times have you heard or tried to answer a simple question, such as 'which programming languages is most efficient, e.g. Java or C#?' As with programming languages, it is not easy to distinguish the advantages of one specific security policy language over another but, unlike programming languages that are used to code an unlimited amount of scenarios, the majority of scenarios covered by security policy languages can usually be modeled and described at an abstract level: *who can access what under which circumstances?*

Taking the above fact into account through our research, we have focused on security policy languages and proposed a framework. The framework which can be controlled by security administrators using an *Abstract Language*, will

facilitate the representation of security policy languages.

Perhaps the very first question comes to mind would challenge the usability of such a framework. In other words do policy language users want a standard? Or why not use one of the current security policy languages as the abstract language instead of introducing one? These are valid points and questions we have reviewed over and over. No doubt generic and academic languages could possibly be considered as an abstract security policy language however we need to bear in mind that the framework has a very specific goal to achieve and that is to provide current and future security policy language users with an abstract and/or standard security policy language. Adopting an existing policy language more generally would probably work for new users however that more likely would not be an acceptable approach for existing security policy language users. We believe our framework can be easily coupled with an existing security infrastructure or can be used as an independent tailored approach for a new project. We have provided even more benefits of the framework in [1].

There are number of similarities between Structured Query Language (SQL) and the *Abstract Language* of our framework. Similar to SQL, which hides the complexity of the underlying database infrastructure (whether it is Oracle, MySQL or another) and provides Database Administrators (DBAs) with a common and abstract language i.e. SQL to communicate with databases, our framework provides security architectures with an abstract security policy language that understands the security domain and hides the complexity thereof. Using this framework eliminates the need to learn how to code security policies in Protune and/or XACML and then code relatively common scenarios like 'who can access what, under which circumstances?' The proposed framework, which understands the security domains, provides users with a much simpler language that maintains the orthogonality of the security system. In other words, we provide a Domain Specific Language (DSL) for security policy languages picture (see Figure 1).

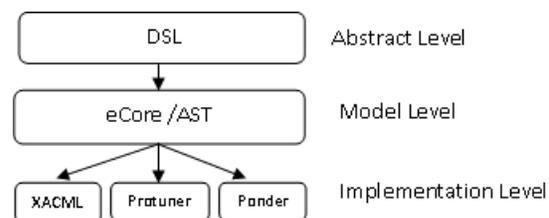


Figure1. Overview of proposed framework.

In [1], we looked at the framework, the main contribution of our research, from different angles. We first described benefits of the framework in details and then re-examined works related to our research and aimed at improving them accordingly. We then theoretically proved that the translation of security policy languages, irrespective of their formalism and specification, is possible and we provided algebra to this effect. In the same paper, we selected and justified our selection of three different policy languages that work in our framework.

In this paper, we aimed to design this framework and to justify our approaches at each individual step. The rest of this paper is organized into three parts. The first part details domain-specific languages, their requirement, patterns and so on. The second part justifies the pattern we have chosen and the third part describes proposed architecture in detail

## II. SECURITY POLICY LANGUAGES

We classified security policy languages according to different categories and nominated one candidate from each category in [1], namely XACML, Ponder and Protune. We will now briefly describe each of these policy languages.

### A. XACML

The eXtensible Access Control Markup Language (XACML) [2] is an OASIS standard that describes both a policy language and an access control decision request/response language (both in XML).

The policy language allows the specification of access control conditions that must be fulfilled by a requester. There are three kinds of top-level elements:

**<Rule>** It is a boolean expression that is not intended to be evaluated in isolation but which can be reused by several policies.

**<Policy>** It is a set of rules and obligations that apply to a request. It contains a set of rules and an algorithm describing how to combine the results of the evaluation.

**<PolicySet>** It contains a set of policy and policy set elements, together with an algorithm describing how to combine the results of the evaluation.

The request/response language allows the sending of queries in order to check whether a specific request should be allowed. There are four different valid values for the answer in the response: Permit, Deny, Indeterminate (a decision could not be made) or Not Applicable (the request cannot be answered by this service).

This provides the basis for the separation of the so-called Policy Enforcement Point (PEP), which is the entity in charge of protecting a resource, and the Policy Decision Point (PDP), which is responsible for checking whether a request is conformant with a given policy. In order to include the execution of actions within the standard, the authors define the *<Obligation>* element.

An obligation is “an action that must be performed in conjunction with the enforcement of an authorization

decision”.

XACML is standard, so it includes many features, among which we highlight the following:

- The language allows the use of attributes in order to perform authorization decisions without relying exclusively on the identity of the requester.
- Different arithmetic formulae, sets and boolean operators, and built-in functions are provided, as well as a method for extending the language with non-standard functions.
- The language includes a *<Target>* element in each rule, policy or policy set in order to allow for indexing and to increase performance.
- Different combinations of algorithms are provided for rule and policy composition: deny-overrides, ordered-deny-overrides, permit-overrides, ordered-permit-overrides, first-applicable and only-one-applicable.
- An *XACML context* is defined in order to provide a canonical form for representing requests and responses. As it is encoded in XML, it is possible to extract information from the context using XPath 2.0.

Following is a security policy example written in XACML.

```
<Policy PolicyId="pol_own_records" RuleCombiningAlgId=
  "urn:oasis:names:tc:xacml:1.0:rule-combining-
  algorithm:permit-overrides">
  <Description> My Policy </Description>
  <Rule RuleId="rul_own_record" Effect="Permit">
    <Description>My Rule </Description>
    <Condition>
    <Apply FunctionId=
      "urn:oasis:names:tc:xacml:1.0:function:string-
      equal">
    <SubjectAttributeDesignator DataType=
      "http://www.w3.org/2001/XMLSchema#string">
      urn:oasis:names:tc:xacml:1.0:subject:subject-id
    </SubjectAttributeDesignator>
    <ResourceAttributeDesignator DataType=
      "http://www.w3.org/2001/XMLSchema#string">
      urn:emc:edn:samples:xacml:resource:resource-
      owner
    </SubjectAttributeDesignator>
    </Apply>
    </Condition>
  </Rule>
</Policy>
```

### B. Ponder

Ponder is a declarative, object-oriented language for specifying security policies with role-based access control, as well as general-purpose management policies for specifying what actions are carried out when specific events occur within the system or what resources to allocate under specific conditions [3]. Unlike many other policy specification notations, Ponder supports typed policy specifications. Policies can be written as parameterised

types, and the types instantiated multiple times with different parameters in order to create new policies. Furthermore, new policy types can be derived from existing policy types, supporting policy extension through inheritance.

Ponder has four basic policy types: authorisations, obligations, refrains and delegations and three composite policy types: roles, relationships and management structures that are used to compose policies [4]. Ponder also has a number of supporting abstractions that are used to define policies: domains for hierarchically grouping managed objects, events for triggering obligation policies, and constraints for controlling the enforcement of policies at runtime. Shown below is an access control policy written in Ponder

```
inst auth+ switchPolicyOps {
  subject /NetworkAdmin;
  target <PolicyT> /Nregion/switches;
  action load(), remove(), enable(), disable(); }
```

### C. Protune

The PROvisional TRust NEgotiation framework PROTUNE [5] aims at combining distributed trust management policies with provisional-style business rules and access-control related actions.

PROTUNE's rule language extends two previous languages: PAPL, which was one of the most complete policy languages for trust negotiation until 2002, and PEERTRUST, which supports distributed credentials and a more flexible policy protection mechanism.

In addition, the framework features a powerful declarative metalanguage for driving some critical negotiation decisions, and integrity constraints for monitoring negotiations and credential disclosure. PROTUNE provides a framework that has:

- Different arithmetic formulae, sets and boolean operators, while built-in functions are also provided as a means of extending the language with non-standard functions.
- Trust management, language-supporting general provisional-style actions (possibly user-defined).
- An extendible declarative metalanguage for driving decisions regarding request formulation, information disclosure and distributed credential collection.
- A parameterised negotiation procedure, which provides semantics for the metalanguage and provably satisfies some desirable properties for all possible metapolicies.
- Integrity constraints for negotiation monitoring and disclosure control.
- General, ontology-based techniques for importing and exporting.

In Protune, a *policy* is a set of rules. The vocabulary of predicates occurring in the rules is partitioned into the following categories:

- *logical* predicates: usual Prolog predicates
- *provisional* predicates: predicates that are meant to

represent actions

- *decision* predicates: predicates used to signal a policy's entry point

The following can be presented as an example of a policy written in Protune:

```
execute(access(resource)):-
  declaration(Uid,Pwd),
  password(Uid,Pwd).
password(uid1,pwd1).
password(uid2,pwd2).
access(_)->type:provisional.
access(_)->ontology:<www.L3S.de/policyFramework#Access>.
access(_)->actor:self.
declaration(_,_)->type:provisional.
declaration(_,_)-
  >ontology:<www.L3S.de/policyFramework#Declaration>.
declaration(_,_)->actor:peer.
password(_,_)->type:logical.
password(Uid,Pwd)->sensitivity:public :-
  ground(Uid),
  ground(Pwd).
```

## III. DOMAIN SPECIFIC LANGUAGES

Previously we have mentioned that we will utilize DSL in our framework. So it would be useful to examine the concept of DSL in more detail.

DSL are not new, having been around for decades. Fowler, in his domain specific language book, called *DSL a new name for an old idea* [6]. DSL, as implied by the name, is a computer programming language that has been explicitly tailored, designed and developed for a specific usage. It provides limited expressiveness and should have a clearly defined domain focus. A domain focus makes a limited language worthwhile [6].

### A. DSL requirement

Before we dive into the design of the framework, let us have a quick look at the requirements, advantages and disadvantages of DSL in order to justify the use of DSL in our framework.

Generally speaking, some of the requirements for general-purpose programming languages apply directly to DSLs. The core requirements for a DSL are as follows:

**Conformity:** the language constructs must correspond to important domain concepts.

**Orthogonality:** each construct in the language is used to represent exactly one distinct concept in the domain.

**Integrability:** the language, and its tools, can be used in concert with other languages and tools with minimal effort.

**Extensibility:** the DSL (and its tools) can be extended to support additional constructs and concepts.

**Longevity:** the DSL should be used and useful for a significant period of time.

**Simplicity:** the language should be as simple as possible in order to express the concepts of interest and to support its users.

**Quality:** the language should provide general mechanisms for building quality systems.

**Supportability:** it is feasible to provide DSL support via tools for typical model and program management, such as creating, deleting, editing, debugging and transforming [7].

### B. Advantages of DSLs

Undoubtedly, the main advantage of using a domain-specific language in the context of our research is that DSL provides a common vocabulary for the domain experts. Using this limited vocabulary results in:

**More accurate communication with Domain experts:** DSLs provide a higher level of abstraction which strips out the low-level details of a programming language, hence allowing programmers to more easily engage with domain experts.

**Efficient system productivity and maintainability:** Domain users are using a limited programming language to communicate with each other. Abstract languages like DSLs are easy to read and understand by all a system's stakeholders and that implies system vulnerabilities will be easier to identify. That in turn would result in better maintainability and productivity of the system. Besides developing a DSL-based application could be considered as cost efficient in long term.

In addition to above, in a large development team where a mixture of experienced and a junior programmer exists, experienced programmers can focus on design and development of the DSL – in collaboration with domain experts - leaving junior members of the team to focus on everything else. This is an efficient use of team resources.

**Validation at domain Level:** A general purpose language compiler does not know anything about domain concepts whereas a DSL can be configured in a way to check validity of domain constraints during the language compilation phase and avoids confusion [6][8].

### C. Disadvantages of DSLs

DSL is effectively another programming language thus:

**Designing a language is hard:** no matter how easy and user friendly is the language, terminology design is a complex and hard task.

**Expandability of DSLs is challenging:** The nature of DSLs is to focus on a specific problem of a domain. DSLs usually evolve iteratively and independently.

**Designing a DSL could be expensive:** Although we have posited the positive aspects of productivity as a benefit of developing a DSL, designing a DSL could be expensive as the task has to be performed by experienced programmers and also involves lots of collaboration and communication with domain experts. The design of a DSL has to be finically justified first. In addition, as DSLs are not multi-purpose languages hence their development could be expensive [8].

### D. DSL Patterns

Generally speaking, there are only two patterns available for implementing DSLs: Internal and External DSLs. Despite this, choosing the right implementation is not as easy as it seems.

At a very high level of definition, internal DSLs are developed on top of an existing *host language* like Java, Ruby or Scala. Therefore, although you are interacting with your DSL, your DSL's commands are parsed and converted to the host language's lines of code behind the scenes. The lines that are generated are then compiled and executed, the outcomes of which will eventually result in your desired action(s). Smart-API, Reflective Meta-Programming, Typed Embedding and many other techniques can be listed as sub-categories of Internal DSL development patterns.

With regard to external DSL development however, the DSL developer is responsible for receiving and parsing the DSL command/script and then generate and execute the code accordingly. In External DSL development, the DSL evolves as it grows. Unlike Internal DSLs, there are not too many patterns available for External DSLs, simply because the coder has to implement everything from scratch.

To put it simply, with an internal language, the developer starts with many of the host language's features and then strips-out/hides those functionalities that are not appropriate to their users. With external DSLs, the developer starts from almost nothing and gradually adds the functionality that is needed for the DSL [6][8].

### E. Internal or External DSL?

When it comes to software design, no universally applicable choice is available. Each different approach can be justified individually, as each has advantages and disadvantages.

In our approach to designing this framework, we decided to use the External DSL design pattern. We justify this approach as follows:

**Expressivity:** Security policies can become extremely complicated; thus, the DSL should be flexible enough to cope with such requirements. As a result, Internal DSLs may not be a good choice, simply because exposing a limited level of the functionality of the host language to DSL users may not be easily achievable with internal DSLs if the correct level of DSL expressiveness is desired.

**Lack of a codebase:** Internal DSLs are often used in scenarios where infrastructure and the knowledge base of a system exist; hence, the DSL can reuse the majority of the existing code and/or infrastructure. DSLs that are usually written to help and engage system domains to test the software can be presented as an example. Although the security policy languages will play an important role in our framework, we do not use/reuse their codebase in our framework.

**Design freedom:** In our framework, we are literally going to design a new abstract language. Thus, we need to be able to tailor the language in such a way that it is more

appropriate for the user. Parsing the DSL, error handling and DSL grammar can all be presented as examples. These, and many more features, cannot be easily achieved using Internal DSLs.

#### IV. PUTTING THEM ALL TOGETHER

##### A. Programming Languages

Choosing the correct programming language was another challenge that we faced. Although most of the available programming languages are sufficiently mature to satisfy a developer in the beginning stages, the variety of these languages makes it difficult to choose the right one. Having said that, we knew that we were going to develop an external DSL from the outset, and that helped us to narrow the options, as seen in Table 1.

Each individual column in the table would have an impact on development of DSLs and specifically on External DSLs. For instance parser combinators are a great feature (library) provided in recent programming likes Scala [9]. By using this library developers would be able to build a structure which gets populated by the defined rules whilst DSL gets parsed by host language's parser.

TABLE 1. PROGRAMMING LANGUAGE TO CODE DSLS

LANGUAGE	STATICALLY TYPED CHECKED	PARSER COMBINATOR	CONTEXT-DRIVEN STRING MANIPULATION
Scala	Yes	Yes	Medium
Groovy	No	No	Easy
C# (.Net)	Yes	Yes	Medium

As it was expected, Table 1 also shows that there is no distinctive winner amongst programming languages. Having said that it seemed Scala provided noticeably more features (libraries) compare to others languages. Admittedly, the knowledge base of our team also had an impact on our selection and resulted to choose Scala as our programming language.

##### B. Proposed Architecture

As mentioned previously in terms of External DSL development, the developer has to build everything from the ground up. Generally speaking, External DSL scripts are parsed by a parser and the system then creates a semantic model from the parsed tokens. The semantic model is then usually integrated into the application.

The development of an External DSL is not very different from designing and developing a completely new programming language. Fortunately, there are existing tools that make language development easier. These tools usually provide a framework that provides the developer with off-the-shelf functionalities, which can be configured to achieve the goal. Xtext [10] is an example of one of these tools.

Xtext is a framework for development of programming languages. It covers all aspects of a complete language

infrastructure, from parsers, over linker, compiler or interpreter to fully-blown Eclipse IDE integration [11]. It comes with good defaults for all these aspects and at the same time every single aspect can be tailored to users' needs.

Briefly put, when using Xtext, the language rules - or in our case, the DSL grammar rules - first need to be defined in Extended Backus Naur Form (EBNF) syntax. Xtext then uses that and utilizes ANTLR to build an e-Core semantic Abstract Syntax Tree (AST) model out of DSL script. Existing code-generators then walk through the semantic model and generate code from it.

Using Xtext, we came up with the following architecture for our framework (see Figure 2). In this architecture, the grey area is provided by the Xtext framework. Having said that, we were able to choose our code generator and, as we had already chosen Scala as our programming language, we decided to use the Oitok framework [12]. Unfortunately, we had so many problems incorporating Oitok into Xtext that we decided to replace it with Xpand[13]. Xpand is a statically-typed template language featuring

- polymorphic template invocation,
- aspect oriented programming,
- functional extensions,
- a flexible type system abstraction,
- model transformation,
- model validation and much more

However, Xpand generates Java (as opposed to Scala) code from the produced AST which effectively forces us towards a diversion from our initial design.

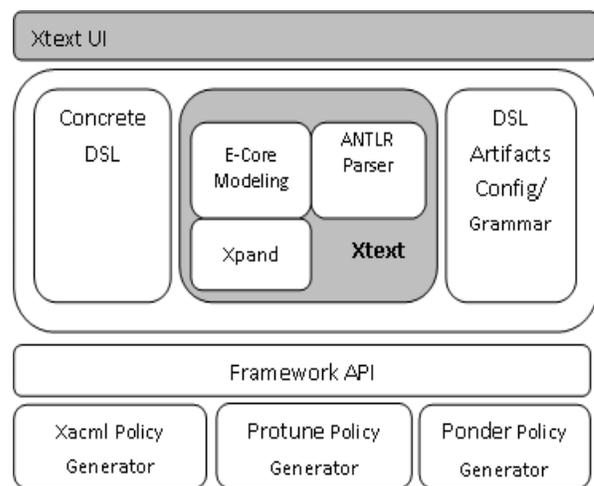


Figure2. Architecture of proposed framework.

In a nutshell, the components used in the above structure are:

**Xtext:** as language development framework uses Eclipse Modelling Framework (EMF); hence, Xtext UI is an Eclipse IDE. **ANTLR** was chosen and **eCore** was used to

model the DSL, but both are internal to Xtext and cannot be easily changed. However, *Xpand* has been chosen as the Java code generator.

**Concrete DSL:** represent concrete semantic model of DSL. Typically these are the DSL Objects/classes which are used by policy generators later on.

**DSL Artifacts:** represents DSL specific configuration like DSL grammar and configuration files.

**Framework API:** is used to connect the external policy generators to DSL framework.

**Policy Generators:** are piece of code which are responsible to generate policy specific code based on the generated concrete DSL and imposed rules enforced by DSL artifacts.

### C. Design Review and Enhancement

Following a user-centric design approach [14], we implemented a proof of concept (PoC) of the proposed architecture and analyzed the code from different angles. The outcome can be summarized as follows.

#### 1) Users' Point of View

The main issue of the design was its dependency on the Eclipse framework. The end user is required to have both Eclipse and the Java Development Kit (JDK) installed on their machines. The look and feel of Eclipse IDE itself to a non-technical user was not very friendly and was the major issue of the design. It imposed another level of unnecessary learning curve to the end user. Not to mention installation of a new piece of software i.e. Eclipse / JDK on a secure domain environment can be pragmatically impossible as well.

#### 2) Developer Team's Point of View

The above design came with a few advantages. As an example Xtext does most of the messy and hard jobs for developers. Providing intellisense/code-assist to developers and even end-user also can be presented as another advantage. However developers also need to learn another language, i.e. EBNF syntax. Imposing such a specific language which can be used in very limited occasions is not always welcomed. Beside this we realized that the EBNF rules could get really complicated and messy when the DSL evolves.

Taking the above issues into account we improved our External DSL implementations and we came up with following design for the framework. In this new architecture we decided to walk away from the Eclipse dependency hence changed the architecture as shown in Figure3.

**User Interface:** Instead of Eclipse IDE, we used *ACE* [15], an iFrame-editable code editor in JavaScript. The main advantage of this component is that the end users can communicate with the system over *http* via their browsers.

**Parser:** As opposed to ANTLR, we used the Scala combinator parser, which is sophisticated, yet easy to use.

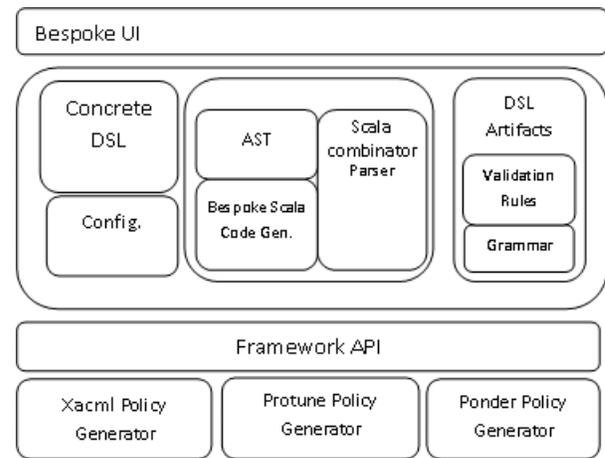


Figure3. Revised architecture of proposed framework

**Code Generator:** We used Scala and developed a code generator from the ground up.

**Validation Rules,** the new architecture again gave us freedom of mix and match various components into the framework. Validation rules was one of the new components we used in our new design. Using the new component security administrator provides the luxury of checking policy rules at compile time instead of runtime.

**Configuration,** Another component which was added to the system was a configuration module. In order to make different parts of system configurable yet decoupled, we have used Scala Cake dependency injection [16].

**AST.:** Finally, we used AST for modeling the parsed DSL script, instead of eCore, which is used by Xtext.

Unlike enhancement of the architecture, the abstract language of our framework is going through ongoing enhancement. We have started from a very basic language and enhanced the grammar of the language through a repetitive process. As it has been mentioned before, at the moment the framework only is capable of translating the abstract language to three security policy languages that we choose in [1]. More likely expanding the framework to accepts more security policy languages would have direct impact on the abstract language.

In Appendix we have provided an example of a policy written in the Abstract language and its corresponding policy written in actual security policy language as it has been generated by the framework.

## V. CONCLUSION

Further to our on-going research on security policy languages, we have proposed and outlined a framework for them in this paper. We have shown how security domain administrators would be able to take advantage of such a framework. In addition, we have shown that such a framework fulfils the requirements of DSL implementation. We then briefly reviewed DSL implementation patterns,

chose an appropriate one and nominated a programming language to implement the proposed architecture. Finally, we showed why and how the architecture had to be evolved in order to be considered fit for the purpose.

## VI. REFERENCES

- [1] A. Aryanpour, S.Y. Yan, S. Davies, A Harmel-Law. 2012. In Proc SAM12. Towards design an interoperability framework for security policy languages. Las Vegas, USA.
- [2] Extensible access control markup language (XACML) version 3.0. Oasis standard, Dec 2013.
- [3] N Damianou, N Dulay, E Lupu, M Sloman. 2000. Ponder: A Language for Specifying Security and Management Policies for Distributed Systems The Language Specification Version 2.3.
- [4] N Dulay, E Lupu, M Sloman, N Damianou. 2001. A Policy Deployment Model for the Ponder Language. Proceedings of the 7th IEEE/IFIP Symposium on Integrated Network Management (IM'01), Seattle USA.
- [5] P. A. Bonatti and D. Olmedilla. 2005. Driving and monitoring provisional trust negotiation with metapolicies. In 6th IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY 2005), Stockholm, Sweden.
- [6] M Fowler. 2010. Domain Specific Languages. Addison-Wesley Professional.
- [7] D. S. Kolovos, R. F. Paige, T. Kelly, F. A. C. Polack. Requirements for Domain-Specific Languages. In Proc. 1st ECOOP Workshop on Domain-Specific Program Development (DSPD 2006)
- [8] D. Ghosh. 2011, DSLs in Action. Manning Publications.
- [9] M Odersky, L Spoon, B Venners. 2010. Armita press. Programming in Scala, Second Edition A comprehensive step-by-step guide. ISBN : 0981531644
- [10] Xtext. Open-source framework for developing programming languages. Version 2.4 September 2013
- [11] Eclipse. An integrated development environment. Version 4.3. June 2013.
- [12] M. Völter. 2008. Model-Driven Development of DSL Interpreters Using Scala and oAW.
- [13] Xpand. A language specialized on code generation based on EMF models. Version 1.4. July 2013
- [14] D Wallach, S Scholz. 2010. User-Centered Design: Why and How to Put Users First in Software Development.
- [15] ACE. An embeddable code editor written in JavaScript. Version 1.1.01. July 2013.
- [16] Cake, Dependency injection for Scala. <http://jonasboner.com/2008/10/06/real-world-scala-dependency-injection-di/>. Visited August 2013

## VII. APPENDIX

The following is an example of a policy written in our proposed abstract security policy language. As it appears from the example we have tried to design our abstract language a human readable language with the hope to provide a common vocabulary to all different parties involved defining a security policy.

```
Protect
  Target "myTarget"
from
  Subjects "mySubject"
for executing
  Actions "action1"
under following
  Subject Conditions "mySubjectCondition1"
```

If we configure the framework to produce XACML policies the above policy will be translated to a XACML policy as follow:

```
<Policy PolicyId="myPolicy"
RuleCombiningAlgId="urn:oasis:names:tc:xacml:1.0:rule-combining-
algorithm:first-applicable"
xmlns="urn:oasis:names:tc:xacml:1.0:policy"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <Description>This is a Auto generated policy.</Description>
  <Target>
    <Subjects>
      <Subject>
        <SubjectMatch MatchId="urn:oasis:names:tc:xacml:1.0:function:string-
equal">
          <AttributeValue
DataType="http://www.w3.org/2001/XMLSchema#string">mySubject</At
tributeValue>
          <SubjectAttributeDesignator AttributeId="ubdRole"
MustBePresent="false"
          DataType="http://www.w3.org/2001/XMLSchema#string"/>
        </SubjectMatch>
      </Subject>
    </Subjects>
    <Resources>
      <ResourceMatch
MatchId="urn:oasis:names:tc:xacml:1.0:function:string-equal">
        <ResourceAttributeDesignator AttributeId="ubd:resource:Type"
          DataType="http://www.w3.org/2001/XMLSchema#string"/>
        <AttributeValue
DataType="http://www.w3.org/2001/XMLSchema#string">MyTarget</At
tributeValue>
      </ResourceMatch>
    </Resources>
    <Actions>
      <Action>
        <ActionMatch MatchId="urn:oasis:names:tc:xacml:1.0:function:string-
equal">
          <AttributeValue
DataType="http://www.w3.org/2001/XMLSchema#string">Action1</Attri
buteValue>
          <ActionAttributeDesignator AttributeId="ubd:action:type"
          DataType="http://www.w3.org/2001/XMLSchema#string"/>
        </ActionMatch>
      </Action>
    </Actions>
  </Target>
  <Rule RuleId="rul_own_record" Effect="Permit">
    <Condition>
      <Apply FunctionId=
"urn:oasis:names:tc:xacml:1.0:function:string-equal">
        <SubjectAttributeDesignator DataType=
"http://www.w3.org/2001/XMLSchema#string">
          mySubjectCondition1
        </SubjectAttributeDesignator>
      </Apply>
    </Condition>
  </Rule>
</Policy>
```

Indeed if we configure the framework to produce Protune or Ponder policies instead, it will generate policies written on those specific security policy languages in accordance to the abstract security policy provided. The generated Protune and Ponder policies for above sample abstract security policy are removed for space conservation.

# The Implementation of BGP Monitoring, Alarming, and Protecting System by a BGP-UPDATE-Based Method using ECOMMUNITY in Real Time

Je-kuk Yun<sup>1</sup>, Beomseok Hong<sup>2</sup>, and Yanggon Kim<sup>3</sup>

<sup>1</sup>Information Technology, Towson University, Towson, MD, U.S.A.

<sup>2</sup>Information Technology, Towson University, Towson, MD, U.S.A.

<sup>3</sup>Computer Science, Towson University, Towson, MD, U.S.A.

**Abstract** - As the number of IP hijacking incidents has increased, many IP hijacking monitoring tools have been implemented. However, none of the monitoring tools can directly control the data plane of BGP routers. Therefore, network administrators should protect their routers by using command line interface when the network administrator receives any warning from BGP hijacking monitoring tools. As the number of routers and prefixes continuously increased, checking the routing information in their routers manually is one of the big burdens on the administrators. In addition, when IP hijacking occurs, it is very important for the administrator to quickly block the bogus prefixes. Otherwise, thousands of traffic will be transferred to the wrong way within a very short moment. In this paper, we extended Quagga-SRx so that the Quagga-SRx can send a BGP update message including an opaque extend community to other iBGP peers for notifying bogus IP prefixes after detecting abnormal IP prefixes. As a result of this, the other iBGP peers can recognize bogus IP prefixes by accepting the BGP update message that includes the opaque extend community, and automatically blocks the bogus prefixes if the iBGP routers have an ability to process the opaque extend community. Therefore, when IP hijacking occurs, the bogus prefixes can be blocked automatically and quickly, which makes the ASes more secure.

**Keywords:** BGP, border gateway protocol, interdomain routing, network security, networks, routing

## 1 Introduction

The BGP is an Inter-domain routing protocol, and is the routing protocol that enables large networks to form a single Internet. The main functionality of BGP is to exchange Network Layer Reachability Information (NLRI) between Autonomous Systems (ASes) so that a BGP speaker can communicate with other BGP routers and ultimately can reach a destination of a certain router [1]. However, when the BGP was designed, its vulnerabilities were hardly considered [2].

Unfortunately, the lack of consideration of BGP vulnerabilities occasionally causes severe failure of Internet service provision. Such a failure called prefix hijacking causes attacks on the routing infrastructure or the control plane of the Internet. The prefix hijacking occurred on the 25th of April in 1997 by a misconfigured router that advertised incorrect prefixes and announced AS 7007 as an origin of them. As a result, it created a routing black hole for almost two hours [3]. Similar events occurred on the 22nd January in 2006, when Con Edison (AS 27506) stole several important prefixes by misconfiguring them [4], and on Christmas Eve 2004 TTNNet in Turkey (AS 9121) advertised the entire prefixes on the Internet so that every route came to them rather than to the correct destinations [5]. The most well-known incident of prefix hijacking was the YouTube hijacking by Pakistan Telecom on the 24th of February in 2008, when in response to government order to block YouTube access within their country, Pakistan Telecom advertised a more specific prefix than the YouTube prefix. One router believes the Pakistan Telecom regarded the incorrect prefix as a more specific prefix and advertised it to the others. As a result, YouTube was not accessible from some ASes for a while [6]. Furthermore, Google was affected by a hijacking on the 9th of July in 2010. The reason of this hijacking was revealed as a copy and paste mistake by a network administrator. Next month Google was affected by another hijacking because AS30890 (EVOLVA Evolva Telecom s.r.l.), a provider from Romania announced the same prefix as Google's prefix which is 8.8.8.0/24 [7]. Such a prefix hijacking happens constantly on the Internet by mis-announcement or intentionally advertising malicious prefixes.

In order to deal with the prefix hijacking and mis-announced prefix in a faster way, several monitoring tools have been developed such as PHAS, BGPmon, Cyclops, and Argus. However, none of the monitoring tools directly handle the data plane of BGP routers. Every time the malicious prefixes are found, the tools could only warn an administrator about the hijacking. In addition, the manual configuration would be a matter. Since the number of BGP routers and prefixes has increased steadily over the past years, we expect that the administrators need to check and manually block prefixes

more than they do now. On the other hand, when IP hijacking occurs, it is very important for the administrator to quickly block the bogus prefixes. Otherwise, thousands of network traffic will be transferred to the wrong way within a very short moment.

In this paper, we extended the previous work [8] to automatically protect their iBGP peers by sending an update message that includes an opaque extended community as well as monitor bogus prefixes. Therefore, iBGP peers can be protected without the network administrator's manual configuration. We describe two main problems of BGP and a solution in Section 3. In section 4, we explain how the BGPMAPS works. In addition, we discuss the implementation of BGPMAPS and the result of BGPMAPS performance test in section 5. Lastly, we conclude the paper in section 6.

## 2 RELATED RESEARCH

### 2.1 Monitoring tools

Prefix Hijack Alert System (PHAS) is a system that detects an attempt to hijack prefixes, owned by other BGP routers, with BGP routing data collected by BGP collectors and it notifies prefix owners of the hijack attempt through a reliable manner. However, PHAS does not guarantee to detect anomaly advertisements [9]. BGPmon is designed to collect a large number of data and provides real-time access with clients for real-time analysis [10]. The collected data is consolidated into a single XML steam that can be extended by clients for their purposes. Cyclops is a system that collects real-time updates of hundreds of routers and displays a graphic view of how the routers are connected to each other [11]. As a result, network administrators can use the tool to detect and diagnose BGP misconfigurations or BGP hijacking. Argus is an agile system that receives real-time updates from BGPmon and daily updates from CAIDA iPlane. Based on the updates, Argus checks whether the update has an anomalous origin according to its local routing information database [12]. After finding the suspicious prefixes, Argus makes a final decision by computing the fingerprint for the suspicious prefixes.

### 2.2 BGP-SRx

BGP-SRx, developed by the National Institute of Standards and Technology (NIST), consists of the SRx Server, the SRx API, and the Quagga SRx [13]. SRx provides a proxy with APIs, which allows the proxy to be embedded on the router and communicate with the SRx Server. The Quagga SRx is a software router on which the proxy is embedded. The SRx Server is connected to the RPKI validation cache, so the SRx Server can validate BGP announcements by comparing the BGP announcements to ROAs in the RPKI validation cache.

## 3 BGP THEATS AND SOLUTION

### 3.1 IP hijacking

Once BGP routers are connected to each other, the BGP routers fully trust each other. If a BGP router intentionally originates a bogus prefix to neighbors, the neighbors that receive the announcement trust the prefix and their traffic is hijacked by the hijacking router.

Fig. 1 shows that AS 500 is trying to hijack the traffic heading for AS 600. AS 600 announces 10.60.0.0/16 to neighbors and traffic heading for 10.60.0.0/16 is transferred to AS 600. However, if AS 500 announces a bogus prefix, 10.60.0.0/16, to AS 100, then the traffic in AS 100 goes to AS 500 because the number of hops to AS 500 is shorter than AS 600's hops. As a result, AS 100 takes AS 500 as the destination for the 10.40.0.0/16.

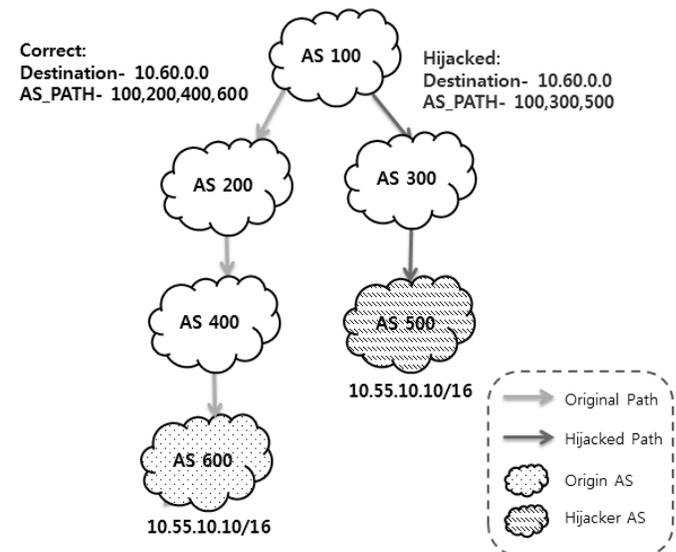


Fig. 1. IP prefix hijacking

### 3.2 Mis-announced route

BGP doesn't validate announcements, so if an AS administrator unintentionally mis-types the length of a prefix or IP address, then any routers that receive the announcements trust the mis-typed length of the prefix or IP address. As a result of this, traffic in the routers that received the wrong announcements goes to the wrong destination.

### 3.3 Origin Validation

Many studies have been conducted to solve IP hijacking, such as Secure BGP (S-BGP) [14], Secure Origin BGP (SO-BGP) [15], Pretty Secure BGP (psBGP) [16], Pretty Good BGP (pgBGP) [17], and so on. Secure Inter-Domain Routing (SIDR) working group completed Resource Public Key Infrastructure (RPKI) [18]. In order to prevent IP hijacking, the only BGP speaker that has been authorized by the IANA

should originate its prefixes. Therefore, BGP speakers should be authorized by the IANA.

The IANA manages an officially verifiable database of the authorized IP prefixes and AS numbers. Therefore, BGP routers need to periodically retrieve the collection of the IP prefixes and AS numbers called Route Origin Authorizations (ROAs) [19] that consist of IP prefix, ASN, and maxLength. When a BGP router originates its prefix, the length of the accompanying prefix must be an integer less than or equal to the maxLength. Otherwise, the prefix is considered invalid. For example, if the IP prefix is 10.30.0.0/16 and the maxLength is 19, then the BGP speaker is authorized to originate 10.30.0.0/17, 10.30.0.0/18, or 10.30.0.0/19, not but 10.30.0.0/20.

## 4 IV. DESIGN OF THE BGP MAPS

### 4.1 Overview

Fig. 2 depicts the simple architecture of the BGP MAPS. The ex-Quagga-SRx is connected to a BGP router with eBGP connection and learns routes from the BGP peers of other ASes. The ex-Quagga-SRx forwards the routes to the internal BGP routers and announces the internal routes of the AS to the BGP peers of other ASes.

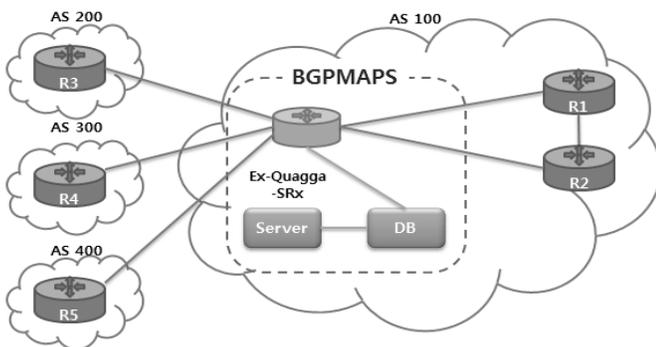


Fig. 2. The simple architecture of the BGP MAPS

In order to learn new routes from eBGP peers, the BGP MAPS is required to establish a BGP connection to the eBGP peers. Whenever the BGP MAPS receives a BGP update message, the ex-Quagga-SRx saves the result of the RPKI validation to the Data Agent as well as the routing table, such as ASN, PREFIX, AS\_PATH, NEXTHOP, etc. In addition, the ex-Quagga-SRx forwards the update message that includes a new attribute of the opaque extended community to the iBGP peers for notifying the iBGP peers of the validation state of the update message.

### 4.2 Operational requirements

In the design of the BGP MAPS, we consider three requirements: (1) it should store a large amount of prefixes' history, (2) it should detect the bogus prefix and notify iBGP

peers of the bogus prefix quickly and automatically when a bogus prefix is forwarded from BGP peers of other ASes, and (3) it should be a long living process.

In the past decade, network administrators were worrying about the growth of the number of routes in the Internet because network routers have a limited memory capacity of storing many prefixes in the routing table. The router may need an extra space to save additional routing information for preventing IP hijacking. As the number of prefixes is increased, the results of the validation are saved in the local database instead of the main memory of the router.

It is very important for network administrators to block bogus prefixes as soon as they detect the bogus prefixes. In case of a big company such as Google, YouTube, Amazon, and so on, a considerable number of Internet packets come and go every single second. However, the network administrators cannot keep paying attention to their routing table. Even though the administrators subscribe IP hijacking alarm service [10] and receive an alarm message through email, it takes time for the network administrators to block the bogus prefixes by using command line interface. In order to detect the bogus prefixes and notify iBGP peers of the bogus prefixes, the opaque extended community attribute, which is including validation state, is added to the update message and the update message is forwarded to iBGP peers.

### 4.3 The architecture of the BGP MAPS

Fig. 3 shows the architecture of the BGP MAPS. The BGP MAPS consists of the Extended Quagga SRx (ex-Quagga-SRx), Data Agent (DA), Alarm Server, and Web Service Agent (WSA). The ex-Quagga-SRx sends an update message to BGP peers and receives update messages from BGP peers through a BGP connection. Then, the ex-Quagga-SRx sends update message information such as ASN, prefix, and maxLength to the SRx Server.

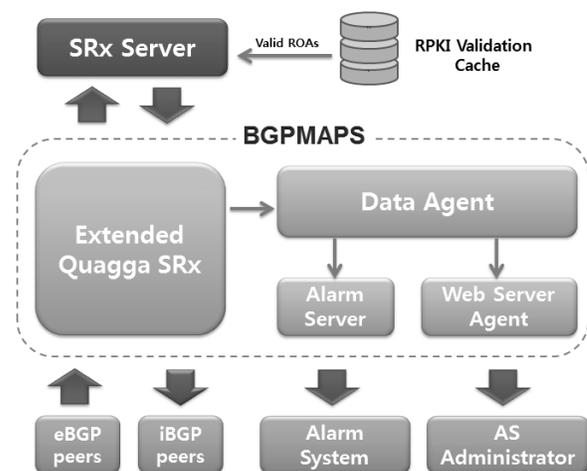


Fig. 3. The architecture of the BGP MAPS.

The SRx compares the update message information to the ROAs and returns the result of validation to the ex-Quagga-SRx. In case of eBGP peers, the Data Agent receives the result of the BGP update message from the ex-BGP-SRx. If the result is invalid, the Alarm Server notifies the Alarm System of the invalid update message. Then, the Alarm System makes an alarm sound, and BGP router administrators can check the invalid update message through the web interface. The Quagga-SRx is connected to a BGP router via iBGP. Once the Quagga-SRx receives update messages, the Quagga-SRx sends a query to the SRx server and receives the result of the update messages. We extended the existing Quagga-SRx so that the Quagga-SRx saves the validation results of the update messages in the database. The Data Agent maintains and manages the validation results of update messages. When an invalid prefix is detected in the database, the Alarm Server makes an alarm sound through the Alarm System to warn AS administrators. The Alarm server sends signal to the Alarm System through TCP by using the IP address and port. The WSA provides AS administrators with a web interface that displays the validation result of update messages. The Alarm System is required to be installed in the AS that wants to receive the alarm service. When an invalid message is detected by the BGP MAPS, the Alarm System receives a signal from the Alarm Server. As a result, the AS administrator doesn't need to monitor a BGP routing table all the time to protect from IP hijacking.

## 5 IMPLEMENTATION AND PERFORMANCE TEST

### 5.1 Creating the Opaque Extended Community

BGP speakers keep updating their routing table by sharing their routing through an update message. Fig. 4 shows the update message consists 'Withdrawn Routes Length,' 'Withdrawn Routes,' 'Total Path Attribute Length,' 'Path Attributes,' and 'Network Layer Reachability Information.' The path attributes includes a number of attributes such as 'ORIGIN,' 'AS\_PATH,' 'NEXT\_HOP,' 'COMMUNITY,' 'EXTENDED COMMUNITY,' and so on. The 'EXTENDED COMMUNITY' attribute will contain opaque extended community to carry the validation state of the update message between iBGP peers. The opaque extended community is a work in progress by Secure Inter-Domain Routing working group "and may be standardized on March 2014."

An update message can include a number of Path Attributes. So if a BGP speaker receives an update message that includes a bogus prefix, then the BGP speaker adds the opaque extended community to the update message and forwards the update message to iBGP peers. The iBGP peers

can know whether the prefix is hijacked by checking the opaque extended community.

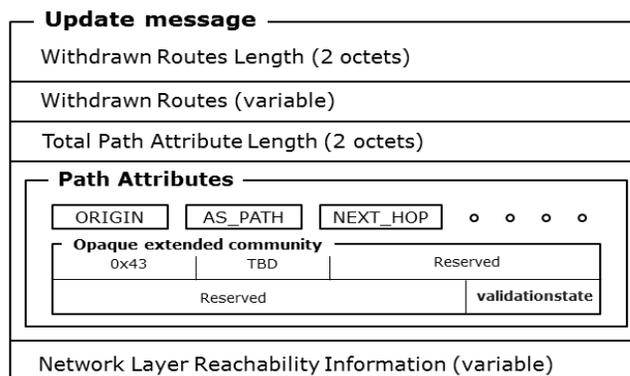


Fig. 4. The update message format

### 5.2 Notifying functions with update message

An operation of notifying function is as follows (see Fig. 5)

- The eBGP peer announces an update message
- The Ex-Quagga SRx sends a query to the SRX Server for an update validation
- The SRX server requests ROAs to the RPKI Validation Cache
- The RPKI Validation Cache provides ROAs
- The SRx Server conducts RPKI validation by comparing a prefix and an ASN to ROAs
- The SRx Server returns the result of validation to the Ex-Quagga SRx
- The Ex-Quagga SRx saves the result of validation to the Data Agent
- The Alarm Server sends notification to the Alarm System according to database in the Data Agent
- The Web Server Agent provides web interface service by using the validation information provided by the Data Agent
- The Ex-Quagga SRx adds the opaque extended community to the update message
- The Ex-Quagga SRx forwards the update message to iBGP peers
- iBGP peers update their control plane according to the update message

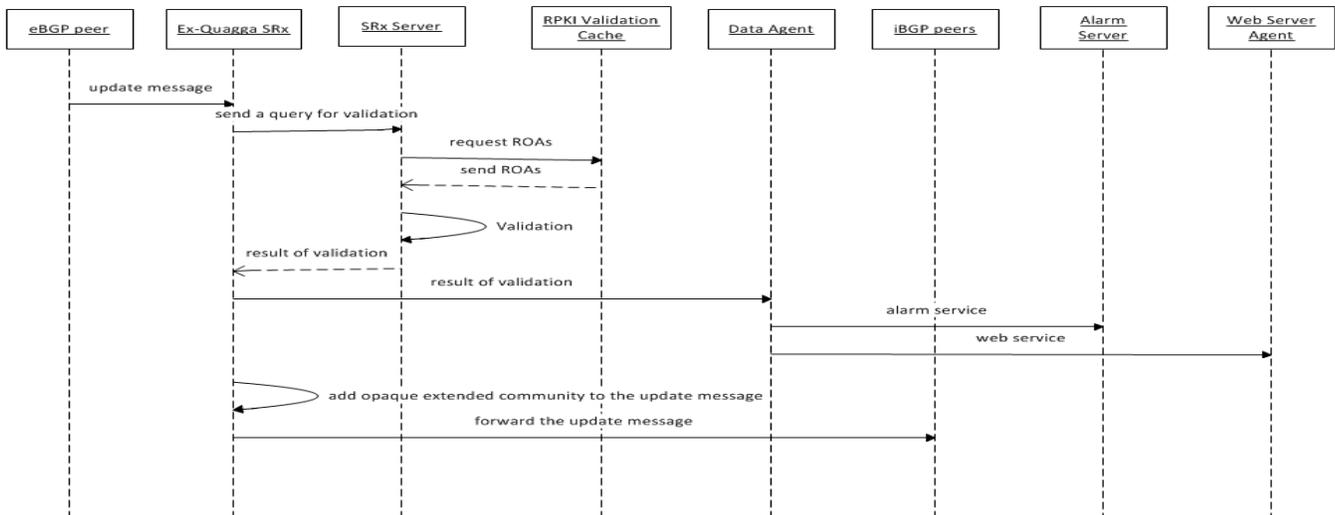


Fig. 5. Sequence of detecting and preventing functions

### 5.3 Simulation of the BGP MAPS

Fig. 6 shows a topology that includes four ASes and 6 BGP routers. AS 200, AS 300, and AS 400 include one router that doesn't have RPKI capability. AS 100 includes two routers that doesn't have RPKI capability and one software router that has RPKI capability, called Ex-Quagga-SRx. Through this topology, we explain how the Ex-Quagga-SRx automatically notifies iBGP peers of an invalid update message when the Ex-Quagga-SRx detects a bogus message.

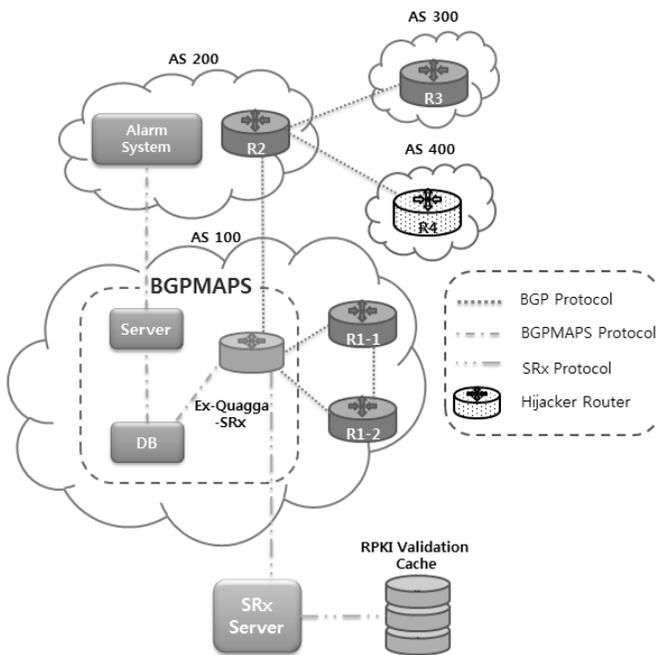


Fig. 6. Topology for simulation

First, RPKI validation Cache includes 3 ROAs which are 10.30.0.0/24 300, 10.30.1.0/24 300, and 10.30.2.0/24 300.

AS 200 originates 10.20.0.0/24 and the announcement is forwarded to neighbors. AS 300 originates 10.30.0.0/24, 10.30.1.0/24, and 10.30.2.0/24. Then, the announcements are forwarded to the neighbors, AS 200, AS 100, and AS 400. Suppose AS 400 hijacks the traffic heading for AS 300 by originating 10.30.0.0/17 to AS 200. Then, Ex-Quagga-SRx receives the bogus announcement and sends a query to the SRx Server to validate the bogus announcement. After detecting the bogus announcement, the Ex-Quagga-SRx creates the opaque extended community attribute, adds the opaque extended community attribute to the update message, and sends the update message to the iBGP peers. Then the iBGP peers can recognize bogus prefixes and will not select the AS 400 as the destination of 10.30.0.0/24. Furthermore, the Ex-Quagga-SRx saves the bogus announcement in the database. The Server monitors the database and automatically makes an alarm sound when the bogus announcement is discovered in the database. Once the sound of the alarm is made in AS 200, the AS 200 network administrator can realize that there is a bogus prefix in the BGP router. In addition, the AS 200 administrator can check the bogus prefix through the webpage that is provided by the BGP MAPS. As a result of this, the AS 200 administrator can ignore the bogus prefix in a short time.

Fig. 7 shows the BGP table on the Ex-Quagga-SRx. The SRxVal shows the results of validation after the SRx Server validates the update messages. We created an eVal column on the table to show the opaque extended community value.

SRxVal	eVal	SRxLP	Status	Network	Next Hop	Metric	LocPrf	Weight	Path
*> u(u,-)	1			10.20.0.0/24	10.55.10.180	0		0	200 i
*> v(v,-)	0			10.30.0.0/24	10.55.10.180			0	200 300 i
*> i(i,-)	2			10.30.0.0/25	10.55.10.180			0	200 400 i
*> v(v,-)	0			10.30.1.0/24	10.55.10.180			0	200 300 i
*> v(v,-)	0			10.30.2.0/24	10.55.10.180			0	200 300 i

Fig. 7. Ex-Quagga-SRx BGP table

Fig. 8 shows the normal BGP table in case that the opaque extended community doesn't exist in the BGP update message. The normal BGP table indicates that the router will select AS 400 as the destination of 10.30.0.0/24 because the prefix 10.30.0.0/25 announced by AS 400 has longer prefix than 10.30.0.0/24 announced by AS 300. In that case, the AS 300 is hijacked and the traffic for 10.30.0.0 will be transferred to the AS 400.

Network	Next Hop	Metric	LocPrf	Weight	Path
* i10.20.0.0/24	10.55.10.180	0	100	0	200 i
* i10.30.0.0/24	10.55.10.180		100	0	200 300 i
* i10.30.0.0/25	10.55.10.180		100	0	200 400 i
* i10.30.1.0/24	10.55.10.180		100	0	200 300 i
* i10.30.2.0/24	10.55.10.180		100	0	200 300 i

Fig. 8. Normal BGP table

Fig. 9 shows R1-1 BGP table after the R1-1 receives the opaque extended community from the Ex-Quagga-SRx. Even though the R1-1 doesn't have a capability of validating update messages, the R1-1 can recognize the invalid prefix by accepting the opaque extended community. The BGP table includes the eVal to show the validation state. As a result, the R1-1 will not select AS 400 as a destination of 10.30.0.0/24.

eVal	Network	Next Hop	Metric	LocPrf	Weight	Path
* i 1	10.20.0.0/24	10.55.10.180	0	100	0	200 i
* i 0	10.30.0.0/24	10.55.10.180		100	0	200 300 i
* i 2	10.30.0.0/25	10.55.10.180		100	0	200 400 i
* i 0	10.30.1.0/24	10.55.10.180		100	0	200 300 i
* i 0	10.30.2.0/24	10.55.10.180		100	0	200 300 i

Fig. 9. R1-1 BGP table

## 5.4 Environment and the result of performance test

We set up a topology that is the same as figure 6 where each router runs on a 3.40 GHz i5-3570 machine with 1 GB of memory running CentOS 6.3 and the SRx Server runs on a 3.40 GHz i5-3570 machine with 1 GB of memory running CentOS 6.3. Because it is important to quickly notify iBGP peers of bogus prefixes, we measured the duration from the moment that Ex-Quagga-SRx receives an update message to the moment that the Ex-Quagga-SRx forwards the update message, which includes the opaque extended community, to iBGP peers. We stored 8000 ROAs in the RPKI validation Cache because the number of ROAs in the RPKI validation Cache affects the performance and there are around 8000 ROAs in the real world. We collected the number of prefixes that originated from each AS the 25th of May in 2013. According to TABLE 1, the BGP router originates, on average, 11.74 prefixes and at most 5230 prefixes.

TABLE I. THE NUMBER OF PREFIXES IN WHOLE AS

	Minimum	Maximum	Average
Prefix(IPv4)	0	4808	11.42
Prefix(IPv6)	0	422	0.32
Total	0	5230	11.74

We varied the number of prefixes and Fig. 10 illustrates the result of the performance test. When a router originates 1000 prefixes and the Ex-Quagga-SRx receives the prefixes, it takes 7.54 seconds. As the number of prefixes is increased, the time it takes increases linearly. Therefore, we can say it takes a reasonable time to notify iBGP peers of invalid prefixes.

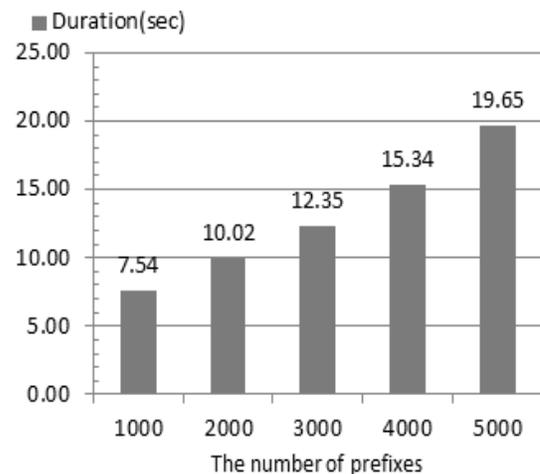


Fig. 10. Result of performance test

## 6 CONCLUSION

As the number of IP hijacking incidents is increased, many IP hijacking monitoring tools are implemented, but network administrators still should pay attention to their routing table to protect their routers by using command line interface when the network administrator receives any warning from BGP hijacking monitoring tools because none of the monitoring tools can directly access the data plane of BGP routers. As the number of ASes and prefixes continuously increase, checking the routing information in their routers manually is one of the big burdens on the administrators. In addition, when IP hijacking occurs, it is very important for the administrator to quickly block the bogus prefixes. Through our performance test, we showed iBGP peers can recognize the invalid prefixes in a reasonable time by accepting the opaque extended community even though the iBGP peer doesn't have a capability of validating update messages. As a result, when IP hijacking occurs, the bogus prefixes can be blocked in a timely manner, which makes the ASes more secure. We automatically prevent IP hijacking within iBGP peers, though eBGP should still be monitored in person.

## 7 References

- [1] Rekhter, Y. 2006. A Border Gateway Protocol 4 (BGP-4). RFC 4271.
- [2] Murphy, S. 2006. BGP Security Vulnerabilities Analysis. RFC 4272.
- [3] "7007 Explanation and Apology," Apr 1997, <http://www.merit.edu/mail.archives/nanog/1997-04/msg00444.html>.
- [4] Rensys Blog, Con-Ed Steals the 'Net. [Online]. Available: [http://www.renysys.com/blog/2006/01/coned\\_steals\\_the\\_net.shtml](http://www.renysys.com/blog/2006/01/coned_steals_the_net.shtml)
- [5] Rensys Blog, Internet-Wide Catastrophe Last Year [Online]. Available: [http://www.renysys.com/blog/2005/12/internetwide\\_nearcatatrophela.shtml](http://www.renysys.com/blog/2005/12/internetwide_nearcatatrophela.shtml)
- [6] Rensys Blog, Pakistan hijacks YouTube [Online]. Available: [http://www.renysys.com/blog/2008/02/pakistan\\_hijacks\\_youtube\\_1.shtml](http://www.renysys.com/blog/2008/02/pakistan_hijacks_youtube_1.shtml)
- [7] BGPmon, Google's services redirected to Romania and Austria [Online]. Available: <http://www.bgpmon.net/googles-services-redirected-to-romania-and-austria>
- [8] J. Yun, B. Hong, Y. Kim, "The BGP Monitoring and Alarming System to Detect and Prevent Anomaly IP Prefix Advertisement", Research in Applied Computation Symposium (RACS 2013), Montreal, QC, Canada, 1-4 October 2013.
- [9] Lad, M., Massey, D., Pei, D., Wu, Y., Zhang, B., and Zhang, L. 2006. PHAS: A prefix hijack alert system. In Proceedings of the 15th conference on USENIX Security Symposium - Volume 15 (USENIX-SS'06), Vol. 15.
- [10] D. Matthews, Y. Chen, H. Yan, and D. Massey. BGP Monitoring System. Available from: <http://bgpmon.netsec.colostate.edu/>.
- [11] Y.-J. Chi, R. Oliveira, and L. Zhang. Cyclops: The AS-level connectivity observatory. ACM SIGCOMM Computer Communication Review, pages 7–16, 2008.
- [12] Xingang Shi, Yang Xiang, Zhiliang Wang, Xia Yin, Jianping Wu. Detecting Prefix Hijackings in the Internet with Argus. In Proc. of ACM IMC 2012.
- [13] BGP Secure Routing Extension (BGP-SRx) by NIST [Online]. Available: <http://www-x.antd.nist.gov/bgpsrx/>
- [14] Kent, S., Lynn, C., and Seo, K. 2000. Secure Border Gateway Protocol (S-BGP). IEEE Journal on Selected Areas in Communications. 18, 4 (Apr. 2000)
- [15] White, R. 2003. Securing BGP through secure origin BGP. Internet Protocol Journal. 6, 3 (September 2003).
- [16] Karlin, J., Forrest, S., and Rexford, J. 2006. Pretty Good BGP: Improving BGP by Cautiously Adopting Routes. In IEEE International Conference on Network Protocols.
- [17] Van Oorschot, P., Wan T., and Kranakis, E. 2007. On Interdomain Routing Security and Pretty Secure BGP (psBGP). ACM Transactions on Information and System Security. 10, 3(July 2007).
- [18] Manderson, T., Vegoda, L., and Kent, S. 2012. Resource Public Key Infrastructure (RPKI) Objects Issued by IANA(Feb. 2012). [Online]. Available: <http://www.rfc-editor.org/rfc/rfc6491.txt>
- [19] Lepinski, M., Kent, S., and Kong, D. 2012. A Profile for Route Origin Authorizations (ROAs). Work in progress (Internet Draft), Feb 2012.

# Designing a Case Research Protocol from NISTIR 7621 Security Guidelines

D. Heier

Dakota State University, Madison, SD, USA

**Abstract** - *This paper proposes a research process to study small business information security. This case study process utilizes the National Institute of Standards and Technology document: Small Business Information Security: The Fundamentals (NISTIR 7621). NISTIR 7621 seeks to be an aid to small businesses in providing guidelines for their security implementation. Small businesses in this study will be evaluated to determine their level of implementation of the NISTIR 7621 guidelines. The study also seeks to understand the constraints that may limit a small business's implementation of security practices.*

**Keywords:** Small Business, Information Security, NISTIR 7621

## 1 Introduction

This paper proposes a process to investigate small business security practices by a case study. This proposed process seeks to improve upon the common security surveys by assessing the accuracy of the survey responses. To better understand the security practices of these businesses, this process incorporates the use of the National Institute of Standards document titled "Small Business Information Security: The Fundamentals" [1]. This document is an interagency report with the intent of providing guidance to small businesses in their security actions. Small businesses are not required to follow any of the guidelines, but are strongly encouraged to in order to meet a basic level of information security [2].

### 1.1 Reviewing NISTIR 7621

NISTIR 7621 was created to be a guideline for small businesses that presented many of the fundamentals of small business security [2]. The document contains four main sections. Section 1 is an introduction that lays out the need for small businesses to take their information security practices seriously. Section 2 describes 10 "absolutely necessary" actions that a small business should take to protect its information systems. These actions are summarized as:

- Protect information/systems/networks from damage by viruses, spyware, and other malicious code.
- Provide security for your Internet.

- Install and activate software firewalls on all your business systems.
- Patch your operating systems and applications.
- Make backup copies of important business data/information.
- Control physical access to your computers and network components.
- Secure your wireless access point and networks.
- Train your employees in basic security principles.
- Require individual user accounts for each employee on business computers and for business applications.
- Limit employee access to data and information, and limit authority to install software.

Section 3 of NISTIR 7621 lays out some highly recommended practices that are suggested to be completed after implementing the absolutely necessary actions described above. Section 4 of NISTIR 7621 describes management related planning considerations for information, computers, and network security. NISTIR 7621 also contains three appendices which provide worksheets that help identify information stored, identify protection needed for the stored information, and estimate expenses that could occur if a security incident occurs.

### 1.2 Small Business Security Concerns

In the United States, the Small Business Administration [3] will, on average, define a small business as having less than 500 employees. These businesses, however, represent over 95% of all businesses and employ over half of non-governmental employees. Small businesses have increasingly become targets for hackers and cyber criminals as larger businesses with more resources have become more secure [2]. The National Cyber Security Alliance reported that in 2012 that 31% of security attacks were on small businesses, an increase from 18% in 2011 [4]. Small businesses typically have their entire technology infrastructure supported by a single employee or small department that is not well informed of security standards [5]. Typical case research questions in this study would be:

- 1) What level of security do small businesses implement when measured against the suggested practices of the NISTIR 7621 guidelines?

2) How do small businesses perceive they could improve their security practices?

## 2 Research Methodology

This proposed analysis of small business security will be based on case study research. Benbasat [6] notes three reasons supporting case study research in information systems. First the researcher can study these information systems in natural settings and then generate theories. Second, the researcher can seek answers to questions of how and why regarding information systems processes. Third, a case is appropriate when few studies are available as the field of study is rapidly changing. Gable [7] notes that case studies are not always completely qualitative and may include a quantitative survey and encourages IS researchers to combine methods as “far as is feasible”. In this particular study, it is anticipated that the initial survey will serve to develop an understanding of the case’s security situation and likely lead to more in depth answers to the interview questions.

Components of research design as suggested by Yin [8] will be utilized in this design as follows:

- *A study’s question.* Yin suggests to clarify the study questions for the appropriate how and why questions. For this study, the survey forms the basis for understanding the business security implementation and answering the first research question. The interview questions could then be developed to gain an understanding of the how and why of the small business responses about the security practice implementation with a goal to answer the second research question.
- *Its propositions, if any.* Yin suggest that propositions direct attention to items that need to be examined further and states that forming propositions will help the research move in the right direction. Based on a review of the literature about the current state of small business security, several propositions could be formed. Propositions could also be developed based on the industry type of the small businesses, geographic location, or demographics that may influence the business practices.  
  
A typical first proposition could be: Lack of resources leads small businesses to fall short of meeting all the absolutely necessary security actions outlined in NISTIR 7621.
- *Its units of analysis.* Yin indicates that the unit of analysis is related to the way the initial research questions were defined. This study protocol suggests defining a geographic area, and specific small business types in this area will be points of analysis with the unit of analysis representing the collection of these cases.

- *The logic linking the data to the propositions.* The interview questions will be measured against the survey questions to determine whether all the propositions have been met for each small business.
- *The criteria for interpreting the findings.* For the survey questions, an overall rating indicating the level to which the small business meets the absolutely necessary security actions outlined in NISTIR 7621 will be calculated. This calculation will help to interpret the responses from the interview questions.

### 2.1 Quantitative data to be collected

A survey will be constructed to collect the following:

- Business demographic information.
- Business information related to hours devoted to IT work.
- Responses to 10 questions based on NISTIR 7621 absolutely necessary actions.

The participants in the survey, when presented with the 10 security practices, are asked to rate their implementation. Table 1 suggests response options for the participant.

**Table 1: Definition of Survey Response Options**

Survey Response Choice	Description
Fully Implemented	If your business completely implements a security practice as described, select Fully Implemented.
Mostly Implemented	If your business implements at least 50% of the practice, select Mostly Implemented.
Partly Implemented	If your business implements some of the practice, but less than 50%, select Partly Implemented.
Not Implemented	If the practice is not implemented at your business, select Not Implemented.
Not Sure	If you are uncertain about how your business implements a security practice, feel free to select 'Not Sure' as your answer.

### 2.2 Qualitative data to be collected

As a follow up to a completed survey, an interview will be conducted with the individual who completed the survey. The first part of the interview will be to discuss each question of the completed survey to determine the individuals understanding of the survey question and accuracy of the response. The second part of the interview will be interview questions as follows:

- 1) What resources do you feel would be needed to meet all the ‘absolutely necessary’ security actions?

- 2) What type of training or guidance would you like to see to help your company better understand computer security concepts?
- 3) Do you feel that there are specific limitations that effect incorporating security practices in your business?

The responses to the three initial questions could determine follow up questions as needed.

### 2.3 Case study protocol

The methodology discussed in this paper leads to a four phase case study protocol as outlined below.

- 1) Preparation Phase
  - Determine area and business types.
  - Research businesses in area.
  - Contact businesses.
- 2) Site Visit Phase
  - Introduce business to research project.
  - Introduce NISTIR 7621.
  - Administer pre-interview survey.
  - Complete interview based survey.
  - Complete Interview questions to determine limitations and constraints.
- 3) Analysis Phase
  - Tabulate quantitative data.
  - Code and categorize qualitative data.
  - Analyze patterns / develop theories.
- 4) Report Phase
  - Present results in case study report.

### 3 Contributions and discussion

One can envision numerous contributions from this research. Results from the survey should provide some insight into the validity of and usefulness of the NISTIR 7621 as a measurement tool for small business security. Insights learned could be used to modify or suggest changes to the absolutely necessary actions that are suggested. The survey results will also be valuable to develop theories based on the relationships that may exist between question items and could form the basis for theoretical models of small business security practices.

This research process will also reduce the limitations of a survey alone method in that it attempts to assess the accuracy of the responses with a repeated survey in interview format. The qualitative data will be very beneficial in understanding how the small businesses view these security practices and how they might envision their options for improving their security practices. Results from the second interview

questions will be helpful in developing methods and strategies that can be shared with small businesses to improve their technical knowledge and application of security concepts.

Future research may involve contributing to the development of practical small business security standards that may improve upon or extend beyond the absolutely necessary standards. This study can also serve as a baseline for developing security training and awareness programs geared towards businesses with limited staffing and financial resources. It is expected that future research from the findings of this case study will follow some of the generalizations suggested by Walsham [9]. These generalizations consist of development of concepts, generation of theory, drawing of specific implications, and contribution of rich insight.

### 4 References

- [1] National Institute of Standards and Technology, "NIST Interagency Reports", Retrieved March, 2013, from <http://csrc.nist.gov/publications/PubsNISTIRs.html>
- [2] R. Kissel, "Small Business Information Security: The Fundamentals", U.S. Dept. of Commerce, National Institute of Standards and Technology, 2009.
- [3] U.S. Small Business Administration, "Small Business Size Standards", Retrieved March, 2013, from <http://www.sba.gov/content/small-business-size-standards>
- [4] National Cyber Security Alliance, "2012 NCSA / Symantec National Small Business Study." Retrieved February, 2013, from <http://www.staysafeonline.org/stay-safe-online/resources/>
- [5] J-Y Park, J. Robles, C-H, Hong, T-H. Kim, & S. Yeo, "IT Security Strategies for SME's". International Journal of Software Engineering and Its Applications (IJSEIA), 2(3), 91-98, 2008.
- [6] I. Benbasat, D.K. Goldstein, &M. Mead, "The Case Research Strategy in Studies of Information Systems" MIS Quarterly, 11(3), 369-369,1987.
- [7] G.G. Gable, "Integrating case study and survey research methods: An example in information systems", European Journal of Information Systems, 3(2), 112-126, 1994.
- [8] R.K. Yin, "Case study research: design and methods (3rd ed.)", Thousand Oaks, Calif.: Sage Publications, 2003.
- [9] G. Walsham, "Interpretive case studies in IS research: nature and method", European Journal of information systems, 4 (2), 74-81,1995.



**SESSION**  
**SECURITY MANAGEMENT II**

**Chair(s)**

**Dr. Diala Abi Haidar**  
**Dar Al Hekma Univ. - Saudi Arabia**



# Security Management of Bring-Your-Own-Devices

Sara Gallotto<sup>1</sup> and Weifeng Chen<sup>2</sup>

<sup>1</sup>Southern New Hampshire University, 2500 North River Road, Manchester, NH, 03106, USA

<sup>2</sup>Department of Math & Computer Science, California University of Pennsylvania, California, PA 15419, USA

**Abstract**—*Many companies today allow their employees to Bring-Your-Own-Devices (BYOD). It is a practical policy that companies could save money by having their employees use personal devices for work. It is also convenient for employees since they could work on the devices they are familiar with. However, there are many risks associated with this practice, and security is definitely a primary concern. In this paper, we describe different security requirements and threats to the BYOD practice and propose potential solutions. This study aims to help companies to choose the most appropriate mobile devices with the consideration of convenience and security. It will also provide supports for companies to deploy BYOD policies safely.*

**Keywords:** Security Management, Bring-Your-Own-Devices, Information Security Strategy

## 1. Introduction

Mobile computing is becoming more and more common in the workplace and in schools. Everywhere you look, there is someone on a smartphone or a tablet. People love the convenience of a small device they can use almost anywhere. Use of mobile devices has grown exponentially in the last few years. According to the International Telecommunication Union [1], there are 6.8 billion mobile subscriptions worldwide, increased from 5.4 billion in 2010. Below is a few of statistics and predictions on mobile devices from mobile industry statistics [2]:

- Nearly all Generation Y consumers owned a mobile phone and 72% owned a smartphone
- Tablets hit 100 million shipments in 2012
- iPad accounts for 69% of tablets
- 81% of U.S. cell users will have smartphones by 2015
- Smartphones and tablets will increase net traffic 26 times for the next 4 years
- There are approximately 4 billion mobile phone users 1.08 billion smartphones and 3.05 billion are SMS enabled
- Android is the dominant operating system for new smartphones sold in 2013 and is forecast to remain so through to 2017
- In 2012, according to International Data Corporation [3]: 68.8 percent of smartphones sold, shipped with Google's free Android OS. That is more than three times the number shipped with Apple's iOS (18.8

percent) and dwarfs BlackBerry OS (4.5 percent) and Nokia's Symbian (3.3 percent)

- According to Strategy Analytics, China overtook the US as the largest smartphone market in Q3 2011 [4].

With the expansion across the globe of mobile devices, today many companies allow Bring-Your-Own-Devices (BYOD) for their employees. BYOD policies have so many benefits. This concept is very practical for a company in that they save money by having employees use personal devices for work. They therefore do not have to purchase tablets, phones, or laptops for every individual. It is also a convenience for employees and allows them to be more productive, which makes management happy.

So what types of mobile devices would be the choice for BYOD? The following convenience considerations could help to pick. The first is the connection to the company network. Windows is aware of this problem and in newer versions has been working to make this a better option for companies to consider. Windows 8 tablets should be able to not only become acceptable for BYOD, but it should be the first choice if a company does implement these policies [5]. With Windows tablets, users can run the same applications as the organization's Windows desktop, which makes support easier, and also improves security. This also helps those users who may not be particularly tech-savvy. They are able to see and do on the go what they see and do while at work, and therefore only need to learn it once. There is less of a learning curve. Another feature of these tablets that may help employee performance is the ability to write on the tablet by hand or with a stylus rather than typing on a virtual keyboard. Some people prefer writing to typing, and are faster and would get more work done with the handwriting recognition capabilities of a tablet.

Compared to the convenience requirements described above, security consideration is more important. That is the focus of this paper. Security vulnerabilities not only impact the business world, but everyday users as well. When deciding which tablet to purchase, it is important to think about what type of work will be done on it, and, in turn, what security features it will need to have. These concerns include:

- Authentication and customizable security features. How does one sign into the device? Can you customize authentication?
- Storage considerations;

- Types of viruses the OS is susceptible to;
- The risks of public Wi-Fi; and
- The portability of the device.

We will describe these concerns in details in the following sections. Different devices (e.g., iPad, Android, Windows tablet) will be compared. Section 2 discusses authentication and access control. Data storage and encryption is presented in Section 3. Section 4 focuses on virus and Section 5 is devoted to public Wi-Fi threats. We will consider portability in Section 6. Possible solutions to address these security concerns will be explained in Section 7. Section 8 concludes this paper.

## 2. Authentication and Access Controls

The way a user logs into and accesses a tablet is extremely important. The more forms of authentication there are, the more secure the device is. One up and coming form of authentication is biometrics. S.I.C. Biometrics has created a biometric security solution for tablets. The iFMID is a fingerprint reader that can be used on devices such as the iPad. The company believes it is “perfect” for use in the public or private sector. It offers identity authentication management on the go, as well as data access control and security. There are several key benefits for getting this device for your tablet. This reader offers important reduction of identity management costs, increased mobile personnel productivity, enhanced environment security, and safe deployment of mobile devices [6].

Although tablets are more difficult to misplace than a phone, it still happens. iPads have many access controls for this reason, including complex alphanumeric passcodes with minimum length and complexity, reuse controls, a maximum number failed attempts allowed, and auto-lock and grace periods. The downside is that if you have multiple users of the iPad, such as for a business, you cannot have different passcodes for different users. The iPad cannot be unlocked by more than one PIN or passcode, and these controls cannot be replaced by third-party controls that may be preferred by the company. One way to go around this barrier is by using a self-authenticating app as a secondary access control for business purposes [7].

Android has similar rules, and also has rules for minimum lower/upper case, digits, and symbols in a passcode that go beyond iOS. Compared to iOS, Android does offer more in terms of access. There are more choices upfront for controlling access to the tablet. It also offers facial recognition, and multi-user tablets are in the works [7].

## 3. Data Storage and Encryption

### 3.1 Data storage

Another factor when choosing a tablet for personal or enterprise use is storage. Especially if the device will be used for school or work, it is important to have plenty of

storage. The important security questions are what kind of storage that is, and where it will be located. There can be physical storage for your data that you control, or it can be cloud storage. The other consideration is where the storage is. If it is on the actual device, could the data be stored on a server on or off the main campus of the school or company? If it is cloud storage, it could be miles or even states away, and not particularly under the company’s control.

The types of storage tablet computers use is relatively expensive compared to that of a laptop. Tablets use flash memory, which is costly, and as a result these devices usually have low amounts of digital storage, and in turn, the need to obtain additional storage. Consumers are moving away from purchasing laptops and are buying these tablets even though they provide little storage. This has prompted the manufacture and sale of other storage mechanisms, such as cloud storage or external disks. Consumers then end up spending even more money. Apple, for example, does not have any way to connect the iPad to an external storage device, so users must rely on the iCloud if they want additional storage. This service comes at a higher cost with lower performance. Hard disk drives, on the other hand, are a much cheaper option. The demand for external storage is growing among users, and options besides the iCloud for iPad users are beginning to appear [8].

Because tablets have very limited storage, they require frequent backups to a desktop or laptop computer, or some other type of external device. If someone wants to use their tablet as a standalone device, they need to pick and choose carefully what items they wish to store on it, and that often requires deleting a lot of items. Music, videos, and photos are not really able to be stored on a tablet because they use so much space, so any professional who must utilize these types of files regularly is not a good candidate for tablet computing. Storage on a tablet is expensive, almost ten times that of an external hard disk. The biggest reason it is used, however, is because it is more durable. Tablets are taken and used everywhere, and are more likely to be dropped or thrown into a backpack. Flash memory is simply tougher and can be put through more wear and tear than a hard disk drive [9].

With businesses and schools incorporating tablets, the demand for more storage is increasing. Desktop computers can store more than a terabyte of data, and laptops are heading in that direction. With each generation of tablet, it is likely we will get closer to this amount of storage. In the meantime, another option is to sync the tablet with a computer. The laptop or desktop computer can be used as external storage, and only some of the files will be stored on the tablet. This is the most common way to use a tablet but still have many large files such as movies and PDFs. Although you can store a lot of data for little money, it is also important to note that the cost per GB of storage for a tablet goes down with higher storage capacity. This means

users can store more on the tablet and rely less on external storage [9].

Some companies have recently developed external hard drives designed specifically for tablets. There is a wireless storage device for iPad, because there is no way to connect anything to this type of tablet. All data is streamed over Wi-Fi, and you can connect up to five devices. Some of these devices do not need an additional power source, which makes them truly portable and ideal for tablet users on the go for work or school. Other tablets can use this device as well, but unlike the iPad, they are also able to use an SD card or hard disk [9].

The other option is cloud-based storage, such as Dropbox or Apple's iCloud. Tablet users can access these services wirelessly. There are paid and free options, which is also appealing to consumers [9].

### 3.2 Data encryption

Data stored on the devices should be encrypted in case it is stolen or otherwise compromised, especially business data. Apple has built-in encryption for the iPad. The device's unique ID and group ID are fused to the CPU and used as encryption keys. This ensures that data removed from the device cannot be read elsewhere. Apple encrypts the entire file system and scrambles email messages, and in the event of theft or loss, encryption keys can be wiped during a remote wipe. If an intruder accesses a jail-broken iPad, however, they can read without a passcode. An additional security measure is to use a third-party app to encrypt stored data [7].

On Android, encryption is optional in newer models, but was not allowed originally. If you use encryption, however, you cannot use swipe, pattern, or face unlocks. Whichever method you use, it is important to remember that on both Apple and Android tablets, encryption won't protect data synced with the user's computer [7].

## 4. Viruses

Once you decide how you are going to tackle data storage, the next consideration for your tablet should be viruses. It is important to know whether viruses are common on the operating system you choose, and also how that operating system is going to handle it if you do stumble upon one.

### 4.1 iOS

Apple, for example, has various security measures that make it very difficult for malware to be developed. This includes a secure boot chain, which verifies the bootloader, OS kernel, and firmware authenticity. Also, it uses system software personalization, which prevents OS downgrade, and application code signing, which prevents execution of code not reviewed and signed by Apple. Last, it has an encrypted file system that maintains strict separation between user and OS partitions, and sandboxing, which stops programs from

accessing each other's code and data. Apple reviews all code submitted to the iTunes AppStore, and users cannot install apps without a certificate from Apple, without "jail-breaking", or, overwriting, their iPad's security system. Apps cannot run as background processes or access privileged resources like contacts or calendars, and the user must grant permission for the app to access emails or text messages. This reduces private data exposure [7].

### 4.2 Android

Android, on the other hand, has the opposite approach. They have created an open OS to run on all different brands of devices. Each manufacturer decides how to protect the device. This results in vulnerabilities that vary between models. They rely on the open source community to report malware. Android (Google) does not regularly review content submitted to the Play Store. Their app certificates can be signed by anyone, even the person who made the app. Users can also install apps from unofficial sources, and this is responsible for spreading most Android malware. Android allows applications to search for and invoke others and run as true background processes. Although this can be useful, it is a feature often abused by trojans. Lastly, although there is a lengthy list of permissions the user must accept, most people just click "ok" and download the app. This can cause problems. People end up downloading and installing applications that they should not trust from a developer that they do not know [7].

### 4.3 Becher's study

In 2011, Becher wrote about researchers expecting a major security event to occur due to the rising number of smartphone users. Because smartphones exhibit increased processing power and memory, increased data transmission capabilities of the mobile phone networks, and open and third-party extensible operating systems, they become an interesting target for attackers. These factors also apply to tablet computers. Small viruses had already occurred, but fortunately no large-scale attacks have happened. Part of that is due to the fact that all of the operating systems seem to be different, so it makes it more difficult to coordinate a widespread attack. Another reason is that the developers of these smartphones and tablets know people are worried about risks to their privacy and security, so they are already made more secure than when they first arrived in the marketplace [10].

There are several types of mobile threats discussed in Becher's research, and several can apply to tablet computers. The most important is a software-centric attack. A common action here is to collect any private data that the hacker can access from your tablet, and forward it to the malware author or users without your knowledge. It may be inconspicuous and innocent seeming in the form of a game or other popular application. It is very difficult to detect an attack on a

device with a more limited processing power than a typical desktop computer. This is a direct violation of privacy and confidentiality [10].

## 5. Public Wi-Fi Threats

Viruses are not the only threat to watch out for. It is important to think about whether people will be using public Wi-Fi, and the dangers that come along with that. Although public wireless internet is convenient, it can and does pose several security threats. A user's personal computer can contain passwords, bank account information, data for their business, and credit card or social security numbers. Files can be "deleted", but most users will never completely get rid of them, and they end up still hidden within the computer. For these reasons, people using public Wi-Fi, whether on their personal or business computers, are a common target for hackers [11].

Identity theft is a very concerning but very real possibility on public Wi-Fi, and the problem that goes with it is that it can be difficult to detect until it's too late. An example used by Sandeep Kale in "The Dangers of Public Wi-Fi Network: Wireless Internet Security" is that although being on a general fundraising website won't give away any information about you, once you click "donate", you'll have to use your bank information or PayPal account, and that's when identity theft can occur [11].

Another threat of public Wi-Fi is that it can actually protect the anonymity of hackers. Anyone can get on the system, because usually there is not much in the way of password protection in a public place such as a mall, airport, or coffee shop. There are also so many users that it's almost impossible to determine who the hacker may be. Another threat is fake public Wi-Fi. This is a scheme where a hacker sets up a fake public Wi-Fi connection, with hopes that people will log on, at which point their every move is tracked. If you're at a coffee shop, for example, and the Wi-Fi name is CoffeeHouse1, a hacker may create a fake connection named CoffeeHouse2 with the hope that someone will try to connect to it. Most people will not be able to tell the difference [11].

Apple supports all common Wi-Fi security methods, but tries to hide most security settings from the user. The user cannot remove or even see Wi-Fi network names to which they have previously connected but are no longer near, so the iPad keeps trying to reconnect, and this can cause it to unintentionally connect to something it shouldn't. Android exposes detailed security parameters to the user, as well as previously-used Wi-Fi connections. This puts the burden on the user to ensure that the list is up to date and any old connections are cleared, however, it can prevent unintended connections and the consequences that go along with them. Another important piece to note is that Apple does not allow or approve third-party VPN clients, but Android does [7].

Because of these risks, it is important to know how to stay safe while using public Wi-Fi. The first step is the make sure that you have the latest security upgrades and software on your computer. Next, only use a public Wi-Fi network if you are sure it is from a real source. To be sure, ask an employee to verify the name and password when connecting. This can prevent connecting to a fake Wi-Fi connection [11].

## 6. Portability

Finally, it's important to think about device access control, and what would happen if the tablet was misplaced. Although tablets are more difficult to misplace than a phone, it still happens. If your tablet is stolen or lost, there may be options to remotely lock or wipe the data. Apple has Mobile Device Management with the "Find My iPhone" App, which can play a sound or display a message on the iPad, display its location to you, or remotely wipe the device. This is also helpful in the office setting, because employers can wipe or reset the iPad of a former employee or reset a jail-broken iPad. Sometimes wiping a former employee's tablet may be taking it too far. There is another option to wipe only company related data. This is called an "enterprise wipe". Only MDM-installed settings and applications will be removed, and any photos, contacts, and iTunes apps will remain untouched [7].

With Android, the user can remotely lock, ask for a PIN, or reset to factory defaults, and can request location of the tablet. Many applications available for Android can request the tablet's location, as long as the user has the location services enabled and the tablet is connected to the internet. There is an "enterprise wipe" command, like Apple has. This is used to remove installed applications, either public or private, email and VPN accounts, any Wi-Fi connections, and certificates. The MDM for Android can also change a tablet's PIN if someone tries to remove device administration from a stolen tablet. Unfortunately, emails and attachments stored on an SD card are not affected by resetting the device. This is not a problem, however, for newer tablets that lack SD storage [7].

## 7. Solutions

### 7.1 BYOD or not

Emerging risks of BYOD include data leakage, loss of control and visibility, and ease of device loss. This concept presents a "conflict between Security and Usability". The questions now is, what can businesses do? The first proposed solution is to prohibit BYOD altogether, but this should be done only if the costs of security breach outweigh productivity gains. Another solution is to adopt a Mobile Device Management (MDM) solution, which is technology that focuses on device and policy management, and has added security measures like authentication and encryption. It is integrated with current IT infrastructure and rules are

added to current policies to manage mobile devices with these programs [12].

Companies prohibiting BYOD can prevent exposure to most of the risks discussed here. Some organizations do retain tight control over what devices employees are allowed to use for work, even in this age of growing BYOD policies and preferences from employees. These companies, however, usually have a situation where the cost of a potential security breach greatly outweighs the possible productivity gains. Although there is almost a guarantee that the security threats would be negligible, changing employee expectations and losing out on the high potential business benefits of BYOD make banning it impractical. Therefore, BYOD policies must be a part of a company's information security (IS) strategy [12].

## 7.2 IS strategies

There is an absolute need in every company to have IS strategy set in place. Technology is changing rapidly, and it is important to stay current and organized. Without planning, security enhancements will begin to lag behind the level of threats, and the gap between necessary IS levels and actual IS levels will widen. Some common established IS strategies include prevention, detection, response, compartmentalization, isolation, deterrence, surveillance, and layering [12].

### 7.2.1 Prevention

Prevention is the most common security strategy implemented by companies. The purpose is to prevent attacks on information assets. The problem with prevention is if a company focuses too much on protecting the information, and end up sacrificing its usability. There are also limitations on space and time. Unfortunately, most companies do not pay attention or plan out prevention strategies until after a significant incident has already occurred [13]. This can be considered to be due to lack of organizational resources and the costs associated with defense [12].

### 7.2.2 Detection

Detection can be considered a preventative measure. Its purpose is to identify unusual behavior or activity. Intrusion detection systems (IDS) are the most common ways companies can identify malicious behavior, misuse, intrusion, and even specific attacks [14]. This system allows a company to continuously scan computers and look for anomalies. The organization still needs to figure out exactly what to do with this information, but it can be extremely helpful in security response. Response strategies make up their own categories in IS strategies. These strategies focus on reaction and recovery. Reaction is the defensive and offensive actions, and recovery refers to restoring normalcy. Response strategy is used to minimize the impact of an incident and prevent any reoccurrence [12].

### 7.2.3 Compartmentalization

Compartmentalization attempts to contain the damage from an attack. This strategy is particularly important when forming policies for mobile devices like tablets. Some companies only allow unauthorized devices to connect to isolated network zones that have been separated from the internal network. This helps control who is on the company's network. There are issues with compartmentalization, as there are with any policy. Organizations may be in a position where they need to sacrifice information assets under attack in one zone in order to save the entire network. Also, security factors must be predefined. Before developing the zones, sensitivity of data, user privilege, and proxy considerations must be determined. Lastly, it is important to consider the cost of creating isolated zones and how long they will last. Creating these zones can also get complex because system and network interdependencies must be known upfront before implementing this creation [12], [15].

### 7.2.4 Isolation

Isolation can be used to trap an attacker. It is also referred to as deception. The goal is to lure and trap the intruder and monitor and analyze their actions. The information gained can be used as part of the company's IDS to improve detection of malicious acts in the future. Isolation allows companies to figure out how to improve their other policies, like prevention, detection, and response [12], [16].

### 7.2.5 Deterrence

The purpose of deterrence is to change the behavior of people in the network to deter them from engaging in any insecure activities. The key to this is a culture that promotes security [17]. There must be policies and procedures in place for non-compliance, and also education should be available for security awareness, computer abuse, moral standards, and self-control. Most companies do not employ use this strategy. The non-technical aspects like education may be difficult for an IS person to comprehend. It also requires a lot of time. The other reason is that this strategy is dependent on employees' levels of self-control and moral beliefs. It is more susceptible to failure because of this. Generally, deterrence programs are not very successful. Employees may simply find another way around the security controls [12].

### 7.2.6 Surveillance

Surveillance is a challenging piece of IS strategy. This aim is to monitor the security environment while maintaining high situational awareness. This allows the company to quickly adapt to risks and threats. It is challenging because both technical and non-technical aspects of the company must be monitored together. The security environment is very complex, and it is difficult to know where it is appropriate to monitor. Also, handling data from monitoring logs

can be difficult due to the time it takes to manage backup, storage, analysis, and deletion of such a large volume of data. Time is money, and this is a very resource-intensive activity [12].

### 7.2.7 Layering

The last piece of IS strategy is layering, or, defense-in-depth. If there is vulnerability in a company's security, a hacker can take advantage of that and get past it with limited resources. Layering, however, requires more resources and knowledge to get past [18]. The perceived cost is so great that it may deter potential offenders. Layers of simple security is a cost effective way for a company to protect itself [12].

## 7.3 Mobile Device Management (MDM)

Aside from the detailed policies of information security, MDM solutions must also be a part of BYOD policies. Modern MDM technology is a great tool to help companies manage employees' devices at work and on the go. There are many security functionalities to allow companies to maintain control over BYOD devices [12].

Risks such as data leakage, loss of organizational control and visibility, and ease of device loss can be reduced with MDM systems. These systems prevent unauthorized or unwanted access to the device. One must be cautious with the amount of security on the device, however, because if it is too much for the user, it's likely the device will hardly be used. The company can also use MDM technology to track the device. This allows them to detect behaviors that go against security policy, such as PIN disabling, for example [12].

In addition to being a detection strategy, MDM systems also provide a response strategy. Companies can use the system to respond to a detected threat by disabling email or VPN access, or even perform a wipe of the device. This is extremely effective in a situation where a tablet gets lost or stolen. The problem that arises is if the tablet is a private device owned by the employee, which is the case in BYOD settings. If this happens, rather than wiping a personal tablet, the company can compartmentalize corporate and personal data upfront with multiple profiles, for example, and decommission the device without harming anyone's personal files [12].

There are so many benefits that make having an MDM system a must in a company that allows BYOD. The next consideration is what type of MDM to use. There are different products out there that companies can use for this purpose, and they all have two primary focuses. The first is device and policy management, and the second is value-add security measures, such as authentication and data encryption as previously discussed [12].

## 7.4 Other requirements

There are several concerns for BYOD compliance, security, and access. The first is IT support and resource availability, without these BYOD is not possible. There is also the concern of setting employees up with the device, and installing the necessary applications for them to be able to use it for work. Choosing an MDM, therefore, should not be based on just security of the system. It must be supported by policy and process. Lastly, it must be able to be integrated with the existing infrastructure and support workflows [12].

## 8. Conclusion

In this paper, we have discussed the security concerns on Bring-Your-Own-Devices (BYOD). Different security requirements and threats were presented. We have also discussed possible solutions to address these problems. The conclusion is that, unfortunately, there is no clear-cut solution to fully protect an individual or company from the risks associated with using a tablet in public for work or even personal use. Companies do, however, need to be aware of the risks of the BYOD trend and create policies addressing them. It is important to incorporate all of the IS strategies and use MDM systems or similar solutions to help prevent these risks [12].

Companies considering implementing a Bring-Your-Own-Devices program need to consider various information security strategies, Mobile Device Management systems, and security risks. There needs to be a policy in place to determine which tablet's operating system is the best fit for what the company needs that employee to use the tablet for. They need to consider authentication and access, customization, encryption, data storage, viruses, public Wi-Fi risks, and portability.

This study aims to help companies to choose the most appropriate mobile devices with the consideration of convenience and security. It will also provide supports for company to deploy BYOD policies safely.

## References

- [1] Brahima Sanou, ICT facts and figures. Printed in Switzerland Geneva, February 2013. Available at <http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013-e.pdf>
- [2] Mobile Industry Statistics. Mobile Commerce and Engagement Stats <http://digby.com/mobile-statistics/>, Retrieved on March 14, 2014.
- [3] IDC. Android and iOS Combine for 91.1% of the Worldwide Smartphone OS Market in 4Q12 and 87.6% for the Year <http://www.idc.com/getdoc.jsp?containerId=prUS23946013\#.UTCOPjd4D1Y>, February 14, 2013.
- [4] L. Sui. China Overtakes United States as World's Largest Smartphone Market in Q3 2011. Strategy Analytics Metrics report. Published on November 21, 2011.
- [5] D. Shinder. Windows 8 Tablets: Secure enough for the Enterprise? [http://www.windowsecurity.com/articles-tutorials/Mobile\\_Device\\_Security/Windows-8-Tablets-Secure-enough-Enterprise.html](http://www.windowsecurity.com/articles-tutorials/Mobile_Device_Security/Windows-8-Tablets-Secure-enough-Enterprise.html), September 26, 2012.

- [6] S.I.C Biometrics. Secure Access Control to Critical Data, Systems and Applications on Tablets  
<http://www.sic.ca/tablets/>, Retrived on March 14, 2014.
- [7] L. Phifer. Battle Royale: iPad and Android Tablet Security Compared. *Tom's IT Pro*. Published on November 15, 2012.
- [8] T. Coughlin. Is 64 GB Big Enough – Why Tablet Computers May Need More Storage?  
<http://www.forbes.com/sites/tomcoughlin/2011/06/06/is-64-gb-big-enough-why-tablet-computers-may-need-more-storage/>, June 6, 2011.
- [9] L. Haber. Tablet Storage Limited, But Options Abound. *Tablet-PCReview*. Published on August 29, 2011.
- [10] M. Becher, F. C. Freiling, J. Hoffmann, T. Holz, S. Uellenbeck, and C. Wolf. Mobile Security Catching Up? Revealing the Nuts and Bolts of the Security of Mobile Devices. In *IEEE Symposium on Security and Privacy 2011*, pages 96-111, Oakland, CA, 2011.
- [11] S. Kale. The Dangers Of Public Wi-Fi network: Wireless Internet Security.  
<http://www.trickswindow.com/networking/public-wi-fi-network-dangers/>, July 24, 2012.
- [12] A. Dedeche, F. Liu, M. Le, and S. Lajami. Emergent BYOD Security Challenges and Mitigation Strategy. Technique report. The University of Melbourne. 2013.
- [13] F. Sveen, J. Torres, and J. Sarriegi. Blind information security strategy. *International Journal of Critical Infrastructure Protection* (2:3), pp 95-109, 1999.
- [14] H. Cavusoglu, B. Mishra, and S. Raghunathan. The value of intrusion detection systems in information technology security architecture. *Information Systems Research* (16:1), pp 28-46, 2005.
- [15] A. Ahmad, S. B. Maynard, and S. Park. Information security strategies: Towards an organizational multi-strategy perspective *Journal of Intelligent Manufacturing* 25:357-370, 2014.
- [16] W. Tirenin, and D. Faatz. A concept for strategic cyber defense Proceedings of *Military Communications Conference (MILCOM)* 1999, pp. 458-463.
- [17] Q. Hu, Z. Xu, T. Dinev, and H. Ling. Does deterrence work in reducing information security policy abuse by employees? *Communications of the ACM* (54:6), pp 54-60.
- [18] R. Anderson. Why information security is hard-an economic perspective. Proceedings of Computer Security Applications Conference (ACSAC) 2001, pp. 358-365.

# SIPPA Approach Towards a Privacy Preserving Voice-based Identity Solution

Bon K. Sy

Computer Science Department, Queens College and University Graduate Center/CUNY, Flushing, NY 11367, U.S.A.

**Abstract** - *The focus of this project is the development of a privacy preserving identity solution. A traditional identity solution associates an individual with a user identity and a user credential for the purpose of authentication, authorization and accounting. One of the underlying assumptions of a traditional identity solution is the ability to secure and to prevent tampering with the association between an individual and the corresponding user information. If this assumption does not hold, one can no longer guarantee the integrity of the system for facilitating authentication, authorization and accounting. The contribution of this project is a novel approach that removes the system reliance on the assumption. Specifically, our approach employs SIPPA to achieve credential regeneration on the fly that eliminates the need for storing such information; thereby avoiding the risk inherent in the assumption.*

**Keywords:** Voice-based key generation; Privacy aware authentication.

## 1 Introduction

The objective of this project is to develop a privacy preserving identity solution based on SIPPA — Secure Information Processing with Privacy Assurance. SIPPA [1,2] is a two-party secure computation method for comparing the private data of two parties without each party disclosing their private data to each other.

In our SIPPA based solution, personal private information or credential information for authentication will not be stored in plain. Private sensitive information will be derived on demand. This eliminates the risk on information leak since no private sensitive information is stored in the first place. Therefore, information privacy is assured. Furthermore, SIPPA protocol execution produces two artifacts; the degree of similarity resulted from the comparison of the private data, which can be used for authentication purpose, and the helper data useful for the information processing needed to regenerate the credentials for authentication/authorization. Since the SIPPA protocol has been analyzed under different security models and situations, the behavior and the security of the identity solution can be derived from that of the SIPPA protocol, and formally analyzed and assessed accordingly.

In this project, a particular embodiment of the proposed identity solution utilizing biometric voice signature and mobile device will be described — although the embodiment could be based on any modalities and devices. The rest of the paper will be organized as the followings. In section 2 we will give a summary on the system architecture of the identity solution, the formulation on the system elements and the information used for authentication. In section 3 an overview on the state-of-the art, and the context under which this project is related to the state-of-the-art, will be given. In section 4 the theory of SIPPA, and the application of SIPPA to realize the SIPPA-based identity solution under real world security model will be presented. In section 5 the system implementation and the experimental result will be detailed, which is then followed by the conclusion that briefly describes our future work.

## 2 Formulation and System Architecture

One of the unique characteristics of SIPPA is to allow one party to reconstruct the private data of the other party when their data are "sufficiently" similar. In the SIPPA reconstruction phase, the *server* party provides helper data for the *client* party to reconstruct server data that preserve perfect accuracy, or an accuracy proportional to the similarity of the private data of both parties. This consequently allows us to realize an identity management workflow not present in a traditional solution, which can be described as below:

- Sensitive credential information for authentication/authorization is encoded by the personal private information of an individual.
- Only the encoded information is stored. Sensitive credential information and personal private information are never stored. But the credential information can be reconstructed during the execution of the SIPPA protocol — when the personal private information presented by an individual is sufficiently similar to that used for encoding the sensitive credential information.

In this project, the identity of an individual is characterized by three facets [3]: (i) *what one knows*, referred to as a UID (Universal ID) — a unique ID generated by the system based on personal information PID such as phone number or birth date, (ii) *what one has*, referred to as a DID (Device ID) such as a personal mobile phone or a device serial number, and (iii) *what one is*, referred to as BID (biometric

ID) such as the biometric voice, face or fingerprint. More specifically, the identity of an individual is a 3-tuple composed of DID, a biometrically encoded encryption key — BID+K, and the decryption of the encrypted hash on UID; where K is a secret key. Formally, an identity is then represented by a 3-tuple:  $\langle \text{DID}, \text{BID}+\text{K}, \text{Dec}(\text{K}, \text{Enc}(\text{K}, \text{Hash}(\text{UID}))\rangle$ .

The architecture of our system consists of 3 components; namely, a voice gateway (VG), an Enrollment Module (EM) comprised of SIPPA server and a local database, and an Identity Storage and Verification Module (ISVM) comprised of SIPPA client and a centralized database.

In our design, the local database of the enrollment module stores the encryption/decryption secret K. The centralized database stores the identity information. For privacy assurance, the EM and the ISVM do not directly share with each other the information in their databases. Furthermore, the two modules do not even have to treat each other as a trust worthy party. This is different from the traditional identity solution where the trustworthiness [4] between the system components similar to that of EM and ISVM is assumed.

### 3 Literature Review

Privacy preserving authentication is an active research topic in many different domains [5-8]. In general, the goal is to minimize disclosure on the identity information of an individual, certain content information about an identity such as phone number or birth date, the linkability of the identity information and its usage, the issuer of the identity [9], and the data matching [10].

The research in this area can broadly be classified into cryptographic based and non-cryptographic based approach. In cryptographic based approach, Public Key Infrastructure (PKI) [16] to issue X.509 certificate with private/public key pair for encryption and message signing [11], one-way hash [12], zero-knowledge proof [13], and commitment scheme are the basic building blocks for developing a privacy preserving identity management solution. Attribute Based Credential for Trust (ABC4Trust) [14] is an exemplary state-of-the-art that allows an individual to use not one public key, but possibly multiple public keys. In addition, certificate is based on the individual's secret key, attributes that may be hidden from the Certificate Authority [15], and proof of knowledge of certificate about identical secret key used in different certificates of the individual. This is different from the conventional Public Key Infrastructure in that a certificate is based on an individual's public key, and the certificate (thus the information in the certificate) is revealed. Although ABC4Trust is an improvement over the traditional approach, the implicit deployment assumption of ABC4Trust is that a secure and trustworthy issuer (typically a Certificate Authority) exists and is always available.

An interesting aspect of non-cryptographic based approach is the idea of Privacy Preserving Data Matching (PPDM) as exemplified by Scannapieco et al [10]. The key idea behind PPDM is the use of an embedding space SparseMap [20] that preserves the similarity distance between two data objects in the metric space. The embedding space is constructed by using a subset of data objects serving as a reference set, and the distance between two data objects is mapped to two distance measures in the metric space; i.e., between a data object and the reference set, and the other data object and the reference set. Through triangular inequality, a lower bound distance measure between the two data objects can be obtained; thus realizing the privacy preserving approximate matching.

Our proposed SIPPA approach towards a privacy preserving voice-based identity solution shares similar characteristics to the research just mentioned. Yet it distinguishes itself with characteristics that are unique and attractive for privacy preserving authentication. In both our approach and ABC4Trust, Public Key Infrastructure (PKI) is required. The main difference lies on the extent that the PKI is used. In ABC4Trust, a key characteristic is to issue every user multiple keys so that privacy protection can be achieved. In our proposed approach, Certificate Authority is only required for the key infrastructure; i.e., the Voice Gateway (VG), Enrollment Module (EM), and Identity Storage and Verification Module (ISVM). Especially in our specific applications of SIPPA approach, it is not clear how a trustworthy environment can be established in order for every user to securely receive the private and public keys needed as in ABC4Trust. With respect to Hash Lock [17], the main difference is the choice of cryptographic primitives. In our SIPPA approach, we require cryptographic primitive to be not only semantically secure, but also to belong to the class of homomorphic encryption [21] for computation over the encrypted domain; e.g., Paillier encryption [22] with homomorphic additive property. By definition, cryptographic primitive that has semantic security property such as Paillier encryption does not encrypt a message to the same cipher text; thus deterring Chosen-Plain text Attack (CPA) [23]. The enforcement of semantic security in Hash Lock, however, will prevent the protocol of Hash Lock to work properly.

In reference to PPDM, our approach also tackles the problem of privacy preserving data comparison through an alternative metric space. However, our approach is completely different from that of PPDM. While PPDM relies on SparseMap for the construction of the embedding space, SIPPA maps the data objects to their Eigen space through the symmetric matrices derived from the data objects. More importantly, PPDM aims at privacy preserving approximate matching. SIPPA, on the other hand, aims at utilizing the mathematical properties implicit in the Eigen space mapping that allows precise reconstruction of the private data based on sufficiently similar data objects.

## 4 Theory, Practice & Security Analysis

An innovation of this project is to develop an identity solution that incorporates privacy assurance with the following properties:

- The identity of an individual is multi-facet and is based on *what one knows* (UID), *what one has* such as a mobile phone, and *what one is* such as biometric voice signature.
- A system that is fail-safe; i.e., it preserves the privacy of personal information — even if the system is compromised.

Our approach towards the development of a fail-safe system is to employ cryptographic key to protect the confidentiality of the UID/DID. The cryptographic key is generated, used and discarded. It is never stored. Only the biometrically encoded encryption key  $K+BID$  is stored; where  $BID$  is a biometric ID as discussed in section 2. The key is regenerated based on the biometrics of an individual whenever it is needed. Given a biometric sample  $S$ , the pre-processing step of the regeneration is a simple cancellation operation; i.e.,  $(K + BID) - S$ .

Note that the cryptographic key  $K$  can be perfectly regenerated in the pre-processing step if  $BID = S$ . However, personal biometrics can seldom be reproduced identically. Therefore, in general  $BID$  and  $S$  are different. When  $BID$  and  $S$  are from the same individual, the error incurred by  $BID-S$  is small. Otherwise  $BID-S$  is relatively large.

### 4.1 SIPPA Theory

SIPPA [1,2] is a 2-party secure computation protocol [24] where a client party can reconstruct source data of a server party under the following conditions:

1. The client party must possess some client data that is a “sufficiently good approximation” of the source data, in order to initiate the SIPPA process.
2. Rather than revealing the source data of the server party to the client party, only some helper data related to the Eigen components of the source data is provided (by the server party) to the client party for reconstructing the source data.

In our case, the SIPPA client retrieves  $K+BID$  from the centralized database, and performs the cancellation  $K+BID-S$ .  $K$  is stored in the local database of the SIPPA server. Through the execution of the SIPPA protocol, the SIPPA client will be able to reconstruct  $K$  if  $(K+BID-S)$  and  $K$  are sufficiently similar. The formulation, the key results of SIPPA summarized as two theorems, and the algorithmic steps are already reported elsewhere [1,2]. Nonetheless, they are introduced to make this paper self-sufficient.

Let  $P1$  and  $P2$  be the SIPPA server and client respectively. Let  $\mathbf{de}$  and  $\mathbf{dv}$  be the column vector representing private data

of  $P1$  and  $P2$  respectively. Let  $(\lambda_{de} \mathbf{v}_{de})$  and  $(\lambda_{dv} \mathbf{v}_{dv})$  be the 2-tuples of the most significant Eigen value and the corresponding unity normalized Eigen vector of the matrices  $\mathbf{de} \cdot \mathbf{de}^T$  and  $\mathbf{dv} \cdot \mathbf{dv}^T$  respectively.

**Theorem 1:** Consider  $(\mathbf{de} \cdot \mathbf{de}^T + \mathbf{dv} \cdot \mathbf{dv}^T)\mathbf{x} = \lambda_{de} \mathbf{v}_{de} + \lambda_{dv} \mathbf{v}_{dv}$ , the solution  $\mathbf{x} = \mathbf{v}$  satisfying  $(\mathbf{de} \cdot \mathbf{de}^T + \mathbf{dv} \cdot \mathbf{dv}^T)\mathbf{v} = \lambda_{de} \mathbf{v}_{de} + \lambda_{dv} \mathbf{v}_{dv}$  has a unity scalar projection onto the unity normalized  $\mathbf{v}_{de}$  and  $\mathbf{v}_{dv}$ , and is a bisector for the interior angle between  $\mathbf{v}_{de}$  and  $\mathbf{v}_{dv}$ .

**Theorem 2:** Consider  $(\mathbf{de} \cdot \mathbf{de}^T + \mathbf{dv} \cdot \mathbf{dv}^T)\mathbf{x} = \lambda_{de} \mathbf{v}_{de} + \lambda_{dv} \mathbf{v}_{dv}$ ,  $\mathbf{de}$  can be efficiently reconstructed — with an accuracy proportional to the closeness between  $\mathbf{v}_{de}$  and  $\mathbf{v}_{dv}$  — by a party with  $\mathbf{dv}$ ,  $\lambda_{dv}$ , and  $\mathbf{v}_{dv}$  when (i) the interior angle between  $\mathbf{v}_{de}$  and  $\mathbf{v}_{dv}$  is less than 90 degree and (ii) the party is given  $\mathbf{x}$  and  $\lambda_{de}/\mathbf{de}^T \cdot \mathbf{x}$ . Specifically,  $\mathbf{de} = (\mathbf{est\_v}_{de}/|\mathbf{est\_v}_{de}|)(\lambda_{de}/\mathbf{de}^T \cdot \mathbf{x})$ ; where

$$\mathbf{est\_v}_{de} = \mathbf{v}_{dv} + [|\mathbf{v}_{dv}| \cdot \tan(2\cos^{-1}(\mathbf{v}_{dv} \cdot \mathbf{x}/(|\mathbf{v}_{dv}| \cdot |\mathbf{x}|)))] \cdot [(\mathbf{x} - \mathbf{v}_{dv})/|\mathbf{x} - \mathbf{v}_{dv}|]$$

Readers interested in the proof of the two theorems above are referred to our other publication elsewhere [1].

### SIPPA Protocol:

**Step 1:** Derive, by the respective party, the most significant eigenvalue and its corresponding unity-normalized eigenvector of  $\mathbf{de} \cdot \mathbf{de}^T$  and  $\mathbf{dv} \cdot \mathbf{dv}^T$ . This step yields  $(\lambda_{de} \mathbf{v}_{de})$  for SIPPA server and  $(\lambda_{dv} \mathbf{v}_{dv})$  for SIPPA client.

**Step 2:** Compute  $\mathbf{x}$  such that  $(\mathbf{de} \cdot \mathbf{de}^T + \mathbf{dv} \cdot \mathbf{dv}^T)\mathbf{x} = \lambda_{de} \mathbf{v}_{de} + \lambda_{dv} \mathbf{v}_{dv}$  utilizing SLSSP. The vector  $\mathbf{x}$  is known to both parties following SLSSP. The details on SLSSP are reported elsewhere [1].

**Step 3:** The party that wishes to determine the deviation between its eigenvector and the other party's eigenvector can do so utilizing  $\mathbf{x}$  (derived in step 2). Suppose that the party with  $\mathbf{v}_{de}$  wishes to determine the angular deviation between  $\mathbf{v}_{de}$  and  $\mathbf{v}_{dv}$ , this can be done by obtaining the angle between  $\mathbf{v}_{de}$  and  $\mathbf{x}$ . i.e.  $\cos^{-1}(\mathbf{v}_{de} \cdot \mathbf{x}/(|\mathbf{v}_{de}| \cdot |\mathbf{x}|))$ . The angular deviation between  $\mathbf{v}_{de}$  and  $\mathbf{v}_{dv}$  is then  $2\cos^{-1}(\mathbf{v}_{de} \cdot \mathbf{x}/(|\mathbf{v}_{de}| \cdot |\mathbf{x}|))$  — due to theorem 1.

**Step 4:** If  $\mathbf{de}$  and  $\mathbf{dv}$  are sufficiently similar as determined by either the angular distance or the Euclidean distance between vectors  $\mathbf{v}_{de}$  and  $\mathbf{v}_{dv}$  as measured by some pre-defined threshold, proceed to send the helper data:  $(\lambda_{de})^{0.5}$  for a perfect reconstruction.

**Step 5:** Derive estimated  $\mathbf{v}_{de}$  -  $\mathbf{est\_v}_{de}$  as stated in theorem 2, and then derive  $\mathbf{de} = (\mathbf{est\_v}_{de}/|\mathbf{est\_v}_{de}|)(\lambda_{de})^{0.5}$  because (1)  $\lambda_{de} = \mathbf{de}^T \cdot \mathbf{de} = |\mathbf{de}|^2$  (from Theorem 1), (2)  $\mathbf{de}/|\mathbf{de}| = \mathbf{v}_{de}$  or  $\mathbf{de} = |\mathbf{de}| \cdot \mathbf{v}_{de}$ , (from Theorem 1), and (3)  $\mathbf{est\_v}_{de}/|\mathbf{est\_v}_{de}| = \mathbf{v}_{de}$  (from Theorem 2).

### 4.2 SIPPA-based Identity Management

In our application of SIPPA, the server data  $\mathbf{de}$  is a vector of  $20 \times 1$  of real numbers in the range  $[0,1]$ . The secret  $K$  stored in the local database of SIPPA server is a  $20 \times 1$  vector of

normalized integer values that are a fixed point representation of the real numbers. During an encryption/decryption, an AES key is generated from the MD5 hash of K. The client data  $\mathbf{dv}$  is also a vector of  $20 \times 1$  of real numbers derived from  $(K+BID)-S$ ; where BID and S each is a normalized  $20 \times 1$  vector representing a biometric voice template of cepstrum coefficient [29] in the frequency range of 0-4KHZ based on Mel scale using triangular filters.

**Protocol for identity enrollment:**

1. An individual established connection through a secure authenticated channel [25] to download client-side software such as applet capable of biometric voice signature extraction and cryptographic key generation.

**Note:** In a secure authenticated channel, messages can be eavesdropped, relayed and replayed, but not altered.

2. The individual submits – through the downloaded client-side software – to the voice gateway his phone number that can be recognized as a caller ID for a call back.

3. If the caller ID is valid and unique, the voice gateway signs the caller ID and calls the individual back. It returns the signed version of the caller ID as the device ID – DID, as well as a token T (e.g. a random number or a timestamp). In addition, the voice gateway also sends T to ISVM.

**Note:** The call back process, and the generation of T and sharing with ISVM complete the commitment scheme.

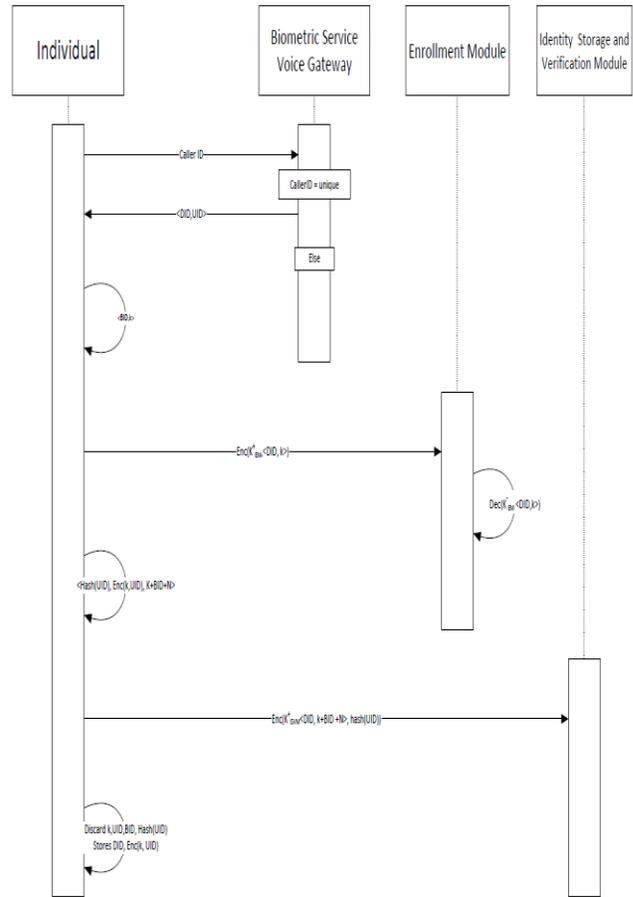
4. The individual records his/her voice sample and uses the downloaded client-side software to extract the individual's voice signature as his biometric ID – BID. The client-side software also generates a cryptographic secret key K.

5. The cryptographic secret key K and DID —  $\langle DID, K \rangle$  — are encrypted (using the public key of EM  $Enc(K_{EM}^+, \langle DID, K \rangle)$ ) and sent to the Enrollment Module (EM) through a secure authenticated channel; and decrypted by EM; i.e.,  $Dec(K_{EM}^-, Enc(K_{EM}^+, \langle DID, K \rangle))$ ; and then stored upon receiving.

6. Three pieces of information is derived by the individual using the client-side software: Generate a UID using some personally known information and the token T, and then hash UID —  $Hash(UID)$ ; Encrypts the hash using K —  $Enc(K, Hash(UID))$ ; Computes  $K+BID+N$  where N is some noise generated by the individual.

7. Three-tuple  $\langle DID, K+BID+N, Hash(UID) \rangle$  is encrypted and sent to the Identity Storage and Verification Module (ISVM) through a secure authenticated channel; and decrypted by ISVM upon receiving.

8. The downloaded client-side software is terminated and discarded. K, UID, BID, and hash(UID) are also discarded. The individual retains only DID, T, and  $Enc(K, Hash(UID))$  (or  $Enc(K, UID)$  if UID is not deem private).



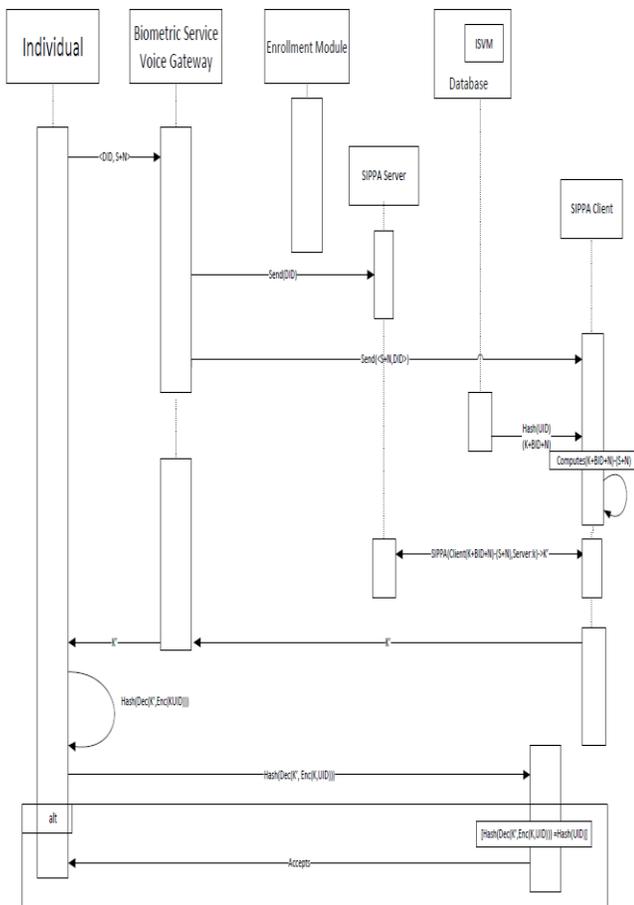
It is noteworthy that the enrollment process described above does not rely on a Certificate Authority to verify the identity of an individual. Instead, the enrollment process above allows an individual to create and self-sign an identity, whereas the process to bind an individual to a unique identity is based on what an individual has (e.g., mobile phone). It does not care the individual information that an individual may specify. It is because the individual information is not relevant to the identity verification process. As such, two individuals could have, for example, the same name but different DID and BID. They will be identified as two different entities as distinguished by different 3-tuples.

**Protocol for identity verification:**

1. An individual presents to voice gateway (VG) his DID and a noise-added biometric sample  $S+N$ .
2. Voice gateway relays DID to SIPPA server, and voice gateway relays  $S+N$  and DID to SIPPA client.
3. Based on DID, SIPPA client retrieves  $Hash(UID)$  from the centralized database. SIPPA client also retrieves  $(K+BID+N)$ , and computes  $(K+BID+N)-(S+N)$ .
4. Execute SIPPA protocol for the SIPPA client to construct a secret  $K'$ ; i.e.,  $SIPPA(\text{client-input: } (K+BID+N)-(S+N), \text{server-inout: } K) \rightarrow (\text{client-result: } K', \text{server-result: similarity})$

between  $K$  and  $K+BID-S$ ); where  $K' = K$  if  $(K+BID+N)-(S+N)$  is sufficiently similar to  $K$ .

5. SIPPA client returns  $K'$  through the voice gateway to the individual for the individual to derive  $Dec(K', Enc(k, Hash(UID)))$  (or  $Dec(K', Enc(k, UID))$ ).
6. Compute  $Dec(K' Enc(k, Hash(UID)))$ , or  $Hash(Dec(K' Enc(k, Hash(UID))))$ , depending upon whether the user stored  $Enc(K, UID)$  or  $Enc(K, Hash(UID))$  (by the individual or SIPPA client).
7. Present the hash of the decrypted UID and the token  $T$  to ISVM for comparing against the  $Hash(UID)$  and  $T$  stored in ISVM; ISVM accepts the claimed identity if the decrypted UID is found identical to  $Hash(UID)$  of ISVM with a matching  $T$  during the authentication.



### 4.3 Security Analysis

The security analysis will begin with a definition of security. The definition of security is based on the composition of the identity solution in terms of the functional components, their interaction relationship, the trustworthiness of the functional components, and the behavior of adversary.

Functional components of the proposed identity solution:

1. Voice Gateway (VG) serving as an interface between a user and the system. In this research we assume the communication is through a secure authenticated channel, which is reasonable and realistic in the real world situation.
2. Enrollment Module (EM) is composed of SIPPA server and a local storage for the cryptographic secret. EM receives from a user during enrollment a DID and a cryptographic secret  $K$  for encryption/decryption. By the principle of separation of duty and need-to-know, no sensitive personal or identity information is stored.
3. Identity Storage and Verification Module (ISVM) is composed of SIPPA client and a centralized database. ISVM is responsible for cryptographic key regeneration based on the helper data provided by the SIPPA server of the Enrollment Module.

The message exchange between the SIPPA client and server during the SIPPA protocol execution is also assumed to be carried out in a secure authenticated channel. In addition, as discussed elsewhere [1] SIPPA protocol is securely usable in the following sense:

- a. The correctness of protocol output on private input data is verifiable through Zero Knowledge Proof.
- b. SIPPA protocol does not assume or relay on honest or semi-honest model. Under the semi-honest model, each party participating in the protocol can retain all the exchanged messages during the protocol execution and can attempt to discover new information. However, the participating parties will not abort or deviate from the protocol.
- c. SIPPA employs Paillier cryptosystem, which is semantically secured, to achieve homomorphic encryption for private data comparison and reconstruction in the encrypted domain.
- d. SIPPA private data comparison could serve as a means for authentication. The accuracy of authentication based on SIPPA private data is comparable to the traditional authentication approaches as measured by AUC (Area Under Curve).

#### Adversary model

In this research we define an adversary model that is realistic in the real world. First, an adversary is assumed to have access to the identity management environment. Therefore, the adversary can enroll himself, impersonate others, or try to influence the behavior of the VG, EM, or ISVM. Furthermore, the adversary is also assumed to possess the following capabilities:

- a. Polynomial bounded computing power.

b. Privilege to initialize a SIPPA protocol execution as either a client or a server. As such, the adversary has access to protocol inputs and outputs.

Finally, the adversary can behave maliciously; i.e., the adversary can abort or deviate from the protocol, and can influence the delivery of messages (without altering them) over the authenticated communication channel. For example, the adversary can corrupt the SIPPA server/client to deliver an incorrect intermediate message (e.g., incorrect  $x$  in step 2) during the SIPPA protocol execution.

#### Analysis walk-through and main result

For simplicity and without the loss of generality, an adversary who can corrupt an individual user or a functional component of the identity management system is considered as a corrupted individual user or a corrupted functional component. This allows us to conduct the security analysis by considering the consequence on the privacy of the individual identity information when an individual, and/or one or more of the functional components are corrupted. The corruptible entities include: Individuals (as imposters), VG, ISVM, EM.

#### Key analysis result:

**Claim 1:** Under the assumption of one-time pre-enrollment key exchange utilizing Public Key Infrastructure among client-side software, Voice Gateway (VG), Enrollment Module (EM), and Identity Storage Verification Module (ISVM), the integrity of enrollment is guaranteed if none of the functional components in the application level is compromised.

**Claim 2:** SIPPA-based IDMS guarantees the privacy of the sensitive identity information — if no more than one entity is compromised; i.e., no information loss on UID and BID. In addition, it is detectable if the integrity of authentication is compromised.

**Claim 3:** The privacy of the sensitive identity information UID and BID is guaranteed even if all the entities are compromised.

#### **@Claim 1:**

In a secure authenticated channel, messages can be eavesdropped, relayed and replayed, but not altered. This can be achieved through message signing process using private key. In reference to client-side software download in step 1 of the enrollment process, the software may be intercepted. But the integrity of the software is guaranteed for the recipient. In reference to step 2, if both the client-side software and the voice gateway are secure, then the only possible attack will be man-in-the-middle attack by redirecting client communication to a malicious voice gateway. Since there is a pre-enrollment key exchange among the parties, the client-side software and the end-point functional components can mutually authenticate each other in the network layer to prevent man-in-the-middle attack [27].

In reference to step 3, the call-back by the voice gateway assures the authenticity of the individual identity as characterized by the individual's device ID. This step also serves as a commitment scheme to bind an individual to his phone number, UID, and DID through the token  $T$ . Furthermore, the encryption of the identity information prior to sending over to the functional components (EM and ISVM) ensures the confidentiality over the secure authenticated channel, while the integrity is assured by the property of the secure authenticated channel. Since adversary is assumed to have only polynomial bounded computing power, the adversary will not be able to reverse engineer the encrypted information into plain text.

With trivial observation, the integrity of enrollment cannot be guaranteed if at least one functional component is comprised. For example, the 3-tuple identity information containing the hash of the UID and biometrically encoded BID will be exposed if the centralized database is compromised.

#### **@Claim 2:**

We now provide a sketch for explaining the situation where only one entity is corrupted:

#### Corrupted voice gateway:

Corrupted voice gateway can still communicate with the individual user and the EM as well as ISVM. However, since all end-to-end communication is under secure authenticated channel, the corrupted voice gateway can at most learn the cryptographic secret during the enrollment, but cannot modify the cryptographic secret to compromise the integrity of the system service for enrollment and authentication.

During authentication, the corrupted voice gateway can learn from the user  $(K+BID)-S$ . Since the uncorrupted EM will never share  $K$  with the corrupted voice gateway, the corrupted voice gateway cannot learn BID from  $(K+BID)-S$ . Even if the corrupted voice gateway will first record the cryptographic secret  $K$  during the enrollment phase, BID still cannot be derived BID from  $(K+BID)-S$  without knowing  $S$ , which is the biometric sample of an individual and is never shared by the individual. Therefore, the privacy of UID and BID is preserved.

Again, since the corrupted voice gateway cannot alter the message in a communication between any two-party in a secure authenticated channel, the content of the hash in steps 5 through 7 of the verification protocol remains the same, thus the integrity of the system service for authentication.

#### Corrupted EM:

Enrollment module is corrupted if either the SIPPA server or the local database storing the cryptographic secret is corrupted. If the local database is corrupted, the cryptographic secret may be arbitrary changed. As a consequence, authentication will always fail. However, this will be detected by test cases grounded on  $BID=S=N=0$  injected into the

centralized database of ISVM and used in the integrity test. In other words, when  $BID=S=N=0$ , SIPPA protocol will return a conclusion where an arbitrary changed  $K$  is not equal to  $(K+BID+N)-(S+N)$  for test cases where  $BID=S=N=0$ .

If SIPPA server is compromised, it can choose to ignore and not to use the  $K$  retrieved from the database. Then the consequence will be the same as before, and is detectable. On the other hand, if SIPPA server is compromised and acts like a malicious user in the SIPPA protocol execution, then the security properties of SIPPA will apply and the followings will result:

- (i) Any attempt to decrypt the cipher text message without the secret key during the protocol execution will fail because the underlying cryptographic scheme Paillier cryptosystem is semantically secure and is not vulnerable to the attack by an adversary with only polynomial bounded computing resources.
- (ii) Any attempt to deviate from the protocol will result in a discrepancy when Zero Knowledge Proof is applied to verify the correctness of  $x$  (derived in step 2 and used in step 3 of the SIPPA protocol).
- (iii) If the help data  $(\lambda_{de})^{0.5}$  is modified before sending to the SIPPA client in step 4 of the protocol, SIPPA client — with verifiable correct  $x$  and  $\mathbf{v}_{de}$  (obtained in step 5 of the SIPPA protocol) — can detect the discrepancy through checking the equality  $(\mathbf{de} \cdot \mathbf{de}^T + \mathbf{dv} \cdot \mathbf{dv}^T)\mathbf{x} = \lambda_{de}\mathbf{v}_{de} + \lambda_{dv}\mathbf{v}_{dv}$ .

#### Corrupted ISVM:

ISVM is corrupted if either the SIPPA client or the centralized database storing is corrupted. If the centralized database is corrupted, 3-tuple identity information may be revealed and arbitrary changed. Since ISVM does not know the cryptographic secret  $K$ , it could not derive  $BID$  from  $K+BID$ . Given polynomial bounded computing power, it cannot reverse engineer  $UID$  from the one-way  $\text{hash}(UID/DID)$ . However, the 3-tuple identity information may be arbitrary changed, resulting in incorrect authentication outcomes. However, this can be detected by test cases grounded on  $BID=S=N=0$  as described before; i.e., SIPPA protocol will return a conclusion where  $K$  is not equal to  $(K+BID)-S$  for test cases where  $BID=S=0$ .

If SIPPA client is compromised, it could obtain  $K+BID$  and  $\text{hash}(UID)$  from the central database. Under the assumption on the polynomial bounded computing power,  $UID$  cannot be reverse engineered from the one-way hash. If  $\text{Enc}(K, UID)$  instead of  $\text{Enc}(K, \text{hash}(UID))$  is stored, then  $UID$  is not deemed private and exposing such information has not privacy leak.

#### Impersonation by individual user:

An imposter can impersonate the identity of others. However, the impostor has no knowledge of  $BID$ . Therefore, the imposter can only make a guess on the biometric sample  $S'$ . When  $BID$  and  $S'$  are not sufficiently similar,  $K+BID-S$

will be rejected by the SIPPA server as being similar to  $K$ . As such, SIPPA server will not provide helper data for SIPPA client to reconstruct  $K$ . Therefore, the privacy of  $UID/BID$  and the biometric template  $BID$  is protected.

#### **@Claim 3**

This is a restrictive case of claim 2 where only the privacy of the sensitive personal information is required and  $UID$  is deemed private. Since impersonator does not have  $\text{Enc}(K, \text{hash}(UID))$ ,  $UID$  cannot be uncovered even if  $EM$  discloses  $K$ . In addition,  $UID$  cannot be reverse engineered from  $\text{hash}(UID)$  even if  $ISVM$  discloses it because adversaries have only polynomial bounded computing power. Similarly, impersonator does not have  $N$ ,  $BID$  cannot be uncovered from  $K+BID+N$  even if  $EM$  discloses  $K$ . Therefore,  $UID$  and  $BID$  are protected even if all parties are compromised.

## 5 Implementation and Experimentation

For proof-of-concept, we conduct an experimental study on a prototype of the proposed identity management system. The objective of the study is to evaluate the usability of the system as measured by the verification accuracy.

The prototype is composed of an Asterisk PBX [28] for accepting up to five simultaneous incoming calls. An Asterisk to Java gateway serves as an interface for Java-based application implemented for speech processing to extract voice signature, and for SIPPA-based privacy preserving comparison.

There are two experimental trials in this study. The first trial is comprised of 90 calls from a pool of a dozen of individuals using three different phone models — with and without enabling the speaker phone mode. All enrollments and verifications were conducted in an environment where the background noise is fairly consistent. The second trial is comprised of over 400 calls from a pool of 20 individuals assuming 60 identities using 20 different phone models with two configurations — with and without headset. In average each individual assumes three different identities (i.e., enrolled three times). Furthermore, there is no restriction on the enrollment and verification in the second trial. For example, enrollments and verifications could be carried out under different noise environments, as well as different phone models and configurations.

Although it is possible to use a digital voice gateway that accepts voice signature directly from a user, in this experimental study a PBX voice gateway was used to accept incoming calls directly from a user. The speech processing for extracting voice signatures and noise injection was handled by the voice gateway instead of the individual users. The reason for this alternation in the protocol for this study is because the list of the phones includes conventional landline phone that has no capability of processing voice in the digital form. As such, a voice gateway such as Asterisk PBX to terminate the

analog PSTN calls is used in this experimentation; even though some other smart phone models such as Android based LG phones used in this study is capable of extracting voice signature and interacting with a digital voice gateway directly.

By shifting the speech processing task to the voice gateway, it exposes an additional vulnerability because a compromise in the voice gateway will leak the private information on the biometric voice signature of the sample S. However, since this experimentation is focused on the verification accuracy, a system with unrealized exploit on the additional vulnerability will result in the same behavior as one where the voice signature extraction is performed by the user.

In addition, one can also argue that there could be additional (telephone) channel noise (e.g.,  $N'$ ) introduced into the voice sample when it is processed on the end of the voice gateway. In this case, it will cause a consistent degradation of the similarity (i.e.,  $(K+BID+N)-(S+N'+N)$ ). The consequence of it is a shift in SIPPA threshold to yield the same result. But with the enrollment performed by the voice gateway, the net effect of the channel noise is roughly cancelled under the assumption of consistent channel noise (i.e.,  $(K+BID+N'+N)-(S+N'+N)$ ).

The result of this study presented as a Receiver Operating Characteristic (ROC) plot on false acceptance (FA) vs false rejection (FR) is shown below. The plots in Fig. 1 are the results of the first trial, detailing the change in the ROC with the speaker phone mode enabled/disabled. The Equal Error Rate (EER) in all cases is about 0.1. The plot in Fig. 2 is the ROC for the entire population without any restriction on the choice of the phone models, operation mode, and the background noise environment. EER is about 0.33.

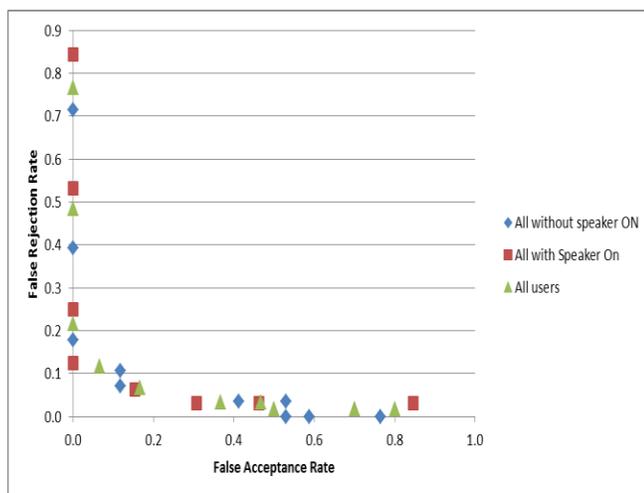


Figure 1. ROC under controlled noise environment

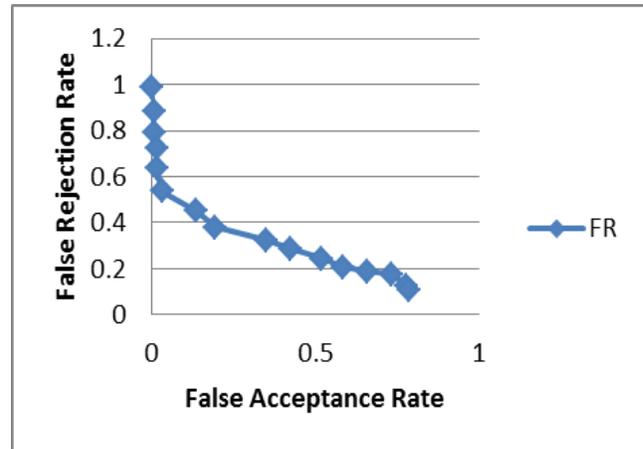


Figure 2. ROC under arbitrary noise environment

## 6 Conclusions

In this paper we present a privacy preserving voice-based identity solution for authentication based on our previous work on SIPPA. For proof-of-concept, we conducted a simulated experimentation to investigate the effectiveness of the prototype system in regard to its performance summarized in the ROC. Our future work will extend on the current research to investigate the effect of noise in the telephony channel on the performance, possible accuracy improvement based on individualized threshold, and the extensibility of its applications.

Acknowledgement: This work is supported in part from a grant by the PSC CUNY Research Award.

## 7 References

- [1] Bon K. Sy & Arun P. Kumara Krishnan, "Generation of Cryptographic Keys from Personal Biometrics: An Illustration based on Fingerprints," *New Trends and Developments in Biometrics*, ISBN 980-953-307-576-6, InTech, 2012.
- [2] Arun P. Kumara Krishnan and Bon K. Sy "SIPPA-2.0 – Secure Information Processing with Privacy Assurance (version 2.0)," *Proc. Of the 9<sup>th</sup> Conf. on PST*, Paris, France, July 2012.
- [3] William E. Burr, Donna F. Dodson, Elaine M. Newton, Ray A. Perlner, W. Timothy Polk, Sarbari Gupta, Emad A. Nabbus, *Electronic Authentication Guideline*; Special Publication 800-63-1; Dec 2011.
- [4] Peter G. Neumann, "System and Network Trustworthiness in Perspective," *CCS 06*, October 30–November 3, 2006, Alexandria, Virginia, USA.
- [5] Kui Ren, Wenjing Lou, Kwangjo Kim, Deng, R., "A novel privacy preserving authentication and access control scheme for pervasive computing environments," *IEEE*

- Transactions on Vehicular Technology*, (Volume:55 , Issue: 4), July 2006.
- [6] Tao Li, Wen Luo, Zhen Mo, Shigang Chen, "Privacy-preserving RFID authentication based on cryptographical encoding," *IEEE Proc. of INFOCOM*, 25-30 Mar 2012, Orlando Florida.
- [7] Elli Androulaki, *A Privacy Preserving Ecommerce Oriented Identity Management Architecture*, Master Thesis, Columbia University, May 2011.
- [8] Mauro Barni, Tiziano Bianchi, Dario Catalano, Mario Di Raimondo, Ruggero Donida Labati, Pierluigi Failla, "Privacy-Preserving Fingercodes Authentication," *MM&Sec'10*, September 9–10, 2010, Roma, Italy.
- [9] Meilof Veenigen, Benne de Weger, Nicola Zannone, "Symbolic Privacy Analysis through Linkability and Detectability," *Trust Management VII: IFIP Advances in Information and Communication Technology*, Volume 401, 2013, pp 1-16.
- [10] Monica Scannapieco, Ilya Figotin, Elisa Bertino, Ahmed K. Elmagarmid, "Privacy Preserving Schema and Data Matching," *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, Jun 2007, Beijing China.
- [11] Ralph Merkle, "A certified digital signature", In Gilles Brassard, ed., *Advances in Cryptology – CRYPTO '89*, vol. 435 of Lecture Notes in Computer Science, pp. 218–238, Springer Verlag, 1990.
- [12] Leslie Lamport, "Constructing digital signatures from a one-way function," Technical Report CSL-98, SRI International, Oct. 1979.
- [13] Jean-Jacques Quisquater, Louis C. Guillou, Thomas A. Berson, "How to Explain Zero-Knowledge Protocols to Your Children," *Advances in Cryptology - CRYPTO '89: Proceedings* 435: 628–631.
- [14] Jan Camenisch, Maria Dubovitskaya, Anja Lehmann, Gregory Neven, Christian Paquin, Franz-Stefan Preiss, "Concepts and Languages for Privacy-Preserving Attribute-Based Authentication," *Policies and Research in Identity Management: IFIP Advances in Information and Communication Technology*, Volume 396, 2013, pp 34-52.
- [15] Denis Trček, *Managing information systems security and privacy*, Birkhauser, p. 69. ISBN 978-3-540-28103-0, 2006.
- [16] Carlisle Adams, Steve Lloyd, *Understanding PKI: concepts, standards, and deployment considerations*, Addison-Wesley Professional. pp. 11–15. ISBN 978-0-672-32391-1, 2003.
- [17] Stephen August Weis, "Security and Privacy in Radio-Frequency Identification Devices," Master Thesis, MIT, May 2003.
- [18] Tassos Dimitriou, "A Secure and Efficient RFID Protocol that could make Big Brother (partially) Obsolete," *Proc. of IEEE PERCOM*, 2006.
- [19] S. Goldwasser and S. Micali, "Probabilistic encryption," *Journal of Computer and System Sciences*, 28:270-299, 1984.
- [20] J. Bourgain, "On Lipschitz Embedding of Finite Metric Spaces in Hilbert Space," *Israel Journal of Mathematics*, 52 (1985), no. 1-2, 46{52.
- [21] Daniele Micciancio, "Technical Perspective: A First Glimpse of Cryptography's Holy Grail," *Communications of the ACM*, Vol. 53 No. 3, Page 96, March 2010.
- [22] P. Paillier, "Public-Key Cryptosystems Based on Composite Degree Residuosity Classes," in *EUROCRYPT*, 1999.
- [23] Jonathan Katz, Yehuda Lindell, *Introduction to Modern Cryptography: Principles and Protocols*, Chapman & Hall/CRC, 2007.
- [24] R. Cramer, I. Damgard, Jesper Buus Nielsen, (<http://www.daimi.au.dk/~ivan/mpc.pdf>) *Multiparty Computation: An Introduction*.
- [25] M. Fitzi, D. Gottesman, M. Hirt, T. Holenstein, A. Smith, "Detectable Byzantine Agreement Secure Against Faulty Majorities," *Proc. of the 21st ACM Symposium on Principles of Distributed Computing (PODC)*, July 2002.
- [26] James A. Hanley, Barbara J. , "A method of comparing the areas under receiver operating characteristic curves derived from the same cases". *Radiology* 148 (3): 839–43. PMID 6878708.
- [27] Jonathan Katz, "Efficient Cryptographic Protocols Preventing Man-in-the-Middle Attacks," Ph.D. thesis, Columbia University, 2002.
- [28] Jim Van Meggelen, Jared Smith, Leif Madsen, *Asterisk: The Future of Telephony*, ISBN-10: 0596009623, Sept 2005.
- [29] Thrasyvoulou T., Benton S.: Speech Parameterization Using the Mel Scale (Part II), (2003).

# Creating a Policy Based Network Intrusion Detection System using Java Platform

Samuel N. John<sup>1</sup>, Charles U. Ndujiuba<sup>2</sup>, Robert E. Okonigene<sup>3</sup>, Chinonso E. Okereke<sup>4</sup>, Miebaka E. Wakama<sup>5</sup>

<sup>1, 2, 4, 5</sup> Department of Electrical and Information Engineering, Covenant University, Ota, Ogun State, Nigeria.

<sup>3</sup> Department of Electrical and Electronics Engineering, Ambrose Alli University, Ekpoma, Edo State, Nigeria.

**Abstract** – Computer network attacks are rapidly increasing on a global scale. Security mechanisms are established and used extensively to counter these security threats knowing their ubiquitous nature and severity. Like in most Financial institutions and big Organizations world wide, in most Nigeria Universities network security threats have become a major challenge and of great concern to the University authorities. Most Nigeria Universities rely solely on their established networks for revenue generation. The Universities have adopted some policies that are peculiar to its needs. Several reported breaches in these Universities network by intruders and sometimes insiders threat have become a major challenge to the authorities. Hence, this paper addressed the security policy based apparatus to stop the havoc caused by these threats. Among the deployable security arsenals is Intrusion Detection System (IDS). There was a need to customized IDS to suit the security policies of specific University network and these security policies vary from network to network. In order to meet this need, a signature-based Network Intrusion Detection System was design to suit the policies of each university network and to accommodate policy changes. This was achieved through Java platform, that is, Jnetpcap Library and Java Expert System Shell (JESS). The results obtained shows that Java's Jnetpcap Library and Expert System Shell provided a way to accommodate dynamic network demands as well as stop specific intrusion. The created Network Intrusion Detection System (NIDS) improved the overall robustness and security of the Network.

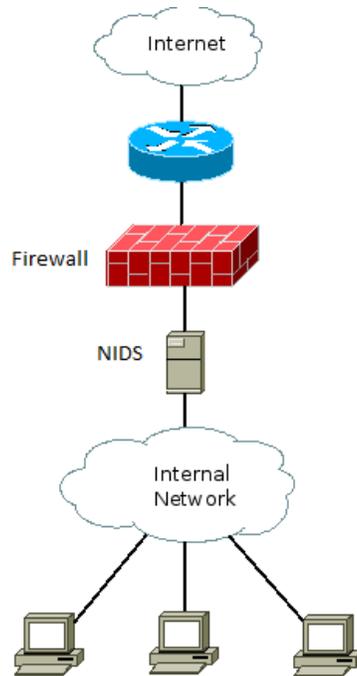
**Keywords:-** Intrusion Detection System, Signature-based, JESS, Jnetpcap Library, Policy Protocols, Security.

## 1 INTRODUCTION

Most Nigeria Universities have adopted the internet to disseminate information about the authorities policies to the immediate communities and the world. The policies involved in online: payment of school fees, registration, display of student results, student payment status, application for admission, verification of basic admission requirements, admission into various degree programmes, obtaining matriculation number,

registration of courses, qualification to write examinations, information about each student status, may defer slightly depending on the University. During our study we observed that due to security breaches in most of the Universities network there are cases of identity theft, use of unauthorized pins for payment, unauthorized access to vital information, unauthorized change of students data. There are several reported cases of increased amount of attacks and sophisticated intrusions developed by intruders to compromise security of these networks. Universities have lost huge amount of revenues as a result of some of these security breaches. The function of an Intrusion Detection System (IDS) is to monitor events occurring in a computer system or network and analyze for signs of possible intrusion. An intrusion can be defined as any set of activities that attempt to compromise the integrity, confidentiality or availability of the network [1, 2]. It can also be seen as actions or traffic not legally allowed on a system or network [3]. Some examples are denial of service, masqueraders, malicious activities such as trojan and viruses. Another definition of intrusion is network policy violation.

Intrusion Detection System is divided into Host-based and Network-based. The former refers to monitoring of a single computer system unit (HIDS) while the later refers to the monitoring and analysis of the traffic of network segments or the whole network (NIDS) [4]. NIDS attempts to detect unauthorized access or intrusion from the analyzed data or information pulled together from the network traffic [5]. It is usually positioned behind the firewall for effectiveness as shown in Figure 1 as it provides a deep inspection of the payload [6]. The Intrusion Detection is usually implemented using the anomaly-based or signature-based methods. In anomaly-based intrusion detection, the baseline for normal behavior and characteristic of the network is gotten and an alert generated when a deviation from the specified baseline occurs [8]. However signature-based intrusion detection method defines patterns from known malicious codes then seeks out such patterns using algorithms to compare packets from the analyzed data and then generates alerts [8].



**Figure 1 Positioning of the NIDS [7]**

Signature-based detection has its advantages in that it is more accurate in identifying an intrusion attempt, provides an easy way of tracking down the cause of alarm due to detailed log files and also, reduced false positives alerts [1]. The term 'intrusion' which is also policy violation depends on the policy setup by the organization using the NIDS. It is therefore important that the NIDS be setup to suit the policy of the organization.

In Covenant University, Otta Nigeria minimum guidelines (policies) were set out for proper, efficient and effective use of ICT (internet) in order to regulate and ensure delivery of qualitative education [9].

Although not every policy is implementable in the NIDS, one of the ICT guidelines (policy) is Web filtering to ensure efficiency and high availability of internet services to all used. The policy requires that MP3 traffic and other high bandwidth intensive services that may not have direct educational or research value are adequately filtered. This is apart from the generic forms of intrusion that will disturb the efficiency of the network such viruses and general malware [9]. NIDS has the ability to perform a deep inspection of the payload of the traffic in the network. This helps to achieve thorough filtering. The thoroughness of the filtering depends on the specified rules or signatures which can be changed from time to time. This paper shows the creation of a signature-based NIDS that suits the specified policy and caters for frequency occurring threats and further support the overall security and integrity of the University Network.

Aijaz Ahmed [10], presented the design and implementation of a Signature-based Network Intrusion Detection System using JESS (Java Expert Shell System) (SNIDJ) which used Snort as a packet-sniffing tool for capturing network traffic and this was manually converted into facts as represented in JESS. Vamshi K Kankanala [11] showed the design and implementation of a Web-based Network Intrusion Detection Expert System (WNIDES) which was implemented using (JESS). The system used Snort program to capture network packets and has a Graphic User interface (GUI) to input new rules (i.e. signature). Chaitanya Chinthreddy [12] presented the design and implementation of an online IDS based on Snort rules as signatures using JESS and tested the system with manually generated packets. The design and implementation of a signature-based NIDS system with Java programmed sniffer using Jnetpcap Library and detection engine written in Java Expert System Shell (JESS) with Snort rules as signature was used. The signature-based NIDS has a GUI that allows users to input signatures apart from the already existing signatures in the detection system. The system was tested on the Covenant University Network. The paper is organized as follows: section 2 shows the overview of the implementation tools used in building the system. Section 3 is the design and methodology of the NIDS, section 4, indicate the implementation of the JNETPCAP library and section 5, shows the system testing and results.

## 2 OVERVIEW OF THE IMPLEMENTATION TOOLS IN NIDS

**In this project the following implementation tools were used to build the NIDS system:**

- Jnetpcap Library
- Java Expert System Shell (JESS)

### 2.1 JNetPcap Library

*JnetPcap* library is an open source java wrapper around *libpcap* and *WinPcap* native libraries. *Libpcap* is an application programming interface (API) for packet capturing from a live network interface. The Windows port of *libpcap* is called *Winpcap*. *Libpcap* and *Winpcap* libraries are used for making the packet capture and filtering engines among other network tools.

*Jnetpcap* expose the functionalities found in the *libpcap* and *Winpcap* library which provides enough tools necessary to carry-out thorough analysis on packets captured from the network..

*Jnetpcap* has an extensive list of protocols which is able to decode and these includes the following:

- Ethernet: 802.3, 802.2, SNAP, SLL, VLAN

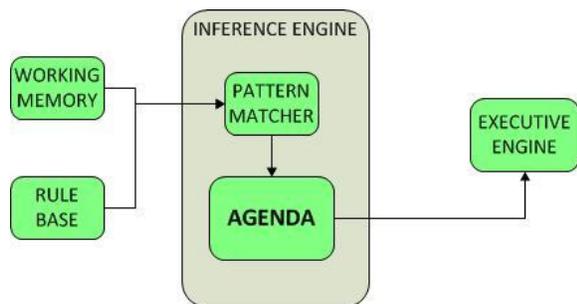
- Network: IP4, IP6, ICMP, ARP, RIP1, RIP2, TCP, UDP, HTTP, HTML, WEBIMAGE
- VOIP: RTP, SDP, SIP
- VPN: L2TP
- WAN: PPP [13]

*Jnetpcap*'s ability to decode such protocol and provide the appropriate header structures for these protocols avails the ability to carry-out adequate analysis and intrusion detective schemes on the protocols as the needs arises.

**2.2 JESS (Java Expert System Shell )**

Java Expert System Shell is a rule engine and scripting environment written entirely in Oracle's Java™ language with a fully developed API used to create rule-based expert system [14,15]. A Rule Based Expert System Architecture is made up of a set of rules called the Rule Base, a memory that analyzes program data known as the working memory and the inference engine which does the decision making and implementation of the decisions made. The Inference Engine is made up of components such as a pattern matcher, an agenda and an execution engine. A pattern matcher refers to the component that scans through the program data and also the rule base to see which rule or rules can be executed on the program data .The agenda on the other hand orders the implementation of the rules in the case way multiple rules must be implemented. The execution engine then 'fires' or executes the selected rules in their specified order [15].

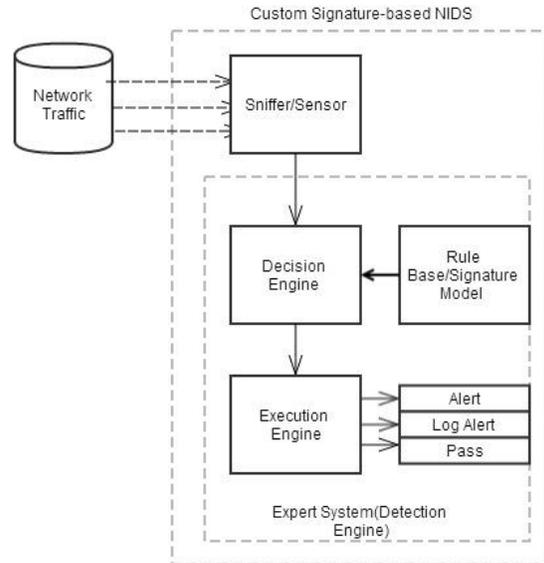
JESS's rule engine uses the Rete matching algorithm which is known to speed the matching of patterns rules against the working memory or program data also known as facts[16,17]. Figure 2 shows the architectural diagram of a Rule Based Expert System with the components that link the system together.



**Figure 2. Architecture Diagram of Rule Based Expert System [15]**

**3 SYSTEM DESIGN & METHODOLOGY**

The overall design model of our Network Intrusion Detection System is based on the working model of Rule Based Expert System Architectural Diagram. The design is as shown in Figure 3.

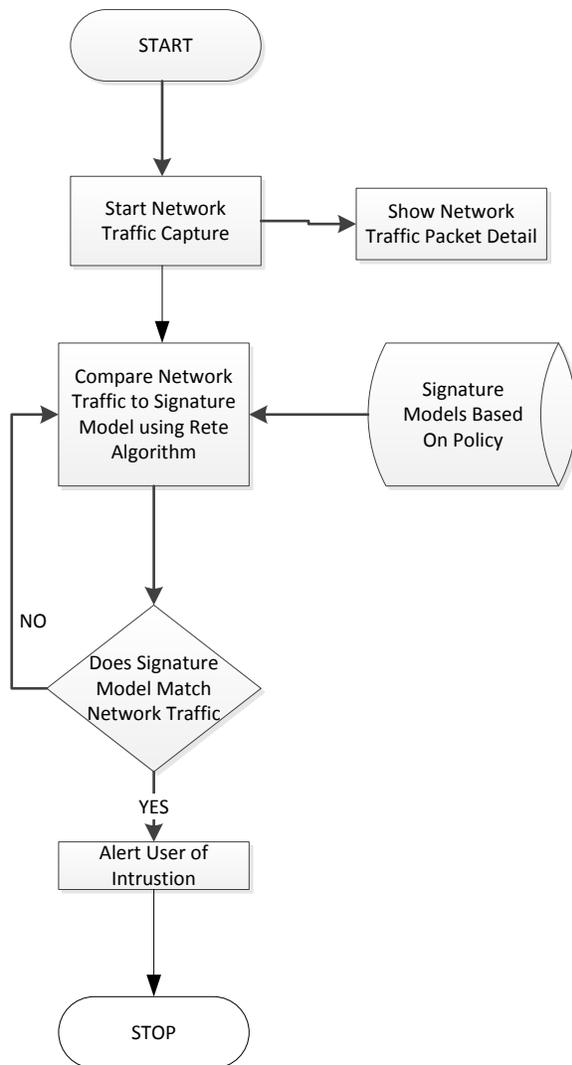


**Figure 3. Overall Design of signature-based NIDS**

It comprises of a sniffer or sensor and the rule based expert system which is the detection engine. Components of the rule based expert system are modified to suit the intrusion detection system i.e. the Rule Base in the developed NIDS is the Signature Model or set of signatures, the Working Memory represents the program data or traffic captured by the sniffer, the Pattern Matcher and the Agenda are represented by the Decision Engine while the Execution Engine does the implementation of the specified signatures. The NIDS also shows details of the network packets and alerts generated in its GUI. The flowchart of the System is shown in the Figure 4.

**Working Principle**

The sniffer/sensor of the policy-based NIDS captures live network traffic in the form of packets. Packets are then interpreted by the sniffer and facts are generated which are sent to the Expert System (JESS). Packets from the network traffic which are now facts are sent to the detection engine where they are compared with existing signature/rules in the Signature Model. The JESS program matches facts in the working memory to rules in the rule base or signature model of our intrusion detection System.



**Figure 4. Flowchart of NIDS Working Principle**

The rules contain function calls that manipulate the fact base. The output is then sent to the decision engine to the execution engine where decision taken is implemented either an alert and log or log and pass on.

#### 4 IMPLEMENTATION OF JNETPCAP LIBRARY

Basically, NIDS was programmed using the Java Platform and JESS. The Jess language can be extended in java and also incorporated into a Java program making it possible to build a single java program, in our case the signature-based Intrusion Detection System and robust GUI. The implementation can be categorized into the following subheadings.

##### 4.1 Packet Capturing, decoding and analysis

The Jnetpcap library allows the sniffer to capture and decode the packets which is to be analyzed by the NIDS in the network. The NIDS analysis the captured packets with their various protocols such as ARP, IP, TCP, UDP and HTTP. Reassembly of fragmented packets will be carried out as detection of signatures strictly relies on complete payload for matching with the rules or else threats or attacks may be missed or not detected.

##### 4.2 Converting Packets to Facts

The properties or details of the packet form the facts that are to be matched with the rules which mostly lie in the payload where most attack-signatures are found. However, certain clients or server addresses, ports may not be allowed and so will also be checked including some other facts that may be of concern. The template of our rules tells what packet data properties are to be checked because these templates form the various properties of the packet which are to be analyzed. This include the Source and Destination addresses and ports, payload of packets, Type of service (TOS), Time to live (TTL), Identification number (ID) and Ack flag. These data are extracted from the packet and converted to their respective valid variable types in the program for comparison with the rules. Some cases may arise when certain information may not be available, for instance, when the last header in the packet is the ARP header, which will leave out the IP, TCP and HTTP header fields. In such a case, the MAC address and payload will be checked as they are most concern in this scenario and other similar cases. Other unavailable field will be set to null. The bottom line in the packet facts are generated from every packet header fields that follows the packet in question.

These facts, after been, generated are then passed on to the rule engine for matching with the rules of the engine which will trigger upon a positive match between the rule and facts that are generated and vice versa.

#### 5. SYSTEM TESTING AND RESULTS

The NIDS system was executed and tested on a Covenant University network. The results are explained below.

At the start of the NIDS a Graphic User Interface pops up giving the details of the packets that move along the network. The details include the source, destination IP address and protocol among others as shown in Figure 5. This also shows the content of the payload message with the time tag (as shown in Figure 6) which is needed for a more in-depth analysis of their content. When patterns matching rules are detected, alerts are shown in the alert window as shown in Figure 7. Apart from been shown in the GUI, details concerning the

packets, payload and alerts are stored and saved in appropriate text file for record purposes and further analysis as the case arises.

No	Time	Source	Destination	Protocol	Info
1	17:55:54	224.0.0.251	192.168.0.105	UDP	
2	17:55:56	1583	1583	UDP	
3	17:55:56	DD:DF:9A:95:48:A	FF:FF:FF:FF:FF:FF	ARP	
4	17:55:56	DD:DF:9A:95:48:A	FF:FF:FF:FF:FF:FF	ARP	
5	17:55:56	DD:DF:9A:95:48:A	33:33:FF:BA:DD:9	ARP	
6	17:55:57	DD:DF:9A:95:48:A	FF:FF:FF:FF:FF:FF	ARP	
7	17:55:57	DD:DF:9A:95:48:A	33:33:FF:BA:DD:9	ARP	
8	17:56:00	192.168.0.255	192.168.0.105	UDP	
9	17:56:00	192.168.0.255	192.168.0.105	UDP	
10	17:56:01	DD:DF:9A:95:48:A	FF:FF:FF:FF:FF:FF	ARP	
11	17:56:01	192.168.0.255	192.168.0.105	UDP	
12	17:56:02	DD:DF:9A:95:48:A	FF:FF:FF:FF:FF:FF	ARP	
13	17:56:03	DD:DF:9A:95:48:A	FF:FF:FF:FF:FF:FF	ARP	
14	17:56:04	DD:DF:9A:95:48:A	FF:FF:FF:FF:FF:FF	ARP	
15	17:56:05	DD:DF:9A:95:48:A	FF:FF:FF:FF:FF:FF	ARP	
16	17:56:06	DD:DF:9A:95:48:A	FF:FF:FF:FF:FF:FF	ARP	
17	17:56:07	DD:DF:9A:95:48:A	33:33:0:0:0:16	ARP	
18	17:56:07	DD:DF:9A:95:48:A	33:33:0:0:0:16	ARP	
19	17:56:07	DD:DF:9A:95:48:A	33:33:0:0:0:16	ARP	
20	17:56:07	DD:DF:9A:95:48:A	33:33:0:0:0:16	ARP	
21	17:56:07	DD:DF:9A:95:48:A	33:33:0:0:0:16	ARP	
22	17:56:07	DD:DF:9A:95:48:A	33:33:FF:BA:DD:9	ARP	

Figure 5. GUI Showing Packet Details

```

0100: 02 00 4e 2f 01 2e 90 30 55 90 4e 50 2f 31 2e 30 208/L0 UDP/110
0200: 20 55 50 4e 50 20 44 65 76 69 63 65 20 48 6f 73 078P-Devlop-Box
0300: 74 22 31 24 30 60 04 63 41 63 60 65 20 23 62 6e 011.0...Com-Box
0400: 74 72 6f 60 3a 60 61 70 20 61 67 65 3a 34 30 08 toolman-agent0.
0500: 0a 0f 0a

Fri Mar 14 17:56:44 WAT 2014
0600: 4e 4f 54 49 46 59 20 2a 20 48 54 54 50 2f 31 2e 30 NOTIFY * HTTP/1.
0700: 31 00 0a 40 4f 70 74 3a 32 53 39 2e 32 30 30 2e 1..Host:239.239.
0800: 32 35 39 2e 32 39 30 3a 31 39 30 30 00 0a 4e 94 281.2013/05..._NET
0900: 3a 75 72 4e 3a 73 63 65 65 60 63 73 20 77 69 66 6a root@ubuntu-virt
1000: 69 61 60 60 69 61 66 63 69 20 42 72 67 3a 61 65 8111anon-org@du
1100: 76 69 63 65 3a 67 64 61 64 65 76 69 63 63 3a 31 7608@DEVServer1
1200: 00 0a 4e 54 53 3a 73 73 64 70 3a 61 60 69 76 65 ..RTS:rapid@live
1300: 00 0a 40 60 60 63 61 76 69 62 6e 3a 68 74 74 70 3a ..Location:htp:
1400: 2f 2f 31 39 32 24 31 36 38 74 91 2e 61 3a 30 30 /192.168.1.1190
1500: 2f 69 67 64 2e 76 60 60 00 0a 55 53 4e 3a 75 75 /sgd.mel..USB:ru
1600: 69 64 3a 30 30 30 30 30 30 30 30 30 20 30 30 30 30 1400000000-0000
    
```

Figure 6. GUI Showing Payload

Fri Mar 14 18:01:42 WAT 2014	Alert, User root kit attempt
Fri Mar 14 18:01:42 WAT 2014	Alert, User root kit attempt
Fri Mar 14 18:01:42 WAT 2014	PORN masturbation
Fri Mar 14 18:01:42 WAT 2014	Alert, User root kit attempt
Fri Mar 14 18:01:42 WAT 2014	PORN erotica
Fri Mar 14 18:01:42 WAT 2014	PORN erotica
Fri Mar 14 18:01:42 WAT 2014	Alert, User root kit attempt
Fri Mar 14 18:01:42 WAT 2014	Alert, User root kit attempt
Fri Mar 14 18:01:42 WAT 2014	PORN masturbation
Fri Mar 14 18:01:42 WAT 2014	PORN masturbation
Fri Mar 14 18:01:42 WAT 2014	PORN masturbation
Fri Mar 14 18:01:42 WAT 2014	Alert, User root kit attempt

Figure 7. GUI Showing Alerts Detected

The NIDS program also provides an interface that includes custom rules/signatures which are not listed in the generic signature base.

In the Figure 8, the window allows the user to include their own rules for the rule engine depending on what are to be detected/tested by the NIDS system, as well as how detection should be handle. Each field allows flexibility in choices such as: any source or destination address or ports, the direction of flow and also choice of to log or alert the packet or both.

This interface enables the policies that guide an organization to be included and thereby making the NIDS system relevant and useful to the network environment of such organization. Covenant University is an example of such organization where this was implemented.

Signature Builder

IDS Mode: Alert

Message: Porn dildo

Source Address: any

Destination Address: any

Source port: 80

Destination port: 537

Direction: unidirection

Content: xxx

HTTP Content: xxx

Submit Clear

Figure 8. GUI Showing Interface to input new rules

## 6 CONCLUSION

This IDS provides a more user friendly interface to the public which makes it very easy for users to get acquainted with, especially when including rules into the rule engine. Not to mention of its platform-independent execution environment as it is being developed as a java program.

Also the use of the JnetPcap library exposes us to a lot of already decoded protocols which we could exploit for various attacks as it may concern the network in question. This library also makes it possible for us to analyze customized protocols for privately owned user networks enabling us to build an IDS for such privately owned networks.

## 7. REFERENCES

- [1] B. Lokesak: "A Comparison Between Signature Based and Anomaly Based Intrusion Detection Systems":[www.iup.edu/WorkArea/DownloadAsset.aspx?id=81109](http://www.iup.edu/WorkArea/DownloadAsset.aspx?id=81109), 2008.
- [2] P. Mell and K. Scarfone: "Guide to Intrusion Detection and Prevention Systems (IDPS)", Recommendation of the National Institution of Standards and Technology (NIST) Special Publication 800-94, 2007.
- [3] H. K. Mbikayi, "An Evolutionary Approach toward Rule-Set Generation for Network Intrusion Detection": *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 2, no. 5, 2012.
- [4] B. M. Beigh and M. A. Peer: "Intrusion Detection and Prevention System: Classification and Quick Review," *ARNP Journal of Science and*

- Technology, vol. 2, no. 7, pp. 662-675, Aug. 2012.
- [5] V. Berk, G. Bakos, "Designing a Framework for Active Worm Detection on Global Networks", Proceedings of the First IEEE International Workshop on Information Assurance, 2003.
  - [6] Chwan-Hwa John Wu, J. D. Irwin: "Introduction to Computer Networks and Cybersecurity" Florida, United States of America, CRC Press Taylor & Francis Grand, 2013.
  - [7] Digital Undercurrents Network and Security Consulting: "Intrusion Detection Systems" <http://www.digitalundercurrents.com/blog/?p=82>, Buffalo, NY, 2011
  - [8] F. B. C. De Ocampo, T. M. L. Del Castillo, and M. A. N. Gomez, "Automated Signature Creator for a Signature Based Intrusion Detection system (PANCAKES)," SDIWC, pp. 198-205, 2013.
  - [9] Covenant University, Covenant University Information and Communication Technology (ICT) Policy. Ota, Ogun State, Nigeria: Covenant University, 2012.
  - [10] A. Ahmed, "Signature-based Network Intrusion Detection System using JESS (SNIDJ)," Texas A&M University, Corpus Christi, 2004.
  - [11] V. K. Kankanala, "Web-based Network Intrusion Detection System," Texas A&M University, Corpus Christis, Graduate Project Technical Report, 2006.
  - [12] C. Chinthireddy, "Using the JESS Expert system tool to implement an Online Intrusion Detection System based on Snort Rules," Texas A&M University, Corpus Christi, Texas, 2011.
  - [13] jNetPCap 1.3 Library overview, Sly Technologies jNetPcap, <http://jnetpcap.com/>, 2014
  - [14] S. N. Laboratories: JESS, the Rule Engine for Java Platform. <http://herzberg.ca.sandia.gov/>, 2013
  - [15] P. Jackson: "Introduction to Expert System"; 3<sup>rd</sup> Edition Addison-Wesley, 1999.
  - [16] Wright and Marshall, Ian Wright and James Marshall. "The execution kernel of rc++: Rete\*, a faster rete with treat as special case". International Journal of Intelligent Games and Simulation, 2(1):36-48, 2003.
  - [17] R. Selvamony, Introduction to Rete Algorithm. India: AP Labs India, Dec. 2010.

# Achieving Web Security by Increasing the Web Application Safety

Maryam Abedi

Dept. of Information

Technology Eng., Shiraz

University, Shiraz, Iran

maryam\_abedy@yahoo.com

Navid Nikmehr

Dept. of Information

Technology Eng., Shiraz

University, Shiraz, Iran

nikmehr@ieee.org

Mohsen Doroodchi

Dept. of Math/Computer

Science, Cardinal Stritch

University, Milwaukee, WI

mdoroodchi@stritch.edu

**Abstract**—As web applications have become an integral part of today's business operations, the concerns about the security of exchanged information on the web have been increasing. Issues such as data breach and leakage of sensitive information is number one concern of businesses for which the web applications are blamed for the most part. Therefore, in addition to the common measures used to secure the communications and transactions on the web, more attention needs to be paid to the preventive measures of integrating security into the development phase. However, for evaluation of effectiveness of such measures, a quantitative method is very essential to calculate the safety of an application against different vulnerabilities. This work presents a new model for measurement of overall safety of web applications. The keyword "safety" is coined to distinguish this measure from the traditional methods.

**Keywords:** quantitative measurement, web application security, safety, vulnerability

## I. INTRODUCTION

Enterprises' critical resources are highly in risk of cyber-attacks due to the vast delivery of enterprise applications with vulnerabilities over the web. Reports of the catastrophic hacking stories reveal that the sensitive data are compromised through web application vulnerabilities. In order to deliver safe applications over the web, such vulnerabilities are required to be studied and understood in depth.

Different forms of injections are reported to be the major concern with web applications. According to OWASP [25] and SANS 2011 Top 25 [41], SQL injection is ranked first among web application vulnerabilities. After that, code and shell injection are introduced as the second security issue in today's web applications. Based on a report by Viega and McGraw, code injection is known as the most challenging security breach as a result of poor input sanitization [39].

Despite considerable research on understanding and managing the security issues, including qualitative aspect of security measurement such as BS7799 [4, 5], ISO17799, NIST SP800-33 [2, 29, 30], there are only few quantitative metrics [3] available for measuring security related issues. These methods are often either not comprehensive enough [19, 22] or are limited only to their specific measurement model which reduces the usability of the model and some are too complicated to be used by developers [26].

This is an undeniable consensus that the capability of measuring, comparing, and contrasting different entities provides the opportunity of a thorough understanding of the underlying concept [22] as Lord Kelvin in 1883 stated: "When you can measure what you are speaking about and express it in numbers you know something about it, but when you cannot

measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind". And security management of web applications through measurement is not an exception.

In this paper, we overlook the concept of enterprise application security in terms of a quantifiable concept we coined as web application safety. The proposed measurements model is using the known vulnerabilities and at the same time is scalable to use the new vulnerabilities. Our aim is to find a practical and universally acceptable quantitative model that can be integrated into the software development life cycle. The proposed model allows the developers to measure the safety of their under application during different development phases.

The foundation of our model is based on measurement of two aspects of standard and best practice prevention methods integrated into the projects. These two aspects are called efficiency and sufficiency of the methods which are explained in details later. In addition, we quantify the effectiveness for each method. At this time, the coefficient of effectiveness is determined subjectively based on experts' perspective.

Since the discovered vulnerabilities are rapidly increasing, the capability of appending additional vulnerabilities to our model along with their corresponding mitigation methods, leads to enhance the flexibility and extensibility of the proposed measurement model. In addition, this extensibility feature provides flexibility to redefine, modify, and improve the proposed quantification metric definitions throughout the development process.

The rest of this paper is organized as follows. Next section covers the proposed model for web application safety measurement followed by proposed metrics used in the overall calculations followed by the experimental results. The last section of this paper is the conclusion.

It is notable to mention that quantified level of safety against specific vulnerabilities for any application is specific to that application and it cannot have a specific scale and range. Therefore, this method is best fit for comparison of the safety level for different versions of the same application.

## II. PROPOSED SAFETY MODEL

As mentioned, this model starts from a list of known vulnerabilities and the corresponding measurement metrics to calculate the overall safety of a web application. To achieve this task, we propose a hierarchical model as illustrated in figure 1. The metrics are categorized based on different vulnerability types as depicted in this figure. Furthermore, in order to evaluate the model we select different metrics to measure the safety of a web application against SQL injection and Shell injection as explained in the following sections. We chose six metrics for each category. Each metric evaluates the sufficiency (and/or) efficiency of

possible preventative method that could have been implemented to raise the overall safety of the application.

The overall safety of an application can be visualized as the root of a tree as shown in Fig. 1 in which the branches are providing the particular safety measurement for a given preventative method.

This approach has a number of benefits. First, the security tester can plan the test using different combination of available vulnerabilities. Second, these metrics reveal the interdependencies of different vulnerabilities to the developers, and consequently the application developer could provide additional isolation between them.

Moreover, this model is scalable and flexible to add new metrics for known or new vulnerabilities. Next section explains each metric in details. For this purpose, we consider the following two parameters for each metric, 1) a name, and 2) a description. The description provides information about the vulnerability and corresponding mitigation methods which can increase safety. It also provides a proposed formula that measures the sufficiency and/or efficiency of mitigation method and returns a numerical value. Each mentioned formula needs some inputs – aka vector of inputs- to return the safety value. We also define two properties for each member of this vector; input name, description and a numerical value.

This numerical value can be entered by the user of the model based on the application or comes from another formula's numerical result of other metrics. Clearly, larger values of the results of each of the formula would contribute directly to a safer application which consequently results in an overall increase of security.

In addition, this model has three more parameters to achieve more accurate value for overall safety. These parameters are listed as follows. The *first* parameter is “*e*”, the effectiveness coefficient, as shown in Fig. 1. It is clear that all the mitigation mechanisms do not have the same contribution toward the application safety improvement. With respect to this fact, this parameter reflects the metric's relative importance and effectiveness in mitigating the overall vulnerability of the application. Obviously, the ones that are more effective have greater weight. We recruit fuzzy logic to determine effectiveness coefficients' value. The process of determining this value is fully explained in section V.

The *second* parameter is “score of vulnerability” as shown in figure 1. It is clear that all types of vulnerabilities do not endanger the safety of a web- application equally; hence this parameter reflects the weight of the vulnerability. In this work, its value equals to the score of the vulnerability in CWE/SANS ranking system, as a reputed reference in this area. The vulnerabilities are prioritized and scored according to their prevalence, importance, and likelihood of exploiting [8].

The *third* parameter is “Phase of lifecycle” as depicted in Fig. 1. Our model is capable of evaluating a web application safety in any of the main three phases of analysis, design, or implementation phase of SDLC. This parameter should be applied as a consistent value throughout the safety evaluation process. Given the fact that implementing of any prevention method in the earlier phases has more effectiveness than postponing them to next phases [13], we assign a greater coefficient to earlier phases. Note that the formulas of metrics for one phase are different with another phase, but the general principle of the model is the same. On the other hand, obviously, the value for all parameters of the model should be obtained from the same phase.

Based on the above explanations, the overall safety against a specified vulnerability is demonstrated as *OSAV* function as follows.

$$OSAV = c * \sum_{i=0}^n (f_i(p)) \tag{2}$$

$$f(p) = a * \sum_{j=0}^n (e_j * p_j) \tag{2}$$

, where *c* is score of vulnerability based on “CWE /SANS” [8], and *a* represents phase of lifecycle of given project, and *e* is effectiveness ∈ [0,1], and *p* represents result value of quantification formula for evaluation of specific preventative method.

As mentioned before, in this work we examine our model and its metrics for the first two highest ranked vulnerabilities as mentioned in the ‘Top 25 Common Weaknesses Enumeration (CWE) database’ [1]. The database that is sponsored by Mitre, is used frequently as a reference by application developers and security engineers to identifying possible weaknesses to attack in software applications. However, it does not mean that this model is restricted to assess safety against this database's vulnerabilities. The proposed associated quantitative formulas for mentioned vulnerabilities are defined in next section.

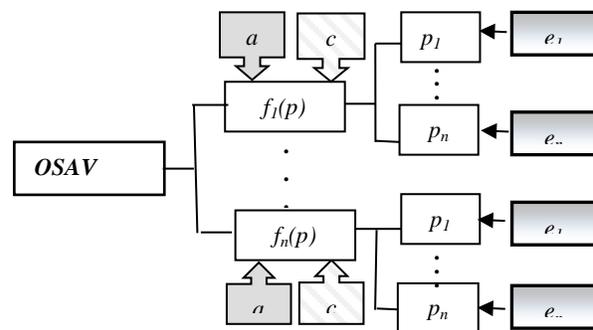


Fig.1.The proposed model

### III. PROPOSED METRICS

In this section, we explain the details of proposed quantitative metric for the top-two web application vulnerabilities from “Top 25 Common Weaknesses Enumeration (CWE) database” [1] to evaluate the sufficiency and/or efficiency of common (and standard) preventive mechanisms in a given application.

#### A. SQL Injection Vulnerability

According to [35], “SQL injection is an attack in which malicious code is inserted into strings that are later passed to an instance of SQL Server for parsing and execution. Any procedure that constructs SQL statements should be reviewed for injection vulnerabilities because SQL Server will execute all syntactically valid queries that it receives.” SQL Injection vulnerabilities represent about 20% of reported vulnerabilities recorded in commonly available vulnerability databases<sup>1</sup> as CWE/SANS assign the score of 93.8 to this vulnerability [8]. Therefore, safety against SQL injection is very critical to web applications. The following list introduces metrics for quantifying safety of an application against SQL injection.

- 1) Type of Inputs used on the Forms
- Name: SQLInj-001- Form Field Type

<sup>1</sup>www2.cenzic.com/downloads/Cenzic\_AppSecTrends\_Q1-Q2-2010.pdf

- **Description:** When textboxes are used to get the user inputs, they can also be used by dynamic SQL queries to boost the danger of SQL injection attacks [37]. Thus, more textboxes for inputs/outputs, more chance of possible dynamic SQL queries which leads to less safety against SQL injection. We propose the efficiency function  $f(n_1, n_2)$  as the ratio of total number of inputs to the number of textboxes as shown below.

$$f(n_1, n_2) = \begin{cases} n_1/n_2 & n_2 > 0 \\ n_1 & n_2 = 0 \end{cases} \quad (3)$$

, where  $n_1$  is total number of form fields such as textboxes, radio buttons, checkboxes, dropdown menus, etc., and  $n_2$  represents total number of textboxes in project's forms that collect and send user's data to dynamic SQL statements.

### 2) Error Presentation Mode

- **Name :** SQLInj-002-Error presentation Mode
- **Description:** due to default database management system behavior of throwing error messages, attackers can potentially expose the structure of databases and It is obvious these error messages help attackers to get a hold of the information which they are looking for (such as the database name, table name, usernames, password hashes etc.). As a mitigation strategy, a particular generic or specific error message should be used in error susceptible cases [36, 18]. To assess the potential database exposure through error messages, we define  $f(n_1, n_2)$  to measure the sufficiency of error exception handling of application as follows.

$$f(n_1, n_2) = \begin{cases} n_1/n_2 & n_2 > 0 \\ n_1 & n_2 = 0 \end{cases} \quad (4)$$

, where  $n_1$  is total number of exception handling mechanisms implemented in the project and  $n_2$  is total number of scenarios that are prone to throw default error message. As we mentioned in previous section, the more value this equation has, the error exception handling -as a mitigation strategy - has been performed the better.

### 3) Input Validation

- **Name:** SQLInj-003 - Input validation
- **Description:** The common weakness that can make an application susceptible to SQL injection is weak input validation. These inputs normally include form data, URL parameters, hidden fields, cookie data, HTTP Headers, and essentially anything in the HTTP request [32].

Constraining input for type, length, format, and range [32], filtering meta characters such as beginning of a comment character, or characters that denote the end of one query or the beginning of a SQL statement [28] are useful strategies to validate data and prevent SQL injection.

As previously mentioned, default error messages may expose the structure of database. An attacker can penetrate into the database by trying particular SQL commands. Accordingly, SQL statements that are used to retrieve or manipulate data are better to be filtered [38].

One of the most common validating strategies to increase security is recruiting range validation technic and also data validation based on matching with a proper regular expression. In other words, we ought to use the approach "Accept Known Good" instead of "Reject Known Bad" [33].

The following function evaluates the sufficiency of validation strategies in an application.

$$f(n_1, n_2) = \begin{cases} (n_1 + n_2 + n_3 + n_4 + n_5 + n_6 + n_7) / m & m > 0 \\ 0 & m = 0 \end{cases} \quad (5)$$

,where  $m$  is total number of any data input that are thrown to dynamic generated SQL statements including form data, URL parameters, hidden fields, cookie data, HTTP headers, and any piece of data in HTTP request [32],  $n_1$  is total number of any *data type* validation,  $n_2$  is total number of any *data format* validation,  $n_3$  is total number of any *data range* validation,  $n_4$  is total number of any SQL meta-characters that has been filtered,  $n_5$  is total number of any SQL commands that has been filtered,  $n_6$  is total numbers of regular expression validators, and  $n_7$  represents total number of range validators that have been recruited in application.

### 4) SQL Statement Generation Mode

- **Name:** SQLInj-004 -SQL statement generation mode
- **Description:** SQL Injection flaws are introduced by utilizing dynamic queries. In this scenario SQL statements are generated based on user's input and each user could be potentially an attacker. Therefore, implementing more dynamic database queries in an application makes it more vulnerable to SQL Injection [34]. The function  $f(n_1, n_2)$  measures the efficiency of SQL statement generation approach, by ratio of total number of SQL statements as numerator and total number of dynamical SQL statements as denominator.

$$f(n_1, n_2) = \begin{cases} n_1/n_2 & n_2 > 0 \\ n_1 & n_2 = 0 \end{cases} \quad (6)$$

,where  $n_1$  is total number of SQL statements including static and dynamic ones, and  $n_2$  is total number of SQL statements that are generated dynamically.

### 5) Efficiency of stored procedure

- **Name:** SQLInj-005-Efficient utilization of stored procedure
- **Description:** A stored procedure is a group of SQL statements that has been created and stored in the database [15]. To boost safety against SQL injection flaws, use of stored procedure is highly recommended as long as they do not include any unsafe dynamic SQL generation [11]. Not only the number of implemented stored procedure increases the application safety, but also their performance contribute to more safety. Thus, here we use the aforementioned input validation sufficiency function as well as the SQL statement generation efficiency function to measure the efficiency of implemented stored procedures. Furthermore, since more stored procedure implementation increases application safety against SQL injection, considering total number of stored procedure is also required. Therefore,  $f(a,b)$  is used to assess the competence of stored procedure engagement in software.

$$f(a,b) = \begin{cases} \sum_{i=0}^n (a_i + b_i) & n > 0 \\ 0 & n = 0 \end{cases} \quad (7)$$

, where  $a_i$  is value of "Input Validation" metric function (SQLInj-003) of the (stored procedure),  $b_i$  is value of the "SQL Statement Generation"

metric function (SQLInj-004) of the (stored procedure), and  $n$  represent total number of stored procedures.

6) Access Restriction

- *Name:* SQLInj-006 -Access restriction
- *Description:* Safety against SQL injection attacks is related to how many users access to how much of data and more exposure of data in term of number users that access it causes application to be more vulnerable.

Hence, to measure the efficiency of access restriction policy in application, it is required to define that each level of access permission is granted to how many users as the *first* variable. Each level's access permission should be specified also to reflect the accessible data through that particular level, as the *second* variable. Our assumption is that the greater value for level of access permission corresponds to higher access permission.

Therefore proposed  $f(n)$  calculates its value by multiplying abovementioned variables for each level. Then summarize the products values of all implemented levels.

Obviously, the greater value of each multiplication (and following that the summarized value) implies the less imposition of access restriction policy which has an adverse effect on safety. Since safety against SQL injection  $\propto 1/\text{data accessibility}$ , we propose  $f(n)$  as follows:

$$f(n)=1/\sum_{i=0}^m(n_i*i) \tag{8}$$

, where  $n_i$  is total number of granted accesses to level  $i$ , and  $m$  is total number of access levels which has been defined in application.

B. OS Command Injection

Briefly, applications are considered vulnerable to the OS command injection also known as shell injection attack if they utilize user input in a system level command. Shell injection attacks lead to execute risky commands on operating system through an application when the attacker does not have direct access to OS. Alternatively, it may make a number of OS restricted commands accessible for attacker when application is privileged [6]. This vulnerability mostly happens when there is an under control procedure in application which needs externally-supplied input arguments to be executed and/or when there is the possibility of getting the externally-supplied procedure or commands calls. Then entire given command has been sent directly to OS for execution [9, 24]. According to CWE/SANS findings, the score for this vulnerability is 83.3 [8].

In this section, a number of common preventative methods to raise safety of a project against shell injection compromise are introduce and then corresponding metrics are proposed to quantify the sufficiency and/or efficiency of thoes methods.

1) Function type generation

- *Name:* ShellInj-001- Function type generation
- *Description:* To boost control on input data, it is recommended to recruit library call policy, instead of using external process [9]. We quantify sufficiency of using library calls by calculating the ratio of number of library calls in the application as numerator, over total number of library call plus external processes as denominator through  $f(n_1, n_2)$ . Obviously, the greater numerator value means more safely provision.

$$f(n_1, n_2)=n_1/n_2+n_1 \tag{9}$$

, where  $n_1$  is total numbers of third-party libraries that are called to generate functions, and  $n_2$  is total number of external processes recruited to generate functions.

2) Jail Or Sandbox Utilization

- *Name:* ShellInj-002- Jail and Sandbox Utilization
- *Description:* It is recommended to enforce strict boundaries between process and operating system. This may restrict which data can be accessed or which commands can be executed by application.

However, this solution may only limits the impact to operating system but rest of the application may still be subject to compromise [9]. A possible solution for such enforcement is to utilize sandbox environment.

A sandbox is a security mechanism for separating running programs and quarantining untrusted running programs. It can be used to execute untested code or untrusted programs from unverified third-parties, suppliers, untrusted users and untrusted websites [16]. Jail sets is a common strategy of sandbox mechanism. Jail is a set of resources limits imposed on programs by operating system kernel (e.g. I/O bandwidth caps and disk quotas) [16,10]. The effectiveness of this method depends on the deterrence capabilities of the particular sandbox or jail. It may only reduce the scope of an attack, such as restricting the violator to execute certain system commands or limiting the data that can be accessed. We demonstrate  $f(a_1, a_2)$  by imposing a logical OR function on those above mentioned strategies recruitment.

$$f(a_1, a_2)= (a_1)OR(a_2) \tag{10}$$

, where  $a_1$  is "code runs in jail sets" as a boolean variable ( $a_1 \in \{0,1\}$ ), and  $a_2$  is "code runs in other forms of sandbox environment" as a boolean variable and  $a_2 \in \{0,1\}$ .

3) Input Validation

- *UniqueID:* ShellInj-003=SQLInjection.SQLInj-003 (Input Validation)
- *Description:* It is highly recommended to validate input since it has a deterrent effect for OS command injection, [21,31]. we quantify its sufficiency and efficiency in the same that has been discussed in SQLInj-003 metric in previous section.

4) Error Presentation Mode

- *Name:* ShellInj-004=SQLInjection.SQLInj-002:Error presentation Mode
- *Description:* Stephanie Reetz considered system default error messages as informative and precious data for adversaries that raises the risk of shell injection compromise, [36]. Therefore the more managed error messages is implemented in application, the risk of penetration will be declined. We quantify sufficiency of implemented error exception handling mechanisms same way that has been discussed in SQLInj-004 metric in previous section.

5) Accounts Isolation

- *Name:* ShellInj-005-Accounts isolation
- *Description:* In order to mitigate shell injection breaches, it is recommended to create role-based access control scheme with restricted privileges in order to be used only for a group of specified tasks and users. By following this strategy, a successful attack will not be accomplished because the rest of application or its environment is not accessible to attacker [40, 23].

We quantify the efficiency of implementation of this policy by means of a linear function.  $f(n_1, n_2, n_3)$  is the ratio of specified roles in

application over the total number of critical tasks plus critical resources. The more specific roles is defined (i.e greater numerator value) to access and excute respectively critical resources and tasks, leads toincrease theefferciency of account isolation policy.

$$f(n_1, n_2, n_3) = n_1 / (n_2 + n_3) \tag{11}$$

, where  $n_1$  is total number of access roles that are specified to access critical resources and execute critical procedures,  $n_2$  is total number of critical resources, and  $n_3$  represents total number of critical procedures

6) Reduction of Attack Surface

- *Name* : ShellInj-06-Attack Surface value
- *Description*: It is obvious that the more resources are available to users, the more exposed to attacks the application is [27, 12]. In [20], the authors argues that the attack surface of an application environment is the sum of the different ways that an attack action can perform through them in order to enter or extract data from an environment. They measure attack surface by means of quantifying the application's interaction with its environment through three types of recourses; entry/exit points, channels, and untrusted data items. According to them, in order to measure the attack surface, we should evaluate the probability that an adversary will use each specific resource in an attack. This evaluation is also accomplished in another work using the ratio of the potential damage-termed *Damage/effort Ratio*(DER) [14], where damage corresponds to possible technical advantages of that resource, and effort is the amount of effort needed to access that. The value for effort in this ratio can be derived from level of access rights that is needed to access that specific resource. The final measurement formula is expressed in a triplet of three DERs of three mentined resouces types includes entry/exit points, channels and untrusted data items[14]. The greater value of this triplet implies more damage for a consistent effort value, which means greater attack surface value as an application's weakness. To utilize this triplet to measure attack surface, we have to slightly modify it for two reasons. First, note that we are about to measure application's safety against OS command injection vularibility not its weakness. Second, appearantly the vector charactericity aspect of this triplet is not practical for our model. Given the above ,the triplet introduced by Gennari, J., and Garlan, D [14], is modified to a linear combination of  $1/(DER_m, DER_c$  and  $DER_i)$  values , as follows:

$$f(DER_m, DER_c, DER_i) = 1 / (DER_m + DER_c + DER_i) \tag{12}$$

, where  $DER_m$  is damage/effort ratio of entry/exit points. These points return to methods which accept or process data that are originated outside of the system and quantify as follows:

$$DER_m = \sum_{i=0}^M ((a)_i / (b)_i) \tag{13}$$

, where  $M$ : Total number of entry/exit points,  $a$  is the level of privilege associated with (Entry/Exit point) $_i$  and  $b$  is the level of rights needed to access (Entry/Exit point) $_i$

$DER_c$  is damage/effortRatio of Channels. Channels are the communication mechanisms used for system interaction with its environment, such as network or inter-process communication mechanisms. Channel damage/effort ratio is measured based on the restrictions imposed on the data that a channel can transmit via their protocols. Less restricted protocols ease compromising for attackers

since they can transmit more types of data, such as executable codes.  $DER$  ratio for channels ( $DER_c$ ) is evaluated in terms of number of data types that are restricted to transmit over channel's protocol as numerator over level of access right needed to access that channel. Hence, the larger numerator values show less restriction on data to transmit over that channel:

$$DER_c = \sum_{i=0}^C (a)_i / (b)_i \tag{14}$$

, where  $C$  is total number of channels,  $a$  is total number of data types that are restricted to transmit over (channel) $_i$  and  $b$  represents the level of rights that is required to access the (channel) $_i$

$DER_i$  is damage/effort ratio of untrusted data items. Untrusted data items are the external exited data stores that application uses them. DER for untrusted data items is measured based on restrictions are put on the data stores.

$$DER_i = \sum_{i=0}^I (a)_i / (b)_i \tag{15}$$

, where  $I$  is total number of external data stores that are utilized by application,  $a$  is total number of untrusted data items, and  $b$  is the level of rights needed to access that data items.

IV. EFFECTIVENESS COEFFICIENT NUMERIC VALUE ATTAINMENT METHOD

In this section, we explain our approach for calculation of the effectiveness coefficient of each metric in mitigating certain vulnerability. Evidently, there are always a number of mitigation strategies that have been recommended to reduce the adverse effects of common vulnerabilities. However, they do not demonstrate comparable effectiveness. For example, suppose that "Input Validation" metric is far more effective than "engaging Stored Procedures" in improving safety against SQL Injection attacks. Therefore, we should consider a greater weight for "Input Validation". One of the common methods to find this parameter is to find a compelling argument from other reliable researches which in our case was not available. Therefore, we picked the alternative method of asking expert developers to fill out questionnaires while using the method.

To answer the question about "effectiveness extent of the preventative method" there are five options on the survey to choose from. We assign  $eff \in \{0, 1, 2, 3, 4\}$  from these options as follows:

- 1)  $eff=0$  for "It has no effectiveness."
- 2)  $eff=1$  for "It has low effectiveness."
- 3)  $eff=2$  for "It is fairly effective."
- 4)  $eff=3$  for "I is highly effective."
- 5)  $eff=4$  for "It is extremely highly effective."

Moreover, as all the respondents were not equally familiar with the subject, we also included a "familiarity weight" parameter as  $fm \in \{0, 1, 2, 3, 4\}$  in calculation of "e". Similarly, we assign  $fm$  value from the survey options as follows:

- 1)  $fm=0$  for "I have never heard about it before."
- 2)  $fm=1$  for "I know this method, but never used it."
- 3)  $fm=2$  for "I rarely use this method."
- 4)  $fm=3$  for "I ferequently use this method."
- 5)  $fm=4$  for "I always use this method."

Another parameter that is involved in our calculation is respondent experience, We interpret value from their answer to correspondent question as follows:

- 1)  $exp=1$  for "Below 1 year."
- 2)  $exp=2$  for "Between 3 to 5 years."
- 3)  $exp=3$  for "More than 5 years"

Moreover, we recruit fuzzy logic to transform the mentioned obtained rational values into numerical ones-  $e \in [0, 1]$ , and exploit them in our final safety calculation formula. In this formula, the product of respondents' "experience" and "familiarity" is considered as weight. The final value is the weighted average of effectiveness values. Then in order to map the result to a number between 0 to 1, we divided the result to 4 as the maximum value, which occurs when all variables have their maximum value, and is calculated as follow.

$$m * [\text{Max}(X) * \text{Max}(W)] / [m * \text{Max}(W)] = m * (12 * 4) / (m * 12) = 4$$

Based on above discussion, the membership function for effectiveness of each metrics defined as follows:

$$(e)_{Metricn} = \left( \sum_{i=0}^m (Xn)i * (Zn)i * (Yn)i \right) / \left( \sum_{i=0}^m (Zn)i * (Yn)i \right) / 4$$

$$\rightarrow (e)_{Metricn} = \left( \sum_{i=0}^m (Xn)i * (Wn)i \right) / \sum_{i=0}^m (W)i / 4$$

(16)

, where  $X:eff \in \{0, 4\}$ ,  $Z:fm \in \{0, 4\}$ ,  $Y: exp \in \{1, 3\}$ ,  $W: exp * fm \in \{0, 12\}$ , and  $m$ : number of respondents

We examined this approach with three respondents ( $m=3$ ). Table 1 contains the generated "e" value for each metrics.

metric	"e" value
SQLinj-001	0.72
SQLinj-002	0.58
SQLinj-003	1.00
SQLinj-004	0.83
SQLinj-005	0.57
SQLinj-006	0.69
Shellinj-001	0.85
Shellinj-002	0.62
Shellinj-003	1.00
Shellinj-004	0.58
Shellinj-005	0.94
Shellinj-006	0.44

Table 1: "e" value for each metrics

V. RESULTS

To examine our model, various developers used the proposed metrics in their projects. The applications included different types of web applications and E-commerce/E-business applications. An Excel worksheet was made and presented to developers to enter the metric parameters to find the overall safety. Once each metric is calculated, the safety against SQL injection and Shell injection vulnerabilities can be found. Table 2 summarizes the detailed results for different tested applications.

Metric name	App1	App2	App3	App4	App5	App6	App7
SQLinj-001	3.00	2.00	4.43	2.75	3.00	1.44	4.00
SQLinj-002	1.00	1.00	0.79	1.25	1.00	0.73	0.93
SQLinj-003	0.00	2.00	0.71	0.00	0.00	2.20	3.00
SQLinj-004	1.00	1.92	1.86	1.25	1.35	1.93	1.30
SQLinj-005	0.00	14.08	8.67	1.00	1.20	9.29	5.67
SQLinj-006	0.00	0.01	0.00	0.25	0.40	0.37	0.04
Shellinj-001	1.00	4.00	1.00	1.00	1.25	4.60	1.23
Shellinj-002	1.00	0.00	0.00	1.50	1.00	0.30	0.02
Shellinj-003	0.00	2.00	0.71	0.00	0.28	2.20	0.83
Shellinj-004	1.00	1.92	1.86	1.25	1.50	2.50	2.87
Shellinj-005	0.57	3.56	1.09	0.60	0.32	4.50	1.09
Shellinj-006	0.13	0.06	0.13	0.25	0.38	0.04	1.23
OSAV	697.49	3439.82	1820.63	827.79	869.44	3978.27	2002.41

Table 2: Results of using metrics in different applications.

Furthermore, the examiners evaluated the usability and functionality of our formulas and metrics by means of another questionnaire. Tables 3 and 4 depict the results.

	Metrics are usable	Metrics increased safety	Definitely will use Metrics in future	Definitely will recommend Metrics to Colleagues	Easy to calculate formulas of metrics
Average score	89.00	68.00	74.71	60.43	94.43

Table 3: Average Results of usability questionnaires for formulas. The number are from 1 to 100.

	All metric's variables are necessary	Easy to find variables in application	No improvement required
Average score	82.29	80.14	64.71

Table 4: Average Results of functionality of metrics. The numbers are from 1-100.

VI. CONCLUSION

Using proper metrics in software engineering has not been very common as opposed to other engineering disciplines due to availability of such metrics. Furthermore, the need for safety and security metrics is probably the most important of all in-demand metrics in software engineering. This work is an attempt to fill out the lack of quantitative metrics in application development and software engineering. In this innovative method, new quantitative model for evaluating the safety of web applications is proposed. The metrics can quantify the overall safety of an application against known vulnerabilities. The main goal in developing this method was to provide an easy-to-use, scalable and flexible model for web application developers. In this way, they can measure the safety at different phases of development. This addresses the issue that web application security has to be looked at as an integrated factor in development and not as an add-on element. In addition to test of the method, different surveys were conducted to evaluate the usability of the formulas and metrics by developers. The feedback from web developers demonstrates that the proposed method is effective to provide a more secure application. This future work would enhance the experiments on the method in real application development.

## VII. REFERENCES

1. CWE, "About CWE", <http://cwe.mitre.org/about/index.html>, n.p., 2011, Last accessed on Sep. 25, 2013.
2. Braungarten, R., "The SMPI model: A stepwise process model to facilitate software measurement process improvement along the measurement paradigms", 2007, PhD Thesis. University of Magdeburg, Germany.
3. Brian, C., "Metrics that matter: Quantifying software security risk", Feb. 2006, Workshop on Software Security Assurance Tools, Techniques, and Metrics, NIST Special Publication 500-265.
4. British Standard Inst., "Information Security Management. Specification for Information Security Management Systems (BS7799-2)", 1999, British Standard Institute, London.
5. British Standard Institute, Information Security Management. Code of Practice for Information Security Management.(BS7799-1)", 1999, British Standard Inst., London.
6. CAPEC, "CAPEC-88: OS Command Injection", <http://capec.mitre.org/data/definitions/88.html>, n.p., June 21, 2013, Last accessed on Sep. 25, 2013.
7. Microsoft, "Create Views", <http://technet.microsoft.com/en-us/library/ms175503.aspx>, n.p., 2013, Last accessed on Sep. 25, 2013.
8. CWE, "CWE/SANS Top 25 Most Dangerous Software Errors", <http://cwe.mitre.org>, n.p., 2011, Web, Last accessed on Sep. 25, 2013.
9. CWE, "CWE-78: Improper Neutralization of Special Elements used in an OS Command (OS Command Injection)", <http://cwe.mitre.org/data/definitions/78.html#Demonstrative%20Examples>, n.p., 2011, Last accessed on Sep. 25, 2013.
10. Deborah R., Gangemi, G.T., "Computer Security Basics", chapter 3 "Computer System Security and Access Controls", 1st Edition, July 1991, O'Reilly Media, ISBN 10:0-937175-71-4.
11. OWASP, "Defense Option 2: Stored Procedures", [https://www.owasp.org/index.php/SQL\\_Injection\\_Prevention\\_Cheat\\_Sheet#Defense\\_Option\\_2:\\_Stored\\_Procedures](https://www.owasp.org/index.php/SQL_Injection_Prevention_Cheat_Sheet#Defense_Option_2:_Stored_Procedures), n.p., Dec. 6, 2012, Last accessed on Sep. 25, 2013.
12. Howard, M., "Fending off Future Attacks by Reducing Attack Surface", <http://msdn.microsoft.com/en-us/library/ms972812.aspx>, Feb. 4, 2003, Last accessed on Sep. 25, 2013.
13. McGraw, G., "Software Security: Building Security In", Feb. 2006, Addison-Wesley, ISBN: 0-321-35670-5.
14. Gennari, J., and Garlan, D., "Measuring attack surface in software architecture", 2011, Tech. Rep. CMU-ISR-11-121, Inst. for Software Research, School of Computer Science, Carnegie-Mellon University.
15. Microsoft, "How To: Protect From SQL Injection in ASP.NET", <http://msdn.microsoft.com/en-us/library/ff648339.aspx>, Last accessed on Sep. 25, 2013.
16. Goldberg, I., Wagner, D., Thomas, R., and Brewer, E., "A Secure Environment for untrusted Helper Applications (Confining the Wily Hacker)", July 1996, Proceedings of the 6<sup>th</sup> USENIX UNIX Security Symposium.
17. J. W. P. Manadhata. Measuring a system's attack surface. Technical Report CMU-CS-04-102, 2004
18. J.D. Meier, Alex Mackman, Blaine Wastell, Prashant Bansode, Andy Wigley, <http://msdn.microsoft.com/en-us/library/ff650175.aspx>, Sep 2005, Web, Access Date : Sep. 25. 2013
19. M.Howard, J.Pincus, J.M.Wing, "Measuring Relative Attack Surfaces", August 2003, Proc. Workshop Advanced Developments in Software and Systems Security
20. Manadhata, P. and Wing, J., "An Attack Surface Metric", Software Eng., IEEE Trans on, Vol:37, Issue: 3, 07 June 2010, pages: 371 – 386. Mark Dowd, John McDonald and Justin Schuh. "The Art of Software Security Assessment". Chapter 8: "Shell Metacharacters", 2006, 1st Edition. Addison Wesley, Page 425.
21. Mazinanian, D., Doroodchi, M., Hassany, M., "WDMES: A Comprehensive Measurement System for Web Application Development" 2012, Telematics and Information Systems (EATIS), 6th Euro American Conf. on, pages: 1 – 8
22. Howard, M. and LeBlanc, D., "Writing Secure Code", Nov. 30, 2009, Microsoft Press, 2nd edition, ASIN: B0043M4ZPC
23. Howard, M., LeBlanc, D., and Viega, J. "24 Deadly Sins of Software Security". "Sin 10: Command Injection." September 3, 2009, McGraw-Hill, ISBN: 0071626751, Page 171.
24. OWASP, "2010 OWASP Top 10", 2010.
25. National Vulnerability Database, "NVD Common Vulnerability Scoring System Support v2". National Institute of Standards and Technology. Last accessed on Sep. 25, 2013.
26. Manadhata, P. K. and Wing, J. M., "Measuring a System's Attack Surface," Jan. 2004, Technical Report CMU-CS-04-102, Carnegie Mellon Univ.
27. Roy, A. K. Singh, and A. S. Sairam, "Analyzing SQL Meta Characters and Preventing SQL Injection Attacks Using Meta Filter", 2011, Int'l Conf. on Information and Electronics Engineering, Singapore
28. S. R. Kumar, T. Alagarsamy K. "A Stake Holder Based Model for Software Security Metrics", 2011, International Journal of Computer Science issues, Vol. 8, Issue 2, ISSN (Online): 1694-0814, Available at: [www.IJCSI.org](http://www.IJCSI.org)
29. Jaquith, A., "Sample Questions for Finding Information Security Weaknesses", CSO, <http://www.csoonline.com/article/221202/sample-questions-for-finding-information-security-weaknesses>, May 18, 2007, Last accessed on Sep. 25, 2013.
30. SANS, "SANS Critical Vulnerability Analysis Archive", <http://www.sans.org>, n.p., March 16, 2007, Last accessed on Sep. 25, 2013.
31. OWASP, "Secure Coding Cheat Sheet", [https://www.owasp.org/index.php/Secure\\_Coding\\_Cheat\\_Sheet](https://www.owasp.org/index.php/Secure_Coding_Cheat_Sheet), n.p., April, 15, 2013, Last accessed on Sep. 25, 2013.
32. Cigital, "Security Issues in Perl Scripts", <http://www.cgisecurity.com/lib/sips.html>, Jordan Dimov, n.d., Last accessed on Sep. 25, 2013.
33. OWASP, "SQL Injection Prevention Cheat Sheet", [https://www.owasp.org/index.php/SQL\\_Injection\\_Prevention\\_Cheat\\_Sheet](https://www.owasp.org/index.php/SQL_Injection_Prevention_Cheat_Sheet), n.p., Dec. 2012, Last accessed on Sep. 25, 2013.
34. Microsoft, "SQL Injection", <http://technet.microsoft.com/en-us/library/ms161953%28v=SQL.105%29.aspx>, n.d., Last accessed on Sep. 25, 2013
35. "SQL injection", MS ISAC, <http://msisac.cisecurity.org/resources/reports/documents/SQLInjection.pdf>, Stephanie Reetz, 23 January 2013, Last accessed on Sep. 25, 2013.
36. Litwin, P., Stop SQL Injection Attacks Before They Stop You", Microsoft, <http://msdn.microsoft.com/en-us/magazine/cc163917.aspx>, 2013, Last accessed on Sep. 25, 2013.

37. Cisco, "UnderstandingSQLInjection", [http://www.cisco.com/web/about/security/intelligence/sql\\_injection.html](http://www.cisco.com/web/about/security/intelligence/sql_injection.html), n.p., n.d, Last accessed on Sep. 25, 2013.
38. Holm, H., Ekstedt, M., Sommestad, T., "Effort estimates on web application vulnerability discovery", 2013, 46th Hawaii International Conference on System Sciences.
39. Viega, J. and McGraw, G., "Building Secure Software: How to Avoid Security Problems the Right Way", 2002, Boston, Addison-Wesley
40. Martin B., Brown M., Paller A., Kriby D., Christey S., "2011 CWE/SANS Top 25 Most Dangerous Software Errors", 2011.

**SESSION**  
**BIOMETRICS AND FORENSICS II**

**Chair(s)**

**Prof. Craig Valli**  
**Edith Cowan Univ. - Australia**



# Using the concepts of 'forensic linguistics,' 'bleasure' and 'motif' to enhance multimedia forensic evidence collection

J. Bishop

Centre for Research into Online Communities and E-Learning Systems  
The European Parliament, Square de Meeus 37, 4th Floor, Brussels B-1000, Belgium

**Abstract** - *Internet trolling has become more widely adopted as a term to describe a range of data misuse and Internet abuse offences. To date there has been no coherent means to interpret online postings for the purpose of forensic collating and reporting of evidence. This paper proposes to use the terms of bleasure and motif, used in French law, in order to classify Internet trolling postings according to the extent their have harmed people (i.e. malum reus) and the extent to which it can be proved such bleasures show actus reus through treating them as motifs as one would in French law. Through investigating the posting of sex-related trolling messages sent to and relating to women on YouTube the study proposes a framework for classifying these messages. These chauvinistic messages are often related to rape, so the paper aims to help crime investigators use multimedia forensics to more easily collect and use evidence in cases of Internet trolling.*

**Keywords:** Multimedia forensics, computer forensics, forensic linguistics, e-dating, rape, tort law, criminal law.

## 1 Introduction

Internet trolling and cyberbullying are cyber security threats that do not usually form part of the formally defined scope of information security (von Solms and van Niekerk 2013, 97). Internet trolling, or Internet abuse in general, will normally be forbidden by the security policy of an information society service provider, meaning it can be interpreted as a breach of information security (Hartel, Junger, and Wieringa 2010). Misuse of information security expertise is serious business and could result in criminal prosecution, bad publicity, personal injury, cyberbullying, suicide, and termination of educational programs, among numerous other negative outcomes (Cook, Conti, and Raymond 2012, 61). Indeed, with new technology comes new demand for means to capture evidence and present it to a court hearing. That is why this paper will look at new terminology and approaches for dealing with Internet-based evidence as well as other electronic evidence. When campaigner Caroline Criado-Perez experienced threats of sexual violence on Twitter, discussed in depth in this paper, the outraged public looked first not to the police, but to the company, and instead of asking for more robust policing, thousands signed a petition asking Twitter for a "report abuse" button (Powell 2013). This could be seen as an improvement over a time when people turned mainly to the government to solve their problems rather than themselves and others (Baum 1993, 31). What is also evident from this is that people do not

trust the police to look after their interests and bring to justice the guilty. This paper will therefore consider the law of evidence not just in relation to criminal law, but concurrently with civil law. Calls to fuse the two forms together (Bishop 2014b, 154; Mugabi and Bishop 2014) may be a long way off being realized. However, as this paper shows, there are a lot of legal principles and precedents that can be used in either a civil or criminal court based on universal elements to the law of evidence.

### 1.1 Bleasure and motif

There is an increasing use in the concepts of 'bleasure' and 'motif' for understanding communications offences and collecting forensic evidence (Bishop 2014a). Bleasure is a term derived from French law and introduced into UK law through following *King v Bristow Helicopters Ltd* [2002] 2002 Scot (D) 3/3, which refers to an imposed injury, whether physical or mental, that has an sustained adverse impact on someone in either the short-term or long-term (Bishop 2014c, 1). The term 'Motif' is also from French law, and could be thought of as a 'smoking gun' – the verifiable proof that a particular act occurred. In terms of Internet trolling, this could be a post on a social networking service that caused a bleasure, such as a person being grossly offended. Court opinions are beginning to surface relating to the admissibility of evidence obtained from social media (Abilmouna 2012, 99), and it is likely to become more and more of an issue, particularly relating to the admissibility of character evidence.

The fact that one has experienced a bleasure does not automatically mean that a fault has occurred. In criminal cases based on public law there is often a high burden of proof to pass for a fault to have been proven to have occurred. In civil cases, such as those based on tort and contract the burden of proof is lower, but that does not automatically mean each bleasure can be considered a fault. It equally does not mean that if a fault has occurred that the person was in fact injured, or rather bleasured. It is a well-rehearsed principle in law that there is no middle ground between civil law and public law, as a specific fault can usually only be heard under either one or the other. But equally this does not mean that there is no common ground between them, even if there are separate procedures for each.

### 1.2 Evidence admissibility

In considering the demeanour of a witness in court, (Stone 1991, 829) argues that many witnesses are intimidated by

testifying publicly in court, as some fear humiliation by a cross-examiner who may attack their powers of observation, memory, or integrity. He argues that such anxiety may be intensified by the number of people in court or the personalities of advocates or the judge. This makes it clear that any computer generated materials used as evidence would have to be generated under clinical conditions outside of the court. Some technologies are capable of being used in court to let counsel know they are upsetting a witness (Bishop 2011), but would be better employed outside of court when being used to understand the character or demeanour of witnesses. Some systems have been successfully shown to be able to detect trauma from victims and perpetrators of sex offences using EEG (Bishop 2012), which could also be used in court. Such evidence showing brain patterns, which can also include MRI scans, as well as the emotions and thoughts they represent, could be submitted as either documentary or video evidence in a court. One might question whether such evidence is admissible when it is generated with computer algorithms and only presented visually. There is in fact already a precedent on this, which would apply to affective computing derived evidence as much as any other form, including print outs. The case of *Sapporo Maru (Owners) v Statue of Liberty (Owners)*, *The Statue of Liberty* [1968] 1 W.L.R. 739 found that there was no distinction to be made between a photograph taken of an mechanical or related event and one taken of a more natural occurrence. The example given was that a photograph of physical rain is as valid as a photo taken of a barometer. Equally, *R v Wood* (1983) 76 Cr. App. R. 23 said that computer generated evidence is valid if it is used to measure external phenomena and not its own computable knowledge. On this basis, one could strongly argue that capturing of human activity electronically can be used as evidence in court. This could include posts to social media as well as emotion capture devices.

### 1.3 Considering actus reus, malum reus, and pertinax reus

One thing is certain in relation to the application of the law in the digital age, and that is that the requirement to prove mens rea is diminishing (Bishop 2010, 299; Corlett 2013, 9). The case of *DPP v Chambers* clearly shows that in the case of communications offenses carried out over the Internet, it is not necessary to prove that a person intended to perform a particular act. Equally under the law of torts it is not necessary to prove that someone intended to perform a particular act - simply that it was reasonably foreseeable that they would. It might therefore be necessary to create general principles to determine that regardless of whether something should be heard under civil or public law, which forms of evidence are best used to prove the degree to which someone has been bleasured.

It could be argued that in relation to any bleasure, whether public or civil, there are three certainties that are required. These are whether a particular act by a particular person was what caused them to be bleasured; whether it was that particular bleasure that caused the detriment to which the person seeks a remedy; and indeed, whether it is likely the

person seeking remedies will be subject to further bleasures in the future through the actions of the same person who made them subject to the one being considered by the court.

These certainties can be found in the emerging rules of *actus reus*, *malum reus* and *pertinax reus*. These can be best understood in relation to computer-related crime through considering cases around Internet trolling. The three main cases are *DPP v Collins*, *DPP v Connolly* and *DPP v Chambers*. These cases considered the definitions of grossly offensive in the case of the first two and the definition of threatening in the case of the last. Regardless of the specific offenses, this section will consider the faults committed in the cases from the point of view they are bleasures, as opposed to public law offences as distinct from the civil law equivalents.

*Actus reus* is a term used to reflect the certainty that a person actually committed the fault of which they are being accused. Whilst this expression is not commonly used in civil law cases it is clear that in order for it to be proven someone suffered a loss that the actions that led to the fault are proven. *Malum reus* is a term that refers to the concept that the fault the person is being accused of actually caused harm to another person (Bishop 2013a, 301). *Pertinax reus* refers to the certainty that the fault the accused carried out that caused harm to another was one which was not out of the ordinary for them (Bishop 2013a, 301).

## 2 An investigation into 'rape' and 'misogyny' discourses online

It is clear from research into motivations behind Internet trolling behaviour that there are sadist and anti-social behavioural reasons behind why someone might want to abuse another online (Bishop 2013b, 28; Buckel, Trapnell, and Paulhus 2014). The period between the 1960s to the 1980s was focused on the link between television news and social and economic power (O'Malley 2010, 519; Wayne et al 2008, 75), and it would seem the current time is focused with how power is shifting from the government and corporations to the people.

### 2.1 Documents and Participants

A multimedia forensic linguistic approach was used in the study to analyse the narratives of acts of Internet trolling (i.e. Motifs) and how they could be seen to harm others (i.e. Bleasures). Understanding the narrative context of evidence is known to be important for establishing the truth (Robins 1995, 201). An annotation coding framework was compiled from extracting replies to commentary by major women associated either with trolling or other controversy in 2013. These included Sally Bercow who was found to have libeled a former Conservative politician, as well as Caroline Craido Perez, who is a feminist activist who called for more women to be on British bank notes. Also considered was Esther McVay, a government minister who had a past of posing in sexually provocative ways, as well as Salma Yaqoob, who is a muslim who has spoken in favour of British troops and men's rights. Some relevant persons were excluded from the investigation. These include Stella Creasy who is a politician that was abused the same time as Caroline Craido Perez, but would simply be

unreliable duplication of her – Creasy as a politician is expected to suffer some abuse as a politician. Mary Beard and Coleen Nolan were both excluded because both received duplicate messages of a bomb-threat sent to all the women in this paragraph (except Salma Yaqoob). With regards to Mary Beard, the facts around her trolling were very similar to Caroline Criado Perez and Salma Yaqoob, so would likely offer no new information.

**2.2 Methodology**

The methodology selected for this study is based on a corpus linguistics approach for extracting Internet trolling messages from posting to major social media platforms and analysed with a qualitative data analysis approach called ‘Framework.’

There is a slight variation to the way a Framework analysis is usually conducted due to limited space, and the fact that only a total of four women were needed. The comments cut across three web-based community genres for 6 types of posting, much of what would appear in the tables, namely the elements/dimensions and categories/classes identified are discussed in the body of the paper, even though the wide ranging ‘data charted’ is contained in the tables and cross-referenced against the person to whom it relates and the type of posting that it reflects, with there being a separate table for each genre.

The study involves the investigation of posts relating to four women who were prominent in the news between 2012 and 2013, where there was a sex-related issue affecting their public persona. Sally Bercow was a political commentator who was found to have defamed a former Conservative Party activist by suggesting he committed sex offences against children. Caroline Criado-Perez was a feminist campaigner subjected to chauvinistic comments and rape threats following winning a campaign to have fewer men on bank notes in favour of more women, namely Jane Austin, who will appear on the new £10 note in Great Britain. Esther McVey, at the time of writing, was a Conservative employment minister who was exposed by the Daily Mail for her past as a model, where she posed for “racy” photographs. Salma Yaqoob made comments in favour of increasing the understanding of male victims of rape, as opposed to the over-representation of women as victims of sex offences. Comments were selected by searching for the name of the person to whom they referred and selecting by judgment those which reflected an attitude towards that person. These were then reduced according to which expressed an opinion connected to a protected characteristic, such as their sex.

**2.3 Results**

The YouTube and weblog pages allowed for the abstraction of content based on the definitions given to the different types of posting described in (Bishop 2013c, 106). These were done for each of the four women and the posts finally selected were those that most accurately matched the definition of the posting type and those which were mostly directed at the four women and of a sex or sexual nature. The terms used to codify the data were interface cues connected directly with online communities and gamification (Kim 2011, ; Kim and Sundar 2011, 599).

**2.3.1 Snacking**

Users who perform snacking offer short bursts of content and consume a lot too (Bishop 2013c, 106). Those trolls taking part in snacking will often make references to their own experiences in the real world, or other events and anecdotes that could be considered relevant to the discussion at hand. It makes no difference whether these are kudos or flames, the nature of them are the same, as the study found and is explained below Table 1.

Table 1 Snacking

Woman	Examples of trolling
Sally Bercow	Sail1948: “The truth re Mcpaedophine will come out after his death. Why did he choose not to sue” Sallywag magazine when they outed him in the mid 1990's?” Learnmore: “For most up-to-date news you have to pay a visit the web and on world-wide-web I found this website as a best web page for newest updates.”
Caroline Criado-Perez	HymenDestroyer: “Frankly I would rather live in a world with minor inconveniences such as "trolls" on websites that have mute, block and ignore features than live in a world where I'm too scared to post anything online because it might be considered abuse.” Graham67626: “Whiny bitch, Twitter doesn't owe you shit.”
Esther Mcvay	Tommy: “Better than that awful Bercow woman !” AE: “This is a self made lady who has had a successful career in business and television before entering politics who should be proud of her achievements and proud of these photographs which are perfectly acceptable. DM A typical no news Saturday so we will create a story”
Salma Yaqoob	QKP2006: “Good on you Salma, they tried to silence you and cast you as hysterical... they succeeded only in humiliating themselves and demonstrating their ignorant bias. You spoke eloquently, with knowledge and intelligence. Well done, keep up the wonderful work!”

The interface cue of ‘group identity’ is very prominent in the snacking posts in Table 1. The reference to Caroline Criado-Perez as being a ‘bitch’ is clearly an indication by the author of dissatisfaction with her belonging to a group of unideal women. As a term, ‘bitch’ generally refers to “*women who ‘aggressively’ act on their own desires in assuming a dominant role in their relations with men*” (Superson 2001, 419). In other words a “bitch” in this context can be seen to be a woman that uses her sex to gain an unfair advantage over others. One might see why this attitude is also reflective in the posts about Caroline Criado-Perez as she was in the minds of many advocating a policy to have fewer men on bank notes solely because they are men. It would appear therefore than women

who ask for more rights for women are prone to attacks from trolls using language connected with the group identity of being a woman.

The interface cues of ‘love’ and ‘fun’ are also easily identifiable in the posts. In this context love means the heart symbol (i.e. ‘♥’), which represents giving something kudos and not the romantic connotations the word is usually subject to (Kim 2011). References to Esther McVey as being “better than that Bercow woman” shows the troller has knowledge outside of the immediate subject favourable to McVey. It also shows a preference for one person over another, in a provocative way. The kudos (i.e. love) shown towards Salma Yaqoob was significant. The comments made by Yaqoob were in support of groups not typically associated with the stereotype of prominent muslim women. For instance she spoke in favour of men victims of sex abuse and in support of British Afghan soldiers in the way one would expect an in-your-face feminist or muslim not to.

**2.3.2 Mobiling**

Mobiling is where users use emotions to either become closer to others or distance from them (Bishop 2013c, 106). The term refers to the use of mobile phones where it is easier to hide one’s actions from others and more immediately post messages that one may later regret. Like with the other types of posting, mobiling can take the form of being either supportive or not so, although it is always dependent on an appeal to emotion. Table 2 shows examples of mobiling based on those posts directed at the four women selected.

Table 2 Mobiling

Woman	Examples of mobiling
Sally Bercow	Peter: “Sally Bercow has hired excellent counsel in William MacCormack QC; she has offered to settle, she has pleaded her innocence; [...] McAlpine is making a big fuss over nothing and he is acting like a *****_** ***** ... [*innocent face*]”
Caroline Criado-Perez	Karaner Karan: “bitch is fucking retarded. you only need the kitchen bitch.”
Esther Mcvay	Landless Peasant: “I could Photoshop a bag ov er her head.”
Salma Yaqoob	bengali289: “Salma why do u give a crap about british troops they are killing muslim u fool!”

One of the key elements of mobiling is the way much of it is based on the interface cues of punishment and rewards, which form an important part of theories around gamification and seduction (Bishop 2013c, 106). The posts identified in Table 2 clearly show that emotional appeals can be both supportive of the person they are directed at or inflammatory. The one most obvious in this regard was the one sent about Esther McVey by Landless Peasant. The comment was clearly meant to be humourous, even though it suggested a punishment of

suffocating McVey. Considering DPP v Chambers there was clearly no credible threat in existence to form a motif, or expect a bleasure to have occurred. Another poster, bengali289, called Yaqoob a “fool,” yet the message suggested the person agreed with the opinion that could be expected of Yaqoob if she was true only to her own self-interests. In the post about Sally Bercow in Table 2, it is clear that the troller, Peter, wished for Lord McAlpine to be punished for making a “big fuss” over the posting of the ‘innocent face’ comment by Bercow. Their reference to Bercow wanting to “settle” suggests the troller is more likely to reward her for the comment.

Turning Caroline Criado-Perez, the comment that she is a “fucking retarded” “bitch” that should “only need the kitchen,” is a clearly chauvinistic comment, but the aggressiveness of it suggests something much deeper. As stated earlier, the term “bitch” is a reference to a woman who uses her status as a woman to try to gain an advantage over others, typically men (Superson 2001, 419). For such emotional comments to be evoked would suggest a total disgust with Criado-Perez’s comments.

It has already been established that Internet trolls post flames and other abusive content to memorial websites because they dislike the insincerity of people who never knew the deceased jumping on the bandwagon (Phillips 2011, ; Walter et al 2011, 12). It is very much likely the same mind-set exists in the abuse that was directed at Criado-Perez, especially as one of the people who was convicted for trolling her was a woman.

Caroline Criado-Perez’s original comments on the BBC World News Discussion Panel might provide some explanation. “*If you want to be appreciated for what you’ve done and recognized publically you had better be a white man.*” she said. “*There are so many great women who have been suggested to us since starting the campaign (to have more women on bank notes), which most people haven’t heard of. Even I, a big feminist campaigner, who has been banging on about it haven’t heard about these women.*”

It could be seen in this context that Criado-Perez is trying to pursue an agenda of making women dominant over men, who might appear on bank notes, for the only reason that they are women, meaning she meets the definition of a “bitch,” as the Twitter user Karaner Karan called her. The presenter of the show challenged Criado-Perez that if women were singled out as deserving special attention on bank notes, then why not other protected characteristics, such as whether a person has a disability. This shows the difficulties with forms of affirmative action that rely on treating a person more favourably because of a protected characteristic that in no way relates to their merits beyond holding a status such as being a woman. The fact that Criado-Perez was a woman seeking to further the rights of women, which in this circumstance would reduce the rights of men, might suggest why trolls – who tend to be against bias, hypocrisy and insincerity – would troll her so aggressively.

**2.3.3 Trolling**

Trolling as a more generic pursuit that seeks to provoke others, sometimes affecting their kudos-points with other users

(Balaban-Salı and Şimşek 2013, ; Bishop 2013c, 106; Thacker and Griffiths 2012a, 17; Thacker and Griffiths 2012b, 17).

Table 3 Trolling

Woman	Examples of trolling
Sally Bercow	Harry: “*innocent face*” hoppinonabronzeleg: “But why so litigious. Has he something to hide. Clearly a lot of web users seem to think so. Remember Jeffrey Archer suing the Star?”
Caroline Criado-Perez	Joebstarsurfer: “Caroline Criado-Perez is a freelance journalist, When a story seems abit odd i like to look at who the people are.Its another twit set again.Gawd.”
Esther Mcvay	TrueBlogge777: “I suppose she had other career options to be fair.”
Salma Yaqoob	IScouserNProud1: “Piers morgan couldnt lick much more arse if he tried, what a twat!”

The cognitive basis on which trolling exists relies very much on extrinsic forms of motivator, often to confirm internal mental states. Levels in video games often show how much one has progressed, which are used in real life, such as progressing through education. Joebstarsurfer’s reference to Caroline Criado-Perez’s career as a ‘freelance’ journalist and in pointing this out suggests that she is not at the appropriate level to be of worth in their mind. The comments in support of Salma Yaqoob which were against Piers Morgan can be seen to be in this category as the person was trying to bring Morgan ‘down a peg or two.’ Suggestions were made by Harry that the person suing Sally Bercow, Lord McAlpine, was affected by numerical factors, such as there were a lots of posts on the Internet saying the opposite to him, and that pursuing countless lawsuits was not appropriate. This clearly shows the role points play in online activities (Kim 2011, ; Kim and Sundar 2011, 599). The reference to Esther McVey as having other “career options” could be considered to refer to the interface cue of ‘learning,’ which is common in online environments (Kim 2011, ; Kim and Sundar 2011, 599).

**2.3.4 Flooding**

Flooding is where trolls get heavily involved with other users by intensive posting that aims to counteract the challenge to their rights or wish to express them (Bishop 2013c, 106). Table 4 presents posts that use flooding, based on the four women selected by the study. Most typical of flooding is that a troller will post the same or similar content to a number of websites in order to ensure their message gets across. Flooding often follows a circumstance where the person has been denied a right to which they believe they were entitled and seek to ensure they enjoy restitution for that by making others fully aware of the situation. The main interface cues indicating flooding is the attempt to assert power and often mastery over a subject, such as where one might not get the recognition one deserves (Kim 2011, ; Kim and Sundar 2011, 599).

Table 4 Flooding

Woman	Examples of flooding
Sally Bercow	Loverat: (Blog A, 1): “What tips the balance clearly against RMPI is the alleged £50K demand from Sally Bercow” (2) “I presume from what you say that you accept Bercow libelled McAlpine-and the point at issue is damages.” (3) “Well, we all know the outcome so probably not alot of point in adding further comment now Sally Bercow has decided to give up.” (Blog B, 1): “Excellent article. Carter Ruck might also take note of this.” (2) “You cannot demand such sums from individuals as though their contribution was the sole or main cause of the damage.”
Caroline Criado-Perez	AWResistance: “Feminism = Socialism Feminism = Collectivism Feminism = Statism Feminism = Delusion Feminism = The opposite of Femininity.” AWResistance: “Patronising and arrogant middle class white females with a superiority complex filling their dull lives with a mission to fuck up common sense and natural behaviour by creating an illusory boogeyman (patriarchy, da evil men) and using it as an excuse to spread their delusional fruitcakery onto society through the use of government.”
Esther Mcvay	Obi Wan Kenobi (Blog A, 1) “Oh I say, minister! The photo shoot rising Tory star Esther McVey might rather forget.” (2) “And was probably refused entry to TV X as she had political ambitions in 1999 – however I don’t see the diffrence4 between TV X and the DWP – They both stand for fuck and suck!” (2) “That would be classed as cruelty to bags!”
Salma Yaqoob	craigowler: “Yep...I watched this programme...my heart went out to Salma for the way she was treated by Dumbleby and the perverted panellists.” craigowler: “Are you aware that without highly inflated bonuses, salaries & recruitment & retention inducements / US Marine Corps would be unable to operate in Iraq & Afghanistan. They rely on unemployed/penniless working class Americans to fill the ranks.”

As can be seen in Table 4 the various trolls posted content over a variety of blogs to get their message across and in many cases multiple times on those blogs. Loverat, whose interest was Sally Bercow, posted comments that were both annoyed with Bercow for giving up and letting Lord McAlpine get damages out of her when she was not the only person involved, who was also targeted. AWResistance’s comments about Caroline Criado-Perez were mainly abusive, but clearly showed a disgust for the power she was trying to assert over men as evidence by their comment about “*patronising and arrogant middle class white females with a superiority*

complex.” In terms of the ‘mastery’ interface cue these were evident in the case of craigowler, who commented in relation to Salma Yaqoob..

**2.3.5 Spamming**

Spamming, often associated with unsolicited mail, is in general the practice of making available ones creative works or changing others to increase the success of meetings one’s goals.

Table 5 Spamming

Woman	Examples of spamming
Sally Bercow	Zarathustra: “The wife of the Commons speaker is not normally someone who I’d go out of my way to admire. When she appeared on Celebrity Big Brother she struck me as somewhat vain and publicityseeking. She’s also a former member of the Oxford University Conservative Association, a group that I’ve been gleefully sarcastic (http://notsobigsociety.wordpress.com/2012/08/10/tv-review-youngbright-and-on-the-right/) about in the past.”
Caroline Criado-Perez	HaggisHunter154: “Remember that female mp on twitter with her racist "voice" but thats ok because shes a woman and black”
Esther Mcvay	Derek Tucker: (1) “when she becomes disabled might understand and she will be disabled just before she passes away.” (2) “likes to be not very nice”
Salma Yaqoob	123hunkyhunk 3: “As a British Muslim myself, I agree with Salma to a large degree. The Wootten Basset march is a very irresponsible and poorly thought out idea by Al-muhajiround.”

In terms of spamming, which can include the mass posting of comments for personal gain, can be seen to link mostly to the interface cues of leader-boards and badges (Kim 2011, ; Kim and Sundar 2011, 599). As can be seen from Table 5 there is strong evidence of these existing in the messages about the four women selected. In the case of Salma Yaqoob, 123hunkyhunk refers to being a ‘British Muslim’ as a badge of honour as a reason for supporting Yaqoob’s comments in support of the British military who were engaged in wars in muslim countries. This comment by 123hunkyhunk is spamming as it can be seen to be self-promotion. The same was the case in relation to Esther McVey. In the discussion about the “racy” images of McVey, Derek Tucker, made off-topic comments about her not understanding disabled people, as her occupation was that of a disability minister presiding over cuts to disability benefits. Other types of ‘badge’ are sometimes less obvious. In terms of the troller, Zarathustra, they suggest that they are significant as they would “not normally” “admire” someone like Sally Bercow, including because she was a member of the Conservative Association at Oxford University. In terms of Caroline Criado-Perez the troller, HaggisHunter154, sought to

be critical of their perceived bias in society in favour of women. Their comment shows clear disgust that in their view women like Caroline Criado-Perez and Diane Abbott can get away with bigoted comments, yet others are unable to, as their badges as women and in the case of Diane Abbott being one of a few Black Women MPs give them an unfair advantage over other groups, which include the average “white man,” which Criado-Perez made reference to. Such opportunism has been criticized recently by veteran politician, Anne Widdecome, who believes those women who seek to both be in the top jobs without taking all that comes with it have defeated the feminist cause.

**3 Towards a model for linking trolling magnitude to sexual abnormalities**

Table 6 presents a synergy of the findings of this paper in addition to other works (Bishop 2012, ; Freud and Freud 2005, ; Power 2003, 379), which are discussed throughout this paper. The first column links the degrees of rape identified with the trolling magnitude scale (Bishop 2013b, 28). The higher the TM then the more effort will be needed to achieve the sexual assault by the perpetrator. The higher the degree of rape, then less is needed in terms of the burden of proof in terms of standard of evidence. In the second column the motivations for seeking sexual relations are displayed along with the associated psychosexual stage in psychoanalysis (Freud and Freud 2005). Much discourse on the gauging of gravity discussed rely on understanding the verbal utterances and internal dialogues of those committing the acts – such as by denying committing an offense they know they have. This may take the form of verbal-textual hostility (VTH), where the person directs their language to the person that abused them and also towards a person they have abused (Asquith 2013). The harm caused by VTH can be enhanced by concurrent violence and no matter the words used, physical and sexual assault causes physical, psychological and social harm well beyond the original incident (Asquith 2013), which can all be seen as bleasures.

Romantic relationships are nearly always motivated by a mutual assumption by one party that the other is perfect in some way. In those at the Genital Stage it is that the other person is the only one in the world for them and at the Latent Stage that the other person will give the most perfect experience not possible elsewhere. In terms of the ‘exploitation’ motivation is for power of a person to try to gain a degree of control not possible elsewhere, such as in a person’s mind. In terms of ‘chivalry’ the respondent will have sought to make the claimant feel helped or provided with a service. It is associated with the Phallic stage, which was created to resemble the sense of power a man has felt over the ages in relation to this. In addition to these, Table 6 also helps to link concepts relating to the types of love that can lead to rape, the belief types that help to support those relationships and what the effects of fixation on the psychosexual stage will result in. Also described in Table 6 are the types of ‘proxy act,’ where are where another person is used in order to gain sexual advantage over another. For instance a person could get a feeling of control over someone by making a vexatious complaint to police authorities

and then getting a sense of sexual satisfaction when the person they complained about are denied liberties and other human rights as a result of this.

Table 6 Associations between trolling magnitude and degrees of rape for understanding sex offences

<b>Trolling Magnitude</b>	<b>Motivation (Stage) [Love Type] {Belief type}</b>	<b>Fixation (Chatroom Bob Type) [Dimensions]</b>
TM1 - Second degree rape (Contact is by chance then mutually escalated)	Perfection (Genital) [Commitment] {Mutual security}	Mature sexual interest (Relation) [Perfected]
TM4 - Fist degree rape (Contact is fast sex talk and action)	Exploitation (Oral) [Passion] (Exchange compliance)	Mouth (Hyper-Sexualised) [Violation]
TM3 - Third degree rape (Contact involves offer of help or service)	Chivalry (Phallic) Intimacy (Dehumanised as object)	Genitalia (Transaction) [Adaptable]
TM4 - First degree rape (Contact is fast sex talk and action)	Exploitation (Anal) Passion	Bowel / Bladder (Violation) [Vitalised]
TM2 - Second degree rape (Contact is tailored escalation)	Perfection (Latent) Commitment (Friendship and love)	Dormant sexual feelings (Relation) [Distorted]

“TM1” is where both parties think they know what they and the other party want without believing one is dominating the other. An action by one party is responded to with a reciprocal gesture from another. Sexual acts are a means of showing trust and the aim of the relationship is to share commonalities. The cost is the time spent with other persons. In terms of proxy acts, another person might be used by the respondent to enhance relationships between them and the claimant, with both holding these roles at different times.

“TM2” is where the actions of the respondent are not reciprocated by the claimant who gave no sign or wanting the sexual act forced on them. The respondent is likely to know the claimant and assaulting them will give a feeling of importance

through a sense of power, motivated because of a lack of appreciation of them by the claimant or others. The cost is the ability of the claimant to have appropriate relationships with others. In proxy acts the respondent will use another person to re-enact the abuse, such as by lying when on trial to ‘rub it in’  
 TM3: This is where the actions of the respondent are met with reciprocal gestures by the claimant due to the nature of the former’s dominance. Sexual acts are a means for the respondent to feel important and in the case of the claimant it is to feel appreciated. The cost is the claimant not getting appropriate support. In terms of proxy acts the sexual gratification comes from the respondent using another person to impose a form of assault on the claimant.

TM3: The actions of the respondent are reciprocated by the claimant but the motivations are different. Sexual acts are a means for the respondent to offset sexual unfulfillment in other relationships, giving them a sense of importance or appreciation that they have lacked. In the case of the claimant, sexual acts are a means to feel loved by someone with whom they share a close connection. The cost is a lack of trust between those who might ‘find out’ and ‘disapprove’ of the actions of the parties. In terms of proxy acts, another person might be used by the respondent to enhance relationships between them and the claimant, with both holding these roles at different times.

TM4: The actions of the respondent are not reciprocated by the claimant who gave no sign or wanting the sexual act forced on them. The respondent is likely to know the claimant and assaulting them will give a feeling of importance through a sense of power, motivated because of a lack of appreciation of them by the claimant or others. The cost is the ability of the claimant to have appropriate relationships with others. In proxy acts the respondent will use another person to re-enact the abuse, such as by lying when on trial to ‘rub it in’

## 4 Discussion

Information security policies have faced a challenge with the increase in cyberbullying and Internet trolling. Also challenging is how to collect evidence to deal with claims of such abuse. This paper has considered multimedia forensics as distinct from computer forensics. Also considered heavily is the role of interface cues in forensic linguistics for the purpose of understanding and supporting evidence collected in relation to issues around sex and sexism as they apply to women. It might be possible to see that the collection of evidence relating to offences committed via computers plus an analysis of it through forensic linguistics is what makes something multimedia evidence. The field of multimedia studies relies on techniques for analyzing digital media-texts. Without forensic linguistics, one might argue that multimedia evidence on its own is as limited as computer evidence, such as that which forms part of computer forensics. The paper has shown that it is possible through using interface cues and standardized posting types to abstract motifs from postings on the Internet as evidence of pleasures. It is quite clear that the reason chauvinistic

comments are used against feminists calling for more rights for women is because their comments appear as being self-interested, biased, hypocritical and similar. One can therefore see that where someone with a protected characteristic calls for a new right for those sharing it to the detriment of those without it, then it can be expected they will receive lots of offensive motifs that could leave them feeling bleasured. A case in point is Lenny Henry, the Afro-Caribbean comic who claimed there were not enough ethnic minorities on television and when a politician criticized Henry, saying that if he wanted to be around more Black people that the best option was to go to a “Black country,” it was the politician that was criticized by the mass media and not Henry, who was calling for ‘positive discrimination’ on the grounds of race, which is illegal in the UK. One can therefore see that in today’s age it is no longer admirable to speak up for people like oneself and in fact quite the opposite. If Caroline Criado-Perez or any other feminist posts on the Internet remarks which could be considered misandrist then they should expect to be held to account for them. Legal jurisdictions might want to strengthen laws on incitement so that where the gynocentrism of women like Caroline Criado-Perez results in expressions of misandry then this is considered as serious as posting messages that are aggravated by racial, religious, disabilist, or homophobic factors, as provided for in the Serious Crimes Act 2007.

## References

- Abilmouna, R. 2012. Social Networking Sites: What an Entangled Web we Weave. *Western State University Law Review* 39:99-270.
- Asquith, Nicole. 2013. *The Role of Verbal-Textual Hostility in Hate Crime Regulation*. London, GB: London Metropolitan Police.
- Balaban-Salı, J. and Şimşek, E. 2013. Abnormality in Virtual Worlds. Paper presented at Proceedings of the International Conference on Communication, Media, Technology and Design. 02-04 May 2013. .
- Baum, Frances E. 1993. Healthy Cities and Change: Social Movement Or Bureaucratic Tool? *Health promotion international* 8:31-40.
- Bishop, Jonathan. 2014a. ‘YouTube if You Want to, the Lady’s Not for Blogging’: Using ‘bleasures’ and ‘motifs’ to Support Multimedia Forensic Analyses of Harassment by Social Media. Paper presented at Oxford Cyber Harassment Research Symposium. Oxford, GB. 27-28 March 2014. .
- . 2014b. Internet Trolling and the 2011 UK Riots: The Need for a Dualist Reform of the Constitutional, Administrative and Security Frameworks in Great Britain. *European Journal of Law Reform* 16:154-67.
- . 2014c. My Click is My Bond: The Role of Contracts, Social Proof, and Gamification for Sysops to Reduce Pseudo-Activism and Internet Trolling. Pages 1-6 in *Gamification for Human Factors Integration: Social, Educational, and Psychological Issues*. Edited by Jonathan Bishop. Hershey, PA: IGI Global.
- . 2013a. The Art of Trolling Law Enforcement: A Review and Model for Implementing ‘Flame Trolling’ Legislation Enacted in Great Britain (1981–2012). *International Review of Law, Computers & Technology* 27:301-18.
- . 2013b. The Effect of Deindividuation of the Internet Troller on Criminal Procedure Implementation: An Interview with a Hater. *International Journal of Cyber Criminology* 7:28-48.
- . 2013c. The Psychology of Trolling and Lurking: The Role of Defriending and Gamification for Increasing Participation in Online Communities using Seductive Narratives. Pages 106-123 in *Examining the Concepts, Issues and Implications of Internet Trolling*. Edited by Jonathan Bishop. Hershey, PA: IGI Global.
- . 2012. Taming the Chatroom Bob: The Role of Brain-Computer Interfaces that Manipulate Prefrontal Cortex Optimization for Increasing Participation of Victims of Traumatic Sex and Other Abuse Online. Paper presented at Proceedings of the 13th International Conference on Bioinformatics and Computational Biology (BIOCOMP’12). USA. 16-19 July 2012.
- . 2011. *Assisting Human Interaction*. Vol. PCT/GB2011/050814 GB: PCT/GB2011/050814.
- . 2010. Tough on Data Misuse, Tough on the Causes of Data Misuse: A Review of New Labour’s Approach to Information Security and Regulating the Misuse of Digital Information (1997–2010). *International Review of Law, Computers & Technology* 24:299-303.
- Buckel, Erin E., Paul D. Trapnell and Delroy L. Paulhus. 2014. Trolls just Want to have Fun. *Personality and Individual Differences* .
- Cook, Thomas, Conti, Gregory and Raymond, David. 2012. When Good Ninjas Turn Bad: Preventing Your Students from Becoming the Threat. Paper presented at Proc. 16th Colloquium for Information System Security Education. .
- Corlett, J. A. 2013. The Problem of Responsibility. Pages 9-23 in *Responsibility and Punishment*. Springer.

- Freud, S. and A. Freud. 2005. *The Essentials of Psycho-Analysis*. London, GB: Vintage Classics.
- Hartel, PH, Marianne Junger and RJ Wieringa. 2010. Cyber-Crime Science= Crime Science Information Security. .
- Kim, Amy J. 2011. Do You Know the Score on Gamification? Paper presented at The Gamification Summit 2010. San Francisco, CA. 12 January 2011. .
- Kim, H. S. and Sundar, S. S. 2011. Using Interface Cues in Online Health Community Boards to Change Impressions and Encourage User Contribution. Paper presented at Proceedings of the 2011 annual conference on Human factors in computing systems. .
- Mugabi, Ivan and Jonathan Bishop. 2014. The Need for a Dualist Application of Public and Private Law in Great Britain Following the use of 'Flame Trolling' during the 2011 UK Riots: A Review and Model. *Transforming Politics and Policy in the Digital Age*. Edited by Jonathan Bishop. Hershey, PA: IGI Global.
- O'Malley, Tom. 2010. Book Review: Television News, Politics and Young People: Generation Disconnected? *Journal of British Cinema and Television* 7:519-21.
- Phillips, Whitney. 2011. LOLing at Tragedy: Facebook Trolls, Memorial Pages and Resistance to Grief Online. *First Monday* 16:.
- Powell, Alison. 2013. Book Review: Regulating Code: Good Governance and Better Regulation in the Information Age. *LSE Review of Books* .
- Power, Helen. 2003. Towards a Redefinition of the Mens Rea of Rape. *Oxford Journal of Legal Studies* 23:379-404.
- Robins, Timothy. 1995. Remembering the Future: The Cultural Study of Memory. Pages 201-213 in *Theorizing Culture: An Interdisciplinary Critique After Postmodernism*. Edited by Barbara Adam and Stuart Allan. London, GB: UCL Press.
- Stone, Marcus. 1991. Instant Lie Detection? Demeanour and Credibility in Criminal Trials. *Criminal Law Review* 1991:829.
- Superson, Anita M. 2001. Amorous Relationships between Faculty and Students. *The Southern journal of philosophy* 39:419-40.
- Thacker, Scott and Mark D. Griffiths. 2012a. An Exploratory Study of Trolling in Online Video Gaming. *International Journal of Cyber Behavior, Psychology and Learning* 2:17-33.
- . 2012b. An Exploratory Study of Trolling in Online Video Gaming. *International Journal of Cyber Behavior, Psychology and Learning* 2:17-33.
- von Solms, Rossouw and Johan van Niekerk. 2013. From Information Security to Cyber Security. *Computers & Security* 38:97-102.
- Walter, T., R. Hourizi, W. Moncur and S. Pitsillides. 2011. Does the Internet Change how we Die and Mourn? an Overview. *Omega: Journal of Death & Dying* 64:12.
- Wayne, Mike, Lesley Henderson, Craig Murray and Julian Petley. 2008. Television News and the Symbolic Criminalisation of Young People. *Journalism studies* 9:75-90.

# A Framework for Leveraging Cloud Computing to Facilitate Biometrics at Large-Scale

John K. Mitchell III and Syed S. Rizvi

Department of Information Sciences and Technology  
 Pennsylvania State University, Altoona, PA 16601  
 Email: {jkm5270, srizvi}@psu.edu

Submitted to 2014 International Conference on Security and Management (SAM'14), **Track:** Biometrics & Forensics

**Abstract**—The cloud computing paradigm provides an efficient environment for the large-scale use of biometric identification systems. However, there is no established standard for the configuration of the cloud or the type of components necessary to achieve optimal biometric performance. In this paper, we present a conceptual framework which addresses the challenges and the requirements of the biometrics at large-scale while dramatically increasing the performance. In this paper, our contributions are three-fold. First, we describe the related work pertaining to biometric integration with cloud computing. Second, we discuss the issues and challenges preventing biometrics from being adopted at large-scale which influenced the design choices of our proposed framework. Finally, we present the proposed framework and highlight the cloud services that should be utilized with the proper configurations to maximize the biometric efficiency. We also discuss the realization of our proposed framework and the components selected. To show the practicality of the framework, we chose Amazon Web Services even though our framework is applicable to any cloud platform that supports the design criteria.

**Keywords**—biometrics; cloud computing; large-scale implementation; authentication; MapReduce; Hadoop

## I. INTRODUCTION

Biometric technologies are replacing traditional password-based methods as a more effective authentication process in the realm of information security. In addition, biometrics are used for visas and passports, voter registration, border control, and credit card transactions [1]. The concept of biometric identification is especially common in the forensic domain. For instance, the Integrated Automated Fingerprint Identification System (IAFIS) [2] is used by the Federal Bureau of Investigations (FBI) to perform crime scene investigations, criminal background checks, or bank employee checks. Furthermore, biometrics are becoming increasingly popular in other areas of the world as well. For example, India has compiled the largest biometric database in the world [3] with the goal of eventually enlisting 1.2 billion citizens. Therefore, biometrics are expected to continually grow exponentially in the next few years. As the acceptance of biometric identification increases, there will be a higher demand from the public and the private sectors. This demand facilitates the need

for large-scale biometric systems. However, there are many challenges and performance issues that prevent large-scale biometric applications from becoming fully realized. As a result, cloud computing has been proposed as a viable solution to the inhibitors associated with the biometrics.

The cloud has many attractive features that could potentially benefit biometric systems and enable large-scale use which include: parallel processing, large storage space, elasticity, real-time support, and automated failover. Moreover, the highly scalable and flexible nature of the cloud reinforces the idea of integration with biometrics to increase performance. When implementing biometrics with cloud computing, there is a steep learning curve to understand the possible cloud services, configurations, and additional components. Our background and experience with the cloud, specifically Amazon Web Services (AWS) [4], allow us to formulate the necessary services and configurations to abstract the challenges and meet the performance requirements. There were several assumptions about the biometric systems which were taken into consideration when designing our framework. We assume that a multimodal system will be utilized since it provides greater reliability than unimodal systems. One example of a multimodal database is the Next Generation Identification (NGI) program [2] that is being developed by the FBI and is scheduled to replace IAFIS in the summer of 2014. A multimodal database significantly increases the quantity of data being stored and computational power required for processing. Additionally, the support of a fused algorithm to compute a decision from multiple modalities may be required. In this research, we design our framework to be user-friendly and simple for non-experts, enable the most efficient integration possible, be applicable to most cloud platforms, and facilitate large-scale use of biometrics.

The rest of the paper is organized as follows: We begin in Section II by describing the related work. In Section III, we identify the challenges and requirements to be addressed in our framework. A comprehensive discussion on our cloud framework will be presented in Section IV with the emphasis on the configurations of the cloud services. In Section V, we provide components that meet the specifications presented in our framework and examine each component individually.

TABLE 1. SIGNIFICANCE OF EACH FRAMEWORK COMPONENT

Critical Factors	Framework Components				
	IaaS	Distributed File System	Distributed Data Processing	Distributed Database	Distributed Database Management
Scalability	X				
Flexibility	X				
Interoperability					X
Data Archiving	X	X		X	
Data Processing	X		X		X
Near Real-time Speed			X	X	X
Fault-tolerance		X		X	

Finally, we conclude the paper and present future directions in Section VI.

## II. RELATED WORK

The existing works pertaining to biometric integration with cloud computing are scarce. Moreover, these works do not provide a comprehensive solution to the challenges and requirements of biometrics in the form of a generic framework. However, there are a few relevant publications that are similar in some aspects to our paper.

For instance, the authors [5] primarily discuss the MapReduce function and present an implementation involving similar components described in our paper. On the contrary, they only focus on the advantages of parallelism over sequential processing of data. They do not discuss other benefits of the cloud or offer a comprehensive framework. Peer et al. [6] present a case study that involves fingerprint recognition for the implementation. However, the authors are more concerned with how to move the biometric platform into the cloud and the challenges that could arise rather than the actual implementation.

Kohlwey et al. [7] present a prototype system that is based on different requirements identified in their research. However, the system proposed by the authors does not incorporate any cloud services or configurations for the components. Instead, they only use the Hadoop software frameworks as the basis for their prototype system. On the other hand, our main focus is the framework and not the actual system used in our implementation. Another work [8] that is similar in scope utilizes the AWS for their biometric and cloud implementation. They do not utilize some of the important services found in our paper such as the Amazon EMR and Hadoop software. In their paper, the authors mention using the MapReduce framework in order to promote more efficient parallelism. We agree with this argument presented by the authors since the Amazon EMR is essential to the data processing capabilities that our framework can offer.

There are several other works that contain similar topics but proceed in different directions. An example of such work

is Gonzales et al. [9] where the authors focus on securing biometric data from open environments using encryption and access control. Similarly, Baniroostam et al. [10] propose a unique method for increasing the security of behavioral biometrics in cloud computing. The main difference between the two works is that Baniroostam et al. uses a method which gains behavioral data over time while Gonzales et al. incorporates advanced cryptography to minimize the exposure of the private key. Another research work that involves encryption is presented in [11] where the entire biometric database is encrypted and outsourced to the cloud.

After reviewing the related literature, we believe that the cloud is a promising solution to the large-scale issues of biometrics. All of the related works agree that the cloud presents an efficient environment for the growth of biometric applications. However, none of these presents a very detailed framework for optimizing the biometric performance and enabling the large-scale adoption.

## III. CHALLENGES AND REQUIREMENTS OF BIOMETRICS

In the current computing world, biometric identification systems are limited by scalability and flexibility constraints as well as several other challenges and performance requirements that cannot be achieved without access to nearly unlimited storage and processing power. To understand the importance of our cloud framework, the biometric issues that motivated our design must be understood as well. Thus, we identify both the known and expected inhibitors and requirements of a large-scale biometric system. The factors that prevent large-scale biometric applications are addressed individually by the components in our framework as shown in Table 1.

### A. Scalability

Conceptually, biometric systems should be able to adapt and accommodate to their own growth while providing higher or lower levels of productivity depending on the scale. On the contrary, the increasing population and demand causes scalability issues that need to be resolved in order to promote large-scale biometric applications. For instance, in previous implementations, the biometric databases have to be entirely redesigned whenever demand increases [12]. As a result, the

modifications are not cost efficient which outweighs the benefits of scaling the application. Furthermore, the procedures and technology for matching and retrieving templates will change over time. The most efficient and effective biometric system must be able to incorporate these design modifications without inducing a sufficient cost upon the consumer.

*B. Flexibility*

Flexibility is another important factor contributing to the scarcity of a highly efficient biometric system. The ability of a system to change quickly based on the growing requirements is directly related to the scalability. As we mentioned in the previous section, the biometric system must be able to adapt new methods and technology (i.e., next generation scanners) in a cost efficient manner. Additionally, the system should support multiple modalities and multimodal databases. The multimodal databases may require fused algorithms to compile the most accurate results which should be accounted for as well.

*C. Interoperability*

The existing biometric systems are developed by a variety of vendors adhering to different national standards. As a result, there is an overabundance of data with different formats and structure which causes an interoperability challenge. Specifically, the proprietary algorithms and variety of hardware utilized present challenges when attempting to share information between different vendors [13].

*D. Data Archiving*

The large volume of data generated by template enrollment presents a challenge when attempting to store the data. The quantity of data depends on the type of modalities used and whether those modalities are used in tandem for a multimodal system. We anticipate that multimodal databases will be commonly employed by the public and private sectors which will result in petabytes of data at their disposal. This requires a database that is capable of supporting a multitude of records ranging from several to hundreds of millions [1].

*E. Data Processing*

As previously discussed, the large amount of data associated with biometrics requires a vast storage capacity. In addition, an extensive volume of computation is required to process identification requests at the ideal speed for optimal performance. The processing power and memory constraints prevent individual servers or computers from processing the data while satisfying the speed and accuracy demands.

*F. Near Real-time Speed*

To meet the performance requirements, the biometric identifications need to be performed in near-real time. One of the performance requirements is that data streams should be continuous and in near real-time for improved data sharing. Furthermore, to increase data consistency, there should be support for near real-time updates among concurrent users [14]. The relative operating characteristic (ROC) curve can be calibrated to accommodate situations that dictate different identification requirements. This can only be achieved by

altering the curve in a near real-time speed in order to effectively balance the security and convenience of the system.

*G. Fault-tolerance*

The possibility of hardware failure for a large-scale application is high and may result in system wide failure. In regards to biometrics, the system must be fault-tolerant to prevent single points of failure and ensure availability for the user. If failure occurs, the system should utilize redundant resources to minimize the interruption of the operation and maintain system reliability [14].

IV. PROPOSED FRAMEWORK

There are multiple cloud services and deployment possibilities that should be considered when implementing biometrics with the cloud. These variables affect the configuration options that are available to the user and ultimately affect the overall performance of the biometric system. In our proposed framework, we describe the cloud services and additional components that should be utilized and the configurations necessary to promote the highest biometric performance achievable. Table 1 shows the significance of each component in regards to meeting the biometric performance requirements.

Fig. 1 shows the relationship among the different components in our framework. The Infrastructure as a Service (IaaS) is used as the hosting environment for the other components as indicated by the arrows. However, the distributed file system extends further than the IaaS level since the distributed file system is used externally for data archiving and backup.

*A. IaaS*

IaaS was chosen as the underlying cloud service since the designer receives more administrative control than Platform as a Service (PaaS) or Software as a Service (SaaS). This allows for a higher degree of customization, which is important when leveraging the cloud to augment existing biometric systems. In addition, the entire biometrics infrastructure can be migrated into the cloud. Thus, a database modification can be performed easily in minutes. This demonstrates the elasticity of the cloud

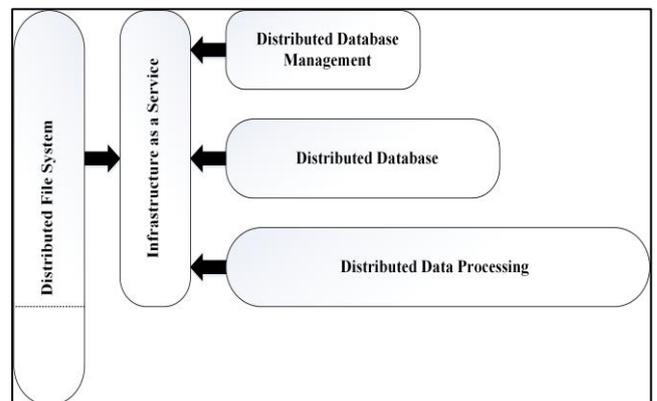


Figure 1: Relationship of framework components

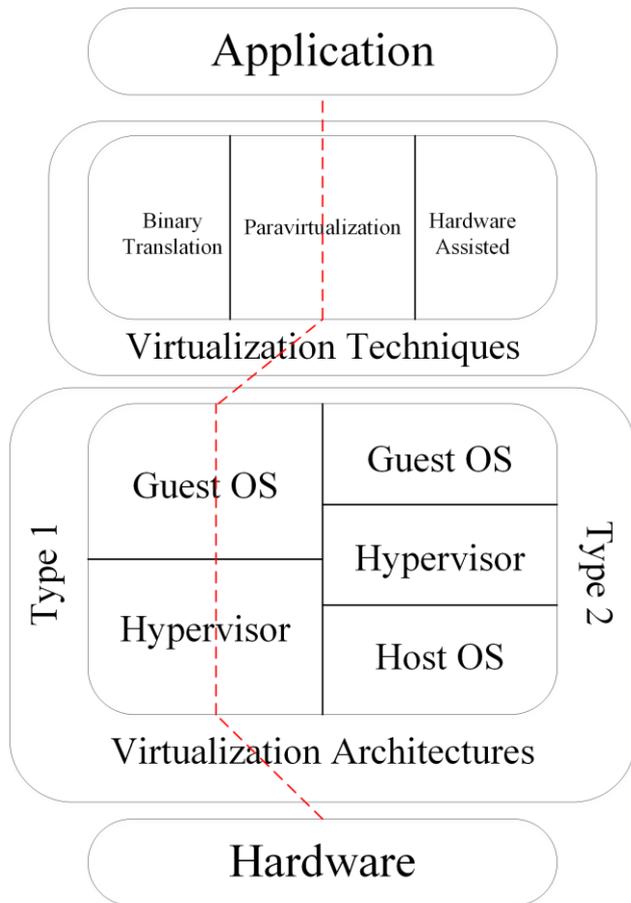


Figure 2: Example of virtualization configurations and framework design

since a biometric application can be scaled quickly based on demand. When accounting for the fixed cost of the cloud service, it is highly cost efficient compared to traditional deployment models.

When examining the virtualization aspect of the cloud, the infrastructure determines the type of hypervisor utilized which affects the performance and capabilities of the system. There are two types of hypervisors known as Type 1 and Type 2, which represent bare-metal and hosted architectures, respectively. Fig. 2 shows the configuration options available to the developer and the design differences between the Type 1 and the Type 2 hypervisors.

In our proposed framework, we chose bare-metal architecture when configuring the virtualization layer since it offers several advantages over hosted architecture. Bare-metal architecture usually provides greater performance due to the direct access of the I/O devices from the virtual machines. Hosted architecture results in less performance since the identification requests would have to be directed through an additional layer which would be the hosted operating system. Furthermore, bare-metal architecture supports real-time

running on operating systems located inside the virtual machines, unlike hosted architecture [15]. The only complication, which would make hosted architecture more desirable, is the overhead involved in the installation of Type 1 hypervisor.

After the type of hypervisor is chosen, the underlying virtualization technique must be taken into consideration. Fig. 2 illustrates the three main virtualization techniques and highlights the design choices of our framework. We selected paravirtualization since it is the most advantageous of the virtualization techniques in regards to optimal biometric performance. The other possibilities are binary translation or hardware assisted virtualization. The performance increase granted from each technique varies according to the coupling of the hypervisors. Paravirtualization enables the greatest performance due to the tightly coupled hypervisors while the loosely coupled hypervisors of binary translation cause significantly less performance. In addition, binary translation and hardware-assisted virtualization can actually result in performance degradation since the hypervisor must routinely interrupt the execution of a virtual machine [15]. Another advantage of paravirtualization is the lack of overhead when compared to the other two techniques.

### B. Distributed File System

A distributed file system for archiving data is essential to meet the capacity demands of a large-scale biometric database. In our framework, the file system would be hosted on the IaaS, which offers nearly unlimited storage capacity and scalability. In addition, the file system must demonstrate functionality since it would be used externally for the immediate retrieval and backup of data. Furthermore, the system must be fault-tolerant and robust which requires exceptional recovery and failover capabilities.

Ideally, we believe the system should incorporate automatic failover. The failure of a node would no longer hinder the operations of the system. Instead, the task of the failed node would be sent to a redundant node with no interruption in the workflow. To the best of our knowledge, there are currently a very limited number of cloud services that include automatic failover in their recovery plans.

### C. Distributed Data Processing

Distributed processing involves multiple computers or processors being utilized to execute applications. Parallelism represents the solution to the data processing challenges of biometrics. In parallel computing, data is partitioned across multiple nodes where computations are solved concurrently. When applied to biometrics, the calculations performed by the algorithm responsible for template matching would be broken down and distributed to several nodes. There are many big data processing tools and frameworks that can be utilized to increase the data processing capabilities of a biometric system. Some of these include MapReduce, Microsoft Dryad, Message Passing Interface (MPI), and Swift [15]. In order to choose the most effective framework or tool for biometrics, the size of the data set and the number of tasks must be taken into consideration. Therefore, we have chosen MapReduce since

the biometric systems require large input data sets and small number of tasks.

MapReduce [15], [16], [17] is a data processing paradigm that enables parallel computations for analyzing and generating large data sets. The input data of a user would be sent into the cloud, partitioned, calculated in parallel, compiled, and sent back to the users in the form of a decision. This increases identification speed while promoting the most accurate results. MapReduce provides organization to the database by moving computations to the location of the data that complies with the data locality principle. As a result, the processing and query latency will decrease.

#### D. Distributed Database

Distributed databases [18] consist of portions of the database being stored in multiple physical locations. These locations are loosely coupled and do not share any physical components. Moreover, a distributed database is dependent on a distributed database management system. Therefore, we included database management as one of the components in our framework. The distributed database is preferable to a relational database for many reasons. The reliability and availability of the biometric system will increase dramatically.

Furthermore, the biometric system expansion and modification is easier and does not affect other related systems. Distribution increases the query processing which improves the biometric performance. In addition, the system becomes more fault-tolerant since offline nodes do not interrupt operations. The distributed database should provide near real-time access to the data located in the biometric database. The record lookups should be in near real-time as well.

#### E. Distributed Database Management System

A distributed database management system controls the distributed database and routinely integrates data in the database to ensure that any change made by the user is updated. Moreover, the distributed database management system supports a distributed metadata repository for near real-time availability. We believe that metadata should be exploited to allow for easier interoperability among existing systems. Metadata can mitigate the problems caused by different template formats by describing the records and allowing for interchange

### V. REALIZATION OF THE FRAMEWORK

In order to demonstrate the effectiveness of our framework, we designed a system based on the previously discussed specifications. We selected several components to be utilized for biometric integration. The AWS are the primary cloud components with several applications being included that are supported by Amazon. The additional applications originate from Apache Hadoop, which is an open source software framework that is synonymous with the big data.

Fig. 3 illustrates the general workflow of the realized components listed in this Section. The input and output data is sent from Amazon Simple Storage (S3) to Amazon Elastic MapReduce (EMR) where parallel computation is performed.

Amazon EMR will have several clusters, based on Hadoop architecture, depending on the modality of the biometric system. The two clusters shown in Fig. 3 indicate that the Apache Hive and the HBase should be run separately to improve performance. The HBase cluster is connected to the Hive cluster by the master nodes. The cluster consists of master and slave nodes. In particular, the slave nodes consist of task and core nodes. The Hadoop Distributed File System (HDFS) is used to store and process data located in the core nodes but does not interact with the task nodes. Finally, the nodes are a product of Amazon Elastic Compute Cloud (EC2).

#### A. Amazon EC2

Amazon EC2 [19] is an IaaS that presents a virtual computing environment to the user. This service grants the ability to create virtual machines called "instances" on Amazon EC2 and exploit them for improved data processing. The number of virtual machines can be increased or decreased according to the scaling of the system. Moreover, Amazon EC2 utilizes the Xen virtualization in its underlying architecture, which has the benefit of allowing hardware assisted or paravirtualization. As stated in the previous section, we believe paravirtualization is the most effective technique for realizing the full potential of biometrics. It is important to note that the Amazon EC2 costs are included when purchasing the Amazon EMR service. This paper includes the Amazon EC2 in the components since it generates the instances used by Amazon EMR.

#### B. Amazon S3

Amazon S3 [20] is a file storage web service that can be used for archiving and analyzing data when implemented with Amazon EC2. We chose Amazon S3 since it is reliable, durable, and supports the other components in our system. Amazon S3 stores and organizes data into "buckets" which are linked to each user account. In addition, Amazon S3 allows users to backup data from the HDFS, Amazon EMR, and the HBase which demonstrates effective recovery of data. The scalability or performance bottlenecks associated with traditional storage databases can be resolved by adding more nodes to increase the system speed, capacity, and throughput.

#### C. Apache HDFS

Apache HDFS [21] is an open source distributed file system that was chosen to be used in collaboration with Amazon S3. The HDFS is highly portable and is capable of storing large files while replicating data across several hosts to achieve data reliability. One notable feature is the data awareness demonstrated by the interaction of the job tracker and task tracker. The job tracker schedules map or reduce jobs to the task tracker depending on the data location. This reduces the amount of network traffic. The HDFS was primarily selected since it is more fault-tolerant than Amazon S3. In addition, the two file systems overcome many of their own design flaws when implemented together. Thus, it is beneficial but not necessary to include both file systems in our system design to achieve maximum biometric efficiency.

There are several advantages for implementing Amazon S3 and HDFS. The HDFS nodes for storing the data are no longer

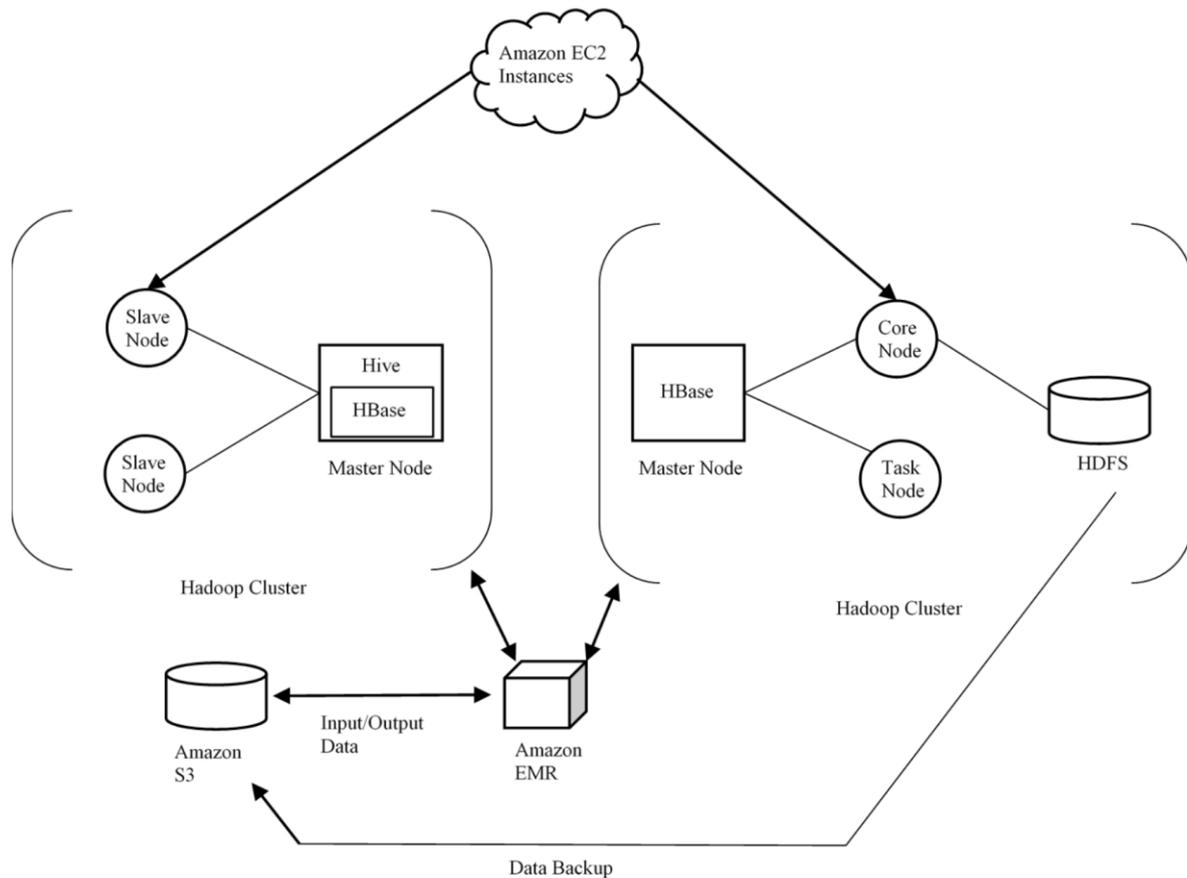


Figure 3: Typical workflow of realized components

required since Amazon S3 provides a massive amount of storage capacity. Additionally, the HDFS data located in an Amazon EMR cluster is lost upon the termination of the cluster. On the other hand, Amazon S3 permanently stores the data ensuring that it is never lost. Another benefit is that HDFS nodes will not become overloaded when using the same data set for multiple jobs [22].

#### D. Amazon EMR

Amazon EMR [22] introduces the MapReduce function to the virtual machines created with Amazon EC2. It is fully integrated with Amazon S3 and HDFS. This component is vital to the performance of biometric systems since the parallel processing speed is dependent on the MapReduce functions. The Amazon EMR cluster consists of master nodes and slaves. There is only one master node in a cluster and the master node is responsible for coordinating the slave nodes. The slave nodes are responsible for running the computations and storing the data.

#### E. Apache HBase

Apache HBase [23] is an open source distributed database that uses column compression to store large quantities of data.

The HBase needs a file system to store its data which is represented in our prototype system by the HDFS. The purpose of the HBase is to augment the HDFS by adding more functionality and addressing several issues. One of these issues is the individual record lookup speed, which is significantly slow. The HBase allows for fast record lookups and access to a small amount of data without the additional latency. Another issue is the batch inserts and deletions. The HBase is an alternative approach that allows users to efficiently perform batch inserts or deletes as well as updates.

In addition, there are several more reasons why we chose the HBase as the distributed database for our system. The HBase utilizes the Bloom filters for improved query when there is an extensive volume of data. There is a tight integration between the HDFS and the HBase, which allows for easier implementation. Furthermore, the regions that contain the HBase tables are split automatically and distributed as the quantity of data increases. Lastly, it is highly fault-tolerant and supports automatic failover for the RegionServers. The RegionServers are responsible for managing the regions in a distributed cluster.

### F. Apache Hive

Apache Hive [24] is an open source data warehouse for managing a distributed database while promoting data query and analysis. The Hive operates using a modified version of the SQL called Hive QL, which enables support for map and reduce functions and different data types. Additionally, unstructured data sources are supported by the Hive and are capable of being processed efficiently. Hive is capable of fabricating a metastore, which is located in MySQL database on the master node of the cluster. The metastore contains the partition names and data types.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we presented a conceptual framework for facilitating the biometric applications at large-scale. We identified the challenges and requirements of current biometric systems to provide the basis for our framework. We then designed the framework to address these issues while increasing the overall biometric performance. At last, we demonstrated the effectiveness of the framework by selecting components for a system that adheres to the criteria we specified. The abstraction of the learning curve will increase the productivity of developers when implementing the biometric with the cloud computing. We anticipate that researchers will be highly motivated to integrate biometric systems with the cloud, especially with a framework available for the enterprise environment.

Since the framework and realization is conceptual, there are considerations and issues that we will address in the future version. The fused result of a multimodal biometric system could present a challenge for the MapReduce component. If an iteration procedure is proven to be more effective for biometric identification, the framework will probably need to be altered slightly to promote maximum efficiency. To address these issues, we will conduct an experiment using our framework for biometric and cloud integration and analyze the results.

### REFERENCES

- [1] Requirements for large-scale biometric systems. (2014) NeuroTechnology. [Online]. Available: <http://www.neurotechnology.com/megamatcher-large-scale-AFIS-and-biometric-identification-systems.html>
- [2] Next Generation Identification. (2014) Federal Bureau of Investigation. [Online]. Available: [http://www.fbi.gov/about-us/cjis/fingerprints\\_biometrics/ngi](http://www.fbi.gov/about-us/cjis/fingerprints_biometrics/ngi)
- [3] Awareness and Communication Strategy Advisory Council. (2010, May). "Aadhar - Communicating to a billion - An Awareness and Communication Report." New Delhi, India. [Online]. Available: [http://uidai.gov.in/images/AADHAAR\\_PDF.pdf](http://uidai.gov.in/images/AADHAAR_PDF.pdf)
- [4] J. Varia and S. Mathew. (2014). "Overview of Amazon Web Services. Amazon Web Services." [Online]. Available: [http://d36cz9buwru1tt.cloudfront.net/AWS\\_Overview.pdf](http://d36cz9buwru1tt.cloudfront.net/AWS_Overview.pdf)
- [5] Shelly and N. S. Raghava. "Iris recognition on Hadoop: A biometrics system implementation on cloud computing." In Proceedings of 2011 IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), pp. 482-485, 15-17 Sept. 2011.
- [6] P. Peer, J. Bule, J. Z. Gros, and V. Struc, "Building Cloud-based Biometric Services," *Informatica*, vol. 37, no. 2, pp. 115-122, June 2013.
- [7] E. Kohlwey, A. Sussman, J. Trost, and A. Maurer, "Leveraging the Cloud for Big Data Biometrics: Meeting the Performance Requirements of the Next Generation Biometric Systems," In Proceedings of IEEE World Congress Services (SERVICES), Washington D.C., pp. 597-601, 4-9 July 2011.
- [8] R. Panchumarthy, R. Subramanian, and S. Sarkar, "Biometric Evaluation on the Cloud: A Case Study with HumanID Gait Challenge," In Proceedings of 2012 IEEE Fifth International Conference on Utility and Cloud Computing (UCC), pp.219-222, 5-8 Nov. 2012.
- [9] D. G. Martinez, F. J. Castano, E. Argones Rua, J. L. Castro, and D. A. Silva, "Secure crypto-biometric system for cloud computing," In Proceedings of 2011 1st International Workshop on Securing Services on the Cloud (IWSSC), pp. 38-45, 6-8 Sept. 2011.
- [10] H. Banirostam, E. Shamsinezhad, and T. Banirostam, "Functional Control of Users by Biometric Behavior Features in Cloud Computing," In Proceedings of 2013 4th International Conference on Intelligent Systems Modelling & Simulation (ISMS), pp. 94-98, 29-31 Jan. 2013.
- [11] J. Yuan and S. Yu, "Efficient privacy-preserving biometric identification in cloud computing," In Proceedings of 2013 IEEE INFOCOM, pp. 2652 - 2660, 14-19 April 2013.
- [12] R. Das. Biometrics in the cloud. *Keesing Journal of Documents and Identity*. No. 42, pp. 21-23, Feb. 2013.
- [13] The National Biometric Interoperability Framework and Capability Requirements. (2014, Feb.). Biometrics Institute. [Online]. Available: <http://www.biometricsinstitute.org/news.php/150/the-national-biometric-interoperability-framework-and-capability-requirements>
- [14] A. Sussman, "Biometrics and Cloud Computing," In Proceedings of Biometrics Consortium Conference, pp. 1-8, 19 Sept. 2012.
- [15] M. Hamdaqa and L. Tahvildari, "Cloud Computing Uncovered: A Research Landscape," *Advances in Computers*, vol. 86, pp. 41-85, 2012.
- [16] D. Gannon and D. Reed, "Parallelism and the Cloud," Microsoft Corp., Redmond, WA., Oct. 2009.
- [17] F. Li, B. C. Ooi, M. T. Özsu, and S. Wu. "Distributed Data Management Using MapReduce," in press.
- [18] Oracle Corp. (1999, Dec.). "Distributed Database Systems." [Online]. Available: [http://docs.oracle.com/cd/A87860\\_01/doc/server.817/a76960.pdf](http://docs.oracle.com/cd/A87860_01/doc/server.817/a76960.pdf)
- [19] Amazon Web Services. (2013, Oct.). "Amazon Elastic Compute Cloud User Guide." [Online]. Available: <http://awsdocs.s3.amazonaws.com/EC2/latest/ec2-ug.pdf>
- [20] Amazon Web Services. (2006, Mar.). "Amazon Simple Storage Service Developer Guide." [Online]. Available: <http://awsdocs.s3.amazonaws.com/S3/latest/s3-dg.pdf>
- [21] Apache Software Foundation. (2014, Feb.). "HDFS Users Guide." [Online]. Available: <http://hadoop.apache.org/docs/r2.3.0/hadoop-project-dist/hadoop-hdfs/HdfsUserGuide.html>
- [22] Amazon Web Services. (2009, Mar.). "Amazon Elastic MapReduce Developer Guide." [Online]. Available: <http://docs.aws.amazon.com/ElasticMapReduce/latest/DeveloperGuide/e-mr-what-is-emr.html>
- [23] Apache Software Foundation. (2014, Feb.). "Apache HBase Reference Guide." [Online]. Available: <https://hbase.apache.org/0.94/book.html>
- [24] Apache Software Foundation. (2013, Dec.). "Apache Hive GettingStarted." [Online]. Available: <https://cwiki.apache.org/confluence/display/Hive/GettingStarted>

# Exploring Digital Forensics Tools in Backtrack 5.0 r3

Ahmad ghafarian<sup>1</sup> and Syed Amin Hosseini Seno<sup>2</sup>

<sup>1</sup>Department of Computer Science, University of North Georgia, Dahlonega, GA USA

<sup>2</sup>Computer Networks Laboratory, Department of Computer Engineering, Ferdowsi University of Mashhad, Iran

**Abstract** - Computer forensics tools are essential part of any computer forensics investigation. The tools can be classified in various ways including, open source vs. proprietary; hardware vs. software; special purpose vs. general purpose, etc. In practice, software tools are more common. Each software tool has its own pros and cons. However, they all have one feature in common, i.e. installation, configuration, and setup. For some tools, the configuration process can be complicated and time consuming. To avoid this, the computer forensics investigators have the option of using the computer forensics tools that are pre installed and configured in Backtrack 5.0 r3. In this paper, we present the results of our experiment with various digital forensics tools that are included in Backtrack 5.0 r3.

**Keywords:** Backtrack, VMware, Computer Forensics Tools

## 1 Introduction

Computer forensics tools play an important role for forensics investigators. Selection of a particular tool depends on the nature of the investigation, reliability, security, and the cost effectiveness. There are many options that digital forensics investigators can choose from. Classifications of computer forensics tools include open source, proprietary, hardware, software, special purpose and general purpose. Each tool has its own advantages and disadvantages. A comprehensive review of the top twenty open source free computer forensics investigation tools can be found in [14]. For a list of proprietary computer forensics tools see [16] & [9]. Brian Career [3] reports on how forensics tools have been viewed historically, i.e. philosophy, security and reliability. He concludes that open source tools are as effective and reliable as proprietary tools. Manson and his team [8] compared one open source tool and two commercial tools. They found that all three tools produced the same results with different degree of difficulty. Backtrack 5.0 r3 has a rich repository of digital forensics tools that support computer forensics specialists to do tasks such as acquisition, analysis, recovery, imaging, vulnerabilities scan, penetration testing, and file interrogation. A survey of Backtrack 5.0 network forensics tools can be found in [7]. The purpose of this research is to study Backtrack 5.0 r3 [2] forensics tools. We examine different categories of computer forensics tools, analyze the types and number of tools in each category, investigate their capabilities, evaluate their effectiveness, and

present the result of our experiment with all the available tools in Backtrack 5 r3. In the next section we discuss the platform for our e.

## 2 Our Virtual Machine Platform

Backtrack is a Linux based operating systems that comes with a rich repository of security and forensics tools [2]. The computer forensics tools are grouped into several categories. We use the forensics tools within the Backtrack.

VMware Workstation is a hypervisor that runs on 64-bit computers [15]. It enables us to set up multiple virtual machines and network them together. Each virtual machine can execute on different distribution of Linux operating system. VMware Workstation is proprietary software but we used the trail version for free. Below are the steps for setting up the platform for our experiment.

1. Install VMware Workstation on a machine
2. Create a virtual machine on the VMware workstation
3. Install Backtrack 5.0 r3 on the virtual machine
4. Launch Backtrack 5.0 r3 from the virtual machine
5. From the list, select forensics and then select a tool

## 3 Forensics Tools Experiment

There are several categories of computer forensics tools in Backtrack. Some categories have more than one tool. In the following subsections, we explore the details of the tools. For each tool, we review its purpose, the syntax for running the tool and the results of executing the tool on our virtual machine platform.

### 3.1 Anti Virus Forensics

The tools in this category include *Chkrootkit*, and *rkhunter*

#### 3.1.1 Chkrootkit

*Chkrootkit* is a program that checks for signs of rootkit infection on a machine during live acquisition. It runs on almost all versions of the Linux. Depending on the option selected by the user, *Chkrootkit* can perform an individual infection scan, *sshd* infection test, as well as full scan. Network administrators can also use it to check for known rootkits. We ran this tool on our virtual machine by issuing the

following command. `/pentest/forensics/chkrootkit -x/-q` where switches, `-x` and `-q` indicate expert mode and quiet mode respectively. It only took a couple of minutes to execute and present the report. It reported no rootkit in our virtual machine as expected.

### 3.1.2 Rkhunter

*Rkhunter* is another utility which can be used in live acquisition to check for signs of rootkits on Linux based systems. It is a rich scanning tool that scans for rootkits, backdoors, local exploits, hidden files and comparing MD5 hashes. We executed this tool on our virtual machine for a full scan using the command: `/pentest/forensics/rkhunter -c -sk`. It examined 163 files and applications and 8 suspects file were identified.

## 3.2 Digital Anti Forensics

*TrueCrypt* is the only tool in this category. It is able to establish and maintain an on-the-fly-encrypted volume. This means that data is automatically encrypted right before it is saved and decrypted right after it is loaded, without any user intervention. To read data from an encrypted volume we must use an encryption key. This tool encrypts the entire file system e.g., file names, folder names, contents of every file, free space, metadata, etc. TrueCrypt currently supports the following hash algorithms: RIPEMD-160, SHA-512 and Whirlpool. This utility is not pre installed on Backtrack 5.0 In our experiment; we installed the TrueCrypt, mounted the volume on VMware virtual machine, and then executed TrueCrypt. To see the effect of TrueCrypt, we saved some Microsoft office files on the mounted volume. When we tried to retrieve the files TrueCrypt asked for the encryption key. Upon entering the encryption key, the files were opened successfully.

## 3.3 Digital Forensics

*Hexedit* is a digital forensics tool which has the capability to view and edit files in hexadecimal or in ASCII format. Some features of *Hexedit* include, reading a device as a file, comparing two files, searching, and statistical calculations on the data of a file. With Hexedit being activated, we used the command `media/ashrafian/test.dd` to examine the content of the `test.dd` file which is saved in *Ashrafian* folder. The result is shown in Figure 1 below.

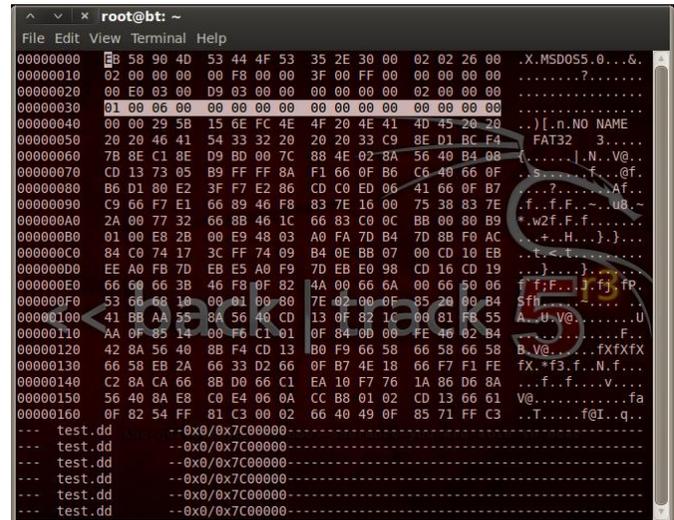


Figure 1- Statistical analysis of test.dd file using Hexedit

## 3.4 Forensics Analysis Tools

The tools in this category include *bulk\_extractor*, *evtparse.pl*, *exiftool*, *misidentify* and *stegdetec*

### 3.4.1 Bulk\_extractor

*Bulk\_extractor* is a program that can scan a disk image to search for personal data such as credit card numbers, email addresses, domain names, urls, telephone numbers, text messages, etc. It can also automatically detect, decompress, and recursively re-process compressed data. It is capable of processing data from various devices such as hard drives, optical media, camera cards, cell phones, network packet dumps, etc. It classifies outputs information in various files such as *ccn.txt* for credit card numbers, *domain.txt* for Internet domains found in the file, *email.txt* for email addresses, *exif.txt* for exif data from media files and *wordlist.txt* for the list of all words extracted from the file.

We used *bulk\_extractor* to scan a USB flash drive. The command to execute it is: `bulk_extractor -o outputdir /media/A.dd` (where *outputdir* is a directory and *A.dd* is a file that contains image of the drive under investigation.) We stored some personal information in the USB to demonstrate the *bulk\_extractor*'s behavior. It turned out that the data that was stored in the USB was retrieved in the corresponding output files. Upon examination of these files, we were able to see the original data that we saved in the USB. The extraction of data was very fast. This speed is attributed to the fact that *bulk\_extractor* can scan different parts of a file in parallel and thus no need for file parsing or any knowledge of the file system. See Figure 2 below.

```

root@bt:~# bulk_extractor -o outfile /media/SHAHBAZI/A.dd
file not found: /media/SHAHBAZI/A.dd
root@bt:~# bulk_extractor -o outfile /media/SHAHBAZI/A.dd
bulk_extractor version:1.2.0
Hostname: bt
Input file: /media/SHAHBAZI/A.dd
Output directory: outfile
Disk Size: 130023424
Threads: 1
Phase 1.
15:41:39 Offset 0MB (0.00%) Done in n/a at 15:41:38
15:41:43 Offset 16MB (12.90%) Done in 0:00:52 at 15:42:35
Warning: Directory Thumbnail, entry 0x0201: Data area exceeds data buffer, ignoring it.
15:41:50 Offset 33MB (25.81%) Done in 0:00:39 at 15:42:29
15:41:56 Offset 56MB (39.71%) Done in 0:00:32 at 15:42:28
15:42:03 Offset 67MB (51.61%) Done in 0:00:25 at 15:42:28
15:42:09 Offset 83MB (64.52%) Done in 0:00:18 at 15:42:27
Warning: JPEG format error, rc = 4
Warning: Directory Thumbnail, entry 0x0201: Data area exceeds data buffer, ignoring it.
15:42:14 Offset 100MB (77.42%) Done in 0:00:11 at 15:42:25
Warning: Directory Thumbnail, entry 0x0201: Data area exceeds data buffer, ignoring it.
Warning: Directory Thumbnail, entry 0x0201: Data area exceeds data buffer, ignoring it.

```

Figure 2- Execution of bulk\_extractor output

### 3.4.2 Evtparse.pl

Evtparse.pl is a Windows event file parser utility. It generates a text output from the event files which may contain useful information. It is a useful tool for work with event files such as Windows log file. We applied this utility on a Windows log file called *A.evtx*. It produced information such as date file was accessed, time, etc. The format we used is: `evtparse.pl -e /media/Shahbazi/A.evtx`, where Shahbazi is a folder that contains *A.evtx* file.

### 3.4.3 Exiftool

Exiftool is a command line utility that allows users to read or write metadata to image files. To retrieve metadata from *A.dd* image, we used the command: `Exiftool -a /media/Shahbazi /A.dd`. As can be seen from Figure 3; Exiftool extracted metadata from the image file. As we can see from the figure, important metadata information are listed.

```

root@bt:~# exiftool -a /media/SHAHBAZI/test.dd
ExifTool Version Number      : 7.89
File Name                    : A.dd
Directory                    : /media/SHAHBAZI
File Size                    : 124 MB
File Modification Date/Time   : 2006:07:30 21:46:50-04:00
File Type                    : MP3
File Type                     : audio/mpeg
MIME Type                    : 1
MPEG Audio Version           : 1
Audio Layer                  : 256000
Audio Bitrate                : 32000
Sample Rate                  : Single Channel
Channel Mode                 : Bands 16-31
Mode Extension               : False
Copyright Flag               : False
Original Media               : 50/15 ms
Emphasis                     : 1:07:43 (approx)

```

Figure 3-The result of running Exiftool on A.dd file

### 3.4.4 Misidentify

This utility can be used to find Windows 32 executable files recursively. We launched this utility to list executable files in a USB drive by using the command `misidentify -r /media/Shahbazi/forensic`. Since our virtual machine is Windows 64, there was no Windows 32 executable file in the forensics folder.

### 3.4.5 Stegdetect

This tool will look for signatures of several well-known steganography embedding programs in order to alert the user that text may be embedded in the image file, such as jpeg. To see if there is steganography embedded message in our *AA.jpg* file in a USB drive, we launched `Stegdetect` by using this command `stegdetect -t /media/Shahbazi/AA.jpg`. The result of executing the utility is shown below

```
stegdetect -t /media/Shahbazi/AA.jpg: negative
```

Where *negative* indicates no message embedded in the *AA.jpg* file.

## 3.5 Forensics carving tools

The tools in this category include *Foremost*, *recoverjpeg*, *safecopy*, *scrounge-ntfs*, & *Testdisk*

### 3.5.1 Foremost

*Foremost* is a popular file carving utility. It takes image files to search for file headers in order to recover files. The carving process utilizes attributes such as unique signatures, file headers, and file footers. Some limitations of *foremost* is when files are fragmented, header is overwritten, it is a common string, or is changed due to actions such as compression. *Foremost* configuration file also allows the forensic examiner to customize the types of files that will be recovered and enables the use of wildcards for pattern matching. *Foremost* opens image file in read-only mode, which is important for maintaining the forensic integrity. It can handle both Windows and Linux file systems. We applied *Foremost* to search the image *Foremost.dd* for jpeg files. The command we used for this process is:

```
Foremost jpeg -o/ root/Desktop/media/Shahbazi/foremost.dd
```

Where switch `-o` specifies the output directory for storing recovered files. As can be seen from Figure 4, the recovered jpeg files are listed. The file size and dates are also displayed.

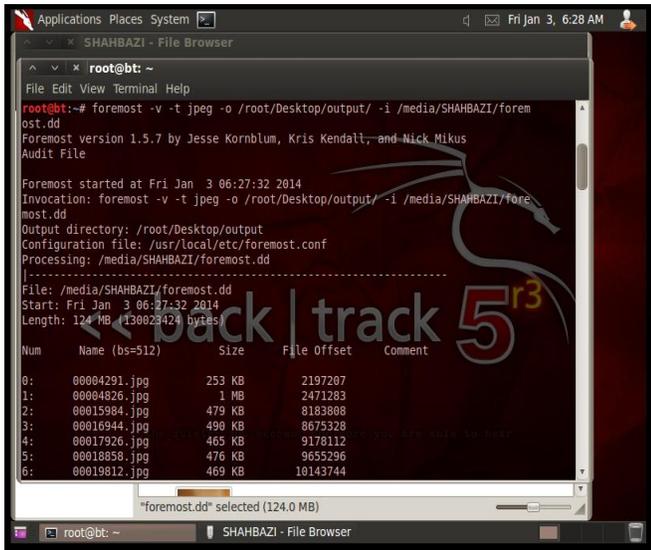


Figure 4-Foremost recovered jpeg files

### 3.5.2 Recoverjpeg

This is another utility for recovering jpeg images from a file system. We also used this tool to recover jpeg images from `foremost.dd` folder on USB drive by issuing this command: `recoverjpeg/media/Shahbazi/foremost.dd`. The same Jpeg files that are shown in Figure 5 were retrieved.

### 3.5.3 Safecopy

This utility can recover as much data as possible from a damaged device, such as a hard drive or a USB drive. Other programs such as `dd`, `cat`, or `cp` will stop reading data once a damaged area is hit, while `Safecopy` will read to a point designated by the user, regardless of damaged areas. It does this by identifying the damaged areas, and skipping around them. To recover data from a damaged USB drive, we used: `Safecopy/media/Shahbazi/root/Desktop/rescue files`. With this tool, we were able to recover files to `rescue` folder that exist on `/root/Desktop`.

### 3.5.4 Testdisk

`Testdisk` is a program that specializes in recovering lost disk partitions and making disks bootable. It has the ability to rebuild partition tables, rebuild boot sectors, fix the Master File Table, and recover deleted partition and files. Our experiment with `Testdisk` reported back no error in our system.

## 3.6 Forensics Hashing Tools

Backtrack 5.0 supports many hashing utility including `Hashdeep`, `MD5deep`, `Shaldeep`, `Sha256deep`, `Tigerdeep`, and `Whirlpooldeep`

All the tools listed above basically do the similar job, i.e. calculating the message digest of an input file. Each utility is a suite of cross platform tools to compute and compare MD5, SHA-1, SHA-256, Tiger, or Whirlpool message digests on an arbitrary number of files. We used `MD5deep` to calculate the hashes of all files in our input folder. The command to do that is: `md5deep /root/Desktop/output/` where the message digests is saved to a file in a directory called `output` (see Figure 5).

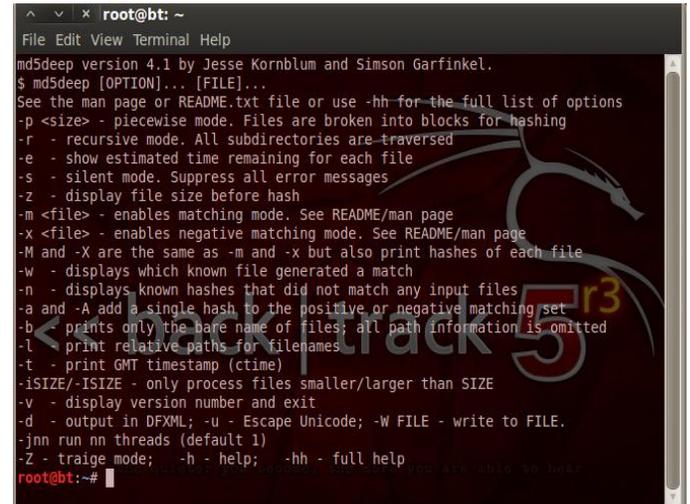


Figure 5-MD5deep running on Backtrack

## 3.7 Forensics Imaging Tools

There are several imaging tools in Backtrack 5.0. These include `Air`, `Dc3dd`, `Ddrescue`, & `Ewfacquire`. In the next subsections we report their performance.

### 3.7.1 AIR

`AIR` (Automated Image and Restore) is a utility which can be used to create forensics bit images from device drives. `AIR` supports MD5/SHAx hashes, SCSI tape drives, imaging over a TCP/IP network, splitting images, and detailed session logging. `AIR` itself is a GUI interface for `dd/dc3dd`. On Backtrack 5.0, when we first selected `AIR`, it downloaded and compiled the necessary components for running the program. Then we followed these steps to create an image of a device:

- Choose USB1 as the source device
- Choose USB2 as the destination device.
- Choose Block size of the source and the destination 512 Byte.
- Choose MD5 as the hash method.

After these steps, when we clicked at the start the imaging begins (see figure 6.) It took several minutes to create an image USB1 in USB2.

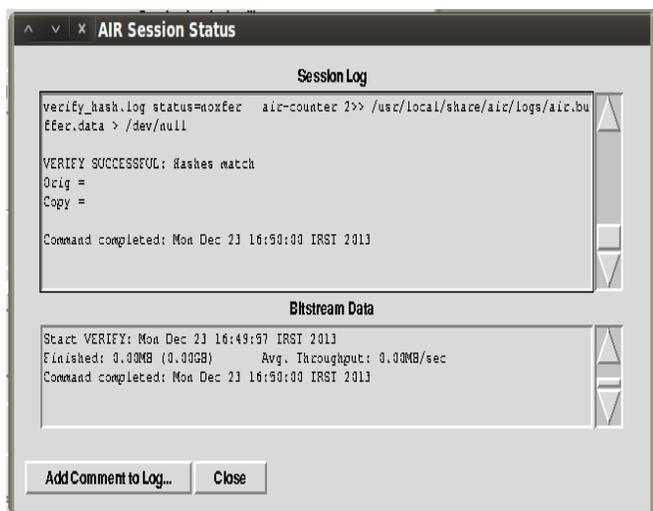


Figure 6-AIR Session Status

### 3.7.2. D3dd

*D3dd* is an altered version of *dd*, that we used to operate low level disk functions. We used *D3dd* to split a large disk image into smaller pieces. For the input file of */dev/sda* it calculated hashes for the individual new files and the original large file was broken into 2 GB pieces with “000” as a suffix in the filename. It also saved logs of all data to */root/Desktop/log.txt*, and output the smaller files to */root/Desktop/images*.

### 3.7.3 Ewfacquire

*Ewfacquire* can be used to create disk images in the EWF format. It includes several message digests including MD5 and SHA1. To create an image of */dev/sdb1* and logging data to */root/Desktop/log.txt*, we obtained the image by issuing this command on Backtrack

```
ewfacquire -d sha1 -l /root/Desktop/log.txt /dev/sdb1
```

## 3.8 Forensics Suites

We applied *DFP* (*Digital Forensics Framework*) to collect, preserve, and reveal digital evidence. In Backtrack 5.0 we can launch it via Forensics Suites menu. In order to run *DFP*, we first loaded an evidence file, i.e. a forensic image that we created using one of the previous tools. *DFP* then processed the evidence file against one of the built-in modules to begin analyzing data. Figure 7 below shows the *DFP* analysis of the evidence.

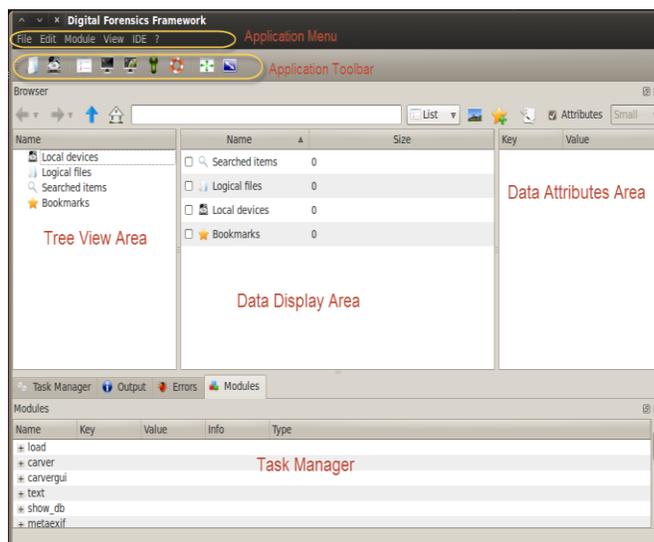


Figure 7 - DFF interface

## 3.9 Network Forensics

Tools in this category include *Driftnet*, *P0f*, *Tcpreplay*, *Xplico*, & *Wireshark*

### 3.9.1 Driftnet

*Driftnet* is a network utility that sniffs traffic for images and other media. Rather than sniffing all traffic using utilities like *Wireshark*, *Driftnet* makes it easier by automatically picking out images and media. We used the command *driftnet -i eth0 -v*, to capture traffic and instruct *driftnet* to be verbose mode in its output. The result produced useful information which is valuable to forensics investigators.

### 3.9.2 P0f

*p0f* is a passive operating systems fingerprinting tool. All the host has to do is connect to the same network or be contacted by another host on the network. The packets generated through these transactions gives *p0f* enough data to guess the system. In our experiment, by issuing the command *p0f -f /etc/p0f -i eth0*, we were able to read fingerprints from */etc/p0f* and listen on *eth0* via *libpcap* application.

### 3.9.3 Xplico

*Xplico* is a Network Forensic Analysis Tool (NFAT) that is capable of extracting application data from packet capture files. It is best suited for offline analysis of PCAP files but it can also analyze live traffic. *Xplico* can extract email, HTTP, VoIP, FTP, and other data directly from the PCAP files. It is able to recognize the protocols with a technique named Port Independent Protocol Identification (PIPI). We executed *Xplico* in Backtrack 5.0 by issuing the following commands:

Start Xplico - /etc/init.d/xplico  
 Go to http://localhost:9876  
 login with default user and password  
 user name: xplico  
 password: xplico  
 Click the new case

In *Xplico* a case is composed of one or more sessions. Figure 8 shows a new case we have created in *Xplico*. This session captured traffic for offline analysis of PCAP files. The captured traffic was analyzed and no unusual activities were found.

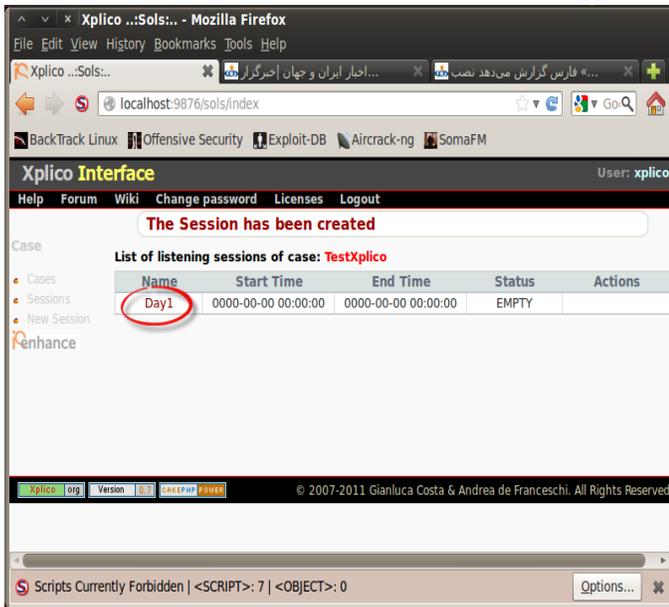


Figure 8 - Session Creation in Xplico

### 3.10 Password Forensics Tools

The tools in this category include *CmosPwd*, *fCrackZip*, and *Samdump*. All these tools are password cracking and or retrieving tools. In the next subsection we describe the result of the application of *FCrackZip*.

#### 3.10.1 FCrackZip

*FCrackZip* is a tool for breaking the password of a password protected zip file with a brute force or dictionary attack [6]. When using this tool in brute force mode we can specify the length, character types, and initial strings for the password. Before we launch this tool we need to upload a password protected zip file to the Backtrack. By default, brute force starts at the given starting password, and successively tries all combinations until they are exhausted. Then, it prints all passwords that it detected. We applied brute force attack for cracking the password protected *srl.zip* file (see Figure 9):

`FCrackZip -b -l 2 -c 'aA1' /root/Desktop/srl.zip`

Where

-b > brute force  
 -c aA1 > char set lower, upper, alphabet  
 -l > length of expected password

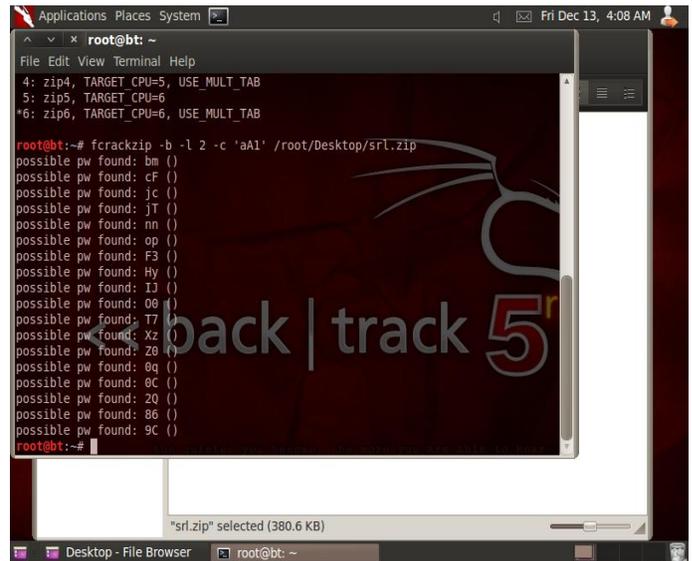


Figure 9- FCrackZip cracking a password protected zip file

We should note, depending on the password combination, this tool may take a long time to process even for a small file.

### 3.11 PDF Forensics Tools

The utilities in this category include *PDFid*, *PDF Parser*, and *Peepdf*

#### 3.11.1 PDFid

*Pdfid* is a utility that can extract useful information from a PDF file. *Pdfid* can show forensics investigators any suspicious activities in the PDF files. It can also scan a PDF file to look for certain PDF keywords such as JavaScript. The execution of *pdfid.py* on file.pdf produced useful information about the file such as header, date, etc.

#### 3.11.2 PDF Parser

PDF Parser is an investigation tool that can be used to examine some of the content within a PDF file. When some characters are discovered that appear to have no meaning, then other tools such as ASCIIhexDecode can be used. We used the command `pdf-parser.py -a /mnt/shared/nw.pdf` to display statistics about the new.pdf file (See Figure 10.)

```

root@bt: /pentest/forensics/pdfid# ./pdfid.py /mnt/ws.pdf
PDFID 0.0.11 /mnt/ws.pdf
PDF Header: %PDF-1.4
obj          55
endobj       55
stream       19
endstream    19
xref         1
trailer      1
startxref    1
/Page       4
/Encrypt     0
/ObjStm     0
/PS         0
/JavaScript 0
/AA         0
/OpenAction 0
/AcroForm   0
/JSRG2Decode 0
/RichMedia  0
/Launch     0
/Colors > 2^24 0

```

Figure 10 - Result of running PDFid on a PDF file

## 4 Conclusions

In this work we have demonstrated the application of various computer forensics tools on Backtrack 5. We showed the syntax for using the tools including applicable switches, and the result of executing the tools on our virtual machine. As it was demonstrated the tools produce consistent results according to their specifications. However, similar results can be obtained by using physical machines. Our results will help the computer forensics investigators on selecting appropriate tool for a specific purpose. It also helps penetration testers to check for signs of vulnerabilities on their system. We showed that Backtrack 5 is a good choice for forensics investigators for several reasons. These include, the tools are free, easy to use, no need for configuration, and produce consistent results.

## 5 Future Research

To extend this research, we intend to gather more detailed instructions of the tools for potential users of the tools. In addition, we plan to install Backtrack on a physical machine and perform the same experiments. We also plan to use selected non-Backtrack open source computer forensics tools, observe their performance and compare the results with the same tools in Backtrack 5.0. Another area which is worth further study is RAM forensics for social networking sites such as Facebook.

## 6 References

- [1] Alex, (Jan 2013), *exiftool-Backtrack 5-Forensics-Digital Forensics Analysis-exiftool*, Question Defense, <http://www.question-defense.com/2013/01/02/exiftool-backtrack-5-forensics-digital-forensics-analysis-exiftool>
- [2] Backtrack 5 r3, <http://www.backtrack-linux.org/backtrack/backtrack-5-r3-released/>
- [3] Carrier, B (Oct 2002), Open Source Digital Forensics Tools: The legal argument. [http://dl.packetstormsecurity.net/papers/IDS/atstake\\_open\\_source\\_forensics.pdf](http://dl.packetstormsecurity.net/papers/IDS/atstake_open_source_forensics.pdf)
- [4] De Smet, D, & Willie L. Pritchett (March 2013), *Backtrack Forensics*, Cookbooks Networking & Telephony Open Source, <http://www.packtpub.com/article/backtrack-forensics>
- [5] Gupta, A, *Digital Forensics Analysis using Backtrack, Part 1 & 2* <http://www.linuxforu.com/2011/03/digital-forensic-analysis-using-backtrack-part-1/>
- [6] Lazzez, Amore (January 2013), A survey About Network Forensics Tool, International Journal of Computer and Information Technology, Vol 2, Issue 1, pages 74-81.
- [7] List of Tools in Backtrack [http://secpedia.net/wiki/List\\_of\\_tools\\_in\\_BackTrack](http://secpedia.net/wiki/List_of_tools_in_BackTrack)
- [8] Manson, D; Carlin, A. ; Ramos, S. ; Gyger, A. ; Kaufman, M. ; and Treichel, J (2007). *Is the Open Way a Better Way? Digital Forensics Using Open Source Tools*, System Sciences. HICSS 2007. 40th Annual Hawaii International Conference on Science, Page 266b-270.
- [9] Mares and Company (2013), Alphabetical list of links to manufacturers, suppliers, and products, [http://www.dmares.com/maresware/linksto\\_forensic\\_tools.htm](http://www.dmares.com/maresware/linksto_forensic_tools.htm)
- [10] Nelson, B; Phillips, A; Stuart, C., (2010), *Guide to Computer Forensics and Investigation*, 4ed, Cengage Learning.
- [11] Nolan, R, Colin O'Sullivan, Jake Branson, Cal Waits, (March 2005), *First Responders Guide to Computer Forensics*
- [12] Rose, M (July 2006) *Brute force cracking*, <http://searchsecurity.techtarget.com/definition/brute-force-cracking>
- [13] Sigh, G, Crack the password protected zip files using fcrackzip-Backtrack, <http://hackthedark.blogspot.com/2012/06/crack-password-protected-zip-files.html>
- [14] Tabona, A. Z. (2002), *Top 20 Free Digital Forensics Investigation Tools for SysAdmins*, <http://www.gfi.com/blog/top-20-free-digital-forensic-investigation-tools-for-sysadmins/>
- [15] VMware Virtualization for Desktop & Server, Application, <http://www.vmware.com/training/>
- [16] Wikipedia, List of Digital Forensics Tools, [http://en.wikipedia.org/wiki/List\\_of\\_digital\\_forensics\\_tools](http://en.wikipedia.org/wiki/List_of_digital_forensics_tools)

# A Method for Authentication using Behavior Biometrics on WEB

Hiroshi Dozono<sup>1</sup>, Naoto Yamasaki<sup>1</sup>, Masanori Nakakuni<sup>2</sup>

<sup>1</sup> Faculty of Science and Engineering, Saga University, 1-Honjyo Saga, 840-8502 JAPAN

<sup>2</sup> Information Technology Center, Fukuoka University, 8-19-1, Nanakuma, Jonan-ku, Fukuoka 814-0180 JAPAN

**Abstract**—Recently, many WEB services are supplied from network, and many cloud systems are operated on the WEB. For accessing these services, authentication should be required, and as the conventional authentication method password authentication with inputting password in the WEB form is mainly used. However password authentication involves some issues. For this problem, we propose a biometric authentication method which can be used on the WEB. In this paper, we propose an authentication framework using HTML5, and implemented the authentication method using the biometrics obtained from handwritten pattern using HTML5.

**Keywords:** Biometrics, Authentication, WEB application, Touch panel, HTML 5

## 1. Introduction

Recently, large amount of data is stored on the network system, and are accessible from WEB because of the growth of the cloud system. And many WEB applications are supplied from internet such as office applications. On the WEB, the password authentication method is mainly used because the conventional authentication system can only handle static data such as password. However, password authentication method involves some issues. For example, password can be stolen easily because the password is a mere string, and may be guessed from personal information like birth day or name of the family.

For this problem, we introduce the biometric authentication method which can be used on the WEB. The biometric authentication method can be classified in two types: Biometric authentication method using biological characteristics and Biometric authentication method using behavior characteristics. In this paper, we use biometric authentication method using behavior characteristics obtained from touch panel because it does not need additional hardware for obtaining biometric features. We proposed the biometric authentication method using touch panel in [1]. This method is implemented as the standalone application on iOS. In this paper, we implemented this system using HTML5, which can be executable on multi-platform system using WEB browser.

## 2. Authentication Methods using Touch Panel

As the popularization of the mobile devices such as smartphones and tablet devices, the devices which equipped with touch panel become widespread. And these devices grow in usage and popularity to the people who is not familiar to the conventional computers. In these situation, a simple authentication method which uses touch panel is desired.

As the authentication method using touch panels, the signature written on the touch panel is often used. However, it is difficult to write identical signature on slippery touch panel especially for capacitive type panel which is recently equipped to almost all mobile devices. As the another method, the authentication method which uses the knowledge factors with selecting the symbols or positions in the image is often used. However this method has a weakness for stealthy glance.

Android devices uses the lock screen which uses the knowledge factor with connecting the points displayed on the touch panel(Figure 1). However, the patterns which can be drawn are not flexible, and users tend to select simple patterns because the points are fixed. Thus, this method also has a weakness for stealthy glance. For avoiding stealthy

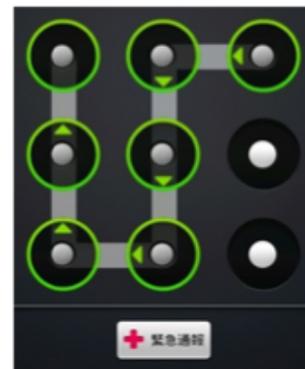


Figure 1: Android lock screen

glances, biometric authentication is effective. Biometric authentication is classified to 2 types; biometrics authentication using biological features and biometric authentications using

behavioral features. As the biological features, the fingerprints, vein patterns and iris patterns are often used. They can achieve high accuracy, however special sensor devices, such as fingerprint reader, are required for implementation. As the behavioral features, keystroke timings[2] and penmanship of handwritten patterns[3] or signature[4] are often used. They can be obtained from conventional input devices, however the accuracy of authentication is lower than that of biological features. At the same time, high accuracy is not always necessary for the devices which is used personally because authentication is used as the lock method just in case the device is stolen or possessed by malicious users. In this paper, the biometric authentication method which uses the biological features obtained from touch panel is proposed. We have proposed an authentication system using the behavior biometrics during drawing the symbol displayed in the touch panel[5]. This system uses pen speed and pen pressure at all sampling times as behavior biometrics, and marks 0.1 as Equal Error Rate(ERR). However, capacitive type panel, which is mostly used recently, can not detect pen pressure, and it requires much computational costs for matching all pen speed and pen pressure at all sampling time. For this problem, we propose an authentication system which generates feature points automatically.

### 3. Authentication System using HTML 5

HTML5 is the advanced version of Hyper Text Markup Language (HTML). HTML5 is the markup language, which consists of HTML, Javascript and CSS. HTML is used to define the contents of WEB pages, CSS is used to define the layouts of WEB pages. Javascripts realize the dynamic WEB pages which are required to implement the authentication system with behavior biometrics. Using these features, the control of animations, storage of data in local system, and socket communication via network can be implemented using HTML 5. Furthermore, a method, which is named Canvas, is provided for drawing shapes in real time, and we use this method for implementing authentication system.

#### 3.1 Implementation of authentication system using HTML 5

In many WEB based system which requires authentication, password authentication is applied. The password is static information, thus it can be compared with registered password on server side which is implemented in server side scripts, such as PHP, Perl. However, it is difficult to implement the authentication method using dynamic data such as behavior biometrics because of the execution speed of server side script or the delay concerning the network. For this problem, the authentication method which uses client side scripts such as HTML5 or Java which can measure the dynamic data in real time is considered to be effective.

For implementing authentication system using HTML5, the security of the data used for authentication should

be considered. The simplest implementation is shown in Fig.2. When the user requests the authentication, the WEB

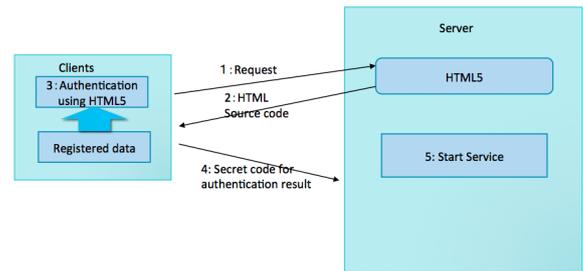


Figure 2: HTML5 implementation of authentication system(1)

application including the authentication program is sent from the server to the user client, and executed on the client. Then, the measured biometrics and the registered biometrics stored in the local clients is compared, and if they are matched, the secret code which notifies the success of authentication is sent to the server, and finally the server starts service to the user. This method is simple, and the authentication data can be stored in the local clients. Considering the leakage of the authentication data from server, it is ideal to store authentication data in local clients. However, this method is critical because HTML 5 source program used for authentication can be easily obtained. If the source code is obtained, it is easy to make dummy program which mimics the successful authentication. If the HTML 5 source code can be encrypted, and can be executed directly in browser using encrypted code, this problem can be cleared, however the encryption of source code is not implemented to HTML 5 yet.

For this problem, the authentication method shown in Fig,3 is proposed. When the user requests the authentication,

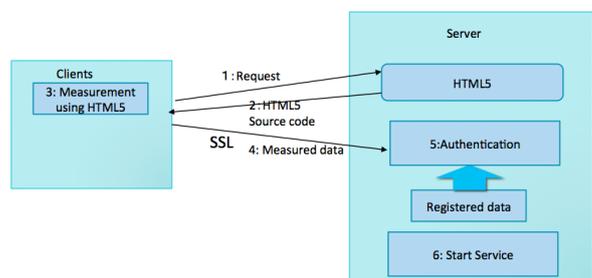


Figure 3: HTML5 implementation of authentication system(2)

the WEB application for measuring the behavior biometrics is sent to the client from server and executed on the client. Then, the biometric data measured on the client is send

to the server via encrypted channel, and it is compared with the registered data stored in the server. If they are matched, the server starts service. Using this method, the WEB application sent from the server just measures the behavior biometrics, and send the data to server, thus the attacker can not mimic the successful authentication without the successful biometric data even if he can get the source code of HTML 5 program. For this system, the special care will be required for registering the biometric data on server. Because the attacker may mimic the user in registering the data. For this problem, the user should be authenticated using another method when he register the data.

Fig.4 shows the another implementation. The difference

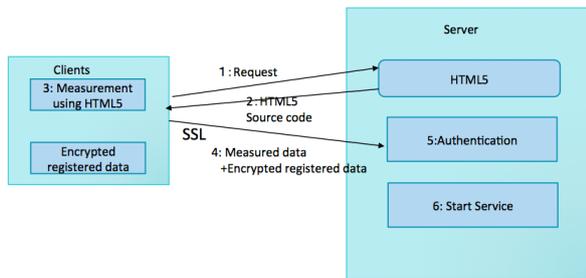


Figure 4: HTML5 implementation of authentication system(3)

from the system shown in Fig.3 is the place for storing registered data. The system in Fig.4 stores the registered data encrypted on the server in local clients, end it to the server, and the authentication is executed on the server. There are some discussions for which is better to store the data on the local clients and the server depending on the characteristics of systems. Our method can be implemented in both type. Of course, the registered data is needed to be encrypted on the server, and needed to be registered after authentication of the user using another method.

### 4. The System Which Generates Feature Points

In [1]. we reported a system which generates feature points. As shown in Figure 5, this system generates feature points automatically from the freehand curving line. These points are displayed on the touch panel during authentication to increase reproducibility of the registered curving line. With displaying the points without curving line, the pattern of connecting the points is used as knowledge factor for authentication. The pen speeds between the points are used as behavior factor for biometric authentication. The feature points are generated on the points which have large feature value. We used 2 factors to detect feature points. The first factor is inner products between the scanned points, and the sharp change of the direction of the line is detected. The

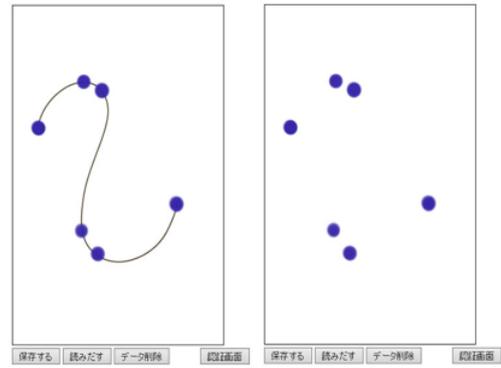


Figure 5: The system which generate feature points

second factor is curvature of the line, and the loose change of the direction is detected. After detecting feature points, the dense feature points are integrated as shown in Fig.6 because too many feature points disturb the smooth drawing of the line.

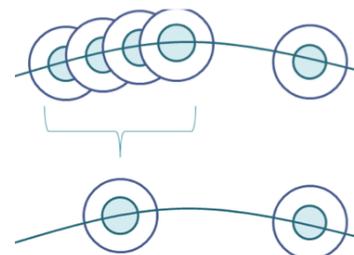


Figure 6: Integration of the feature points

The advantage of this method is small number of parameters compared with conventional method. The conventional method for matching handwritten patterns with registered data uses all scanned points. Thus, it needs complex computations, and needs the transmission of large data for authenticating on the server. Our method uses some parameters which represent the pen speed between feature points, thus the computation becomes simple, and needs to transmit small number of data.

### 5. Experimental Results

#### 5.1 Experimental results using HTML 5

To examine the detection of feature points using HTML5, the experiments are conducted. In this experiments, the server-client model is not used. The experiments of detecting the feature points and authentications are conducted in the client using HTML 5.

As mentioned in the previous section, the strength of the security of the connecting pattern of feature points as knowledge factor and the security of the behavior biometrics of the pen speeds between the feature points are examined with the experiments of authentications. Both of them are obtained clients and authenticated using HTML 5.

As the integrated environment for developing HTML 5 program, we used ApranaStudio 3. And we specified Chrome as default browser because the operation of HTML 5 is different for each browser.

In the experiments, one user, who is one of the authors, registered his handwritten pattern, and the registered user and another 3 users ( User A, User B and User C) examined the authentication. To compare the strength of security of the combinations of knowledge factor and behavior biometrics with that of only using behavior biometrics, the experiments are conducted without notifying connection pattern, and with notifying connection pattern. Figure 7 shows the registered patterns. For Pattern 1, pattern 2 and pattern 3, four, six and

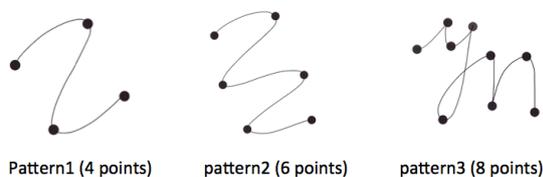


Figure 7: Registered patterns and generated feature points

nine points are generated, and the number of feature values used for authentication are three, five and seven respectively. The difference of the authentication results with number of feature points is examined.

Table 1 shows the results using both of the knowledge factor of connection patterns and biometric features of pen speeds without notifying connecting patterns. For all

Table 1: Result of authentication without notifying connecting patterns

	Registered user	User A	User B	User C
Pattern1 (4 points)	1.0	0.0	0.0	0.0
Pattern2 (6 points)	0.8	0.0	0.0	0.0
Pattern3 (9points)	0.5	0.0	0.0	0.0

user, the authentications are repeated in ten times. Without notifying connecting patterns, all of non-registered users could not find right connecting patterns, so none of them can successfully authenticated for all patterns. However, the rates of successful authentication decreases with increasing the number of feature points.

Table 2 shows the results using biometric features only with notifying connecting patterns. Without knowledge factor, the strength of security for the non-registered users

Table 2: Result of authentication with notifying connecting patterns

	Registered user	User A	User B	User C
Pattern1 (4 points)	1.0	0.5	0.3	0.7
Pattern2 (6 points)	0.8	0.2	0.0	0.4
Pattern3 (9points)	0.5	0.0	0.0	0.0

becomes worse. Especially for Pattern 1 with small number of feature points, almost half of authentication of non-registered users are successful, because too small number of feature values (3 values) is considered to be useless. With increasing the number of feature points, non-registered users become hard to be authenticated. However, for pattern3, the registered user can authenticated only half of trials. Among the non-registered users, the rates of User C is higher than User A and User C, because User C was authenticated just after seeing the successful authentication of registered user.

### 5.2 Experimental results using HSP

For the comparison, we conducted the authentication experiments using HSP scripting language. HSP is the scripting language which is implemented to multi-platforms such as Windows, Linux, MacOS, and it can be used on the mobile devices such as android and iOS using mobile version of HSPDISH. HSP have many kind of instruction set, and it is relatively easy to make the application which uses system level operations. HSP can even obtain the pen pressures if the device is equipped with the touch panel which can detect pen pressure. However, HSP is basically executed in the local system as application, thus it can not be used for implementing WEB application. As one of script interpreter, we compared the results of HTML with that of HSP.

Table 3 shows the results using both of the knowledge factor of connection patterns and biometric features of pen speeds without notifying connecting patterns using HSP scripts. The users and registered patterns are not identical

Table 3: Result of authentication without notifying connecting patterns

	Registered user	User A	User B	User C
Pattern1 (6 points)	1.0	0.0	0.0	0.0
Pattern2 (7 points)	0.9	0.0	0.0	0.0
Pattern3 (8 points)	0.7	0.0	0.0	0.0

to those used in the previous experiments. Without notifying connecting patterns, non-registered user can not found connecting patterns as same as in the previous experiments. The rate of acceptance of registered user was better than that of previous experiments.

Table 4 shows the results using biometric features of pen speeds only with notifying connecting patterns using HSP scripts. There is not clear difference between the results using HTML 5 and HSP, however the results of

Table 4: Result of authentication without notifying connecting patterns

	Registered user	User A	User B	User C
Pattern1 (6 points)	1.0	0.3	0.1	0.4
Pattern2 (7 points)	0.9	0.2	0.0	0.3
Pattern3 (8points)	0.7	0.0	0.0	0.0

HSP is considered to be a little biased to acceptance of authentication. The difference comes from the tuning of the parameters. Considering this point, the performance of HTML5 and HSP is considered to be almost same.

## 6. Conclusion

In this paper, we proposed the biometric authentication method using behavior biometrics in WEB Application. The authentication system generates the feature points automatically, and both of the knowledge factor of the connecting pattern of feature points and the biometric features of the pen speed between feature points are used for authentication. For this system, we proposed the authentication system which comprised of the authentication server and client which measures the biometric feature using HTML 5 in real time. We examined the performance of generating feature points using HTML 5, and conducted authentication experiments. The performance of the authentication is almost same as those of HSP which is implemented natively to the computer. From the experiments, non-registered users can not be authenticated without notifying connecting pattern, and they can be partially authenticated with notifying pattern. However this method may be the abuse to the hacker even if the connecting pattern is known,

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, we use pen speeds only as behavior biometrics. Using another biometrics, such as curvature between feature points, may help to improve the accuracy.

## References

- [1] Hiroshi Dozono, Yuuki Inaba and Masanori Nakakuni: Biometric Authentication System That Automatically Generates Feature Points, Proceedings of the International Conference on Security and Management 2013(2013)
- [2] F. Monrose and A.D. Rubin: Keystroke Dynamics as a Biometric for Authentication, Future Generation Computer Systems, March(2000).
- [3] 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)
- [4] 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)
- [5] 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)

**SESSION**  
**SECURITY APPLICATIONS**

**Chair(s)**

**Dr. Hiroshi Dozono**  
**Saga University - Japan**

**Dr. Yun Bai**  
**Univ. of Western Sydney - Australia**



# An Approach and Its Implementation for Cloud Computing Security

Yun Bai<sup>1</sup>, Khaled M. Khan<sup>2</sup> and Chunsheng Yang<sup>3</sup>

<sup>1</sup>School of Computing, Engineering and Mathematics  
University of Western Sydney, Australia

Email: ybai@scem.uws.edu.au

<sup>2</sup>Department of Computer Science and Engineering, Qatar University, Qatar

Email: k.khan@qu.edu.qa

<sup>3</sup>National Research Council Canada, Ottawa, ON, Canada

Email: Chunsheng.Yang@nrc-cnrc.gc.ca

**Abstract**—With the increasing popularity of the datacenter provided by Cloud computing, the datacenter security is becoming an important issue. Also there are other issues to be concerned about such as its speed and standard, but data security is the biggest one. When an organization uses a remote datacenter provided by Cloud, it needs to maintain the confidentiality of the outsourced data and ensure only authorized user or client is allowed to access its data resource. In this paper, we investigate the security issue related to datacenter of cloud computing and propose a security approach for it. We use a formal logical method to specify the data and employ intelligent agents to enforce appropriate security rules on it. We outline the authentication mechanism and present a detailed authorization or access control approach. The implementation of the proposed approach will be discussed and investigated. The authentication and the authorization mechanisms work together to prevent any unauthorized attempt to data stored in the datacenter.

**Keywords:** Access Control, Knowledge Representation, Knowledge-Based Systems, Cloud Computing security, Formal Specification <sup>1</sup>

## 1. Introduction

Cloud computing is a structure that provides efficient and cost-effective service, allows an organization to access applications, services or data that resides in a remote site. In this way, the organization can save cost in terms of infrastructure management, software development and the datacenter maintenance. Some major organizations such as Amazon, Google, Microsoft now offer cloud services to the public. Amazon offers virtual machines and extra CPU cycles as well as storage service. Google offers database, online documents, and other online software. Microsoft

provides the organizations with the Window applications, datacenters, database services [20].

However, the introduction of the cloud computing also brings major security concern about the organization's data hosted in a non-local datacenter. Generally, the success of cloud computing depends on three key issues: data security, fast Internet access and standardization [17]. Among the three issues, the biggest concern is data security. When an organization uses a remote datacenter provided by Cloud, it needs to maintain the confidentiality of the outsourced data and ensure only authorized user or client can access its data resource.

This paper is to address the security issue of the data hosted in the datacenter. We propose an approach to protect the datacenter security by using an authentication and authorization mechanisms. Authentication is a mechanism to identify users legitimacy and to ensure that only the legitimate users are allowed to access the data stored in the datacenter. Authorization or access control is a mechanism to control legitimate users' access on the data items, it ensures that the users are only allowed to access the data and to perform operations on the data according to the security policy and rules. The implementation of the approach will be investigated and discussed.

Many works such as [4], [25],[24] etc. studied authorizations or access control extensively and a variety of authorization specification approaches such as access matrix [6], [8], role-based access control [5], access control in database systems [3], authorization delegation [14], multiple access control policies [13], procedural and logical specifications [2] have been investigated. Since logic based approaches provide a powerful expressiveness [9] [16] as well as flexibility for capturing a variety of system security requirements, increasing work has been focusing on this aspect. However, how these approaches can apply to cloud computing environment is an area yet to be explored.

There are certain research works have been done in Cloud computing security. [12] discussed Cloud computing

<sup>1</sup>This publication was made possible by a grant from the Qatar National Research Fund under its NPRP Grant No. 09-079-1-013. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Qatar National Research Fund.

security requirements from a systematic point of view. It attempted to provide a roadmap for researchers on the topic of cloud computing security requirements and solutions. Several criteria and factors were discussed and investigated in achieving cloud computing security. [10] presented a design for a file system to secure file system storage service for Web 2.0 application. [15] proposed approaches to provide catch based security and performance isolation for cloud computing environment. In [22], a hierarchical attribute-based encryption for fine-grained access control in cloud storage services was proposed. This approach provides a method to help enterprise to efficiently share confidential data on cloud server. [21] also discussed a hierarchical attribute-based solution for access control in cloud computing. Even there are many cloud computing security research has been carried out, as far as we know, logic-based formal approaches in this area has not been extensively studied. This paper discusses cloud security issue from formal logic point of view.

In this paper, we investigate cloud computing security by using formal approaches. In section 2, we study the features of the cloud computing and propose a structure for secure datacenter with authentication and access control functions provided, the authentication function is also outlined; in section 3, we present a detailed authorization mechanism by using a formal logical specification, the function and evaluation of authorization request are discussed; in section 4 we discuss the implementation of the authorization function; conclusion and some future work are outlined in section 5.

## 2. Securing Datacenter Structure

In a cloud computing structure, the organizations using the cloud services are normally referred to as clients. These clients can be located geographically differently. Cloud computing provides various services to the clients [11] such as Software as a service (SaaS), Platform as a service (PaaS). Datacenter is an important service provided by cloud computing. It is a collection of data and servers which the clients subscribe. As an organization using datacenter service, its data and application are located on servers geographically different from its local site. As discussed previously, when a client's data is housed in a datacenter, it saves the cost of maintaining it. When client from different location needs to access its data, the data seems to be just located locally.

However, since the data is located in a non local site, ensuring the security of the data is not as simple as if housed locally [17]. It is not feasible to enforce the same local security measurement. To ensure safe access to the datacenter, it needs a coordinated security measurement between the datacenters and the clients accessing the datacenter.

For this purpose, we propose an approach to ensure the security of a system such as a datacenter hosted by cloud computing: authentication and authorization. Authentication

is used to authenticate the identities of the users, it controls who can access the datacenter. On the other hand, authorization is used to control that the legitimate user only performs legitimate operations on the data once it has been successfully authenticated. The two mechanisms work together to effectively provide the datacenter with secure accesses.

We use  $C_1, C_2, \dots, C_n$  to represent the client servers, each of them acts on behalf of a group of users. The users access the datacenter through their respective client servers. We also use  $AeS$  to represent the authentication server and use  $AoS$  to act as the authorization server. All users need to register with the  $AeS$  server, their information such as password, identification number, IP address, etc. are stored in a database which is managed by  $AeS$  server. This ensures that only the registered, legitimate users are allowed to access the datacenter. All users access right to certain data is also stored in a database which is managed by  $AoS$  server. This controls that the authenticated users only perform legitimate operations on the allowed data of the datacenter.

Here is the process when a user needs to access certain data of the datacenter. The user requests to its client server, the client server passes the information to  $AeS$ .  $AeS$  then checks the user's credential and request with its database if they match or not. If matches, the authentication is successful, it then passes on to the  $AoS$ .  $AoS$  checks the user requested operation and data with its database if they match or not. If the user request is within its specified right, the request is granted, otherwise, it is denied.

We will employ a kind of Kerberos [19] system to carry out authentication function. we will only illustrate the authentication process for one client server  $CS$ . All other client servers follow the same procedure for their user authentication.

To protect the messages between  $AeS$  and  $AoS$  servers, we assign both servers a secret key for encrypting and decrypting the messages between them. When a user  $U_1$  requests to access the data of the datacenter, it needs to be authenticated by  $AeS$  server then subsequently authorized by  $AoS$  server.

$U_1$  first sends its request to  $CS$ , the  $CS$  carries out a basic checkup, then sends the request to  $AeS$  server,  $AeS$  checks the information  $CS$  supplied about  $U_1$  against its database information about  $U_1$  if everything matches correctly. If matches, the authentication is successful, then  $U_1$  is a legitimate user and is issued a ticket encrypted by the secret key shared by  $AeS$  and  $AoS$  servers.  $U_1$  does not possess the secret key, hence cannot alter the ticket. It can only pass it to  $AoS$  server for requesting access to the datacenter.

To ensure that each data is only accessed by legitimate users and is manipulated by legitimate operations,  $AoS$  server manages the database about the users entitled access

rights to each data of the datacenter. When the *AoS* server receives the ticket from  $U_1$ , it decrypts the ticket by its secret key shared with the *AeS* server. Then compares the user's request to its database record about what operations the user can perform on which data. If the request is within the specification of  $U_1$ 's entitled rights, then  $U_1$  can access the datacenter as it requested. Otherwise the request is denied.

### 3. A Formal Method for *AoS*

Now, we present the detailed specification, function and evaluation of the *AoS* server.

In our approach, each client server provides data access service for a group of users. These users access the datacenter generally located at a remote site. We assume that the Internet on which the system relies is safe and sound. For the authorization mechanism, assume each agent manages one client server, a master agent takes overall control of all the agents. We concentrate on the investigation of a single agent by proposing a logic model for its specification and evaluation. All the other agents follow the same model.

We introduce a formal logic model for representing *AoS* security rules and policies based on a first order language. We explain syntactic and semantic descriptions for this policy base model.

$\mathcal{L}$  is used to represent a sorted first order language with equality, with four disjoint sorts for *legitimate object* with object constants  $O, O_1, O_2, \dots$ , and object variables  $o, o_1, o_2, \dots$ , *legitimate user* with user constants  $U, U_1, U_2, \dots$ , and user variables  $u, u_1, u_2, \dots$ , and *group legitimate user* with group user constants  $GU, GU_1, GU_2, \dots$ , and user variables  $gu, gu_1, gu_2, \dots$ , and *legitimate operation*  $OP, OP_1, OP_2, \dots$  to represent *read*, *write*, *update*, etc. respectively.  $\mathcal{L}$  also includes a ternary predicate symbol *request* which takes arguments as *legitimate user* or *group legitimate user*, *legitimate object* and *legitimate operation* respectively; A ternary predicate symbol *can* which takes arguments as *legitimate user* or *group legitimate user*, *legitimate object* and *legitimate operation* respectively; A binary predicate symbol  $\in$  which takes arguments as *legitimate user* and *group legitimate user* respectively; A binary predicate symbol  $\subseteq$  whose both arguments are *group legitimate user*; Logical connectives and punctuations: as usual, including equality.

Using language  $\mathcal{L}$ , a legitimate user  $U$  can *READ* legitimate object  $O$  of the datacenter is represented by a ground formula  $can(U, O, READ)$ . We generally use capital letters for constants and lower case letter s for variables. A ground formula is a formula without any variables. A user  $U$  requests access to object  $O$  of the datacenter by means of  $OP$  is represented by a ground formula  $request(U, O, OP)$ .

Now we explain the group membership concept. If  $GU$  is a group constant representing a specific group users called *DIRECTOR*. " $U$  is a director" means  $U$  is a member of the group  $GU$ , this can be represented using the formula  $U \in$

$GU$ . We can also represent inclusion relationships between different user groups such as  $GU_1 \subseteq GU_2$  which means that all members of  $GU_1$  is also a member of  $GU_2$ . Furthermore, we can represent constraints among users' authorizations. Certain access rights are defined by the roles such as a general staff can access general information, a group leader can access his group members personal information, etc. For example, the rule stating that "a director can update confidential file F", "Alice is a director", "manager can access confidential file F" can be represented as follows. We use  $D$  to represent the group director,  $M$  to represent the group manager.

$$\forall u.u \in D \supset can(u, F, UPDATE), \quad (1)$$

$$Alice \in D, \quad (2)$$

$$\forall u.u \in M \supset can(u, F, ACCESS), \quad (3)$$

When access rights are assigned to a group, that means all group members are entitled the access rights the group assigned unless otherwise specified. We define that if a group entitles access to certain object, then all the members of the group can access the same object unless otherwise specified. This is called the *inheritance* property of authorizations. This can be represented as:

$$\forall u.u \in GU \wedge can(GU, O, OP) \supset can(u, O, OP). \quad (4)$$

Where  $u$  represents any member of the group  $GU$  and  $O$  is a data object of the datacenter that  $GU$  can access,  $OP$  is the operation performed on  $O$ .

Here is the formal definition of the security rule base of the *AoS* by using language  $\mathcal{L}$ .

*Definition 1:* A security rule base *SRB* is a quintuple of  $(LU, LO, LOP, F, C)$  where  $LU$  is a finite set of legitimate users;  $LO$  is a finite set of legitimate data objects of the datacenter;  $LOP$  is a set of legitimate operations performed on the data;  $F$  is a finite set of ground literals and  $C$  is a finite set of closed first order formulas.

We define a formula with no free variables as a closed formula. In our formalism, both *facts* and *rule constraints* are represented by closed formulas of  $\mathcal{L}$ . For example,  $can(Alice, F, WRITE)$ ,  $Alice \in D$  are facts.  $can(Alice, F, READ) \supset can(Amy, F, READ)$ ,  $\forall u.u \in M \supset can(u, F, ACCESS)$ , are rule constraints. These rule constraints are viewed as access constraints which should be always satisfied. We refer to a fact as a ground formula, a ground literal, or an atom.

A *model* of a security rule base is the assignment of a truth value to every formula of the security rule base in such a way that all formulas of the security rule base are satisfied [7]. Formally, we give the following definition.

*Definition 2:* A *model* of a security rule base  $SRB = (LU, LO, LOP, F, C)$  is defined to be a Herbrand model [7] of  $LU \cup LO \cup LOP \cup F \cup C$ . *SRB* is said to be *consistent*

if there exists some model of  $SRB$ . The set of all models of  $SRB$  is denoted as  $Models(SRB)$ . A formula  $\psi$  is a consequence of  $SRB$ , denoted as  $SRB \models \psi$ , if  $LU \cup LO \cup LOP \cup F \cup C \models \psi$ . In this case, we also say  $\psi$  is satisfied in  $SRB$ .

The following examples show how the security rule base work.

*Example 1:* Assume both  $U_1$  and  $U_2$  are legitimate users,  $D$ -records is a legitimate data record of the datacenter, and  $OP$  is a legitimate operation. Both  $U_1$  and  $U_2$  belong to a group called  $STAFF$ . This group can perform  $OP$  on  $D$ -records. The constraint states that if someone belongs to a group then he/she inherits the group's access rights. In our security rule base, this situation can be specified as  $SRB = (LU, LO, LOP, F, C)$ , where

$$\begin{aligned} LU &= \{U_1 \in LU, U_2 \in LU\}, \\ LO &= \{D\text{-records} \in LO\}, \\ LOP &= \{OP\}, \\ F &= \{U_1 \in STAFF, U_2 \in STAFF, \\ &\quad can(STAFF, D\text{-records}), OP\}, \text{ and} \\ C &= \{\forall u, g, o, op. u \in g \wedge can(g, o, op) \supset \\ &\quad can(u, o, op)\}. \end{aligned}$$

It is not difficult to see that facts  $can(U_1, D\text{-records}, OP)$  and  $can(U_2, D\text{-records}, OP)$  are consequences of  $SRB$ , and  $SRB$  has a unique model  $m$  where:

$$\begin{aligned} m &= \{U_1 \in LU, U_2 \in LU, D\text{-records} \in LO, \\ &\quad OP \in LOP, U_1 \in STAFF, U_2 \in STAFF, \\ &\quad can(STAFF, P\text{-records}, OP), \\ &\quad can(U_1, D\text{-records}, OP), can(U_2, D\text{-records}, OP)\}. \end{aligned}$$

*Example 2:* Suppose one more constraint is added to example 1 as “ $U_1$  and  $U_2$  are claimed to be conflict of interest, they are not allowed to access the same data object”. In this case, the security rule base is specified as:  $SRB = (LU, LO, LOP, F, C)$ , where

$$\begin{aligned} LU &= \{U_1 \in LU, U_2 \in LU\}, \\ LO &= \{D\text{-records} \in LO\}, \\ LOP &= \{OP\}, F = \{U_1 \in STAFF, U_2 \in \\ &\quad STAFF, can(STAFF, D\text{-records}), OP\}, \text{ and} \\ C &= \{\forall u, g, o, op. u \in g \wedge can(g, o, op) \supset \\ &\quad can(u, o, op), \\ &\quad \forall o, op. can(U_1, o, op) \supset \neg can(U_2, o, op)\}. \end{aligned}$$

Two models will yield as:

$$\begin{aligned} m1 &= \{U_1 \in LU, U_2 \in LU, D\text{-records} \in LO, \\ &\quad OP \in LOP, U_1 \in STAFF, U_2 \in STAFF, \\ &\quad can(STAFF, P\text{-records}, OP), \\ &\quad can(U_1, D\text{-records}, OP), \neg can(U_2, D\text{-records}, OP)\}. \end{aligned}$$

or

$$\begin{aligned} m2 &= \{U_1 \in LU, U_2 \in LU, D\text{-records} \in LO, \\ &\quad OP \in LOP, U_1 \in STAFF, U_2 \in STAFF, \\ &\quad can(STAFF, P\text{-records}, OP), \\ &\quad can(U_2, D\text{-records}, OP), \neg can(U_1, D\text{-records}, OP)\}. \end{aligned}$$

Now we discuss the evaluation of  $SRB$ . When a user  $U$  requests access to  $D$  of the datacenter by certain operation, the task of the  $AoS$  is to evaluate such a request and determine either to grant or deny the request.

For a request  $request(U, O, OP)$ , Generally, the  $AoS$  will first check its corresponding  $SRB$  to find out if  $U$ ,  $O$  and  $OP$  are legitimate user, data object and operation or not. If yes, it then checks the facts of the  $SRB$ , if  $can(U, O, OP)$  presents,  $request(U, O, OP)$  is explicitly granted. Otherwise, it does reasoning about the related facts and rules, calculates the model of the  $SRB$ . If  $can(U, O, OP)$  is entailed in the model, then  $can(U, O, OP)$  can be deduced, hence the request is implicitly granted; otherwise, the request is denied.

*Definition 3:* For an access request  $request(U, O, OP)$ , the  $AoS$  evaluates the  $SRB = (LU, LO, LOP, F, C)$  by calculating its model  $m$ . If  $can(U, O, OP) \in m$ , or  $SRB \models can(U, O, OP)$ ,  $request(U, O, OP)$  is to be granted; otherwise, it is to be denied.

*Example 3:* The  $SRB$  is as described as in Example 1. In addition,  $U_3$  is also a legitimate user. The access requests are:  $request(U_1, D\text{-records}, OP)$  and  $request(U_3, D\text{-records}, OP)$ . In this case, the  $SRB = (LU, LO, LOP, F, C)$ , where

$$\begin{aligned} LU &= \{U_1 \in LU, U_2 \in LU, U_3 \in LU\}, \\ LO &= \{D\text{-records} \in LO\}, \\ LOP &= \{OP\}, \\ F &= \{U_1 \in STAFF, U_2 \in STAFF, \\ &\quad can(STAFF, D\text{-records}, OP)\}, \text{ and} \\ C &= \{\forall u, g, o, op. u \in g \wedge can(g, o, op) \supset \\ &\quad can(u, o, op)\}. \end{aligned}$$

Again, the unique model  $m$  is:

$$\begin{aligned} m &= \{U_1 \in LU, U_2 \in LU, U_3 \in LU, \\ &\quad P\text{-records} \in LO, OP \in LOP, U_1 \in STAFF, \\ &\quad U_2 \in STAFF, can(STAFF, P\text{-records}, OP), \\ &\quad can(U_1, D\text{-records}, OP), can(U_2, D\text{-records}, OP)\}. \end{aligned}$$

Obviously,  $SRB \models can(U_1, D\text{-records}, request(U_1, D\text{-records}, OP))$  is granted;  $SRB \not\models can(U_3, D\text{-records}, OP)$  does not hold, so  $request(U_3, D\text{-records}, OP)$  is denied.

## 4. The Implementation of $AoS$

From the above evaluation process, we can see that the major implementation issue is the model generation. For a given  $SRB$ , once its set of model is computed, any access request will be answered just by checking the model set.

Since the sets of  $LU$ ,  $LO$  and  $LOP$  represent the set of legitimate users, the set of legitimate data objects and the set of legitimate operations. Their truth values are specified by the security agents at the beginning and can be verified easily. In the following discussion, we will skip the verification of these two sets and concentrate on the sets of  $F$  and  $C$ .

For a security rule base  $SRB = (LU, LO, LOP, F, C)$ , generally the constraints in  $C$  may include universally quantified variables<sup>2</sup>. From the implementation consideration, we need to *ground* each constraint containing variables in  $C$  to all of its propositional instances. This technique is often used in the implementation of first order dynamic systems, eg.[23].

In Example 1,  $C$  contains one constraint:

$$\forall u, g, o, op. u \in g \wedge can(g, o, op) \supset can(u, o, op).$$

During implementation, this constraint needs to be replaced by its two ground instances:

$$\begin{aligned} U_1 &\in STAFF \wedge can(STAFF, D-records, OP) \\ &\supset can(U_1, D-records, OP), \text{ and} \\ U_2 &\in G \wedge can(STAFF, D-records, OP) \supset \\ &can(U_2, D-records, OP). \end{aligned}$$

In the rest description, when we refer to the security rule base  $SRB = (LU, LO, LOP, F, C)$ , we assume that the set of  $C$  only consists constraints without variable occurrence.

From the implementation point of view, we need to define some additional concepts that will be used in our algorithms. For a security rule base  $SRB = (LU, LO, LOP, F, C)$ , an *inconsistency* is a set of literals whose conjunction is inconsistent with  $C$ . A *minimal inconsistency* is an inconsistency which has no subset that is also an inconsistency.

Let  $L$  be the set of all ground literals of the language defined. To get the set of models of  $SRB = (LU, LO, LOP, F, C)$ , first we need to find out the set  $\mathcal{I}$  of minimal inconsistencies between  $L$  and  $C$ . This can be achieved using an inference engine. Given  $L$  and  $C$ , we can use the resolution proof to find out all of the minimum-length proofs which lead to empty clauses, and the required inconsistencies can be directly read off from these proofs.

For instance, let us consider Example 1 and see how one can obtain the set  $\mathcal{I}$  of minimal inconsistencies. To simplify the problem, let  $a$  stand for  $U_1 \in STAFF$ ,  $b$  for  $U_2 \in STAFF$ ,  $c$  for  $can(T, D-records, OP)$ ,  $d$  for  $can(U_1, D-records, OP)$  and  $e$  for  $can(U_2, D-records, OP)$ . Then this policy base can be viewed as  $SRB = (LU, LO, LOP, F, C)$  where  $F = \{a, b, c\}$  and  $C = \{a \wedge c \supset d, b \wedge c \supset e\}$ . Furthermore,  $a \wedge c \supset d$  is equivalent to  $\neg a \vee \neg c \vee d$  and  $b \wedge c \supset e$  is equivalent

to  $\neg b \vee \neg c \vee e$ . Here  $L = \{a, \neg a, b, \neg b, c, \neg c, d, \neg d, e, \neg e\}$ . The resolution proof of Figure 1 shows the procedures to obtain the set  $\mathcal{I} = \{\{a, c, \neg d\}, \{b, c, \neg e\}\}$  of minimal inconsistencies.

Once the set  $\mathcal{I}$  of minimal inconsistencies between  $L$  and  $C$  is obtained, a model  $m$  of  $SRB$  can be achieved by a maximal subset of  $L$  which contains  $F$  but does not contain any minimal inconsistency.

For the above example, considering the first inconsistency  $\{a, c, \neg d\}$ , we get models  $\{a, b, c, d\}$  and  $\{a, b, c, d, e\}$ . Taking the second inconsistency  $\{b, c, \neg e\}$  into account, leaves us the only model  $\{a, b, c, d, e\}$ . That is,

$$\{U_1 \in STAFF, U_2 \in STAFF,$$

$can(STAFF, D-records, OP),$

$can(U_1, D-records, OP), can(U_2, D-records, OP)\}$ .

Figure 1 illustrates the steps for the resolution proofs.

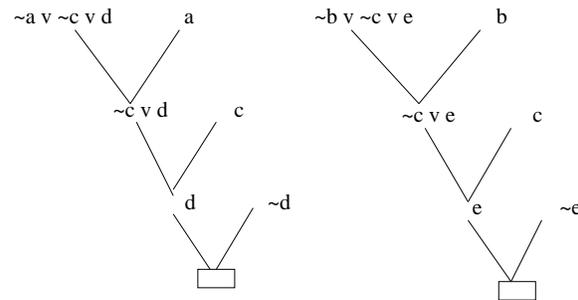


Fig. 1: Resolution Proofs

In summary, the algorithm is as follows:

### Model Generator Algorithm

*Input:* a finite set  $F$  of ground literals, a finite set  $C$  of ground formulas and a finite set  $L$  of ground literals over  $F \cup C$ .

*Output:* a finite set of models of  $F \cup C$ .

- 1) Use resolution proof to find the set  $\mathcal{I}$  of minimal inconsistencies between  $L$  and  $C$ .
- 2) Find the maximal subset of  $L$  which contains  $F$  but does not contain any inconsistency of  $\mathcal{I}$ .
- 3) From the set of all such maximal subsets of  $L$  to form the set of models of  $F \cup C$ , i.e.,  $Models(F \cup C)$ .

In the above algorithm, we achieve step 1 using a theorem prover OTTER [18]. In fact, step 1 can be pre-computed as a separate procedure for finding the set of minimal inconsistencies between  $L$  and  $C$ .

## 5. Conclusions

In this paper, we have examined the security issue of the datacenter in cloud computing environment. We proposed

<sup>2</sup>Technically, an existential quantifier in a formula can be eliminated by introducing Skolem function [7].

a structure to authenticate the users and to control their accesses in order to protect the datacenter from malicious attempt. We have sketched a framework for the authentication process and introduced a detailed formal approach for the access control mechanism. We investigated a logic approach for representing authorization rules and evaluating user's access request.

This paper is the extension of our previous work [1] where *legitimate operation* has been added and two binary predicates *request* and *can* have been augmented to ternary predicates in order to better capture the essence of the security rules and policies. And also the implementation issue has been considered and discussed. The resolution proof is employed into the implementation process. However, more detailed access rights, different operations on data object need to be investigated. This part will be investigated in our future work.

## References

- [1] Y. Bai and S. Policarpio, On cloud computing security. *Proceedings of the International Workshop on Communications Security & Information Assurance*, pp388–396, 2011.
- [2] E. Bertino, B. Catania, E. Ferrari and P. Perlasca, A logical framework for reasoning about access control models, *ACM Transactions on Information and System Security*, Vol.6, No.1, pp71–127, 2003.
- [3] E. Bertino, S. Jajodia and P. Samarati, Supporting multiple access control policies in database systems, *Proceedings of IEEE Symposium on Research in Security and Privacy*, pp94–107, 1996.
- [4] J. Chomicki, J. Lobo and S. Naqvi, A logical programming approach to conflict resolution in policy management, *Proceedings of International Conference on Principles of Knowledge Representation and Reasoning*, pp121–132, 2000.
- [5] J. Crampton and H. Khambhammettu, Delegation in role-based access control. *International Journal of Information Security*, Vol.7, pp123–136, 2008.
- [6] M. Dacier and Y. Deswarte, Privilege graph: an extension to the typed access matrix model, *Proceedings of European Symposium on Research in Computer Security*, pp319–334, 1994.
- [7] S.K. Das, *Deductive Databases and Logic Programming*, Addison-Wesley Publishing Company, UK, 1992
- [8] D.E. Denning, A lattice model of secure information flow, *Communication of ACM*, Vol.19, pp236–243, 1976.
- [9] R. Fagin, J.Y. Halpern, Y. Moses and M.Y. Vardi, *Reasoning about knowledge*. MIT Press, 1995.
- [10] F. Hsu and H. Chen, Secure File System Services for Web 2.0 Application. *ACM Cloud Computing Security Workshop*, pp11–17, 2009
- [11] J. Hurwitz, R. Bloor, M. Kaufman, F. Halper, *Cloud Computing for Dummies*, Wiley Publishing Inc., 2010.
- [12] I. Iankoulova, M. Daneva, Cloud Computing Security Requirements: a Systematic Review. *Proceedings of International Conference on Research Challenges in Information Science*, pp1–7, 2012.
- [13] S. Jajodia, P. Samarati, M.L. Sapino and V.S. Subrahmanian, Flexible support for multiple access control policies. *ACM Transactions on Database Systems*, Vol.29, No.2, pp214–260, 2001.
- [14] T. Murray and D. Grove, Non-delegatable authorities in capability systems. *Journal of Computer Security*, Vol.16, pp743–759, 2008.
- [15] H. Raj, R. Nathuji, A. Singh and P. England Resource management for Isolation Enhanced Cloud Services. *ACM Cloud Computing Security Workshop*, pp77–84, 2009
- [16] R. Reiter, A logic for default reasoning. *Artificial Intelligence*, **13** (1980) 81–132.
- [17] J. w. Rittinghouse, J. F. Ransome, *Cloud Computing, Implementation, management, and Security*, CRC Press, 2010.
- [18] S. Russell and P. Norrig, *Artificial Intelligence - A Modern Approach*. Prentice Hall, 1995.
- [19] W. Stallings, *Cryptography and Network Security - principles and Practice*, 5th edition, Pearson, 2006.
- [20] A. T. Velte, T. J. Velte, R. Elsenpeter, *Cloud Computing - A Practical Approach*, McGraw Hill, 2010.
- [21] Z. Wan, J. Liu, R. Deng, HASBE: A Hierarchical Attribute-Based Solution for Flexible and Scalable Access Control in Cloud Computing. *IEEE Transactions on Information Forensics and Security*, Vol.7, No.2, pp743–754, 2012.
- [22] , G. Wang, Q. Liu, J. Wu, Hierarchical Attribute-Based Encryption for Fine-Grained Access Control in Cloud Storage Services, *Proceedings of ACM Conference on Computer and Communication Security*, pp735–737, 2010.
- [23] M. Winslett, *Updating Logical Databases*. Cambridge University Press, New York, 1990.
- [24] T.Y.C. Woo and S.S. Lam, Authorization in distributed systems: A formal approach, *Proceedings of IEEE Symposium on Research in Security and Privacy*, pp33–50, 1992.
- [25] J. Zhou and J. Alves-Foss, Security policy refinement and enforcement for the design of multi-level secure systems, *Journal of Computer Security*, Vol.16, pp107–131, 2008.

# Research on the Security of OAuth-Based Single Sign-On Service

R. Zhu<sup>1,2</sup>, J. Xiang<sup>1,2</sup>, and D. Zha<sup>3</sup>

<sup>1</sup>Data Assurance and Communication Security Research Center, CAS, Beijing, China

<sup>2</sup>State Key Laboratory of Information Security, Institute of Information Engineering, CAS, Beijing, China

<sup>3</sup>University of Chinese Academy of Sciences, Beijing, China

**Abstract**—*OAuth 2.0 is an open standard for authorization, and provides a method for third-party applications to access users' resources on the resource servers without sharing their login credentials. It is widely used in Single Sign-On (SSO) service due to its simple implementation and compatibility with a diversity of the third-party applications. It has been proved secure in several formal methods, but some vulnerabilities are exposed in practice. In this paper, we propose a general approach to analyze the security of OAuth-based SSO service. From the perspective of the parameters and flows defined in the protocol, we conduct firstly a careful analysis of its security and design five potential attacks. Then, we examine the typical identity providers (including qq.com, weibo.com, baidu.com and renren.com) and 50 relying parties (such as dianping.com, juemei.com). The results indicate several problems existing in the implementation details, such as the access token leakage. In conclusion, we come up with six recommendations in order to improve the OAuth-based SSO Service.*

**Keywords:** Single Sign-On; OAuth 2.0; Security Detection

## 1. Introduction

With the development of the Internet, especially the wide use of web 2.0, web applications (such as online shopping, instant messaging and wiki) have become an important part of our lives.

Meanwhile, due to the requirement of user management, most of web applications need the authentication and authority. At present, most applications identify a user by password, which results in the increasing passwords and memory load. Besides, it is tiresome for users to enter the password frequently. In order to improve the situation, Single Sign-On (SSO) service emerges. In the SSO, users are allowed to access several Relying Parties (RP for short, such as *dianping.com*) while logging into Identity Provider (IdP for short, such as *weibo.com*) only once, which is obviously convenient for users.

OAuth 2.0 [1], OAuth for short, is widely used in SSO service due to its simple implementation and compatibility with a diversity of third-party applications compared with other protocols (such as OpenID [2] and SAML [3]). As an open standard for authorization, OAuth provides a method for third-party applications to access users' resources on the resource servers without sharing their login credentials.

Although OAuth has been proved secure in several formal methods, the formal analyses were conducted in abstract models but not in the implementations. And several empirical studies have revealed a few vulnerabilities the implementation details, but these empirical studies focused on the case studies, which can not cover all the implementations. Hence, it is still in urgent need of a general approach to detect the security of SSO services.

In this paper, we propose a general approach to analyse the security of OAuth-based SSO service. Firstly, we elaborate the parameters and the flows defined in the protocol, and design five attacks in some presuppositions. Then the approach containing nine detection items is provided to check the presuppositions. Lastly, we conduct an experiment on some typical IdPs and RPs. The results indicate that the approach is available and easy-to-use.

## 2. Background and Related Work

In OAuth-based SSO service, a resource sever is viewed as an IdP to manage users' identities and to identify the user. While a third-party application is viewed as a RP to serve users, and authenticate the user by user's profile gotten from IdP. In brief, the user's profile hosted on IdP is authorized and shared for RP to identify the user. Note that IdP and RP are relative concepts, namely, one could be either IdP or RP in different scenarios.

### 2.1 Authentication Flows in OAuth

OAuth-based SSO service is implemented by browser redirection. When a user logs into RP with an account on IdP, RP requests IdP to authenticate the user by redirecting user's browser to IdP. And IdP identifies the user before returning the authentication result to RP. During the identification, if the user authorizes RP to access his/her resource, IdP will issue RP an access token which is used to make IdP's API calls for accessing or handling user's resource hosted on IdP. In other words, RP authenticats the user with the user's profile.

Two SSO flows are defined in OAuth: the server flow (the "Authorization Code Grant" in the specification) and the client flow (the "Implicit Grant"). They are identified by the parameter *response\_type*. For the server flow with *response\_type = code*, illustrated in Fig. 1, RP gets an access token from IdP directly by presenting the shared secret with IdP.

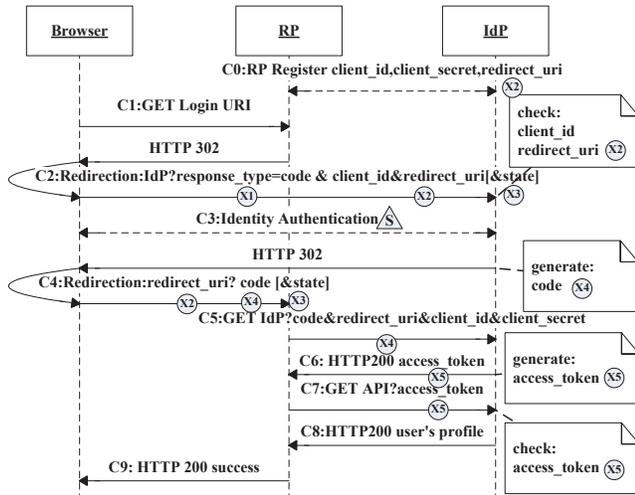


Fig. 1: The Server Flow in OAuth.

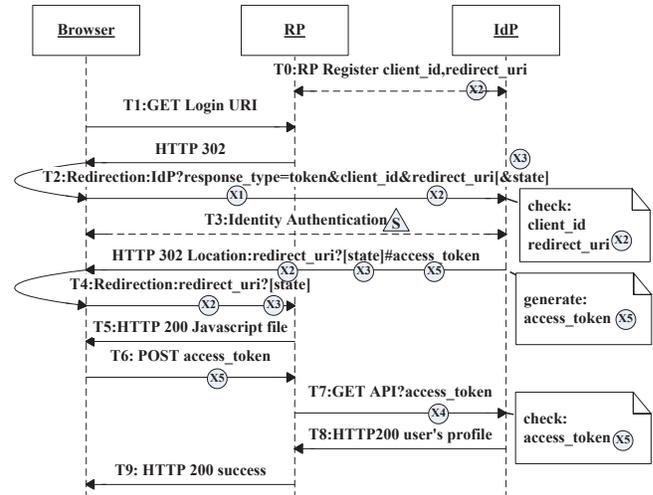


Fig. 2: The Client Flow in OAuth.

- C0: **RP Register:** RP submits its profile including a redirection URI *redirect\_uri* to IdP for register, and receives a client ID *client\_id* and a shared secret *client\_secret* from IdP.
- C1: **Login Request:** Browser (or a user) sends a HTTP request to RP after the user clicks the login button.
- C2: **Authentication Request:** RP responds a redirection message with the parameters to IdP including *response\_type=code*, *client\_id*, *redirect\_uri* and an optional parameter *state*.
- C3: **Identity Authentication:** After checking *client\_id* and *redirect\_uri* against its own local storage, IdP responds a login page to identify the user. And then an authentication session is created, namely session *S*. This step could be omitted if the session has been created.
- C4: **Issuing Authorization Code:** IdP responds a redirection message with authorization code *code* and *state* (if presented) to the server identified by *redirect\_uri*.
- C5: **Access Token Request:** RP sends an access token request with the parameters including *code*, *redirect\_uri*, *client\_id* and *client\_secret*.
- C6: **Access Token Response:** After validating these parameters, IdP responds an access token *access\_token* to RP.
- C7: **User's Profile Request:** RP makes an IdP's API call with *access\_token*.
- C8: **User's Profile Response:** IdP validates *access\_token* and returns user's profile to RP.
- C9: **Login Response:** RP creates an authentication session.

For the client flow with *response\_type = token*, IdP sends directly an access token to user's browser after identifying the user. RP gets the access token from the browser using a script (e.g. a javascript file). Fig. 2 illustrates the client flow.

- T0: **RP Register:** RP submits its profile including a redirect URI *redirect\_uri* to IdP for register, but receives only a

- client ID *client\_id* from IdP.
- T1: **Login Request:** The same as C1.
- T2: **Authentication Request:** The same as C2 except *response\_type = token*.
- T3: **Identity Authentication:** The same as C3.
- T4: **Issuing Access Token:** IdP responds a redirection message with *access\_token* appended as an URI fragment of *redirect\_uri*. Note that the request to RP does not contain *access\_token* in the URI fragment.
- T5: **Extracting Access Token:** RP responds a script to extract *access\_token* in the fragment.
- T6: **Submitting Access Token:** The extracted *access\_token* is submitted to RP by the script automatically.
- T7-T9: The same as C7-C9 respectively.

## 2.2 Comparison of the Flows

In the both flows, the browser creates authentication sessions with both IdP and RP respectively in step C3/T3 and C9/T9. As long as possessing an access token, RP is able to make web API calls to access or handle the user's resource in step C7 or T7.

IdP could validate RP with the shared secret in the server flow, while IdP sends an access token appended as an URI fragment to user's browser without authenticating RP in the client flow, thus the latter is more vulnerable. What is worse, the user can hardly tell the difference of the both flows with the naked eye.

## 2.3 Related Work

In theory, several formal approaches have been used to examine the OAuth, such as Alloy framework used by Pai et al. [4], universally composable security framework used by Chari et al. [5] and Muiphi used by Slack et al. [6], and all the results were included in the official OAuth security guide [7]. Thus the implementation following the guide is secure

in theory. On empirical analysis, Wang et al. [8] focused on the actual web traffic going through the browser, and discovered several logic flaws in some SSO services (e.g. Google ID, Facebook), which are used by attackers to tamper with the authentication messages. Sun et al. [9] examined the implementation details of three major OAuth-based IdPs including Facebook, Microsoft and Google, and uncovered several vulnerabilities that allowed attackers to gain access to the user's profile and to impersonate the user on RP.

Actually, all the existing approaches are deficient. The formal analyses were conducted in the abstract models, and some implementation details could be left out, which has been proved by the empirical studies; while the existing empirical studies exposed the vulnerabilities in the case analyses, which could omit some proofs.

In OAuth 2.0, the cryptography is taken out, and its transport security depends on the protocol SSL/TLS or HTTPS. However, HTTPS protection has no effect on the messages hosted on the end-points (issuer, browser and receiver). Meanwhile, session management is a vital task during the authentication, especially the undetectable session *S* (cf. step C3 or T3). In this paper, we focus on the overall security of OAuth including transport security, end-point security and session security, and propose a general approach to detect the security of SSO. On protocol analysis, we examine the flows and five variable parameters, and summarize their usages, requirements and potential risks. On empirical study, 5 attacks based on protocol analysis are proposed, and 10 IdPs (including *qq.com*, *weibo.com*, *renren.com* and *baidu.com*) and 136 login flows using by 50 RPs (such as *dianping.com*, *jumei.com* etc.) are checked to validate the availability of the approach.

## 3. Methodology

### 3.1 Attack Model

It is assumed that 1) use's browser and computer are invulnerable, 2) both RP and IdP are not malicious, and 3) the direct communications between RP and IdP are secure. However, due to the universality of web vulnerabilities [10], IdP or RP may have some vulnerabilities, such as Cross-Site Scripting (XSS), Cross-Site Request Forgery (CSRF). The purpose of an attacker is to gain an unauthorized access to user's resources hosted on IdP or RP. In brief, the abilities of the attacker are limited as follows:

- The attacker is a web user who could set up a website, issue some links through the blog, microblog and email, and even launch an attack to a vulnerable website.
- The attacker is also a sniffer who could sniff the unencrypted traffic in the Internet, and tamper simply with the traffic.

### 3.2 Proposed Approach to Detect the Security of OAuth

Our analytic framework is to provide a general approach to detect the security of OAuth-based SSO service. It is a two-stage process including protocol analysis and empirical analysis, as illuminated in Fig. 3.

In the stage of protocol analysis, we review the protocol carefully to reveal the usages, requirements and potential risks of the five variable parameters and session *S* (*S* and *X1-X5* in Fig. 1 and Fig. 2).

Five attacks (identified by *A1* to *A5*) based on the former are provided in the latter stage, all of which have been proved available in the following experiment. And the analysis on the presuppositions of the five attacks reveals that we need only to execute a detection on item *I1* to *I9* as follows to evaluate the security of an SSO service:

- I1*: Whether HTTPS protection is deployed by RP.
- I2*: Whether an unpredictable state parameter is used by RP.
- I3*: Whether RP is prevented against CSRF attack.
- I4*: Whether RP stores the access token in the cookie or URI.
- I5*: Whether the code is cross in use.
- I6*: Whether IdP supports the two response types simultaneously and indiscriminately.
- I7*: Whether any redirection URI in the realm of RP could pass IdP's checking.
- I8*: Whether the token is a bearer token.
- I9*: Whether a mechanism to end the session *S* is provided.

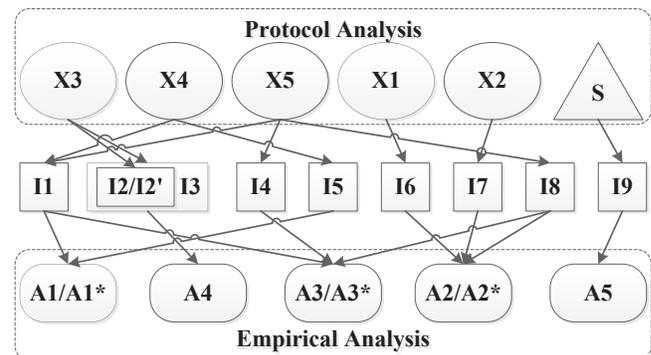


Fig. 3: Analytic Framework of Detecting the Security on Parameters and Session *S*.

Based on the analytic framework, the proposed approach can be described with the pseudo-code as follows:

```
AttackList Detection() {
    // Parameters X[1-5] and S are defined in Section 3.3;
    // Attacks A[1-5] are defined in Section 3.4;
    // The symbol * means the attack has greater damage.
    AttackList attacks = NULL; // all the potential attacks.

    if(CHECK(X4.I1: HTTPS is employed by RP) = TRUE) {
```

```

if(CHECK(X4.I5: The code is cross in use) = TRUE)
    attacks.Add(A1*);
else
    attacks.Add(A1);
}
if(CHECK(X1.I6: Two response types are supported by IdP)
    = TRUE AND
    CHECK(X5.I8: The token is a bearer token) = TRUE) {
    if(CHECK(X2.I7: The realm URIs are checked out by IdP)
        = TRUE)
        attacks.Add(A2*);
    else
        attacks.Add(A2);
}
if(CHECK(X5.I4: The access token is stored incorrectly
    by RP) = TRUE AND
    CHECK(X5.I8: The token is a bearer token) = TRUE) {
    if (CHECK(X5.I1: HTTPS is employed by RP) = FALSE)
        attacks.Add(A3*);
    else
        attacks.Add(A3);
}
if(CHECK(X3.I3:Protection against CSRF is employed by RP)
    = FALSE) {
    attacks.Add(A4);
    // check Item X3.I2' : Whether state is invalid.
    if(CHECK(X3.I2: The state parameter is used by RP)
        = TRUE)
        Warning(The state parameter(X3) is INVALID);
}
if(CHECK(S.I9:Ending session S is provided by RP)= FALSE)
    attacks.Add(A5);
return attacks;
}

```

### 3.3 Protocol Analysis on Parameters and Session

In the authentication flows, five variable parameters are defined, and two authentication sessions are created. In the rest part of this section, we discuss the usages, requirements and potential risks of the parameters and the session S.

#### X1: Response Type (*response\_type*)

**Usage:** It is set by RP and used to select which flow to be used in the following sequence, that is, to decide the way for RP to gain an access token.

**Risk:** As mentioned in 2.2, the both flows are similar in user's view, so attackers could set *response\_type* = *token*, and gain the access token from the browser through a script without user's awareness.

#### X2: Redirection URI (*redirect\_uri*)

**Usage:** It is set by RP and used to decide the destination of request C6 or T6, which receives the code or access token.

**Requirement:** IdP should check it strictly to avoid sending the code or access token to other receivers except RP.

**Risk:** The receiver, besides an attacker, of the code or access token could log into RP as the victim, and gain the victim's profile.

#### X3: State Parameter (*state*)

**Usage:** It is generated by RP for tracing the session between browser and RP to prevent against CSRF attack.

**Requirement:** It should be an unpredictable value bound to the session with RP.

**Risk:** It is an optional parameter. If it is lost, another countermeasure against CSRF attack **MUST** be taken specified in the protocol. In practice, CSRF attack is often neglected by developers.

#### X4: Authorization Code (*code*)

**Usage:** It is sent to RP in the server flow and used for RP to request an access token from IdP.

**Requirement:** It **MUST** be an expiring and one-time token, and be transported in a channel with HTTPS protection specified in the protocol.

**Risk:** If gaining it and using it before the victim, the attacker could impersonate the victim and log into RP, which has a greater risk if the code could be used for multiple RPs, e.g. the code sent to RP1 could be used to log into RP2.

#### X5: Access Token (*access\_token*)

**Usage:** It is generated by IdP and used for RP to make web API calls to gain or handle user's profile. RP use the received profile to authenticate the user.

**Requirement:** It should be an unpredictable value and not be sent to the others except RP.

**Risk:** Due to the decryptographic design of OAuth, the access token is often implemented as a bearer token. That is to say, any bearer of the token, e.g. an attacker, could access or handle user's profile.

#### S: Session with IdP

**Usage:** It is created by IdP in step C3 or T3 and used to authenticate the user for the latter login.

**Requirement:** It is often maintained through the cookie technology.

**Risk:** After created, the session keeps alive and is only bound to cookies invisible to the user, so other mechanisms should be employed to clear the session.

In conclusion, the misuse of each of X1 to X5 and S could result in insecure factors, thus understanding the strict requirements is a precondition for implementing an invulnerable OAuth-based SSO service.

### 3.4 Empirical Analysis

The former results reveal that the overall security requires competent developers to implement the SSO service. In practice, the simplicity of OAuth misleads the developers, and they often fail to make the implementations meet the requirements. Thus in this section, according to all the requirements, 10 IdPs and 136 authentications used by 50 RPs are examined, and the results indicate that risks lay in the carefulness though the OAuth-based SSO service is widely used.

**A1: Stealing and Embezzling Authorization Code**

In the server flow, HTTPS protection is used for all the communications with IdP, but not for ones submitting the authorization code to 90% of RPs (cf. Table 2) in step C4. Thus an attacker is able to steal the authorization code and makes the victim fail to submit it (e.g. by modifying the authorization code in the traffic). It allows the attacker to log into RP as the victim via the code.

Cross use of the code is not detected in our study, but it is reported that the authorization code generated by *weibo.com* for RP1 is allowed to be used to log into RP2 [11]. In this situation, the risk is of greater damage.

Interestingly, 30% of RPs protect their own login page with HTTPS, but only 9.56% of RPs provide HTTPS protection for the authorization code.

**A2: Stealing Access Token via XSS**

All the detected IdPs maintain the both flows and do not limit their RPs to select which flow to use. In this way, XSS vulnerability in *redirect\_uri* page on RPs using the client flow, as Sun et al.[9] put it, allows attackers to launch the client-flow sequence and to exact the access token via script in step T4.

Meanwhile, for RP with XSS vulnerability, the attacker could construct a link with *response\_type = token* and trick victims to click it, and the similar way could be used to gain the access tokens after victims enter their passwords.

What is worse, most of IdPs only check whether the redirection URI is in the realm of RP, which causes that this attack could be launched in case any page on RP has XSS vulnerability, that is to say, the attack surface is broadened.

**A3: Eavesdropping or Stealing Access Token**

The bearer token is deployed in all the detected IdPs. It is the only identity for RP to access user's profile, that is to say, the one who possesses the token could be viewed as the right RP to access user's profile. So RP is bound to guarantee its confidentiality.

In fact, a few RPs (nearly 9%, cf. Table 2) still keep the access token in the cookie or URI without any protections, hence it is likely to be stolen from the browser, e.g. by using script to read it from cookie without *httponly* attributes, or by examining the history of the browser. For RPs without HTTPS protection, even an attacker could eavesdrop on the open traffic to get the access tokens.

**A4: CSRF Attack**

It is specified in OAuth that CSRF attack, ranking 8 in Top-10 List issued by OWASP [10], must be handled. And the state parameter is recommended to handle it. Without a protection, an attacker could launch the attack as follows:

(a) The attacker starts a server-flow sequence with

his/her own account, but stops before step C4 and saves the request URI named *csrf\_uri*.

(b) *csrf\_uri* is issued in the Internet via blog or microblog, and a victim clicking the *csrf\_uri* will log into RP as the attacker.

(c) Due to the victim logs in as the attacker, the attacker may be able to record victim's information.

On the surface, the authorization code, after all, is for one-time use, so CSRF attack is inefficient and negligible. Nevertheless, binding with a local RP account is required in some RPs (e.g. *dianping.com*), so the victim may bind one's own RP account to the IdP account (or the attacker's account), in other words, the attacker may control victim's RP account via the IdP account.

The results manifest that only 44.12% of RPs take CSRF attack into consideration, and 39.71% adapt the recommendation using the state parameter. It is worth noting that a small part of RPs set the state parameter as a fixed value, and that about 25% of RPs do not check the parameter strictly and accept the request without the parameter after step C4, all of which fall flat against CSRF attack.

**A5: Risk of the Implicit Session with IdP**

As the session S (cf. Fig. 1 and Fig. 2) lays in the browser implicitly, the user can't end it by closing some page after login, which may initiate the following risk: The user logs into *dianping.com* with the account on *renren.com* on a common computer, as shown in Fig. 4, then logs out of *dianping.com* but not *renren.com* when leaving, that is, the session with *renren.com* is alive. The latter user on the same computer can log in to *dianping.com*, even other RPs, e.g. *jumei.com*, as the identity of the former user by just clicking a button.

For this reason, logging out of the IdP is a recommendation for a complete solution to refrain from the aforementioned risk. Nevertheless, none of the detected RPs maintains a mechanism to end session S though parts of IdPs offer an interface to log out.

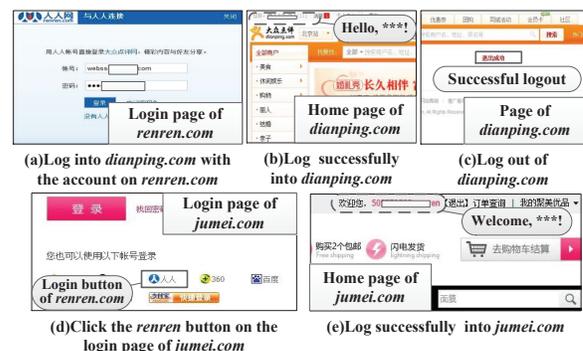


Fig. 4: Risk of the Implicit Session with IdP.

## 4. Experiment

### 4.1 Experimental Environment

An investigation to a portal web site containing 211 links shows that there are 196 RPs using the SSO service and 28 IdPs providing it, and the top four of the most popular IdPs are *qq.com*, *weibo.com*, *baidu.com* and *renren.com*.

In our experiment, we analyze 136 authentication sequences used by 50 RPs with the four IdPs, and 10 IdPs using OAuth.

### 4.2 Methods to Detect the Items

The 9 items could be divided into two categories: static items, including I1, I2 and I4, which could be checked by going through the communications in the sequences for "https", "state", "access token" and other related words; and dynamic items, including the rest of the items, which could only be checked by constructing or tampering with HTTP requests. Therefore, our detection contains three steps:

Table 1: Detection Methods of Each Dynamic Items.

Item	Detection Method
I3	(1) Log into RP using Firefox. (2) Stop in step C4, and save the request URI. (3) Send the request URI using IE. (4) Conclusion: If login succeeds using IE, NO protection against CSRF is deployed. * If the request URI contains the state parameter, continue the detection as follows: (5) Repeat step 1 and 2. (6) Send the request URI using IE without the state parameter. (7) Conclusion: If login succeeds, the state parameter is INVALID.
I5	(1) Log into RP using Firefox with <i>account1</i> . (2) Stop in step C4, and save the code in the request URI, named <i>code1</i> . (3) Log into RP using IE with <i>account2</i> . (4) Stop in step C4, and replace the code in the request URI with <i>code1</i> . (5) Send the constructed request URI. (6) Conclusion: If login succeeds, the code is cross in use.
I6	* Only use for RP deploying the server flow. (1) Construct the request C2 with <i>response_type = token</i> . (2) Send the request, and enter the username and password. (3) Conclusion: If an access token is returned, attack A2 is likely available.
I7	(1) Construct the request C2 with another redirection URI in the realm of RP. (2) Send the request. (3) Conclusion: If a login page is returned, the URI in the realm of RP could pass IdP's checking.
I8	* The access token could be gained by the way similar to I6 or by registering a RP on the test platform provided by IdP. (1) Review the API document offered by IdP. (2) Make an API call with the access token using the browser. (3) Conclusion: If the correct information is return, the token is a bearer token.
I9	(1) Click the login button on the page of RP, and complete the sequence. (2) Click the logout button(s) on the page of RP. (3) Re-click the login button on the page of RP. (4) Conclusion: If login succeeds without entering the password, session S is not ended after logging out of RP.

- 1) **Data Collection:** Log into RPs with an account on IdPs using Firefox, and record all the communications with Live HTTP Header plug-in.
- 2) **Data Analysis:** Depict the actual login sequence according to the communications, then review whether HTTPS is deployed by RP to submit the code, whether the state parameter is contained in the authentication request (I2), especially whether the state parameter is a fixed value, and whether the words, such as "access token", "access\_token" or other similar words, are contained in the cookie or URI (I4).
- 3) **Dynamic Detection:** Construct special requests conforming with the actual sequence to conduct the dynamic detection. Each detection methods are listed in Table 1.

## 4.3 Results

### 4.3.1 Result of Detection Items on RPs

50 RPs are checked on item I1 to item I4 and item I9 in the RP detection. The result is demonstrated in Table 2, which indicates 44 RPs with *qq.com*, 43 RPs with *weibo.com*, 25 RPs with *baidu.com* and 24 RPs with *renren.com*. We also find that 15 in 50 RPs (30%) deploy the HTTPS protection for their own login pages.

Table 2: Result of Detection Items on RPs

IdP	OAuth	HTTPS(I1)	State(I2)	Invalid State(I2')
<i>qq</i>	44	9.09%	59.09%	46.15%
<i>weibo</i>	43	9.30%	25.58%	9.09%
<i>baidu</i>	25	8.00%	52.00%	0.00%
<i>renren</i>	24	12.50%	16.67%	0.00%
Total	136	9.56%	39.71%	9.56%
IdP	CSRF(I3)	Token(I4)	Session(I9)	
<i>qq</i>	54.55%	2.27%	0	
<i>weibo</i>	25.58%	13.95%	0	
<i>baidu</i>	60.00%	8.00%	0	
<i>renren</i>	41.67%	12.50%	0	
Total	44.12%	8.82%	0	

HTTPS(I1): The percentage of RPs without deploying HTTPS;

State(I2): The percentage of RPs using the state parameter;

Invalid State(I2'): The percentage of RPs with the invalid state in State(I2);

CSRF(I3): The percentage of RPs with the protection against CSRF;

Token(I4): The percentage of RPs storing the token in the cookie or URI;

Session(I9): The percentage of RPs with a mechanism to end the session S is provided.

### 4.3.2 Result of Detection Items on IdPs

We only detect ten typical IdPs on item I5 to item I8 in the IdP detection. The result is enumerated in Table 3. Out

Table 3: Result of Detection Items on IdPs

Total	Code(I5)	response_type(I6)	rediert_uri(I7)	Bearer(I8)
10	0	10	10	10

Code (I5): The count of IdPs using the cross-use code;

Response type (I6): The count supporting the two response types simultaneously;

Redirection type (I7): The count s without checking the URI strictly;

Bearer (I8): The count using bearer token.

of our expectation, except item I6, all the others could be used by attacker to launch an attack.

## 5. Conclusion

In this paper, we introduce briefly two flows defined in OAuth. And based on the analysis of the parameters and session management, five attacks are carried out. Above all, a general approach to detect the security of OAuth-based SSO service is provided, and detections for both RPs and IdPs are conducted. Nevertheless, an automatic detection tool is unavailable, which will be the next focus.

Meanwhile, in order to protect the users' accounts, the security of the authentication is vital. However, the aforementioned analyses reveal that the implementations are vulnerable. As a result, several recommendations are proposed in order to improve the SSO service:

- 1) The confidentiality of the communications must be brought to the forefront. On one hand, HTTPS protection could be deployed. On the other hand, an alternative could be provided, e.g. during registration, a shared secret could be created between IdP and RP to encrypt the messages.
- 2) The protection against CSRF attack must be employed by RP. The state parameter must work if used.
- 3) Both RP and IdP should store the access token in a secure way, especially not expose it in the cookie or URI.
- 4) RP registers the response type and a specific redirection URI on IdP, while IdP validates the parameters strictly.
- 5) An interface should be provided by IdP to end the authentication session S, and RP should call the API when users log out.
- 6) The authorization code generated by IdP should be specialized to the right RP.

## 6. Acknowledgement

This work was supported by National Natural Science Foundation of China grant 70890084/G021102 and 61003274, Strategy Pilot Project of Chinese Academy of Sciences sub-project XDA06010702, and National High Technology Research and Development Program of China (863 Program, No. 2013AA01A214 and 2012AA013104).

## References

- [1] *The OAuth 2.0 authorization framework*, IETF Std. RFC6749, 2012.
- [2] F. Brad, R. David, H. Dick, and H. Josh. (2013) OpenID authentication 2.0-final. [Online]. Available: [http://openid.net/specs/openid-authentication-2\\_0.html](http://openid.net/specs/openid-authentication-2_0.html)
- [3] OASIS. (2013) SAML specifications. [Online]. Available: <https://wiki.oasis-open.org/security/FrontPage>
- [4] S. Pai, Y. Sharma, S. Kumar, R.M. Pai, and S. Singh, "Formal verification of OAuth 2.0 using Alloy framework," in *Proc. CSNT'11*, 2011, p. 655-659.
- [5] S. Chari, C. Jutla, and A. Roy. (2011) Universally composable security analysis of OAuth v2.0. [Online]. Available: <http://eprint.iacr.org/2011/526.pdf>
- [6] Q. Slack, and R. Frostig. (2011) OAuth 2.0 implicit grant flow analysis using Murphi. [Online]. Available: <http://www.stanford.edu/class/cs259/WWW11/>
- [7] *OAuth 2.0 Threat Model and Security Considerations*, IETF Std. RFC6819, 2013.
- [8] R. Wang, S. Chen, and X. Wang, "Signing me onto your accounts through Facebook and Google: A traffic-guided security study of commercially deployed single sign-on web services," in *Proc. SP'12*, 2012, p. 365-379.
- [9] S.T. Sun, and K. Beznosov, "The devil is in the (implementation) details: an empirical analysis of OAuth SSO systems," in *Proc. CCS'12*, 2012, p. 378-390.
- [10] OWASP. (2013) Open web application security project top ten project. [Online]. Available: [https://www.owasp.org/index.php/Category:OWASP\\_Top\\_Ten\\_Project](https://www.owasp.org/index.php/Category:OWASP_Top_Ten_Project)
- [11] lhshaoren. (2010) A vulnerability on *weibo.com* (*sina.com*). [Online]. Available: <http://www.wooyun.org/bugs/wooyun-2010-039727>

# Evaluating Gesture-Based Password And Impact of Input Devices

Lakshmidevi Sreeramareddy, Atcharawan Janprasert, and Jinjuan HeidiFeng

Computer and Information Sciences Department  
Towson University  
Towson, MD, USA

**Abstract**—Traditional alphanumeric passwords have both security and usability challenges. Passwords that are more secure are usually more difficult to remember. We propose a gesture-based password approach that may serve as a useful alternative authentication method. This password is potentially easier to remember than traditional alphanumeric passwords due to the picture superiority effect. This method also has potential security benefits over most recognition-based graphical passwords because it supports larger password space, which makes it more robust to dictionary attacks. In addition, this method requires shorter time to create and login than those reported for recognition-based graphical passwords. We expect this approach to be particularly beneficial on mobile devices with touch-screen input. We conducted a preliminary user study to evaluate the efficacy of this method using two input techniques: the mouse and the touch-screen. The preliminary result confirms the potential of the gesture password method.

**Keywords**—*security-application; Authentication; Gesture-based passwords; Touch screen*

## I. INTRODUCTION

Effective authentication mechanism is crucial in the Web environment and a variety of other domains. To date, the most commonly used authentication method is alphanumeric passwords, which have substantial challenges regarding security and usability [1]. Randomized alphanumeric passwords have larger password space and higher security, but are difficult to remember. In order to remember their passwords, people tend to use patterns, names, or English words in the passwords, making them vulnerable to dictionary attacks. We propose a gesture-based password approach that aims to reduce the memory load of the authentication process. This approach allows the users to draw an image freely without any guides or reference points as their password. In addition to the image drawn, we also examine additional measures such as stroke length, image size, angles, and drawing speed to authenticate the user [2].

We are planning a series of studies to evaluate the effectiveness of this method with different user populations and under various task scenarios. In this paper, we report the result of a preliminary user study that provides initial insight to the usability and security of the proposed method and the impact of different input devices on performance when creating and reproducing the password.

## II. RELATED RESEARCH

Present-day authentication systems primarily use alphanumeric passwords, which consist of numbers, alphabets and special characters. Alphanumeric passwords can provide stronger security if users choose randomly generated, longer passwords. However, randomly generated, longer passwords are hard to remember. Thus, users often tend to choose easy-to-remember passwords such as variations of their own names or meaningful words in the English dictionary [1]. Easy to remember passwords are also easy to guess, making many alphanumeric passwords vulnerable to dictionary attacks [1].

Graphical-based passwords are proposed as possible alternatives to alphanumeric passwords. They were developed aiming to achieve better usability while maintaining acceptable level of security. Graphical passwords vary by nature according to the type of memory retrieval leveraged in the password scheme. One category is recognition based passwords that provide a collection of image or images to the user. Examples include Déjà Vu [3], Passfaces [4], and VIP [5]. The other is recall based passwords that requires the users to recreate a pre-set drawing or click on points on an image. Most of the recall-based passwords that require drawing do not provide any cues, such as the DAS [6] and the PassShape [7] applications. The recall-base passwords that use points and clicks usually provide cues to assist the recall process. Examples include PassPoints that asks the user to click on points on an image [8]. More recently, PassTiles [9] was developed that supports combination of features of recognition-based passwords, completely recalled-based passwords, and cued recalled-based passwords.

Graphical passwords are believed to be easier to remember because they take advantage of the picture superiority effect [10], which shows that the humans can remember images better than they remember textual information. Studies suggest that graphical-based passwords may provide good usability and security support to the user [11]. Recently, graphical passwords have been increasingly adopted in the mobile environment, such as smartphones (e.g., Android Pattern Unlock). The disadvantage of many graphical passwords, especially recognition-based passwords, is the smaller password space, making them more vulnerable to brute force dictionary attack. Recognition-based graphical passwords also tend to require longer login times [9]. In addition, graphical passwords are also highly susceptible to shoulder surfing because the

password image is displayed on screen and can be easily seen by another person [12, 13].

The DAS (Draw A Secret) scheme developed by Jermyn et al. [6] is closely related to the proposed gesture-based password. A DAS password is a picture drawn on a grid mapped to a sequence of coordinate pairs using the cells through which the drawing passes. The DAS scheme was thought to be promising for two reasons. First, a DAS password based on a 5x5 grid offers a password space larger than that associated with typical textual passwords, thus offering the potential for more security than textual passwords. Second, DAS passwords are easily repeatable because users do not need to reproduce the exact same stroke. Instead, they can draw strokes that may differ visually as long as they enter the same cells in the same order. However, Nali and Thorpe [14] found that, similar to textual passwords, users tend to underutilize the password space made available through the DAS scheme. More specifically, users tend to draw DAS passwords using predictable characteristics that users find easy to recall (e.g., symmetry, centering within the grid) which greatly reduces the actual password space utilized, making the passwords more vulnerable to attack. This study also revealed a significant usability problem with DAS passwords. According to the original design, DAS passwords cannot follow the grid lines or cross grid corners, but 29% of the passwords users tried to create contained one of these two features and were therefore considered invalid. Rejecting one of four passwords a user tries to create is likely to lead to frustration, confusion, and the use of second-choice passwords. Alternatives to DAS proposed later have shown improvements but the two problems still exist [15, 16]

Two gesture password prototypes have been developed that take one or multiple strokes as passwords. The Passdoodles method [17] considers the shape of the stroke and the movement speed. The gesture-based touchpad system [18] considers the shape of the strokes and the length of pause between strokes. Neither system was formally evaluated via empirical user studies.

More recently, researchers started to explore the possibility of using behavioral measures to enhance the authentication process. For example, De Luca et al. used pressure, coordinates, size, speed, and time as an additional layer to authenticate a user on a touch screen smartphone [19]. Sae-Bae et al. examined fingertip dynamics (e.g., all fingertip moving or partial fingertip moving) as an additional authentication component and tested the idea using an iPad [20]. Preliminary studies suggest that the use of behavioral factors is promising. However, both studies required pre-defined shapes or gestures or the strokes being created in reference to specific locations on the screen.

We developed a drawing (recall) based gesture password application that allows users to draw their passwords freely on a canvas using any input devices. Different from many of the existing recall-based graphical passwords such as DAS, YAGP, and PassShape, our application does not require the use of any guides or reference points (e.g., a grid-based canvas). Further, different from PassShape, our application does not limit the password images to specific shapes or gestures. In this

initial study, we aimed to evaluate the gesture password method and examine whether different input techniques affect performance. We focused on two input devices that are commonly used by both the general public and people with cognitive disabilities: the mouse and the touch-screen [21].

### III. APPLICATION DESIGN

#### A. Algorithms

The password creation and re-entry method is a template free approach that allows the user to draw any image based on their imagination. The users do not need to limit the strokes to pre-defined shapes such as a line or a circle.

We use the \$N recognizer algorithm to determine how similar the newly entered password is to the original password [22]. The \$N recognizer was built on top of \$1 recognizer [23] to support multi-stroke password. \$N recognizer was built using simple trigonometric and geometric functions to perform the template matching between template and candidate entries. Since our solution is template free, we removed the entire existing template. A template is added when a user creates a password and recognition is done against the saved password template when the user enters the same password to login. The similarity between the original passwords and the re-entries is measured through a confidence score (ranging between 0 and 1).

#### B. Behavioral measures

In addition to the similarity between the images of the original password and the re-entries, we also captured and considered the behavioral measures to enhance the gesture-based authentication application. The measures that we initially examined include:

- stroke length,
- password size (the area of the bounding box around the password image),
- speed of movement,
- pause between strokes,
- and movement angles. We computed two types of angles: start angles and end angles. The start angle is the angle formed between the initial line of a gesture (point 0 to point 7) and the horizontal line. The end angle is the angle formed between the ending line of a gesture and the horizontal line.

These behavioral measures contain useful cues regarding the drawing habit of the user as well as how well the user remembers the password. For example, a user may draw a password at a lower speed if they have just created that password. They may also pause longer between the strokes if they are drawing a new password. The behavioral measures may also help detecting shoulder surfing attacks. If the attacker saw the password image on screen, he may be able to draw a similar image, but it would be much more difficult to observe and imitate similar speed of drawing or length of pauses between strokes.

### C. User Interface

The proposed gesture password is implemented as a web application that supports client and server architecture. We implemented the password drawing function using JavaScript. Because one major group of the target users are people with limited cognitive capabilities, the user interface for the gesture-based password application adopts a simple, user-friendly design. We have two slightly different user interfaces for password creation and login. The user interface for the gesture password creation application is demonstrated in Fig. 1. The 'gray' area in Fig. 1 is the canvas that users use to draw their password. The yellow box on top displays text feedback to the user. The text box below the canvas allow user to give a name to the password. This user-given name is for the purpose of login and matching. It also helps us understand how the users construct the password and accurately document the concept or object used in the password image. Once a user successfully draws a password, he can click the 'submit' button to submit the password. The 'cancel' button can be used when the user is unsatisfied with the drawing and wants to erase the image.



Fig. 1. Sample password creation page

## IV. EVALUATION STUDY DESIGN

We conducted a controlled empirical study to investigate how users interact with the password application and whether different input devices have any significant impact on the password quality or performance.

### A. Participants

Twenty five neurotypical participants took part in the study. 13 participants were females and 12 were males. The average age of the participants was 31 (Stdev. =10.5). All participants have previous experience using computers and the Web. All participants have Web accounts and are familiar with the traditional password authentication process.

### B. Tasks and Procedure

A within-subject design was adopted and each participant completed two sessions. In each session, participants created and re-entered passwords using either a mouse or a touch-screen device (iPad). The order of the input devices was counter-balanced to control the learning effect. At the

beginning of the study, the researcher explained and demonstrated the gesture passwords to the participants. During each session, participants first tried to create one password and re-enter it multiple times. Then they started the formal session and created 6 gesture passwords and re-entered each password ten times. Under the iPad condition, participants used their finger to draw the password directly on the screen. At the end of the study, participants completed a paper-based demographic questionnaire and a satisfaction survey.

## V. RESULTS

All participants easily mastered the method and were comfortable drawing the password according to observation of the researchers and the satisfaction survey.

### A. Content of Passwords

Participants created a total of 294 different passwords. We analyzed the password images that the participants created. The content of the passwords varies dramatically and covers a broad spectrum. The major categories include:

- vegetables, fruits and other food,
- trees, flowers and other plants,
- animals,
- human faces and emotions,
- buildings,
- scenery,
- vehicles,
- characters or words from participant's native language,
- mathematical shapes,
- English words or letters, and
- other objects.

Among the 294 passwords, 269 (91%) were drawing of objects or concept or sceneries. Only 25 (9%) were based on English letters or words. This is encouraging because if users tend to create the graphical passwords using English letters under the influence of the traditional alphanumeric passwords, the password space will be much smaller than the theoretical space and therefore, the method will be vulnerable to brutal force dictionary attacks. Fig. 2 demonstrates three sample passwords created in the study.

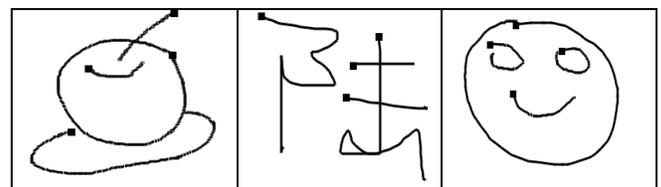


Fig. 2. Sample passwords. (from left to right: an apple on a plate, a Chinese character, a smiling face).

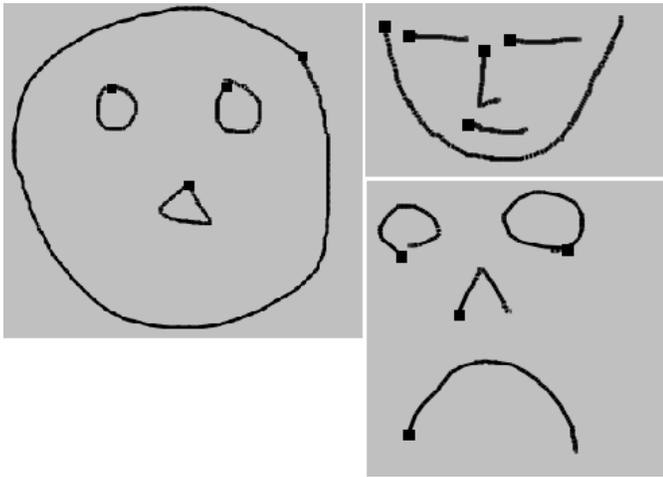


Fig. 3 Sample password images about human faces

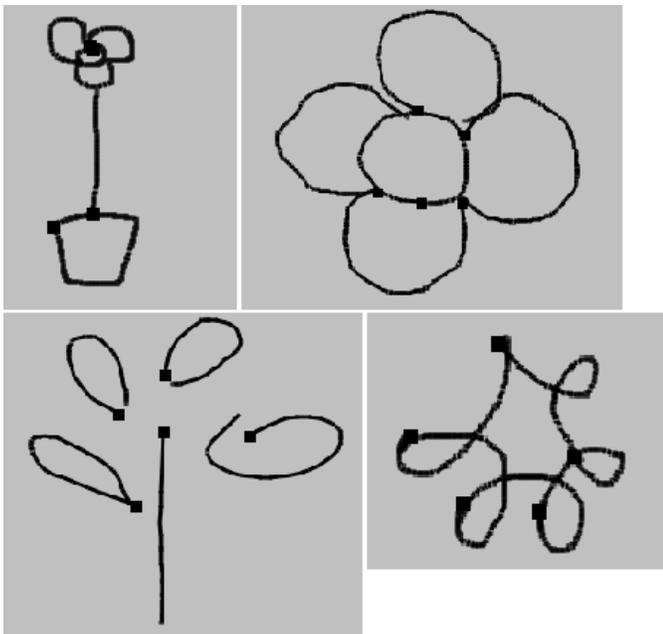


Fig. 4 Sample password images about flowers

It should also be noted that, even though different users may use the same concept or object for their passwords, the drawings are substantially different, suggesting that the actual password space is still quite large. As demonstrated in Fig. 3, the password images that participants created about human faces differ substantially at both the image level and the stroke level. Fig. 4 demonstrates four sample password images about flowers, which also differ dramatically at both the image level and the stroke level. This result suggests that the free draw password application may provide larger password space compared to applications that require pre-defined shapes.

### B. Accuracy of password re-entry

Given that users were asked to free draw without any guides or reference points such as the grid provided in DAS,

one of the key questions that needs to be investigated is whether the users could reproduce those passwords accurately. The confidence scores directly measure how accurate the reproduced password images match the image of the original password. The distribution of the confidence scores for the reproduced passwords under both input conditions are illustrated in Fig. 5. The confidence scores of more than 65% of the passwords reproduced under both conditions are higher than 0.8. 90% of the passwords reproduced under both conditions had confidence scores higher than 0.7. These preliminary results suggest that participants could reproduce the password with considerable level of accuracy.

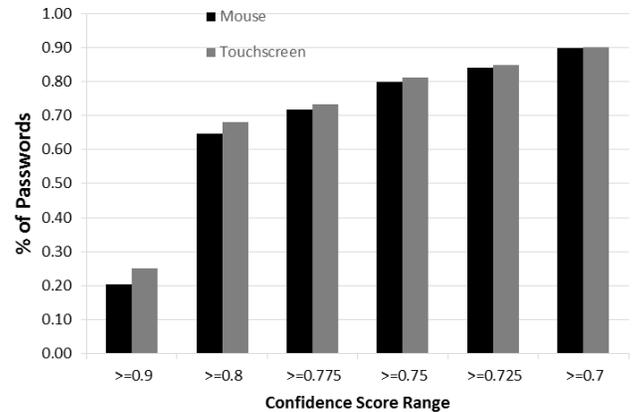


Fig. 5 Distribution of confidence scores under both input conditions.

### C. Comparison between mouse and touch-screen

Repeated ANOVA tests were used to compare the performance measures under the two input device conditions.

1) *Confidence scores and number of strokes:* As discussed in the previous section, the accuracy of password re-entry is obtained from confidence scores (CS). No significant difference was observed between the two devices regarding CS of the passwords ( $F(1, 22) = 0.06$ , n. s.; mean 0.81 for mouse, 0.82 for iPad) and the number of strokes in the passwords ( $F(1, 22) = 3.27$ , n. s.; mean 4.4 for mouse, 4.3 for iPad).

2) *Time to create and re-enter password:* The ANOVA test suggests that input devices had no significant impact on the time spent to create a password ( $F(1, 19) = 0.04$ , n. s.; 11.7 seconds for mouse, 13.2 seconds for iPad). However, input device had significant impact on the re-entry time for passwords ( $F(1, 22) = 8.35$ ,  $p < 0.01$ ). Participants spent longer time to enter the passwords using the mouse than the iPad (15.1 seconds vs. 6.5 seconds). (Fig. 6). *It should be noted that both the password creation time and the re-entry time are substantially shorter than those reported for the recognition-based graphical passwords [9, 24].*

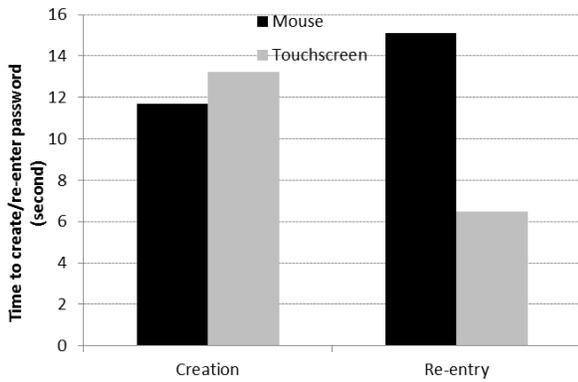


Fig. 6. Average time spent to create or re-enter passwords (seconds).

3) *Percentage of pause time*: The ANOVA test result suggests that input devices had significant impact on the percentage of pause in total task time both during creation ( $F(1, 18) = 18.74, p < 0.01$ ) and re-entry ( $F(1, 22) = 8.50, p < 0.01$ ). As shown in Fig. 7, when creating passwords, participants spent higher percentage of time in pauses using the iPad than the mouse. In contrast, when re-entering the passwords, they spent lower percentage in pauses using the iPad than the mouse. Since the percentage of pauses for the mouse remained the same between creation and re-entry, the difference is due to the dramatic decrease in pauses when re-entering passwords using the iPad.

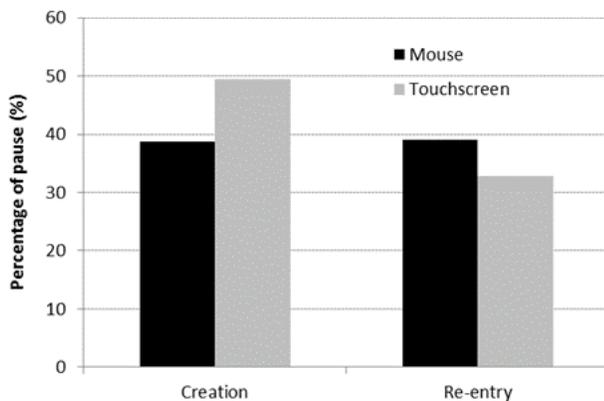


Fig. 7. Average percentage of pauses during creation and re-entry.

4) *Length of strokes*: The ANOVA test result suggests that input devices had significant impact on the length of strokes both during creation ( $F(1, 20) = 77.52, p < 0.01$ ) and re-entry ( $F(1, 21) = 55.91, p < 0.01$ ). As shown in Fig. 8, participants drew longer strokes when using the iPad than when using the mouse.

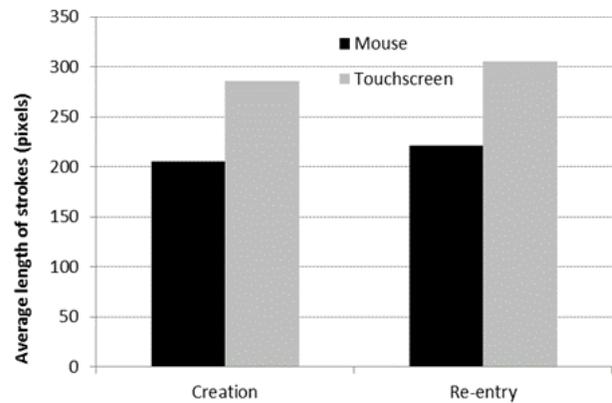


Fig. 8. Average length of strokes during creation and re-entry (pixels).

5) *Drawing speed*: The ANOVA test result suggests that input devices had significant impact on the drawing speed both during creation ( $F(1, 21) = 58.47, p < 0.01$ ) and re-entry ( $F(1, 22) = 92.36, p < 0.01$ ). As shown in Fig. 9, when using the iPad, participants drew more than twice as fast as when using the mouse.

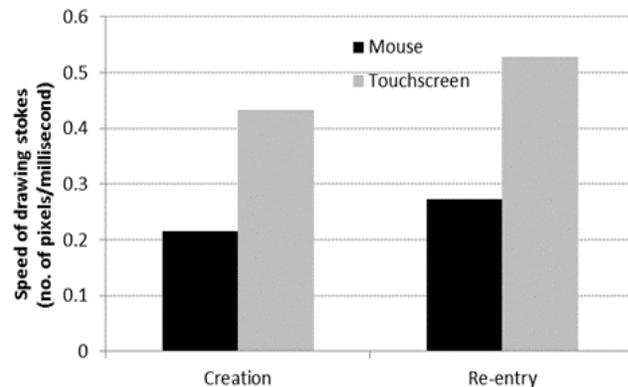


Fig. 9. Average drawing speed during creation and re-entry (number of pixels per millisecond).

#### D. User Preference

Among 25 participants, 20 (80%) preferred to use the iPad to create and enter the gesture passwords. 12 (48%) thought the gesture passwords are easier to remember than the alphanumeric passwords and 14 (56%) thought the gesture passwords are more secure than the alphanumeric passwords. Consider that all the participants are expert users for the alphanumeric passwords but novices for the gesture passwords, it is encouraging that almost half of the participants perceived the gesture passwords to be easier to remember than the alphanumeric passwords.

## VI. DISCUSSIONS AND FUTURE WORK

This study is a preliminary investigation that we conducted to justify more extensive empirical evaluations. From the

security perspective, the study provides initial insight about how the users construct the passwords and what types of concepts or objects they might use in their passwords. The collection of passwords that participants created helps us understand the actual password space. Encouragingly, the content of the passwords seems to be quite diversified and the influence of the traditional alphanumeric passwords seems to be limited. However, the actual password space of the gesture password method needs to be further investigated. As suggested by [25], users may also tend to choose memorable passwords and adopt specific strategies when creating recall-based passwords. For example, they may tend to create symmetric images or images that contain small number of components. We need to collect a much larger password corpus in order to thoroughly study the actual password space.

The results demonstrate that there are differences in a number of performance measures when participants create and enter graphical passwords using touch-screen vs. mice. In contrast to alphanumeric passwords, which many people found difficult to enter using a touch-screen device [21], the proposed gesture passwords seem to be easier to use with a touch-screen device than a mouse. Using the iPad, participants spent shorter time entering the gesture passwords. It seems that the touch-screen device allowed easier and smoother transition between strokes, as indicated by the dramatically lower percentage of time pausing between strokes during the re-entries. Participants also drew longer strokes in higher speed when using the iPad than when using the mouse. Finally, participants overwhelmingly prefer to use the iPad to enter the gesture passwords. These results suggest that the proposed gesture password method has the potential to be adopted in mobile devices due to its ease of use and efficiency.

As a preliminary investigation, the study has raised more questions than it has answered. There are many aspects regarding the password usability and security that need to be examined. We are planning a series of future studies to investigate the following issues:

- Could users easily memorize their passwords? Could they remember their passwords for a long period of time? Further, how does the use of multiple gesture passwords affect memorability of the passwords? A series of longitudinal studies are needed to answer these questions.
- How error-prone is the authentication? How does the addition of the behavioral measures affect the authentication accuracy? We are using machine learning techniques to identify appropriate thresholds for different measures in order to improve the authentication accuracy.
- How does the proposed gesture password method respond to shoulder surfing attacks?
- How do different input devices affect the use of the same password? For example, if a user creates a new password on his touchscreen tablet, can he accurately enter the same password using a mouse?
- One major motivation to develop this gesture-based password method is to provide an easier to remember

authentication method for users with limited cognitive capabilities (e.g., people with Down syndrome, seniors). We are planning a longitudinal study to examine how senior users with declining memory adopt the gesture password method.

## VII. CONCLUSIONS

This study suggests that the gesture password method may have the potential to serve as a usable alternative authentication method. As the confidence scores demonstrate, participants were able to reproduce the passwords with accuracy. The content of the passwords drawn is diversified, which has positive implications for security. The password creation time and login time are substantially shorter than those reported for recognition-based passwords. Future studies will be conducted to investigate the actual password space, the memorability of the password, how vulnerable is the password to shoulder surfing attacks, and whether this method would benefit people with limited cognitive capabilities.

## REFERENCES

- [1] D. Florencio and C. Herley. 2007. A large-scale study of web password habits. In WWW'07: Proceedings of the 16<sup>th</sup> International Conference on World Wide Web, New York, USA, 2007, ACM.
- [2] L. Sreeramareddy, J. Feng, and A. Sears 2012. Preliminary Investigation of Gesture-Based Password: Integrating Additional User Behavioral Features. Symposium On Usable Privacy and Security (SOUPS) 2012. Washington, D.C.
- [3] R. Dhamija and A. Perrig 2000. Deja Vu: A User Study Using Images for Authentication. In Proceedings of the 9<sup>th</sup> USENIX Security Symposium (Denver, Colorado, USA, August 14-17, 2000), pp4-19.
- [4] Real User Corporation, 2005. Passfaces: Two Factor Authentication for the Enterprise, 2005.
- [5] D. Angeli, M. Coutts, L. Coventry, G. I. Johnson, D. Cameron, and M. H. Fischer. 2002. VIP: A Visual Approach to User Authentication. In Proceedings of the Working Conference on Advanced Visual Interfaces (Trento, Italy, May 22-24, 2002). ACM Press, New York, NY, 316-323.
- [6] I. Jermyn, A. Mayer, F. Monrose, M. Reiter, and A. Rubin 1999. The design and analysis of graphical passwords. In Proceedings of the 8<sup>th</sup> USENIX Security Symposium. 1999.
- [7] A. De Luca, R. Weiss, H. Hussmann 2007. PassShape: Stroke based shape passwords. Proceedings of OzCHI. 2007. 239-240.
- [8] S. Wiedenbeck, J. Waters, J.-C. Birget, A. Brodskiy, and N. Memon. 2005. PassPoints: Design and Longitudinal Evaluation of a Graphical Password System. International Journal of Human-Computer Studies. 63, 102-127.
- [9] E. Stobert and R. Biddle 2013. Memory retrieval and graphical passwords. Proceedings of the Symposium on Usable Privacy and Security (SOUPS) 2013.
- [10] A. Paivio, T. Rogers, and P. C. Smythe. Why are pictures easier to recall than words? Psychonomic Science, 11(4):137-138, 1968
- [11] R. Biddle, S. Chiasson, and P. C. van Oorschot. Graphical Passwords: Learning from the First Twelve Years. ACM Computing Surveys, 44(4), 2012
- [12] X. Suo, Y. Zhu, and G. S. Owen. 2005. Graphical Passwords: A Survey. In Proceedings of the 21<sup>st</sup> Annual Computer Security Applications Conference (Tucson, Arizona, USA, December, 5-9, 2005), IEEE Computer Society 463-472.
- [13] N. H. Zakaria, D. Griffiths, S. Brostoff, and J. Yan 2011. Shoulder surfing defence for recall-based graphical passwords. Proceedings of the Symposium on Usable Privacy and Security (SOUPS) 2011.

- [14] D. Nali, and J. Thorpe 2004. Analyzing user choice in graphical passwords. Technical Report TR-04-01, School of Computer Science, Carleton University, Canada, 2004.
- [15] P. Dunphy and J. Yan 2007. Do background images improve “Draw a Secret” graphical passwords? Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS).
- [16] H. Gao, X. Guo, X. Chen, L. Wang, X. Liu 2008. Yagp: yet another graphical password strategy. Proceedings of the Annual Computer Security Applications Conference, 2008.
- [17] C. Varenhorst 2012. Passdoodles: A Lightweight Authentication Method. MIT Research Science Institute. Last Retrieved March 10, 2014 from [http://people.csail.mit.edu/emax/public\\_html/papers/varenhorst.pdf](http://people.csail.mit.edu/emax/public_html/papers/varenhorst.pdf)
- [18] D. Mejia and J. Doose 2010. Gesture based touchpad security system. Last retrieved March 10, 2014 from [http://people.ece.cornell.edu/land/courses/ece4760/FinalProjects/s2010/jkd27\\_dm472/MyPage/index.html](http://people.ece.cornell.edu/land/courses/ece4760/FinalProjects/s2010/jkd27_dm472/MyPage/index.html)
- [19] A. De Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann 2012. Touch me once and I know it's you! Implicit authentication based on touch screen pattern. Proceedings of ACM CHI Conference 2012. 987-996.
- [20] N. Sae-Bae, K. Ahmed, K. Isbister, and N. Memon 2012. Biometric-rich gestures: A novel approach to authentication on multi-touch devices. Proceedings of ACM CHI Conference 2012. 977-986.
- [21] L. Kumin, J. Lazar, J. Feng, B. Wentz, and N. Ekedebe 2012. A usability evaluation of workplace-related tasks on a Multi-Touch tablet computer by adults with Down syndrome, Journal of Usability studies. vol. 7, no. 4, 118-142.
- [22] L. Anthony and J. Wobbrock 2014. A Lightweight Multistroke Recognizer for User Interface Prototypes. Technical paper. Last retrieved March 10, 2014 from <http://faculty.washington.edu/wobbrock/pubs/gi-10.2.pdf>
- [23] J. O. Wobbrock, A. D. Wilson, and Y. Li. 2007. Gestures without libraries, toolkits or training: a \$1 recognizer for user interface prototypes. In Proceedings of the 20th annual ACM symposium on User interface software and technology (UIST '07). ACM, New York, NY, USA, 159-168.
- [24] Y. Ma, J. Feng, L., Kumin, and J. Lazar 2013. Investigating user behavior for authentication methods: A comparison between individuals with Down syndrome and neurotypical users. ACM Transactions on Accessible Computing. Vol. 4(4) 1-27.
- [25] P. C. van Oorschot and J. Thorpe. 2008. On predictive models and user-drawn graphical passwords. ACM Transactions on Information and System Security. 10(4).

# Reversible image watermarking scheme with perfect watermark and host restoration after a content replacement attack

Alejandra Menéndez-Ortiz\*, Claudia Feregrino-Uribe\* and José Juan García-Hernández†

\*Coordinación de Ciencias Computacionales, INAOE

Luis Enrique Erro #1, Sta. Ma. Tonantzintla, Puebla, CP. 72840, México

{m.menendez, cferegrino}@inaoep.mx

†Laboratorio de Tecnologías de Información, CINVESTAV

Parque Científico y Tecnológico TECNOTAM - Km. 5.5 carretera Cd. Victoria-Solo La Marina

Cd. Victoria, Tamps. CP. 87130, México

jjuan@tamps.cinvestav.mx

**Abstract**—Robust reversible watermarking schemes (RWS) were designed to increase the robustness of RWS; when no attacks occur they are able to reconstruct the host signal and to extract a watermark; under attacks they only extract the watermark or reconstruct the host signal. This work proposes a robust RWS that can both extract a watermark and reconstruct an image even under attacks. The solution chains a fragile reversible stage with a self-recovery stage in order to achieve robustness for both watermark and image. The results obtained with this construction show that it can correctly extract the watermarks and perfectly reconstruct the images after modifications, so the scheme is robust against content replacement attack. Future efforts are aimed towards the reduction of perceptual impact and robustness improvement. This investigation sets the foundation on a first robust RWS that can restore both watermark and image after attacks.

**Keywords**—Reversible watermarking, perfect reconstruction, robustness, content replacement attack

## I. INTRODUCTION

Digital watermarking schemes insert a secret watermark into a host signal in such a way that it is imperceptible for a human observer but can be recovered using an extraction algorithm. These schemes can be the solution for specific application scenarios, like authentication, copyright protection, fingerprinting, and the like. Nonetheless, watermarking insertion produces irreversible modifications. In application scenarios where sensitive imagery is treated (such as deep space exploration, military investigation and recognition, and medical diagnosis [1]), these irreversible modifications cannot be acceptable. *Reversible* watermarking techniques, also known as *invertible* or *lossless* [1] can insert a secret watermark within a carrier signal and later the modifications suffered during insertion can be reversed to obtain the host signal. However, reversible watermarking schemes (RWS) can only reconstruct the host signal if the watermarked version does not suffer attacks. In presence of attacks the reversibility of the scheme is lost, *i.e.* these schemes are fragile by nature.

Robust RWS have been developed to compensate the fragility of reversible watermarking techniques, focusing on two aspects 1) robustness of the watermark and 2) robustness

of the signal. Works like those designed by Hossinger *et al.* [2] and De Vleeschouwer *et al.* [3] can reconstruct the original host signal and the secret watermark if no attacks occur; in case of attacks these schemes can extract the secret watermark, losing the host signal. Another type of robust RWS, known in the literature as self-recovery schemes, are the ones presented by Zhang *et al.* [4, 5] and Bravo-Solorio *et al.* [6]. In this type of robust RWS, algorithms are able to reconstruct the host image regardless of attacks but they cannot insert a secret watermark. The embedding capacity of these schemes is used for embedding control data necessary to compensate attacks, but no space is left for secret information (payload). Figure 1 shows a classification for the robustness that current RWS present. To date, robust RWS found in the literature can only extract the watermark when attacks occur; self-recovery watermarking schemes can just reconstruct the host image when the watermarked image is submitted to attacks; the proposed scheme focuses on robustness for both the watermark and the signal and, to our knowledge, there are no other RWS that comply these characteristics.

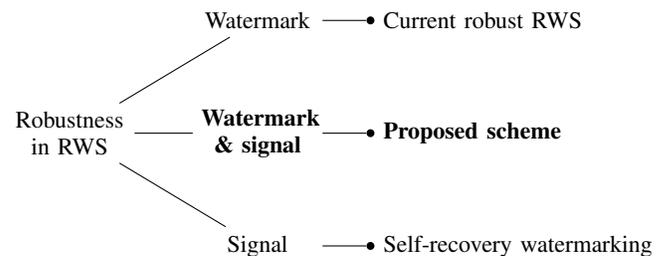


Fig. 1. Classification of schemes with robustness in reversible watermarking.

The problem of host image and secret watermark perfect reconstruction, even under attacks, has found a solution from the communications area. Kalker and Willems [7], among others, designed systems where a host signal and a secret watermark can be reconstructed regardless of modifications suffered in the transmission channel. This problem has been theoretically solved only for binary signals and a binary channel; however, there are no implementations for practical scenarios and these models cannot be used for practical watermarking applications.

This investigation presents a first approach on a robust RWS that allows both the extraction of a hidden watermark with low error probability and the perfect reconstruction of a host image even under attacks. A scheme like the one proposed in this work could be used for covert communications in military environments, because it ensures the recovery of the original host image and the extraction of the secret information regardless of the modifications imposed by the transmission channel.

The rest of the document is organized as follows: section II details the problem addressed and explains the proposed solution, the results obtained with it are presented in section III, a discussion of the work is given in section IV and the final remarks are summarized in section V.

## II. ROBUST REVERSIBLE WATERMARKING

A robust reversible watermarking scheme, like the one given in this work, allows to extract a hidden watermark and to restore a host signal even in the presence of attacks and has the following characteristics: the *sender* side must embed a secret message ( $M$ ) into a host signal ( $X$ ). This message is related to and embedded into the signal, the watermarked signal must be transmitted through a non-binary channel,  $M$  must be embedded into  $X$  in such a way that  $X$  suffers a distortion lower than a threshold  $\alpha$ . The *transmission channel* is noisy and non-binary, where a set of attacks may occur and cause a distortion  $\beta$  to the watermarked signal, resulting in an attacked signal ( $\hat{Y}$ ). From  $\hat{Y}$ , the *receiver* side should be able to extract  $M'$  and counteract the distortions  $\alpha$  and  $\beta$  in order to reconstruct the host signal  $X$ . Figure 2 shows the general robust reversible watermarking scheme.

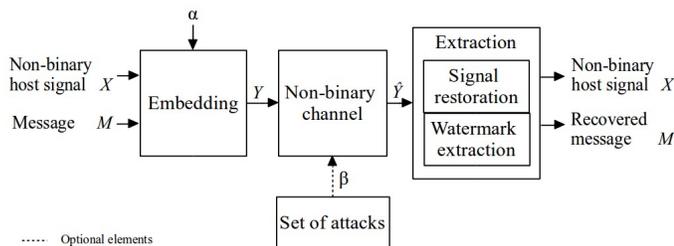


Fig. 2. Elements in a robust reversible watermarking scheme.

The strategy followed to construct a scheme with these characteristics is presented below.

### A. Fragile reversible stage and self-recovery chaining

In order to construct a robust reversible watermarking scheme, the chaining of a fragile reversible scheme and a self-recovery scheme is posed and the solution is shown in Figure 3. The fragile scheme embeds a watermark and control data to reconstruct the host image, the self-recovery scheme embeds information to counteract the modifications caused by the attacks. The scheme by Sachnev *et al.* [8] was selected as the fragile reversible one because it remains as one of the most efficient reversible watermarking schemes in the literature. The scheme by Zhang and Wang [4] was selected as the self-recovery one because it can perfectly reconstruct a host image after attacks. The scheme consists of two processes:

embedding and extraction/restoration. These processes along with an intermediate stage of attacks is described next.

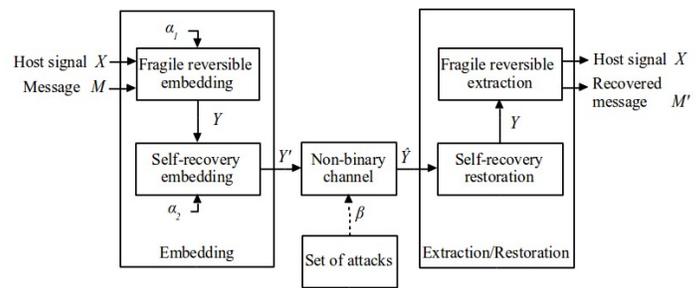


Fig. 3. Elements in the proposed construction.

1) *Embedding*: The embedding process inserts a watermark ( $M$ ) into  $X$  in the fragile reversible stage, producing a first-stage watermarked image ( $Y$ ); the distortion caused by this stage must be lower than a threshold  $\alpha_1$ . The self-recovery stage inserts into  $Y$  information to counteract the modifications of the attacks, producing a second-stage watermarked image ( $Y'$ ), the distortion must be lower than a threshold  $\alpha_2$ .

*Fragile reversible embedding*. In this stage the embedding algorithm described in [8] is used to embed the watermark into the host image. This algorithm constructs the payload to be embedded by concatenating the bits of the watermark with the least significant bits (LSB) from the first 30 pixels in the image. The payload is hidden using the histogram shifting algorithm detailed in [8]. A header is constructed with parameters necessary for the decoder ( $T_p$ , positive threshold;  $T_n$ , negative threshold; and  $i$ , index of last marked pixel) and their binary representation is embedded into the LSB of the first 30 pixels. The embedding algorithm produces a location map that contains information about the overflow and underflow cases, that location map is necessary to restore the host image.

*Self-recovery embedding*. This stage embeds data into the image itself to detect the regions where tampering took place (check bits) and data to restore the original values of the pixels from the tampered areas (reference bits). This stage uses the watermark embedding procedure described in [4], where the reference bits are a compressed version of the binary image representation and the check bits for tamper-detection are the result of hashing every block in the image. Both the reference bits and the check bits are pseudo-randomly permuted prior embedding in order to be dispersed through the image during the embedding procedure, increasing in this way the robustness.

2) *Attacks*: Because of the scheme's characteristics, it inherits the robustness of the work by Zhang and Wang. This scheme only resists the content replacement attack when the tampered area is less than 3.2% of the image and so does the proposed solution. Section III details this attack.

3) *Extraction/Restoration*: In order to perfectly reconstruct the host image and to extract the hidden watermark, the self-recovery stage must be carried out first, followed by the fragile reversible stage. The self-recovery stage enables the detection of the tampered areas of the image and provides data to compensate the  $\beta$  and  $\alpha_2$  distortions, so the first-stage watermarked image ( $Y$ ) is recovered. From  $Y$  the fragile

reversible stage is capable of extracting the hidden watermark ( $M'$ ) and restoring the host image ( $X$ ).

*Self-recovery restoration.* This stage uses the image restoration procedure described in [4], where reference bits and check bits are extracted from the image itself. A set of calculated check bits are obtained from the values of  $\hat{Y}$  in the same fashion as the extracted check bits were calculated. To identify the tampered areas, extracted check bits are compared against calculated check bits. If the differences between them exceed a threshold, the block is considered tampered. Pixel values and reference bits extracted from tampered blocks are not reliable, so these missing bits are calculated from non-tampered blocks, using reference bits and the pixels' binary representation.

*Fragile reversible extraction.* This stage employs the decoder algorithm given in [8]. It takes  $Y$  and extracts the first 30 pixel's LSB to construct parameters for data extraction ( $T_p$ , positive threshold;  $T_n$ , negative threshold; and  $i$ , index of last marked pixel). With these parameters along with the location map, the histogram shifting decoding procedure extracts the hidden payload and reconstructs the original image pixel values. The 30 bits that were appended to the watermark and hidden in the embedding process are utilized to restore  $X$ 's first 30 pixel values.

The constructed scheme aims to provide robustness for both the image and the watermark against malicious attacks.

### III. EXPERIMENTAL RESULTS

In this section, it is studied the possibility of correctly extract a watermark and to perfectly reconstruct a host image in the presence of attacks. The objective of these experiments is to validate the robustness of the embedded watermark and the restored image against the content replacement attack. They also help to corroborate that the reversibility is maintained.

#### A. Embedding

A binary image is embedded as a watermark into grayscale test images. Figure 4 shows the watermark image of  $210 \times 210$  pixels, a logo proposed by our research team. The test set contains nine grayscale host images of  $512 \times 512$  pixels, shown in the first column of Table I. The differences between images are measured by using the peak signal-to-noise ratio (PSNR) given in (1). A  $\text{PSNR}_{\text{FSW}}$  value is measured between the host image and the first-stage watermarked (FSW) image; a  $\text{PSNR}_{\text{SSW}}$  value is measured between the first-stage watermarked image and the second-stage watermarked (SSW) image; and a  $\text{PSNR}_F$  value is calculated between the host image and the second-stage watermarked image. The results are shown in Table I.

$$\text{PSNR} = 20 \log_{10} \frac{255}{\sqrt{\text{MSE}}} \quad (1)$$

where MSE is the mean square error given by:

$$\text{MSE} = \frac{\sum_{m=1}^M \sum_{n=1}^N (X_{m,n} - Y_{m,n})^2}{M \times N} \quad (2)$$

where  $X_{m,n}$  is the pixel value at the  $m^{\text{th}}$  row and  $n^{\text{th}}$  column of the host image,  $Y_{m,n}$  is the pixel value at the  $m^{\text{th}}$  row and  $n^{\text{th}}$  column of the modified image,  $M$  is the number of rows and  $N$  is the number of columns in the images.



Fig. 4. Binary image inserted as a watermark.

#### B. Attacks

To test the robustness of the scheme presented in this work, the second-stage watermarked images were subjected to a content replacement attack, where some pixels of the image were changed by a region of pixels from another image. The content replacement attack was selected to perform these tests to corroborate that the proposed scheme has the same robustness as the one by Zhang and Wang [4]. The distortions caused by this attack were measured in terms of PSNR values and the results are presented in Table II.

#### C. Extraction/Restoration

These experiments take an attacked image and reconstruct the first-stage watermarked version ( $Y$ ), then the host signal is restored and a watermark is extracted, both from  $Y$ . Two PSNR values from these steps are collected, a  $\text{PSNR}_{\text{FSW}}$  value is measured between the restored first-stage watermarked image and the host image; a  $\text{PSNR}_{\text{FR}}$  value is measured between the final restored image and the host one. The differences between an original watermark and an extracted one are measured with the bit error rate (BER) given in (3). The collected PSNR values and BER values are presented in Table III.

$$\text{BER} = \frac{\text{Errors}}{N_{\text{bits}}} \quad (3)$$

where  $N_{\text{bits}}$  is the total number of bits and Errors is given by:

$$\text{Errors} = \sum_{n=1}^{N_{\text{bits}}} \begin{cases} 1, & W'_n \neq W_n \\ 0, & W'_n = W_n \end{cases} \quad (4)$$

$W'_n$  is the  $n^{\text{th}}$  bit from the extracted watermark and  $W_n$  is the  $n^{\text{th}}$  bit from the original watermark.

### IV. DISCUSSION AND FUTURE WORK

From Table I it can be seen that the embedding process has an average distortion of 26.63 decibels (dB), which is lower than PSNR value of 30 dB or higher for well reconstructed images [9]. However, the scheme could be used in applications with a higher tolerance for distortions, like in watermarking of compressed images [10, 11]. Nonetheless, if the scheme is required in other type of applications, it is necessary to

TABLE I. MEASURED PSNR VALUES IN DB AFTER WATERMARK EMBEDDING.

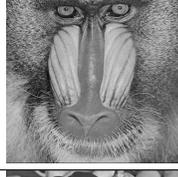
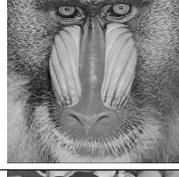
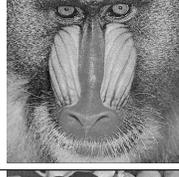
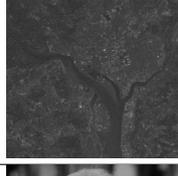
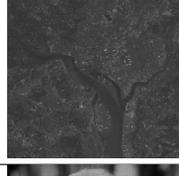
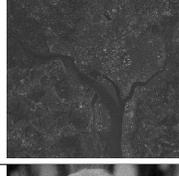
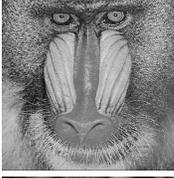
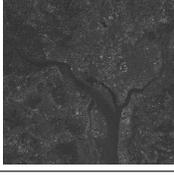
Host image	1st-stage watermarked image	PSNR <sub>F<sub>SW</sub></sub>	2nd-stage watermarked image	PSNR <sub>SSW</sub>	PSNR <sub>F</sub>
		41.23		23.60	23.32
		43.55		27.29	26.99
		43.25		27.89	27.53
		40.98		25.01	24.56
		42.22		28.98	28.37
		36.76		20.79	20.15
		41.74		28.45	27.77
		41.84		31.27	30.19
		42.38		31.88	30.81
Average		41.55	—	27.24	<b>26.63</b>

TABLE II. MEASURED PSNR IN DB AFTER CONTENT REPLACEMENT ATTACK.

2nd-stage watermarked image	Attacked image	PSNR	2nd-stage watermarked image	Attacked image	PSNR	2nd-stage watermarked image	Attacked image	PSNR
		22.60			24.64			24.91
		23.57			25.69			17.52
		25.66			27.95			25.05

reduce the perceptual impact that the scheme imposes over the watermarked images. The  $PSNR_{FSW}$  values are the same for the first-stage watermarked images in the embedding and in the extraction/restoration processes, which means that the self-recovery stage perfectly reconstructed  $Y$  even after a content replacement attack. The  $PSNR_{FR}$  values of  $\infty$  in Table III mean that the restored images have exactly the same values as the host ones and the BER results show that there are no errors in the extracted watermarks.

The results indicate that our scheme has a degree of robustness against content replacement attack and maintains the reversibility property even after modifications, although under the same circumstances as in the scheme by Zhang and Wang [4]. The drawbacks of both schemes are that the tampering cannot be corrected if the tampered area is greater than 3.2% of the image, because the reconstruction of the missing pixels depends on finding the solution of a binary linear equation system. In case of a modification greater than 3.2% of the image, the restoration process cannot reconstruct the first-stage watermarked signal ( $Y$ ) and therefore, the extraction/restoration process does not restore the host image neither extracts the hidden watermark ( $M'$ ). The robustness of the proposed scheme is given by the robustness of the self-recovery scheme employed. Because the robustness of the scheme in [4] is limited to very specific attacks (only content replacement), our scheme is also limited to content replacement attack. These experiments are only a concept test to verify the validity of the proposed construction, so in this stage of the research a thorough evaluation of attacks is not essential.

Future efforts are aimed to reduce the perceptual impact imposed by the embedding process and to resist other attacks. Moreover, increasing the severity of the content replacement attack while maintaining at least current robustness is an additional goal.

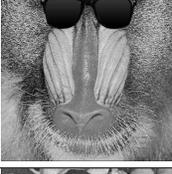
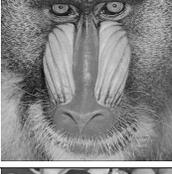
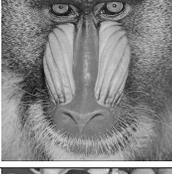
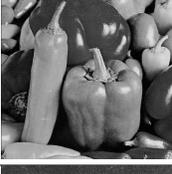
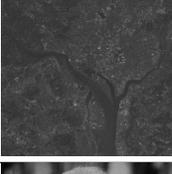
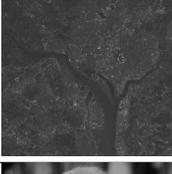
## V. CONCLUSIONS

In this work, a first approach to construct a robust reversible watermarking scheme is proposed by concatenating a fragile reversible watermarking scheme with a self-recovery watermarking scheme. The proposed scheme aims at the perfect reconstruction of a host image and the extraction of a hidden watermark with low error probability even under attacks. The results obtained with this chaining show that the scheme has a degree of robustness against the content replacement attack, the hidden watermarks can be extracted and the host images can be restored after modifications. Although the scheme produces distortions, it has the capability for perfect reconstruction of the host images after the extraction/restoration process is applied. The results also suggest that it is possible to design a robust reversible watermarking scheme for practical watermarking applications and this construction sets the foundation as a first approach towards the design of a robust reversible watermarking scheme that can both restore a host image and extract a watermark even under attacks. As far as we know, this investigation is the first attempt to propose a robust reversible watermarking scheme with perfect watermark and host restoration.

## REFERENCES

- [1] R. Caldelli, F. Filippini, and R. Becarelli, "Reversible Watermarking Techniques: An Overview and a Classification," *EURASIP Journal of Information Security*, vol. 2010, no. 2, pp. 1–19, Jan. 2010. [Online]. Available: <http://dx.doi.org/10.1155/2010/134546>
- [2] C. W. Honsinger, P. W. Jones, M. Rabbani, and J. C. Stoffel, "Lossless recovery of an original image containing embedded data," US Patent, August 2001, uS Patent 6,278,791.
- [3] C. De Vleeschouwer, J.-F. Delaigle, and B. Macq, "Circular interpretation of bijective transformations in lossless watermarking for media asset management," *Multimedia, IEEE Transactions on*, vol. 5, no. 1, pp. 97–105, 2003.

TABLE III. MEASURED PSNR (dB) AND BER (%) VALUES AFTER EXTRACTION/RESTORATION.

Attacked image	1st-stage watermarked image	PSNR <sub>FSW</sub>	Restored image	PSNR <sub>FR</sub>	Original watermark	Extracted watermark	BER
		41.23		$\infty$			0.0
		43.55		$\infty$			0.0
		43.25		$\infty$			0.0
		40.98		$\infty$			0.0
		42.22		$\infty$			0.0
		36.76		$\infty$			0.0
		47.74		$\infty$			0.0
		41.84		$\infty$			0.0
		42.38		$\infty$			0.0

- [4] X. Zhang and S. Wang, "Fragile Watermarking With Error-Free Restoration Capability," *IEEE Transactions on Multimedia*, vol. 10, no. 8, pp. 1490–1499, 2008.
- [5] X. Zhang, S. Wang, Z. Qian, and G. Feng, "Reference Sharing Mechanism for Watermark Self-Embedding," *IEEE Transactions on Image Processing*, vol. 20, no. 2, pp. 485–495, 2011.
- [6] S. Bravo-Solorio, C.-T. Li, and A. Nandi, "Watermarking method with exact self-propagating restoration capabilities," in *IEEE International Workshop on Information Forensics and Security (WIFS), 2012*, 2012, pp. 217–222.
- [7] T. Kalker and F. M. Willems, "Capacity bounds and constructions for reversible data-hiding," in *Security and Watermarking of Multimedia Contents V*, vol. 5020, 2003, pp. 604–611. [Online]. Available: <http://dx.doi.org/10.1117/12.473164>
- [8] V. Sachnev, H. Kim, S. Suresh, and Y. Shi, "Reversible Watermarking Algorithm with Distortion Compensation," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 1, p. 316820, 2010. [Online]. Available: <http://asp.eurasipjournals.com/content/2010/1/316820>
- [9] D. S. Taubman and M. W. Marcellin, *JPEG2000: Image compression fundamentals, standards and practice*. Kluwer Academic Publishers, 2004.
- [10] S. Emmanuel, H. Kiang, and A. Das, "A Reversible Watermarking Scheme for JPEG-2000 Compressed Images," in *IEEE International Conference on Multimedia and Expo, 2005. ICME 2005.*, Jul. 2005, pp. 69–72.
- [11] A. V. Subramanyam, S. Emmanuel, and M. Kankanhalli, "Robust Watermarking of Compressed and Encrypted JPEG2000 Images," *IEEE Transactions on Multimedia*, vol. 14, no. 3, pp. 703–716, June 2012.

# Component Rejuvenation for Security for Cloud Services

Chen-Yu Lee<sup>1</sup>, Krishna M. Kavi<sup>1</sup>, and Mahadevan Gomathisankaran<sup>1</sup>

<sup>1</sup>Department of Computer Science and Engineering, University of North Texas, Denton, TX 76203, USA

**Abstract**—Software rejuvenation has been used to improve reliability of systems by periodically checkpointing and restarting them. In this paper, we propose to use rejuvenation as a mechanism to enhance the security of Cloud infrastructure; in addition to intrusion detection, scanning for hidden threats such as malware, viruses, etc. These threats can also be eliminated by continuous and periodic component rejuvenation. In this paper, we compare the costs associated with rejuvenation with scanning.

**Keywords:** Rejuvenation, malware, security, computer virus

## 1. Introduction

Computer viruses have been evolving into more complex malware and the detection and elimination of such threats is becoming very expensive in large IT operations. Figure 1 shows the number of new types of malware detected over the past ten years—the number has increased very rapidly since 2010.

Software Rejuvenation technology was first proposed by Lin in 1993 [1]. The author observed that system performance degrades with time, and failure rates also increase with time. This phenomenon was termed software *aging*. A proactive solution to this problem is *software rejuvenation* to gracefully terminate an application or a system and restart it in a clean internal state [2]. Rejuvenation technology was originally used for software fault tolerance [3] [4]. In this paper, we propose to apply rejuvenation as an approach to enhance software security in Cloud computing environments. At present most systems rely on scanning for viruses and malware. We feel that some of these threats can also be eliminated by periodic component rejuvenation. Rejuvenation can either be used in place of scanning or in addition to scanning. In this paper, we develop a model to compare the cost of rejuvenation with scanning for malware. We emphasize that Rejuvenation is more than just restarting of systems, it also includes checkpointing software applications and systems in clean states, and periodically rolling back the software to known clean states. We also permit updating of checkpoints to include the application of safe and authorized updates to the software.

The rest of the paper is organized as follows: Section 2 introduces how rejuvenation can be used to enhance security. This section also introduces a model for estimating the cost of rejuvenation. Section 3 compares rejuvenation with scanning approaches. Section 4 provides our conclusions and future work.

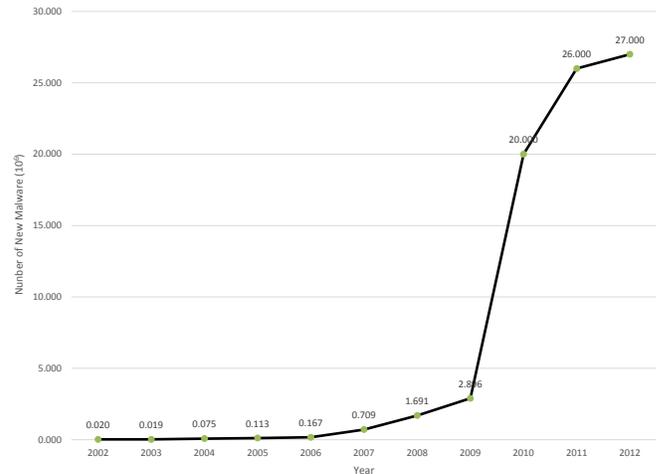


Fig. 1: The number of new malware each year [5] [6]

## 2. Rejuvenation for Security

### 2.1 Environment description

The proposed rejuvenation is applied to Cloud computing environments to enhance security and stability of the systems. Many commercial operations rely on Cloud computing and in such applications, maintaining low mean time to repair (MTTR) and the cost of repair are essential to the profitability of the operations. Therefore, they normally use software patches to fix problems, instead of completely updating their systems. The patches include system patches, software patches, malware/virus signatures, firewall rules, etc.

### 2.2 Work Flow of Rejuvenation for Security

We propose to use rejuvenation (i.e., checkpoint, rollback, recover and restart) to improve security and reliability of components, and the rejuvenation is applied periodically. The rejuvenation can be applied modularly to minimize the downtime of the system. Each module is restored (or rejuvenated) to a clean checkpoint and reconnected with other related modules. Patches can also be applied to modules during a rejuvenation to reduce their vulnerabilities and to eliminate detected malware. The patches should be verified as clean and distributed by authorized providers, to assure that patched modules are clean. The work flows is shown in Figure 2, and the main processes are described below:

- **Checkpoint:** When a new software module is tested, verified, and ready to go online, it is assumed to be clean and a checkpoint of the module is made. Periodically, the module is rolled back to the clean checkpoint to scrub the module of any infections. If any design fixes or other patches are made available to the module since its original release (and the patches are verified as trusted and clean), the module is upgraded during the rejuvenation period, and the checkpoint image is updated to the new clean state.
- **Recover:** All modules of a system go through a rejuvenation process (checkpoint-recovery) periodically, where the periodicity is determined based on the cost of rejuvenation and the frequency of new malware introductions. The process eliminates not only software aging and soft or intermittent faults, but some potential malware. In addition, the rejuvenation is performed when an abnormal condition or a suspected security threat is detected, to restore the checkpoint image.
- **Restart:** The module always restarts after each recovery. This eliminates software aging and some common security threats, including denial of service (DoS) and others.

### 2.3 Cost Model

To evaluate the applicability of rejuvenation to enhance security, it is necessary to compare the cost of rejuvenation with other known defense mechanisms. This section discusses the cost of performing rejuvenation compared with malware scanning.

Malware scanning software (e.g., anti-malware software) is usually performed as a daemon, scanning all the stored files, executing processes, the kernel and other system software continuously. Scanning may detect more security threats than that can be eliminated using only rejuvenation. However, scanning for malware consumes computational resources and thus the following model can be used for estimating the cost of malware scanning ( $CoMS$ ).

- **Instance size( $V$ ):** The cost of scanning is proportional to the size of the system being scanned. In addition to scanning of the system at startup, malware scanning occurs continuously and is invoked when changes to the system are detected (such as file updates, internet downloads, mail attachments or other changes to the system state). In this paper, we relate the cost of scanning to the average volume of new information that must be scanned over a given period of time. The period of time and the volume of data scanned are compared with the rejuvenation period and the volume of information involved in the rejuvenation process.
- **Scan speed( $SS$ ):** This is the rate at which a system can be scanned to detect malware or virus signatures.
- **Cloud computing fee ( $CCF$ ):** The fee charged by Cloud providers (whether the computing is used for

scanning or for providing services).

The total cost involved with malware scanning  $C_{MS}$  for size  $V$  over a chosen time period  $T$  is

$$C_{MS}(V, T) = V \times SS^{-1} \times CCF \quad (1)$$

As an example, if it is assumed that the scanning speed is 26.58 MB/sec [7] and the computing price charged by Amazon EC2 is \$0.176, \$0.351, \$0.702, and \$1.404 per hour for instance size 4, 32, 40, and 80GB [8], the cost of performing one malware scan on a Cloud environment with 10 GB size data would be \$0.007, \$0.117, \$0.293, and \$1.173 respectively.

In this paper, we assume two types of rejuvenation: a regular, periodic rejuvenation at fixed periods, and ad hoc rejuvenations when anomalies or threats are detected. Thus, factors that contribute to the cost of rejuvenation are divided into two parts. One is the cost involved with rejuvenation ( $CoR$ ), and the other is involved with monitoring ( $CoM$ ). Rejuvenation makes some modules unavailable (downtime) during the process of restoration. The following are the various factors that influence the cost of rejuvenation.

- **Downtime ( $DT$ ):** While performing the rejuvenation, some modules will be unavailable and the downtime can range from a few seconds to a few minutes.
- **Number of transactions lost ( $TL$ ):** The number of the transactions lost during the downtime.
- **Potential revenue associated with each transaction ( $PR$ ):** Each successful transaction would generate revenue, so the potential revenues lost are included in the cost of rejuvenation.
- **Version storage fee ( $SF$ ):** Since clean modules and checkpointed states must be saved, we include the cost of storage with rejuvenation. In some cases, we may need to save  $m$  snapshots or checkpoints to fully recover the system to a clean state. Thus we include the total cost of storage needed for checkpointing. This can be compared with the volume scanned by malware scanners.
- **Data transfer fee ( $TF$ ):** We assume that the checkpoints are stored in a backup or archival facility and this information has to be transferred to executing environments during restoration. We include the data transfer costs for transferring  $n$  bytes of data transferred between an execution environment and backup facility.

The total cost involved with rejuvenation  $CoR_{periodic}$  for size  $V$ , with a rejuvenation period of  $T$ , is

$$CoR_{periodic}(V, T) = CoR_1 + CoR_2 + CoR_3 \quad (2)$$

, where three major  $CoR$  terms are:

- $CoR_1 = DT \times TL \times PR$

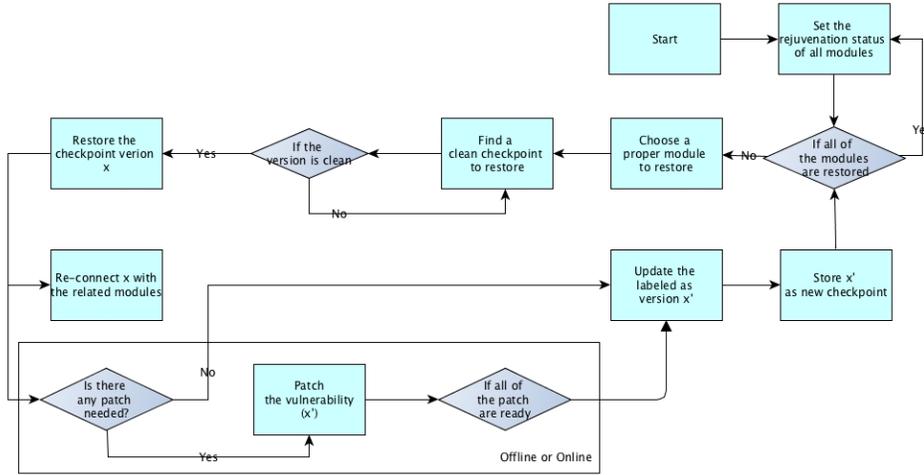


Fig. 2: The workflow of secure rejuvenation mechanism

- $CoR_2 = V \times SF$
- $CoR_3 = V \times TF$

In addition to periodic rejuvenation, ad hoc rejuvenation (rollback to clean checkpoint and restart) is also applied when an abnormal condition or a security violation is detected. The detection may be based on monitoring system performance or other indicators. For example, performance indicators, including memory allocations, CPU usage, network traffic, disk writes, may indicate abnormal behavior of applications. In general, one should collect a large number of performance indicators and apply Principle Component Analysis (PCA) to determine key indicators. We will include the cost of monitoring the system to identify abnormal conditions in the cost of rejuvenation. The cost depends on the volume  $V$  of information monitored.

$$CoR_{adhoc}(V, t) = CoM(V, t) + CoR(V, t) \quad (3)$$

We will use  $CoM(V, t)$  to represent the monitoring cost at the time a rejuvenation is warranted, and  $CoR(V, t)$  to represent the cost of restarting the module that exhibited an abnormal behavior. The second component can be assumed to be similar in cost to that of the periodic rejuvenation shown above.

$$CoR_{adhoc}(V, t) = CoM(V, t) + CoR_{periodic}(V, t) \quad (4)$$

Since the ad hoc rejuvenation can take place at any time between scheduled periodic rejuvenation, we will use a probability distribution that associates the probabilities of detecting an abnormal condition over this period of time. We can now compute the expected cost of rejuvenation that includes both ad hoc and periodic rejuvenation as follows.

$$CoR_{total}(V, T) = \int_0^T f(t)(CoM(V, t) + CoR_{periodic}(V, t)) dt \quad (5)$$

Here  $f(t)$  is the probability density function that reflects the probabilities of detecting abnormal behaviors.  $T$  is the scheduled rejuvenation period.

Consider for example that it takes 17 seconds to rejuvenation a system (i.e., the downtime is 17 seconds [9]), the average number of transactions lost in a year is 355.72 [9], the average potential revenue of a transaction lost is \$100,000, and the storage fee charged by Amazon is \$0.095 per GB-month and data transfer fee is \$0.12 per GB, the cost of performing each periodic rejuvenation is \$19.16 for the 10 GB cloud instance [10]. If we assume that in addition to hourly scheduled rejuvenation, ad hoc rejuvenations are warranted with a probability of 10% in between scheduled rejuvenations and if we assume that monitoring consumes 0.1% of CPU time, the total cost of rejuvenation can be estimated as

$$CoR_{total}(10GB, 1hr) = 1.9161401 + 19.16 \quad (6)$$

It should be noted that the above examples are just for illustration only. We are currently testing our models in a controlled environment, both to validate our cost models, and also to understand the benefits of rejuvenation in eliminating or mitigating security threats.

### 3. Analyss and Comparison of Rejuvenation for Security

In this section we describe the capabilities in terms of defense against various security threats and cost associated with rejuvenation and malware scanning techniques.

### 3.1 Characteristics comparison

Rejuvenation has been used as a fault-tolerant approach in software systems. In a similar manner, rejuvenation can be applied as a defense against security threats. By restoring components to clean or healthy states, rejuvenation can make the system less prone to catastrophic failures. Without rejuvenation, survivability of systems is in jeopardy because of hidden failures and software aging. In Table 1, we compare the capabilities of rejuvenation with malware scanning when applied to survivable systems. While facing DoS [11] or low-rate DoS attacks [12], rejuvenation can reboot the system and recover quickly. Current approaches perform log analyses to detect DoS attacks, which can be costly and slow to recover. Secure rejuvenation, like survivable system, can eliminate or mitigate the effects of several types of malware. Some malware that cannot be eliminated using rejuvenation only include trapdoors and salami attacks: the former is eliminated by compiler-based code checker and detected by resource monitors, and the other is detected by semantic code analyzers.

Rejuvenation can eliminate the 10 most common malware, based on how widely the malware is distributed [13]; these are listed in Table 2. Some of these are browser related malware which redirects the user to advertisement pages or embeds the advertisement plugins on the user's browser. Still other advertising malware may change the system registry. These malware often push the advertisement windows, ask the user to click while the user is executing the infected applications. Malware scanning can also eliminate these threats.

Table 1: The comparison of the features, and the abilities of treat elimination between rejuvenation and malware scanning for security.

Feature	Rejuvenation	Malware Scanning
Fault avoidance	Partial	No
Fault tolerance	Yes	Yes
Denial of Service(DoS) or Low-rate DoS	Reboot	Log analysis
Worm elimination	Restore to checkpoint	Scanning
Logic bombs elimination	Restore to checkpoint	Scanning
Trapdoors elimination	No	No
Virus elimination	Restore to checkpoint	Scanning
Trojan horse elimination	Restore to checkpoint	Scanning
Salami Attacks elimination	No	No
Automated software-patching	Yes	Yes
Intrusion dection	No	Yes

- Log analysis of survivable system could be performed by multivariate correction analysis, feed-absed control analysis, or other mechanisms.
- Secure rejuvenation can avoid partial faults, but not for all.

### 3.2 Cost comparison

Figure 3 shows the cost of rejuvenation performed periodically for different frequencies: four hours, six hours, 12

Table 2: Top 10 malware and mobile malware in 2013. [14]

Malware	Mobile malware
Adware.Relevant.CC	DangerousObject.Multi.Generic
Adware.NewNextMe.A	Trojan-SMS.AndroidOS.OpFake.bo
Win32.Application.-SearchProtect.O	AdWare.AndroidOS.Ganlet.a
Gen:Variant.Graftor.739	Trojan-SMS.AndroidOS.FakeInst.a
Gen:Adware.Plus.1	RiskTool.AndroidOS.SMSreg.cw
Gen:Variant.Adware.-Graftor.125313	Trojan-SMS.AndroidOS.Agent.u
Script.Packed.IFrame.K@gen	Trojan-SMS.AndroidOS.OpFake.a
Gen:Variant.Graftor.82095	Trojan.AndroidOS.Plangton.a
Win32.Application.Somoto.C	Trojan.AndroidOS.MTK.a
NSIS.Adware.-OneClickDownloader.B	AdWare.AndroidOS.Hamob.a

hours, and 24 hours over a year. The cost of rejuvenation over a year depends on the frequency of rejuvenation and the cost of each rejuvenation. In Figure 3 we did not include monitoring and ad hoc rejuvenation costs, since these costs depend on the probability of detecting an abnormal condition. The figure also shows the cost of scanning for malware. The cost of scanning depends on the size of the system being scanned (for example, scanning a 32 GB estimated at \$0.351/hr [8]). Although systems continuously scan for malware, in Figure 3 we have shown the cost of malware scanning under different scenarios. The red horizontal lines represent the cost of scanning continuously. We also show the cost of malware scanning when scanning takes place at four, six, 12 and 24 hour periods - similar to rejuvenation. Systems need only to scan new information generated during the period and we assume that the amount of new information generated is proportional to the length of the period. It can be seen that the cost of rejuvenation decreases with the decrease in frequency (less frequent rejuvenation). The cost of continuously scanning for malware (see the three dash lines) is higher than rejuvenation at certain rejuvenation periodicities. If one assumes that systems scan for malware at fixed internals (such as every four hours), rejuvenation costs are higher except when it takes place once every 80-100 hours. Based on our assumptions and cost models, rejuvenation once every 24 hours appears to be a reasonable choice for different system sizes.

However, the data in this figure is not fair to rejuvenation. We anticipate rejuvenation taking place one module at a time, instead of rejuvenating the entire system with 32, 40, or 80 GB, thus only a small number of modules become unavailable at a time. But Figure 3 shows data assuming that the entire system is rejuvenated at the same time. The analysis in this paper is based on our model and published costs of computing and storage. We are currently working on collecting experimental data to more accurately compare rejuvenation with scanning for malware.

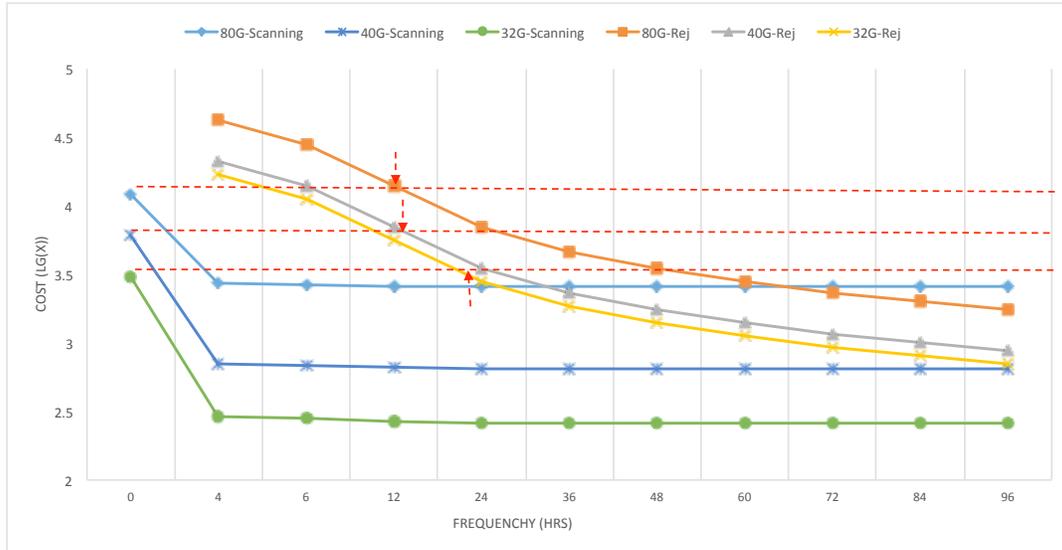


Fig. 3: Cost comparison of secure rejuvenation versus malware scanning

### 3.3 Other Benefits of Rejuvenation

Malware is getting more sophisticated and the sophistication is increasing in recent years. McAfee's report shows that there are over 100,000 new malware instances detected in a given day; that is 69 new threats every minute [15]. There are three phases in the detection and elimination of malware. The first is the undetected phase in which the malware strain cannot be detected in the system. The second is the identification phase in which the malware strain is detected as a malicious code pattern used to generate its signature. Finally, the malware strain enters the detected phase after its signature is updated. A study by Damballa demonstrated that the typical gap between malware release and detection using anti-malware is 54 days, almost 8 weeks [16]. Nearly half of the 100,000 malware samples go undetected on the zeroth day (first testing day), and there were at least 15% of the samples remained undetected even after 180 days. This means that the system may suffer from undetected malware for a long period of time.

Suppose a system component is infected with 100K malware strains every day since it is released, the number of potential malware strains hidden may increase over the next several weeks before some strains are detected. On average it will be 9 weeks when detected malware signatures are released, and the number of hidden malware will be reduced as shown in the Figure 4.

In contrast, the proposed rejuvenation mechanism periodically restores the component as a "clean" version (checkpoint), thus the exposure of the system to new malware introduction is the time between rejuvenations. Assuming that the component is rejuvenated once a day, it remains in "clean" status at the beginning of each day. After the 9th week, some malware strains are eliminated because of the

signatures, thus the potential malware strains lurking may decrease as long as the backup version is not infected.

Consider the recently publicized *heartbeat* vulnerability (CVE-2014-0160) with SSLs [17]. Rejuvenation can at least mitigate the amount of information leaked since rejuvenation will restore the system to a clean state and memory is refreshed with information.

### 3.4 Application of Cloud Services

Our rejuvenation mechanism can be applied on VM-based environments such as cloud services, as well as mobile devices (eg., Android). For cloud services, the services can be rejuvenated one service at a time such that the impact of rejuvenation is not felt by the entire system. And the speed of checkpointing and restoring services to a clean checkpoint is increasing in most virtual systems, while the cost of storage and transfer of checkpointed data is becoming less expensive [9]. Furthermore, any patches or upgrades to services can be done separately from a running system.

### 3.5 Application of Mobile Device

Kaspersky Lab's report shows that there are 143,211 malicious programs detected in 2013 and they detected approximately 10,000,000 unique malicious installation packages in 2012-2013 [14]. Sometimes malware resists the anti-malware protection because the following reasons:

- Obfuscation: Confusing the anti-malware code such that analysis becomes difficult. The more complex the obfuscation, the longer it will take for anti-malware software to neutralize the malicious code.
- Android vulnerability: Malware uses the vulnerabilities to bypass the code check, enhance the privilege to

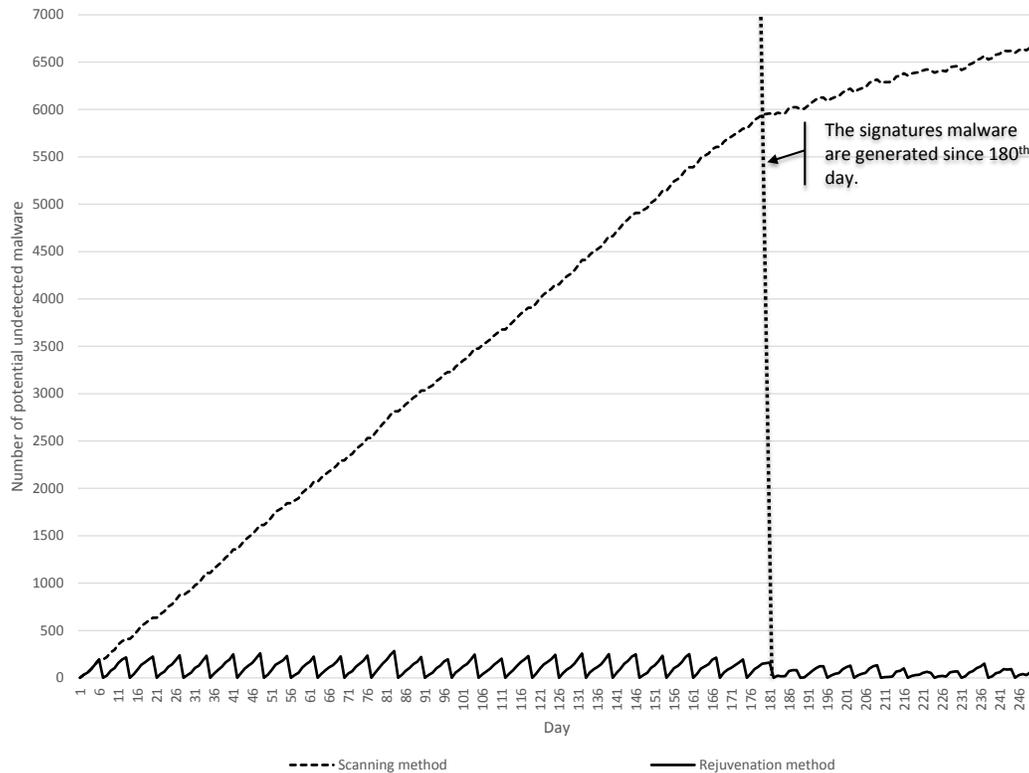


Fig. 4: The potential number of malware remaining in a system after use of scanning versus rejuvenation

extend their capabilities, and make it more difficult to be removed, like Trojan-SMS.AndroidOS.Svpeng.a.

Therefore, it is difficult for normal users to remove malware, since most of the malware is embedded in the legitimate software and get the administrator privilege during the installation. There are only two options for users. One is to reset the system to factory settings, but some malware could obstruct this reset. The other is to apply anti-malware software to continuously scan, analyze, and eliminate it; but this consumes processing and thus the battery life of the device.

Our rejuvenation mechanism restores the checkpointed image from either the storage of the device or from some external storage, or may rely on trusted zones to bring the system to a clean or consistent state. If the rejuvenation is performed while the device is connected to a power source, the battery life is not affected. Rejuvenation can be performed on a regular basis, similar to checking periodically for software patches and upgrades. Rejuvenation mechanism, therefore, is more suitable in a mobile environment, than malware scanning techniques.

## 4. Conclusion

The cybersecurity of Cloud-based computing systems are becoming critical to modern society as we are becoming

ever more dependent on information infrastructures. Balancing system reliability, availability and security is complex. Malware and other security threats are becoming more sophisticated. Thus a multipronged approach is necessary to improve security as well as system survivability. We feel that software rejuvenation, which has been successfully employed as a fault-tolerant mechanism, can also be used as a defense against security threats. In this paper, we introduced a model that can be used to compare the costs associated with rejuvenation and malware scanning so that one can determine the rejuvenation frequencies that lead to cost-effective defense against hidden threats. While we compared rejuvenation as an alternative to scanning in this paper, they should be used together. We are currently testing our models in a controlled environment, both to validate our cost models, and also to understand the benefits of rejuvenation in eliminating or overcoming the security threats.

## Acknowledgment

This research is supported in part by the NSF Net-centric and Cloud Software and Systems Industry/University Cooperative Research Center and NSF award 1128344.

## References

- [1] F. Lin, "Re-engineering option analysis for managing software rejuvenation," *Information and Software Technology*, vol. 35, no. 8, pp. 462–467, Aug. 1993.
- [2] Y. Huang, C. Kintala, N. Kolettis, and N. D. Fulton, "Software rejuvenation: analysis, module and applications," in *Proc. of Twenty-Fifth International Symposium on Fault-Tolerant Computing, 1995. (FTCS-25)*, Pasadena, USA, 1995, pp. 381–390.
- [3] R. Agepati, N. Gundala, and S. V. Amari, "Optimal software rejuvenation policies," in *Proc. of 2013 Reliability and Maintainability Symposium (RAMS)*, Orlando, USA, Jan. 2013, pp. 1–7.
- [4] S. Oikawa, "Independent kernel/process checkpointing on non-volatile main memory for quick kernel rejuvenation," in *Proc. of Architecture of Computing Systems (ARCS 2014)*, ser. LNCS. Springer, 2014, vol. 8350, pp. 233–244.
- [5] "Pandalabs annual report," PandaLabs, Tech. Rep.
- [6] "Symantec global internet security threat report." Symantec Corp, Tech. Rep., 2008.
- [7] "Scan speeds for 2011/2012 antivirus software," Antivirus Ware, 2011. [Online]. Available: <http://www.antivirusware.com/testing/scan-speed/>
- [8] A. W. Services, "Amazon ec2 price," 2013. [Online]. Available: <http://aws.amazon.com/ec2/pricing/>
- [9] F. Machidaa, D. S. Kim, and K. S. Trivedi, "Modeling and analysis of software rejuvenation in a server virtualized system with live vm migration," *Performance Evaluation*, vol. 70, pp. 212–230, 2013.
- [10] "Scan speeds for 2011/2012 antivirus software," Antivirus Ware, Tech. Rep., Sep. 2011. [Online]. Available: <http://www.antivirusware.com/testing/scan-speed/>
- [11] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R. P. Liu, "A system for denial-of-service attack detection based on multivariate correlation analysis," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 2, pp. 447–456, Feb 2014.
- [12] Y. Tang, X. Luo, Q. Hui, and R. Chang, "Modeling the vulnerability of feedback-control based internet services to low-rate dos attacks," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 3, pp. 339–353, March 2014.
- [13] "Malware distribution by percentage within the top 10," G Data Software AG, 2013. [Online]. Available: <https://www.gdatasoftware.co.uk/security-labs/statistics/top10-malware.html>
- [14] V. Chebyshev and R. Unuchek, "Mobile malware evolution: 2013," Kaspersky Lab ZAO, 2013. [Online]. Available: [http://www.securelist.com/en/analysis/204792326/Mobile\\_Malware\\_Evolution\\_2013](http://www.securelist.com/en/analysis/204792326/Mobile_Malware_Evolution_2013)
- [15] "Infographic: The state of malware 2013," McAfee, Inc., Tech. Rep., Apr. 2013. [Online]. Available: <http://www.mcafee.com/us/security-awareness/articles/state-of-malware-2013.aspx>
- [16] "3% to 5% of enterprise assets are compromised by bot-driven targeted attack malware," Damballa, Inc., Tech. Rep., Mar. 2008.
- [17] "Tls heartbeat read overrun (cve-2014-0160)," OpenSSL Security Advisory, 2013. [Online]. Available: [https://www.openssl.org/news/secadv\\_20140407.txt](https://www.openssl.org/news/secadv_20140407.txt)

# Ontology-based Privacy Setting Transfer Scheme on Social Networking Systems

Chen-Yu Lee<sup>1</sup>, Krishna M. Kavi<sup>1</sup>, and Mahadevan Gomathisankaran<sup>1</sup>

<sup>1</sup>Department of Computer Science and Engineering, University of North Texas, Denton, TX 76203, USA

**Abstract**—*In this age of eternal connectedness, using various social networks, privacy is an important problem that needs to be transparent. Even though social networks provide a user the ability to control their privacy settings, it is often difficult to get similar privacy settings across different social network systems. This is because: (a) privacy settings rely on complicated privacy rules which define the access control to different elements by different groups; (b) the terminology used varies in different social networking systems; (c) for a typical user, it is hard to set all the privacy settings as desired due to the complicated navigation through the social networking sites. Our goal is to design a framework that enables the transfer of privacy settings among social networking systems. We collect a user's privacy settings for many social services and store them in an ontology database. When a user registers on a new social network, our system provides recommendations of settings for it based on the user preferences as indicated by the settings in other services. Our framework covers personal privacy settings and the settings of the relationships of groups/tags and other elements.*

**Keywords:** social networking systems, privacy settings, ontology, social networking systems, privacy settings, ontologies

## 1. Introduction

Social networking system and services have become an important part of on-line world for most people. Social networks allow users to share information among friends, groups, and companies instantaneously around the world. While sharing information is an important social phenomenon, the risk of losing privacy increases. Furthermore, the unsuspecting users may be prone to identity theft<sup>1</sup>, password disclosure<sup>2</sup>, account cracking<sup>3</sup>, and so on.

Facebook<sup>4</sup> is a popular social networking service that has more than 1.19 billion users, as of September 2013, and is still growing [1]. For the past few years, Facebook

allowed users to control their privacy settings in terms of sharing personal information, digital objects and other information, with other uses and third party services. In June 2012 Consumer Reports magazine reported that at least 13 million users had never set, or knew about Facebook's privacy tools and 28% of users shared all, or almost all of their wall posts with a wider public than their friends and sometimes to the entire public [2]. In general, we feel that users who have accounts on multiple social networks would like to have similar privacy settings. It is often difficult to get similar settings across different social network systems for three reasons: (a) privacy settings rely on complicated rules which control the access to different elements by different groups; (b) the terminology used varies greatly from one social network to another; (c) for a typical user, it is hard to set all the privacy settings as desired due to the complex navigation through the social networking sites. In this paper, we propose a framework that enables the transfer of privacy settings between social networking systems. We collect a user's privacy settings on many social services and store them in an ontology database. When the user registers on a new social service, our system provides recommendations for settings for the new one based on their setting on other services. Our framework covers personal privacy settings and the settings of the relationships of groups/tags and other elements.

The rest of the paper is organized as follows. Section 2 discusses research that is closely related to ours. Section 3 introduces our ontology model of security and privacy on social network systems and the privacy permission model is described in section 4. Section 5 presents the privacy transfer scheme and our experimental prototype is explained in section 6.

## 2. Related Works

Research on ontology-based privacy control started with development of ontologies for access control on social networking services. Kruk et al. proposed a Friend-of-a-Friend (FOAF) model which describes the social relationship as a directed graph [3]. FOAF-Realm is one of the earliest schemes to apply the FOAF ontology model for making decisions on resource access control according to the friendship levels. In Carminati's model [4], each authorization rule is designed subject to the type, depth, and trust level of the

<sup>1</sup>Social Thievery: Will Your Tweets Get You Robbed? Available: <http://mashable.com/2011/11/01/social-thievery-infographic/>

<sup>2</sup>Facebook ID Can Be Hack by Stealing Security Question - Answer, Available: <http://hackworm.blogspot.com/2013/03/facebook-id-can-be-hack-by-stealing.html>

<sup>3</sup>The hacker who broke into Mark Zuckerberg's Facebook page will get a \$12,000 reward from online donors. Available: <http://www.dailymail.co.uk>

<sup>4</sup><https://www.facebook.com/>

relationship which is represented in OWL. Villegas et al. [5] proposed a personal data access control (PDAC) scheme to classify the community of users into three parts: acceptance, attestation, and rejection using the "trusted distance" measure. The trusted distance is measured by the relation hops between users and the other experiential information. Finin and Elahi [6] relied on role-based access control (RBAC) policies in social environment using OWL [7]. In 2009, Carminati's framework defined access control policy, filtering policy, and admin policy encoded in Semantic Web Rule Language (SWRL) [8]. Masoumzadeh's OSNAC system supports both user and system level authorization. It defined three main concepts: DigitalObject, Person, and Event in user level authorization [9]. Li et al. proposed a SPAC system, which extracts the privacy configuration patterns from each user's profile and privacy settings using a semantics-enhanced K-Nearest Neighbors (K-NN) classification algorithm. The system predicts the privacy setting for new friends based on the patterns [10]. Shehab et al. presented an access control framework which enables users to specify shared data attributes and use the shared data as to manage third party applications in social networking services in 2012 [11].

In 2013, Masoumzadeh proposed a policy analysis framework to theoretically reason the missing pieces of policies and controls [12]. Kayes et al. proposed an ontology-based social ecosystem data model to generate platform-independent default privacy policy settings for each relationship group of a user according to multiple types of social interactions captured from various sources on a user's devices [13]. Masoumzadeh and Kayes' works inspired us to explore the development of a framework that allows transfer of privacy settings from one social network system to another.

It is not uncommon for people to have memberships in multiple social networks, and there are many services providing the ability to manage profile settings and to integrate the social medium on multiple social network sites such as Atomkeep<sup>5</sup>. However, the previous works are used to perform the access control in a social site, but they cannot solve the problem of porting privacy settings from one social site to another. A typical user does not want to learn how to manage settings in every new social network. Our framework aims to address this need.

### 3. Modeling Social Networking System Information

Our model of a generic social networking service consists of three parts: user, digital objects and provider. Our model is expressed in the Web Ontology Language (OWL2) in which each concept in the social service is modelled by an abstract object *class*. The relationships between classes are captured

<sup>5</sup><http://atomkeep.com/>

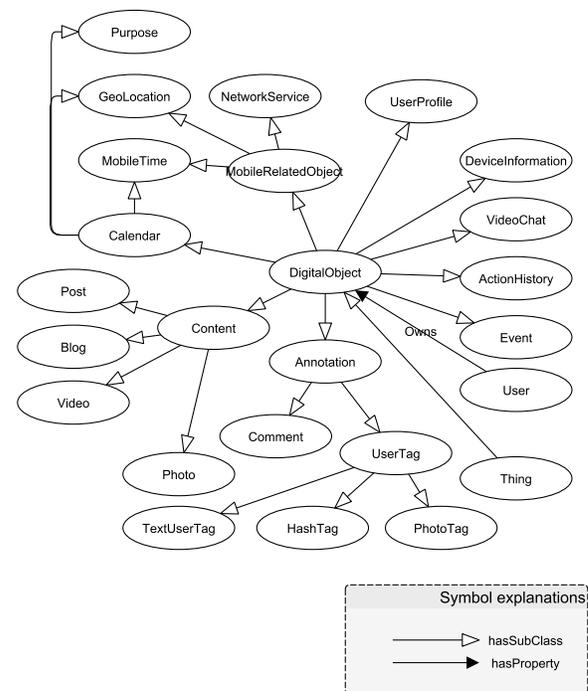


Fig. 1: The DigitalObject class in our ontology model

by *object properties*, and the relationship between classes and data values are captured by *data properties*.

#### 3.1 Digital Object Model

To model a social system, it is important to model its content because the content is the target of all the access controls. In our OWL model, we take a DigitalObject class to capture the users' content in the social networking service as shown in Fig. 1. The content of a user can be divided mainly into two parts: user activities and user profiles. In general, people post their messages, photos, or videos on sites and leave some comments, or place tags on friends' posts. We model the former post as Content which contains Blog, photo, video, and post; the latter is modeled as Annotation which has Comment, TextUserTag, PhotoUserTag, HashTag, and URI subclasses. People also like to arrange schedules of activities which captured as Calendar composed of GeoLocation, MobileTime and Purpose. In addition, some information may be saved by social sites such as footprints of life, and activity history, which are modeled as MobileRelatedObject, ActionHistory in our model.

#### 3.2 User Profile Model

While exploring the issues of personal privacy, in addition to the content on the social networking site, a

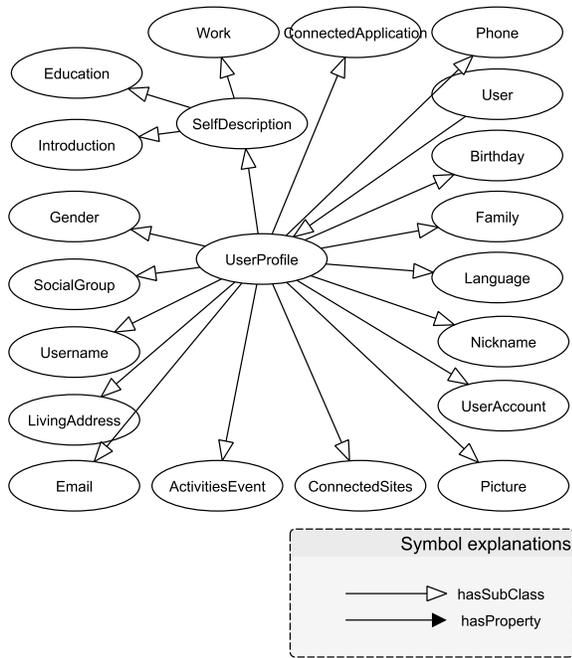


Fig. 2: The User class in our ontology model

user's personal information is critical. Personal information, or user profile, is a special content in the system because it is an important reference for people to know with whom they are interacting in social networks. Therefore, the publicity of the user profile should be controlled carefully. In general, the user profile contains many pieces of personal information such as photo, email, birthday, gender, phone, address, and other information related to the user's background like education, language, and work experiences. To capture all of a **UserProfile**, we model **UserAccount**, **UserName**, **Picture**, **Email**, **BirthDay**, **Nickname**, **Phone**, **Gender**, **Language** and **LivingAddress** which is different from the concept of location. We also design some classes for personal experience like **Work**, **Education**, and **SocialGroup**. The user profile model is shown in Fig.2.

### 4. Modeling Privacy Sensitive Permissions

Personal privacy is guaranteed by access control policies which determine what information is revealed to whom. In our framework by default the data is deemed private to anyone except its owner. We design some access control properties to present the sharing status of each content. For example,

- 1 Alice\_Email isShared Bob
- 2 Alice\_Email.email:"Alice@abc.com"

where Alice\_Email is an instance of Email address of Alice. The statement allows Bob to share Alice's email, Alice\_Email. In the ontology, the statement does not reveal if Bob also shares other emails of Alice or if Alice permitted sharing of any other type information with Bob. Our goal is to capture these types of information. To do this, we define the following properties with access control.

- **Searchable**: Indicates that the object is granted searchable access in a service provider's search engine.
- **isShared**: Denotes an object shared with a user, i.e., a user is granted access to the shared information. As in the previous example, the statement shows that Alice's mail is shared with Bob.
- **Shareable**: Shareable property holds when the object is granted shared access to at least one user.

$$1 \quad \text{Shareable: } \exists \text{ isShared.User}$$

- **isTagged**: Tag-adding is a popular interaction between users in social sites. They like to add tags to posts, photos, videos, and any other content objects. "The isTagged property implies that the object permits tagging by other users. For example, a photo owned by Alice, which is tagged by Bob.

- 1 Alice Owns Photo.df0075d2-b26e-4f6d-bbef-6df56ae8d653
- 2 Bob Creates PhotoUserTag.5c934274-c53e-4f22-8bcb-2d8fa793a9ec
- 3 Photo.df0075d2-b26e-4f6d-bbef-6df56ae8d653 isTagged PhotoUserTag.5c934274-c53e-4f22-8bcb-2d8fa793a9ec
- 4 PhotoUserTag.5c934274-c53e-4f22-8bcb-2d8fa793a9ec.User: Bob

where "Photo.df0075d2-b26e-4f6d-bbef-6df56ae8d653" is an instance of user photo, "PhotoUserTag.5c934274-c53e-4f22-8bcb-2d8fa793a9ec" is an instance of PhotoUserTag created by Bob, and

$$1 \quad \text{PhotoUserTag, TextUserTag, HashTag } \subseteq \text{ UserTag}$$

- **Taggable**: An object with the Taggable property implies that users may add tags to the object.
- 1 Taggable:  $\exists$  isTagged.UserTag
- **isLinked**: People often add photos on their web pages or blogs to make them rich. The property indicates that the content object in the system is linked by a URI which is a local or a foreign link. For example, Alice's photo is linked.

- 1 Alice Owns Photo.df0075d2-b26e-4f6d-bbef-6df56ae8d653
- 2 Alice Owns URI.809bb690-6656-4bbf-b28b-c3c7ce86be0e
- 3 URI.809bb690-6656-4bbf-b28b-c3c7ce86be0e.uri: "www.csrl.edu/photo/alice"

4 Photo.df0075d2-b26e-4f6d-bbef-6df56ae8d653 isLinked

- Linkable: The property holds if the object is granted linking capability.

1 Linkable:  $\exists$  isLinked

- IsCommented: Making comments on a friend's post, photos, or videos is a common operation in social systems, allowing the ability to chat asynchronously, express feelings, thinking, and other actions as they would say to each other in face-to-face meetings. But people may also decide not to permit comments on their posts. The property defines if commenting access is granted to the object or not. For example, Alice allows her posts to be commented by the public.

1 Alice Owns Post.6a299789-1c78-4b4d-8746-044b4234c225

2 Post.6a299789-1c78-4b4d-8746-044b4234c225 isCommented.User PublicUser

where "Post.6a299789-1c78-4b4d-8746-044b4234c225" is an instance of Post owned by Alice and

1 PublicUser:  $\forall$  User

- Commentable: The property holds if the object has granted commenting capability to other users.

1 Commentable:  $\exists$  isCommented.User

When a social network system is widely adopted, the account service would be designed to be a public identity service for signal sign-on to provide the federated identity for authentication mechanism like OpenID<sup>6</sup>. In addition to those properties used for the user's inner social system, some properties are designed to restrict access by other applications or sites.

- is3rdPartyAPPShareable: The social networking system will build a platform for third-party application developers who may provide useful services to the users in the social system. The property allows access of the object by third party applications running inside the social system. The following statement, for example, allows a social game, The Sims, access to Alice's email information.

1 "Alice@abc.com" is3rdPartyAPPShareable TheSims

- is3rdPartySiteShareable: The property is similar to is3rdPartyAPPShareable, but the subject is third party sites which are built as standalone services rather than services built on the inner social platform.
- is3rdPartySiteLoginable: This property is also like the previous two, but it focuses on user login. The user can determine whether the account in the social system

<sup>6</sup><http://openid.net/>

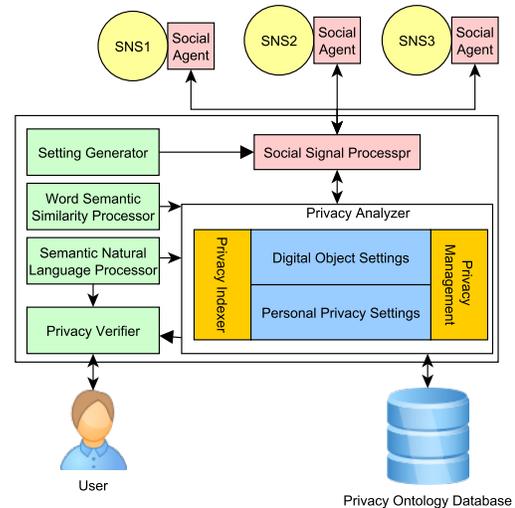


Fig. 3: The system architecture of our scheme.

is used as a federated identity for authentication of accounts on other sites. In the case of consent, the user can login to the target site using the social network account in the system.

## 5. Privacy Transfer Scheme

We propose a privacy transfer scheme which assists users in handling the privacy settings on social networking services with comparable preferences. The scheme is mainly divided into two parts: extraction and setting processes. The former extracts the privacy settings from one or more social networking services, and the latter sets the settings on a new service.

### 5.1 System Architecture

To realize our privacy transfer scheme, we designed several modules to extract information, analyse privacy settings, format the settings in the required manner, and recognize the similarity of terms used in different social networking services. Our system architecture is shown in Fig. 3 and each module is described below.

- Social agent: A third party application that extracts privacy information about a user and sends it back to our social signal processor. It works under the requirements: (1) the user registers the application and (2) allows access to privacy information. A user may share some information with friends and applications (including our social agent) or the user uses default settings (except for allowing our social agent). Our agent being a third application, it can gather the sharing information equal to that shared with friends or applications.
- Social signal processor: This module gathers privacy information sent from social agents executing on different

social networking services and then sends them to the privacy analyzer. It is also responsible for placing the generated privacy setting on a new social networking service.

- **Privacy analyser:** The default setting of each index in the database is private, not shared with any object (friend or application). On the other hand, if privacy index is PUBLIC, the information is available to all objects. The privacy information is grouped into three parts: digital object, personal privacy, and access control settings. The analyser works by analysing the settings using our ontology model, setting the privacy indexes, and managing the user instances. Finally, the module saves user's instances in the privacy ontology database. For example, Alice shares her living address with Bob, the ontology database stores

- 1 Alice Owns Alice\_LivingAddress
- 2 Alice\_LivingAddress:1 Oak St. #1, Denton, Texas"
- 3 Alice\_LivingAddress isShared Bob

where "1 Oak St. #1, Denton, Texas" is an instance of `LivingAddress`. Our system analyses the shared information, including personal privacy, to provide privacy setting recommendations.

- **Setting generator:** This module generates privacy setting scripts to be sent to the social signal processor for uploading them to a new social networking service.
- **Word semantic similarity processor:** As stated previously, the terms used by different social networking services are different, and sometimes they may use different terms to mean the same thing. We use word similarity between the terms and indexes [14].
- **Semantic natural language processor:** This module provides the ability to generate user friendly explanations for each generated privacy rule so that the user can approve selected settings.
- **Privacy verifier:** This module asks the user to verify the settings generated by setting generator. The settings can be uploaded to social networking services only if they are verified.

## 5.2 Extract Privacy Setting

Assume that the user already has a privacy setting on one social networking service, and he/she wants to transfer the setting to another one. The user has to register our application plug-in on the first service. Then the application begins the privacy setting extraction process following the workflow illustrated in Fig. 4.

- Step 1 The **social agent**, installed on the social networking service as a third-party plug-in, extracts privacy information.
- Step 2 **Social signal processor** gathers the social information from multiple social sites, if available, and sends them to privacy analyser for analysis.

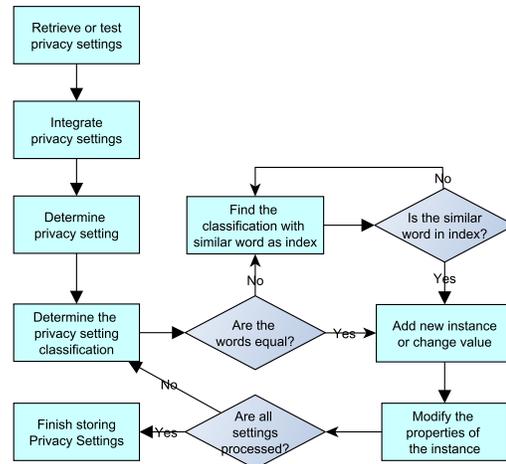


Fig. 4: The workflow of extracting the privacy setting from social networking services.

Step 3 **Privacy analyser** first distinguishes the privacy settings from received information and analyses the privacy setting according to the ontology indexes such as `Birthday`, `Gender`, etc. If it matches, the analyser adds a new instance of the classes or changes the value of the existing instances. If it does not match, the analyser finds a similar word as an index and then adds an instance to the index class. Usually, a pair of instances are added/modified for each privacy rule.

Step 4 Store the instances and properties in our ontology database. If all the privacy settings are not processed completely, go back to **Step 3**.

## 5.3 Set Privacy Setting

Suppose a user registers with a social networking service and installs our plugin application. Our scheme can then transfer his/her privacy settings to a new social networking service as described in the workflow shown in Fig. 6.

- Step 1 When the social signal processor receives the request to transfer privacy settings onto the target social networking service, it asks the privacy analyzer to collect the settings from the database.
- Step 2 The privacy analyzer collects the privacy information of the user for the target service and then sends the information to the privacy verifier for verification.
- Step 3 The **privacy verifier** adds annotations for each privacy setting for the user to verify.
- Step 4 The user can read the annotations to understand the corresponding settings and determine if they fit his/her preferences. For example, the annotations detail which privacy information is shared with which friend, group, or application, or if

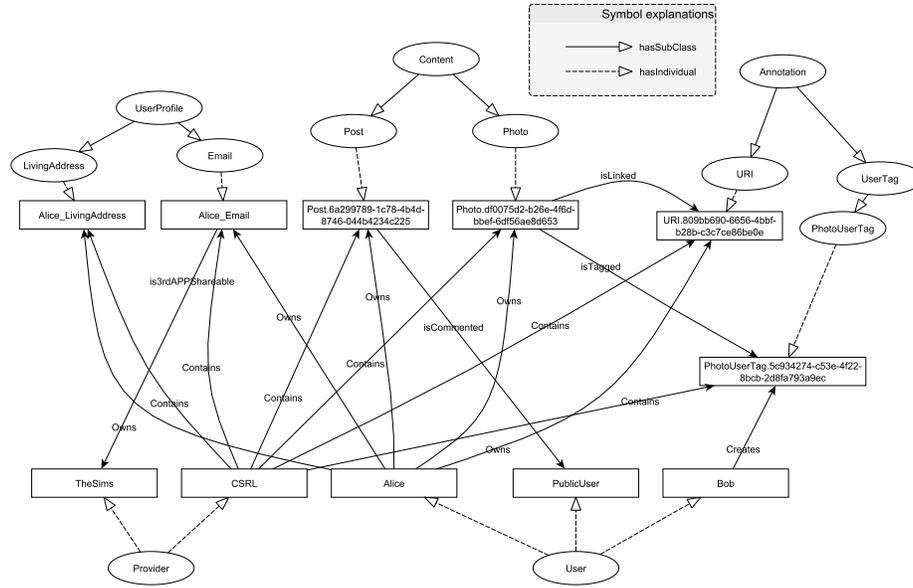


Fig. 5: A part of the privacy instances in our system which is described in section 4.

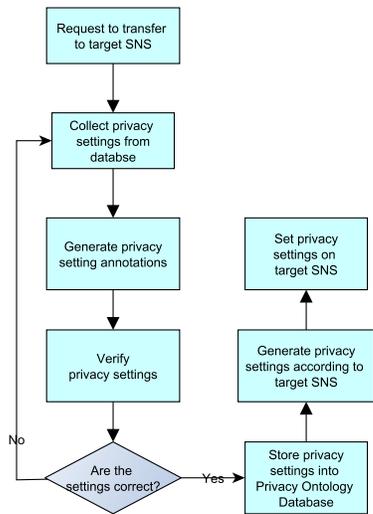


Fig. 6: The workflow of setting the privacy setting on new social networking services.

equivalent privacy settings to be set on the target social networking service.

## 6. Implementation

### 6.1 Prototype

Our ontology-based scheme is implemented by the Protégé and Drupal platform to examine the performance of transferring a user’s privacy settings. Drupal is an open source content management system that provides modules to build a social networking system. It also provides an application programming interface (API) for developers to create third party applications on the constructed social network system.

Fig. 5 shows a part of the privacy instances in our system which is described in section 4. Each ellipse is a class and each rectangle is an instance of the corresponding class. The arrow with a solid line indicates a subclass: e.g., Content has two subclasses Post and Photo. The arrow with a dashed line indicates individuals: e.g., User class has three individuals Alice, Bob, and PublicUser. The prototype stores the privacy settings of users on CSRL<sup>7</sup> as a service provider in our prototype. Because of a large number of individuals for some classes, especially Content and Annotation, they are named with a Universally Unique Identifier (UUID) to prevent name conflicts: e.g., Post.6a299789-1c78-4b4d-8746-044b4234c225.

<sup>7</sup>CSRL stands for Computer Systems Research Laboratory.

the information is set public. If there are errors in the privacy settings, the user can correct them manually and go back to **Step 3**.

- Step 5 The privacy analyser stores the correct privacy settings into the database.
- Step 6 The privacy analyser sends the settings to the social signal processor to generate privacy settings for the target social networking service.
- Step 7 The social signal processor sends the generated privacy script to the social agent which causes the

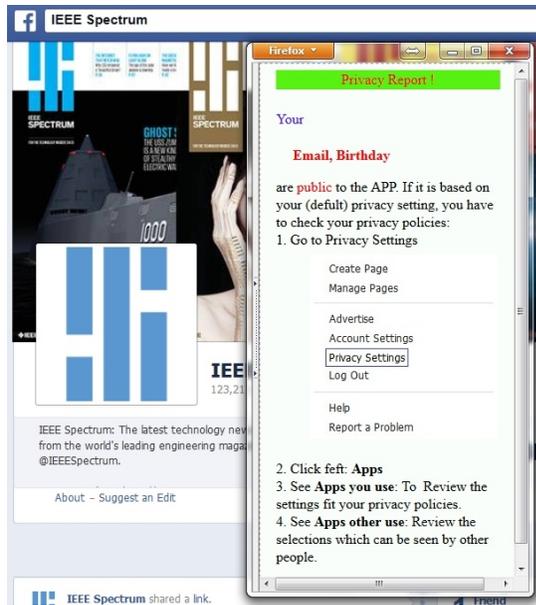


Fig. 7: An experiment for Facebook. The prototype is built on Facebook platform to notice users' privacy settings.

## 6.2 The Restriction on Public Social Networking Services

Our goal is to apply our scheme on real social networking services such as Facebook, Twitter, Google+, and so on, but at this time this is not feasible due to these restrictions:

- Third party application restriction: The third party application on public social networking services can only obtain limited information. Due to the privacy management and protection, only limited general personal information (which may include user name, friend list, location, time, etc.) is accessible to third party applications.
- Independent application restriction: The API of the service provider normally does not provide any information from one application to other applications, since each application is assumed to be independent on the platform. Thus it is difficult to understand which user information is shared with another application, and further to provide some privacy recommendations.

For these reasons, our scheme applied in a real social networking service, such as Facebook, can only extract the privacy information that is shared by users. However, this is adequate to demonstrate our framework. Our App on Facebook can access the user's email and birthday because the default privacy setting in Facebook is different from the setting stored in our system. In this case, our app produces a pop-up window as shown in Fig. 7. The user can learn how to change the setting by following our instructions step by step. We believe, therefore, users can then discovery the correct place to easily make the equivalent privacy setting.

## 7. Conclusion

Since Facebook became popular in social networking, there are more companies providing their own social networking services, including Google+, Qzone, Tumblr, and so on. It is unreasonable to expect users to acquaint themselves with every service's specific process for making various privacy settings.. We proposed a privacy transfer scheme to alleviate this problem. Our scheme not only provides recommendations to users on selecting their privacy settings, it also provides the ability for users to store and manage these settings. Users may have different privacy settings in different social networking services for different purposes, but if they desire to accept the similar settings, our scheme provides some recommendations and step-by-step guidance for them. In the future, we will extend our scheme to privacy management on mobile devices where large amounts of personal information are most commonly stored. Our goal is to develop a novel way of extracting and migrating privacy settings among public social networking services by overcoming existing, site-specific barriers to the process.

## Acknowledgment

This research is supported in part by the NSF Network-centric and Cloud Software and Systems Industry/University Cooperative Research Center and NSF award 1128344.

## References

- [1] "Facebook reports third quarter 2013 results. facebook," 2013.
- [2] "Facebook & your privacy: Who sees the data you share on the biggest social network?" *COMM. REP. MAG.*, 2012.
- [3] S. R. Kruk, "Foaf-realm: control your friends' access to resources," in *Proc. FOAF Workshop*, 2004.
- [4] B. Carminati, E. Ferrari, and A. Perego, "Rule-based access control for social networks," in *Proc. OTM'06*, 2006, pp. 1743–1744.
- [5] W. Villegas, B. Ali, and M. Maheswaran, "An access control scheme for protecting personal data," in *Proc. PST'08*, 2008, pp. 24–35.
- [6] A. J. T. Finin, L. Kagal, R. S. J. Niu, W. Winsborough, and B. Thuraisingham, "Rowlbac - representing role based access control in owl," in *Proc. SACMAT'08*, 2008, pp. 73–82.
- [7] N. Elahi, M. M. R. Chowdhury, and J. Noll, "Semantic access control in web based communities," in *Proc. ICCGI'08*, 2008, pp. 131–136.
- [8] B. Carminati, E. Ferrari, and A. Perego, "Enforcing access control in web-based social networks," *ACM T. INFORM. SYST. SE.*, vol. 13, no. 1, pp. 6:1–6:38, 2009.
- [9] A. Masoumzadeh and J. Joshi, "Ontology-based access control for social network systems," *INT J. INF. PRIV. SECUR. INTEGRITY*, vol. 1, no. 1, pp. 59–78, 2011.
- [10] Q. Li, J. Li, H. Wang, and A. Ginjala, "Semantics-enhanced privacy recommendation for social networking sites," in *Proc. TrustCom'11*, 2011, pp. 226–233.
- [11] M. Shehab, A. Squicciarini, G.-J. Ahn, and I. Kokkinou, "Access control for online social networks third party applications," *COMPUT. SECUR.*, vol. 31, no. 8, pp. 897–911, 2012.
- [12] A. Masoumzadeh and J. Joshi, "Privacy settings in social networking systems: What you cannot control," in *Proc. ASIA CCS'13*, 2013, pp. 149–154.
- [13] I. Kayes and A. Iamnitchi, "Out of the wild: On generating default policies in social ecosystems," in *Proc. IEEE ICC'13*, 2013.
- [14] L. Han, T. Finin, P. McNamee, A. Joshi, and Y. Yesha, "Improving word similarity by augmenting pmi with estimates of word polysemy," *IEEE T. KNOWL. DATA. EN.*, vol. 25, no. 6, pp. 1307–1322, 2013.



**SESSION**  
**POSTERS**

**Chair(s)**

**TBA**



# Steganography through Block IO

Joe Jevnik and Leonidas Deligiannidis

Wentworth Institute of Technology  
 Computer Science  
 550 Huntington Av.  
 Boston, MA, 02115, USA  
 {jevnikj | deligiannidis1}@wit.edu

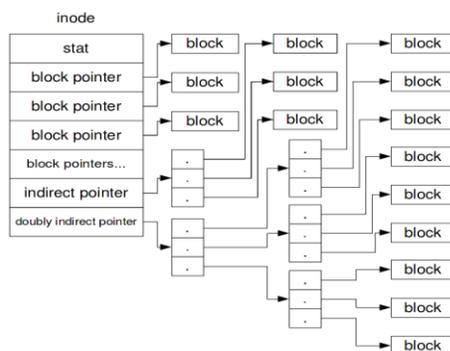
## Abstract

We present an implementation of an information concealing technique on file systems. Files are stored as bytes, but they are allocated in multiples of blocks (normally chunks of 4 Kbytes). Unused portions (bytes), if any, of the last block of a file remains allocated for it. The unused portion of the file is not counted towards its file-size. We demonstrate our implementation where we can store / hide a message in this unused portion of the last block of the file without affecting its file size. Normal reads of the file will not reveal the hidden information as according to the operating system this hidden information does not exist.

**Keywords:** Steganography, Block IO

## 1. Introduction

The practice of concealing information so that only the communicating parties know about the existence of the information is called Steganography [1]; by concealing we mean hiding the existence of a message and not the encryption of a message. Different techniques exist where one can hide information within images [2], or after the logical end of a file [3]. Certain applications simply ignore data appearing after the point that they miss determine that they reached the end of the file; the rest of the data could be loaded in memory but never presented to the user.



**Figure 1.** Diagram of an i-node.

One can hide information within an image by storing information in the least significant bit of each pixel of an

image. Visually, the altered image is indistinguishable from the original. Its message digest however, will be affected.

The system call `mmap()` [4][5] is used to map addresses of memory objects into the calling process's virtual address space; commonly these objects are files.

In Unix-styled file systems, a file is a pointer to an i-node (index node). The i-node is a structure that contains a section with the status of the file, commonly accessed through the `stat` command or through the `stat()` system call, followed by pointers to blocks that hold the file's data. There is a number of direct block pointers [6][7][8], and a single, a double and a triply indirect pointer. These pointers are used when the file does not fit in the direct pointed blocks.

The blocks are fixed size, and thus all data stored in any file will occupy some number of blocks. A file  $n$  bytes wide, with a block size of  $b$  will then occupy:

$$\text{Ceil}(n / b) \text{ blocks.}$$

With this information, we may try to use this block style IO to hide information.

## 2. Method

The method used to hide information results from the fact that although the physical space allocated for a file is a multiple of the block size, the file may contain any non-negative size up to the limit of the file system. For example, a file that contains only the single character 'a' will be 1 byte wide (0x61); however, when the file is saved, it will occupy an entire block. The concept is to write data into this extra space that is allocated for the file, but otherwise unused.

## 3. Implementation

When one creates a new file, it has no blocks associated with it. As soon as you try to write data to it, it must allocate at least one block. We start by writing any non-null strings to the file we wish to hide information in. This forces our file to acquire some space. We may calculate exactly how much space we have left to work with using the function 1.

This is because if one writes a multiple of the block size of bytes, it will not allocate a new block resulting in 0 free

space, and otherwise, we will only have the amount of unused space in the last block.

$$f: \mathbb{Z} \rightarrow \mathbb{Z}$$

$$f(n) = \begin{cases} 0, & \text{if } n \text{ is a multiple of } b \\ b - (n \bmod b), & \text{otherwise} \end{cases}$$

**Function 1.** Calculating space to hide a message.



**Figure 2.** The last block of data; grey denotes space filled with data bytes. The white part of the figure is the unused portion of the block and that is the place where we can conceal a message.

Therefore, we start by writing the visible bytes to the file. Afterwards, we must map this file to memory using `mmap()`. This now gives us a pointer that points to this file. To write the hidden data, we must only write to the locations in white as shown in figure 2. If we attempt to write to locations  $n$  where  $n$  is less than  $m$ , we will be overwriting the visible data, and if we write past the end of the block, we will experience undefined behavior, most likely a SIGBUS error, as we will be addressing random memory that is not associated with our file.

To read this hidden data, one must perform a very similar task. Simply load the file into addressable memory with `mmap()` and begin reading that array from indices that are after the length of the file.

#### 4. Discussion

The software we developed performs very and we are able to hide a message in the unused portion of the last block without changing the size of the file. One major limitation, however, is that it is difficult to distribute these files with the hidden data; in [9] the authors hide information in executables that they can distribute over the internet. The internet does not work with block IO. When we upload a file 1 byte wide, it would only upload that file's data, not the whole block. Complementary, when downloading one byte of data, we would only receive that byte and not the block it is stored in on the server. This makes transfer via the internet impossible with common transfer means. Another difficulty is that file copying is not normally done with blocks either. Using the standard GNU `cp` program to copy these files will result in a loss of the hidden data, as it will be equivalent to allocating the same amount of blocks, and then `memcpy`'ing the blocks up the length of the file. This means the only way to make a local copy is with a special copy program that would once again call `mmap()` and re-write the data that was there.

This does not affect hard links though, as hard links point to the same i-node, so they will be sharing the same blocks. This is a limitation and an advantage in the sense

that it makes it very difficult for someone to copy this hidden data unknowingly and distribute it, or to steal it. One means that could be used is to copy onto some external physical storage device, such as a USB mass storage device using the special copy program, and then using that same program to move it off the device onto another machine.

Another limitation is the amount of data that can be hidden in a single file. Because the unused portion of the last block of the file is so limited, one alternative would be to write the location of a place to find a larger message, such as a URL. Corollary, hiding a compiled binary may be difficult, but one may hide an interpreted program there. For example, one could hide a bash script that curl'ed a binary, called `chmod +x`, and then executed it. This allows for smaller amounts of data to result in more actions. We were able to fit a "hello world" program into the hidden section, but not much else, as full elf64's are commonly larger than 4095 bytes.

#### 5. Example Programs

An example implementation of this concept that includes writing data, reading data, and executing data is freely available for download at:

[https://github.com/llllllllll/information\\_hiding](https://github.com/llllllllll/information_hiding)

This example is provided under the terms of GNU public license version 2.

#### References

- [1] Kipper G. Investigator's guide to steganography. Print ISBN: 978-0-8493-2433-8 eBook ISBN: 978-0-203-50476-5. Auerbach Publications; 2004.
- [2] Ptzmann B. Information hiding terminology. First Workshop of Information Hiding Proceedings. Lecture Notes in Computer Science. 1996 May 30-Jun 1; Cambridge, UK. Springer-Verlag. 996;1174:347-350.
- [3] Leonidas Deligiannidis, Charlie Wiseman, Mira Yun, and Hamid R. Arabnia, "Security Projects for Systems and Networking Professionals". Emerging Trends in Computer Science & Applied Computing. Emerging Trends in ICT Security. Editors: Babak Akhgar & H. R. Arabnia. ISSN: 978-0-12-411474-6, Elsevier Inc. pp111-22., Nov. '13.
- [4] MMAP(3P) , The POSIX Programmer's Manual, accessed April 15, 2014
- [5] MMAP(2), The Linux Programmer 's Manual, accessed April 15, 2014
- [6] Operating Systems A Design-Oriented Approach by Charles Crowley. Publisher Irwin Book Team ISBN: 0-256-15151-2 1997
- [7] Operating Systems Design and Implementation Second Edition by Andrew S. Tanenbaum and Albert S. Woodhull. Prentice-Hall Inc. ISBN: 0-13-638677-6 1997
- [8] Operating System Concepts Essentials by A. Silberschatz, P. Galvin, and G. Gagne. John Wiley & Sons. Inc. ISBN: 978-0-470-88920-6 2011.
- [9] A.A.Zaidan, B.B.Zaidan, Hamid.A.Jalab, "A New System for Hiding Data within (Unused Area Two + Image Page) of Portable Executable File using Statistical Technique and Advance Encryption Standard", International Journal of Computer Theory and Engineering, Vol. 2, No. 2 April, 2010,1793-8201

# A formal data flow model to study network protocol encapsulation inconsistencies

François Barrère, Romain Laborde, Hicham Elkhoury, Abdelmalek Benzekri

IRIT Laboratory - Toulouse University, Route de Narbonne 118, 31062 Toulouse, France,

(barrere, laborde, benzekri}@irit.fr hichamelkhoury@gmail.com

**Abstract-** Here is presented a way to detect configuration inconsistencies that may affect a communication when multiple intermediate systems are crossed and modify data units they receive to match a protocol or a security policy.

## 1. INTRODUCTION

Setting up a communication between systems connected to a computer network often implies to cross many equipment belonging to many third parties. Each device receive either packets, frames and execute specific functions like retransmission, routing, firewalling, blocking, modifying some fields, etc.. to be compliant with a policy design by a user or an administrator. Keeping a global view of all treatments executed by these devices is inherently a distributed and complex task [4]. Vendors are providing devices and generally a specific language is provided to create a file containing multiple rules. Inconsistencies in the configurations of network equipment are frequently encountered due to misconfiguration: errors are sometime voluntary, majorly not, but may cause serious security problems. Our objective is to bring a modelling system and a tool to help users, operators (administrators) and designers (engineers) to be sure that crossing a network will not be a source of problem. Our modelling system is based on a data flow representation, aim to study the feasibility of specific configuration and make the right decision in analysing the impact of each security device. (Clearly is it possible to combine different transformations inside a single or multiple systems). For example on a single system, two rules or mechanism be may opposite (e.g. one filtering rule allows a data flow to cross an interface while another one forbid it). An end to end communication requires the use of a specific TCP port number, but datagram are encrypted and while crossing a firewall the port number cannot be recognize.

## 2. DATAFLOW REPRESENTATION

### 2.1. Data flow founding principles

In the OSI, IEEE, TCP/IP model, a frame is the result of a set of protocol encapsulation chain. Each protocol add new fields and affect values to those fields. One may analyse a frame and give all the different fields inside the frame. The data flow is composed by the set of protocol data unit that are transmitted from the source system to the destination one crossing multiple intermediate systems. Bridge, routeur, firewall... are transforming the data flow they receive into a new one depending on the crossed-system functionality and protocols that are run.

Protocol data unit (PDU) which are delivered to a system can be characterized using different views. A list of the protocol that were used to produce the PDU may be given, the list of fields accessible in the PDU, the list of fields for which the value contain an authenticated value, the list of fields that cannot be modified without leading to an error, the list of

algorithm used to encrypt, an array that link the fields that are protected by an algorithm....

### 2.2. Formal Data flow model

The basics of the formal data flow model have been introduced in[1] except  $\mathcal{L}$  set :

- $\mathcal{A}$  is the set of possible attributes. When an attribute  $a \in \mathcal{A}$ , it mean it exist a couple  $\langle \text{name}, \text{value} \rangle$  where name is a field that can be found while a protocol is executed, and value is its content. For example an ip adress is an attribute linked to IP protocol.
- $\mathcal{P}$  is the set of protocols, i.e., the set of logical blocks. Each protocol need to be identified and claim one ore more fields to be updated during its execution. Thus an instance of protocol  $p \in \mathcal{P}$  can be defined as a couple  $\langle \text{protoid}, \text{attributes} \rangle$  where a)  $\text{protoid} = \langle \text{name}, \text{id} \rangle$  is the name of the protocol and a unique identifier, b) attributes are defined on the Power-set of  $\mathcal{A}$ , i.e.,  $\text{attributes} \in \mathbb{P}(\mathcal{A})$ ,
- $\mathcal{S}$  is the set of security algorithms that can be run during the encapsulation chain of protocols (for instance, DES, 3DES, HMAC-SHA1, etc.)
- $\mathcal{L} = \{ \text{all}, \text{val} \}$  has been added into[2] in order to determine what is possible or not regarding an attribute. It give the state of an attribute that has been encrypted. If the attribute is completely encrypted (tag *all*), then it is not possible to get the attribute. When only its value is encrypted (tag *val*), the attribute can be accessed but its value cannot be retrieved except if the associated secret/keys are reachable.

Technically, an incorrect execution of a chain of protocols often implies that one or more fields contained inside a PDU are missing, wrong, suspect, impossible to retrieve, or to be replaced by a new one. Some protocol claim that payload and protocol control information must be encrypted, other one claim to discard any change of a fields because some of they are protected, encrypted and/or must not be replace or retrieve. This analysis led us to modelize a dataflow as :

$\mathcal{F} \subseteq \mathcal{E} \times \text{AUTHN} \times \text{CONF}$  where :

- $\mathcal{E}$  is the encapsulation chain of protocols. For example  $\langle \text{HTTP}, \text{TCP}, \text{IP}, \text{CSMA/CA} \rangle$ ,
- $\text{AUTHN} \subseteq (\mathcal{A} \times \mathcal{P} \times \mathcal{A} \times \mathcal{P} \times \mathcal{S})$  contains the attributes of the data flow that have been authenticated. If  $(a_1, p_1, a_2, p_2, s) \in \text{AUTHN}$  then it means that attribute  $a_1$  of protocol  $p_1$  guarantees the integrity of attribute  $a_2$  of protocol  $p_2$  via the security algorithm  $s$ . For example (FCS\_field, IEEE802.3, ipdest\_field, IP\_protocol, CRC\_32) means that frame check sequence field generated using IEEE 802.3 protocol guarantees the integrity of IP destination fill in during IP protocol execution, while using CRC\_32 algorithm implemented in IEEE802.3 solution.

- $CONF \subseteq BAG(\mathcal{A} \times \mathcal{P} \times \mathcal{S} \times \mathcal{L})$  represents the attributes of the data flow that have been encrypted, such that:  $(a, p, s, all) \in CONF$  stand for attribute  $a$  of protocol  $p$  is completely encrypted via the security algorithm  $s$ .

2.3. Data flow Operators

Operators must be provided to handle information contained inside the protocol list AUTHN, CONF sets assigned to a dataflow  $f$ . Here, they are (given into an intuitive form) :

- $proto \leftarrow Get\_Protocol(f, protoid)$
- $attribute \leftarrow Get\_Attribute(f, protoid, attName)$
- $flow \leftarrow Modify\_Attribute(f, protoid, attribute)$   
Modifies an attribute belonging to a specific protocol. Before modifying an attribute it must belong to a protocol and must be readable.
- $flow \leftarrow Add\_Proto(f, proto, protoid)$
- $flow \leftarrow Delete\_Proto(f, protoid)$
- $flow \leftarrow Add\_AUTHN(f, att1, proto1, att2, protoid2, algo)$
- $flow \leftarrow Delete\_AUTHN(f, att1, proto1, att2, protoid2, algo)$
- $flow \leftarrow Add\_CONF(f, attribute, protoid, algo, level)$
- $flow \leftarrow Delete\_CONF(f, attribute, protoid, algo, level)$

3. CONFLICT DETECTION USING PETRI NETS SAMPL

Colored Petri Nets [3] are a formal specification language consisting of a set of tokens whose type is represented by a color, a set of transitions, and a set of places with a domain (which defines the types of tokens that can be stored in that place). They are well-known for their graphical and analytical capabilities for the specification and verification of concurrent, asynchronous, distributed, parallel and nondeterministic systems.

IPsec [5] and NAT are well known protocols that are not applicable one following the other [6]. Using the dataflow specification, we can detect conflicts while studying the impact of the transformation chain  $tf_{NAPT} \circ tf_{AH}^{tunnel}(f)$  on the given data flow  $f = (< ip_1, tcp_1 >, \{\}, \{\})$ .

The IPsec/AH (Authentication Header) is designed to ensure integrity and authenticity of IP datagrams without data encryption. The integrity is guaranteed by the Authentication Data (AD) field. The AH protocol has two modes: transport and tunnel. In transport mode, AH is inserted between the IP header and the next layer. It protects the entire IP packet except for the mutable fields (i.e. the fields DSCP, ECN, Flags, Offset, TTL, Header Checksum). In tunnel mode, the inner IP header carries the ultimate IP source and destination addresses, while an outer IP header contains the addresses of the IPsec peers. It protects the entire inner IP packet. In AH Tunnel mode, the entire original IP header and data become the “payload” for the new packet protected by AD field. Figure 1 present a sample of a Petri that is built. Data flow is transformed into

$f' = f_{AH}^{tunnel}(f)(< ip_2, ah, ip_1, tcp_1 >, AUTHN, \{\})$  First, as the dataflow contain a classical IP packet  $ip_1$ , before its

transformation, then the attributes can be retrieved, and tunnel mode can operate:  $action_{AH\_Tunnel}$  is implemented using a function  $ApplyIPsec(f, proto-ipsec, mode, gw, algo)$  with the parameters: (1) the data flow  $f$ , (2) the protocol, (3) the mode, (4) the gateway and (5) the cryptographic algorithm hmac-md5.

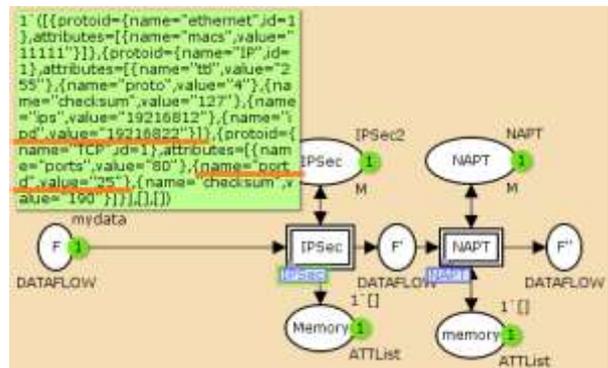


FIGURE 1

Next, the NAPT mechanism attempt to perform  $Action_{NAPT}$ , which is an expression calling  $Modify\_Attribute()$  in order to change the  $dest\_add$ , the  $dest\_port$  and the checksum. But when the NAPT operates, the token representing data flow  $f'$  will stay blocked inside it. Two different explanation explain this situation: a) Basic NAPT refers to a situation where the payload field can't be examined at all. The flow  $f'$  can't be transformed because the protocol encapsulated directly after IP1 is AH and that protocol does not contain any attribute called ports b) in Advanced NAPT configuration the former IP1, TCP1 attributes are considered to be accessible. Port value may be retrieved using  $Get\_attribute$ . Consequently they could also be modified but since the attribute “port” belongs to and is protected by the set AUTHN, the modify attribute could not be executed:  $f''$  will not be produced and the error will then be detected.

4. CONCLUSION

Network security requires the coordination of various heterogeneous and interdependent devices but conflicts may occur while these devices are crossed. Here has been presented the basics of a formal approach using a data flow oriented-framework.. If only one example has been given here, it was applied to other well know protocols. Conflicts have been detected without requiring any a priori knowledge or experience. An other ongoing work is to design a generic model for network security mechanisms.

5. REFERENCES

- [1] El Khoury & al “A Generic data flow model Safeconfig 2011 October 2011 Washington DC
- [2] ‘A Formal Data Flow-Oriented Model For Distributed Network Security Conflicts Detection’ in ICNS 2012: St. Maarten,
- [3] Ratzer and al “CPN Tools for Editing, Simulating, and Analysing Colour Petri net
- [4] Al-Shaer, E. and Hamed, H. (2004), ‘Discovery of Policy Anomalies in Distributed Firewall’ in *INFOCOM 2004*
- [5] RFC 4301 Security Architecture for the Internet Protocol
- [6] Adoba, Dixon “IPsec NAT compatibility requirements 2004

## **SESSION**

# **II: SPECIAL TRACK ON IOT AND SCADA CYBERSECURITY EDUCATION**

## **Chair(s)**

**Prof. George Markowsky**



# Teaching Cybersecurity to Wide Audiences with Table-Top Games

Tadhg Fendt

Department of Mathematical Sciences  
Lewis & Clark College  
tfendt@lclark.edu

Jens Mache

Department of Mathematical Sciences  
Lewis & Clark College  
jmache@lclark.edu

## Abstract

Cybersecurity is a field of growing importance. A particular challenge is that there is an ever-growing base of technology that needs securing, coupled with a shortage of security specialists. This creates an important role for security education. Security education is considered difficult, especially with non-technical students, because the field is so broad. Table-top gaming has been suggested as an educational starting point to make a wide audience aware of the issues and to foster curiosity and enthusiasm for the field. In this paper we examine two such games, *Control-Alt Hack* and *[d0x3d!]*, compare their strengths and weaknesses and feasibility in the undergraduate classroom. In conclusion, *[d0x3d!]* seems preferable for use in the classroom.

## 1 Introduction

Given the challenges that accompany security education, we believe that more tools and activities are needed for instructors to effectively teach it. Table-top games provide a learning experience that is appropriately non-technical as a starting point and for students without much computer science background, yet still hands-on and thought provoking. In this paper, we examine two games, *Control-Alt Hack*, and *[d0x3d!]*. We review the basic mechanics and logistics of both games as well as the security concepts and the methods used to introduce them. Particular differences of interest to us are: the use of reading and language in the games,

competitive vs. cooperative game-play, and the use of dynamic modeling. We discuss these differences, strengths and weaknesses of the games as well as their feasibility in an undergraduate class.

### 1.1 Motivation

This paper is directly motivated by a growing need for conceptually rich, non-technical resources for students without much computer science background. During the 2014-2015 school year, Lewis & Clark College will be offering an interdisciplinary perspectives in cybersecurity class in collaboration with the International Affairs department. As time in the classroom is at a premium, we want to be sure that if we decided to use one or both of these games that we could do so effectively and efficiently.

## 2 Background

*Control-Alt Hack* and *[d0x3d!]* have the same basic driving principle behind them: exposing “non-experts” to concepts in security with the aim of increasing awareness [1,2]. The designers readily admit that the games do not provide in-depth or technical instruction of security. However, they make the convincing argument that this is not necessary for the stated goals of outreach and exposure [1,2].

In *Control-Alt Hack* [1] each player becomes a white-hat hacker in a security consulting firm. Players have character cards that give them a

certain set of skills to help them complete security audits and other missions. Players compete to gain the most “hacker cred” and the most successful hacker eventually becomes CEO of their own security firm.

In *[d0x3d!]* [2], the players work as a team to recover personal data that has been stolen and hidden on a computer network. The players take on different roles (i.e. wardriver, cryptanalyst, etc.) that give them special abilities to complete the mission. The team must infiltrate the network and recover the stolen data, all while the administrators patch, decommission, and possibly detect intrusions.. Both games are turn-based, card-driven games though *Control-Alt Hack* additionally uses dice rolls to resolve mission attempts.

### 3 Reading and Language

The vernacular of a game being used for educational purposes is a very important consideration for that game’s effectiveness. This may seem rather strange at first because after all, you don’t read games, you play them. But it turns out that it actually depends to a large degree on the type of game, which brings us to the first big difference between the two: *Control-Alt Hack* is much more text dependent than *[d0x3d!]*. The driving game mechanism in *Control-Alt Hack* is the mission card. The cards have a title, a description of the overall task, and lastly several sub components to the mission that are specific to one or more of the “hacker skills.” In short, there is a lot of writing and it is used as the main way in which information is conveyed to the players. On the other hand in *[d0x3d!]* the cards make much better use of pictures and any writing is usually one or two words.

In a game that uses text as the main conduit for information, the clarity and efficiency of the words becomes even more important.

Unfortunately, in addition to being more text-heavy overall, we feel that *Control-Alt Hack’s* use of language is less effective in communicating security concepts for three reasons.

Firstly, it seems that *Control-Alt Hack* is attempting to get as much information into the game as possible. In one sense, this is good because it shows just how broad the field is and also breaks down the stereotype of security people as always feverishly typing on the command line. The downside, however, is that too much information can overwhelm students and not really stick with them. If you have one hour to play a game in class, less information can mean more focus.

Secondly, the vocabulary itself is sometimes quite vague. Mission cards address topics such as: wireless connection protocols, weaponized exploits, and software vulnerabilities. For teaching non-computer science students, it seems that these words are less effective than the more specific ones found in *d0x3d!*: honeypot, integer overflow, logic bomb, etc. The latter group of short, specific terms can be easily looked up and researched independently and later incorporated into class.

Finally, we also find several of the missions and much of the text to be superfluous and of questionable relevance to any computer security curriculum. In many cases, it seems these are included for humor which is certainly not a detraction in games generally. However, many of the jokes and comedic situations are only funny to those with knowledge of computer science or the tech industry and thus of little value to many students.

However, we think that *Control-Alt Hack* was successful with its use of language in the implementation of the “hacker skills” system. In

the game, almost all mission are resolved under one or more of the five abilities the designers consider to be essential: hardware hacking, software wizardry, network ninja, social engineering, and cryptanalysis. While these terms are still somewhat vague, they subtly and effectively inject a very important question into the entire game: what is cybersecurity? It is actually a fairly hard question to answer. Security is a broad and multifaceted topic and these categories get that idea across.

This categorization also struck us as a good mental exercise for instructors in relation to curriculum design and for time allocation. Assuming that we agree with the rough categories, which of them should we be focusing on in our teaching efforts? In our security class for computer science students this past year, we focused roughly 40% each on software issues and network skills while the remaining time split between social engineering, crypto and hardware. Whether this is an optimal mix is certainly a pending question. There are many factors that contribute to what the syllabus will ultimately look like for a security course, including available tools, infrastructure and resources but in any case this provides interesting food for thought for those endeavoring to teach security.

#### **4 Competitive vs. Cooperative Game-play**

Another difference worthy of note is the nature of play in both games. *Control-Alt Hack* is a competitive game with players vying for the top CEO spot, whereas [*d0x3d!*] has all players cooperatively trying to recover the stolen data. Does a game being either competitive or cooperative make a difference for its educational outcomes? There are reasonable arguments on both sides. Competitive games are often seen as being more fun because there is the possibility of being a unique winner. Games that are more fun might be more readily played by students.

On the other hand, some research has shown that women, relative to men, are less likely to want to play a competitive game [2]. This suggests that cooperative games may be more inclusive and can even help combat the severe gender gap in computer security. Further studies have shown cooperative games result in higher levels of interaction between players [3]. This could potentially lead to greater inter-player discussion and analysis as students review their play and adjust strategy together. On balance, we give the nod to the cooperative game because while it may not maximize fun, it certainly does not preclude it, and many other benefits can be conferred in an educational setting.

#### **5 Games as Models**

In general, games are usually trying to model something and the better the model, the better the game. Models are also a very good way of teaching. This is because they allow for the abstraction of complex systems so they can be examined and conceptually understood without the overhead and information overload. *Control-Alt Hack* and *d0x3d!* are no exception to this rule, and both attempt to model a different aspect of security with varying success.

*D0x3d!* takes the approach of actually modeling a computer network on which stolen data is hidden. The various pieces of infrastructure that make up the network are represented by tiles that the players can compromise and move through. This is also a dynamic model because of the actions of the system administrators, which are built into the game with pseudo-random card draws. These “patch” card draws can lead to the securing or decommission of compromised network infrastructure, constraining the players by changing their environment. In addition to being dynamic during the course of a single game, *d0x3d!* also allows for changes to be made and thus encourages experimentation and playing the

game many times. At the beginning of the game, the players are free to “configure” the network topology however they want. This allows players to incorporate their own ideas and learning into the game to make play more interesting or challenging. Most importantly, with a few simple rules this model exposes players to the concept of navigating computer networks -- a real-world task that is a significant part of security from our experience with competitions such as CCDC [4] -- all without having to master details of secure shell, protocols or port numbers. Students could play the game first and then actually attempt some of the network traversal they were doing on a local lab or in the cloud.

*Control-Alt Hack* alternatively models the much more general concept of working as a security professional. Players are given a character with various skills that can be improved over time. The characters carry out what can be described as contracts to elevate their career until such a point that they can win the game by being the top hacker. This model too is dynamic because the players can interact in the game and take actions that affect one another. We feel that this model is less successful because the system it tries to emulate is complex and inexact relative to a simple computer network. It would also be hard to try to make direct links from actions in the game to activities students could actually attempt. A mission card that has the player complete a security audit would be hard to relate to for a non-expert.

## 6 Classroom Feasibility

When evaluating something like a game for use in class, it is important to consider certain logistical aspects of implementation. From experience, it may be unrealistic to have all the students play the game outside of class. This means playing the game in-class, heightening

the need for efficiency and ease of use. While both games have supporting websites that offer suggestions to educators planning on using the game, we find that [d0x3d!] has two additional aspects that make a difference. First, the [d0xed!] website includes videos that concisely and effectively explain the rules of the game. While it may be unreasonable to expect students to play the game outside of class, given them a ten-minute online video to watch before coming to class is pretty low cost to even the least enthused students. Second, [d0x3d!] is open source and everything needed to play the game can be retrieved online and printed out for free.

## 7 Conclusions

In this paper we have reviewed the strengths and weaknesses of two table-top games, *Control-Alt Hack* and [d0x3d!], as well as their viability as teaching tools in the undergraduate classroom. We discussed in particular the role of language, competitive vs. cooperative game-play and the role of dynamic modeling. We conclude that [d0x3d!] is preferable as an educational tool to be used in-class and we will be attempting to use it in a perspectives course that includes non-computer science students during the 2014-2015 school year. Future work includes observations and results from the use of the table-top games in undergraduate courses.

## 8 Acknowledgements

This work is partially supported by NSF grant DUE-1141314, by the John S. Rogers Science Research Program of Lewis & Clark College, and by the James F. and Marion L. Miller Foundation. We would like to thank Richard Weiss for useful discussions. We further thank Alicia Kirkland, Miles Crabhill, Jon Poley, James Josephson, Christian Dicker, Sam Kelly and John Sibandze for their feedback and play-testing.

## 9. References

- [1] Tamara Denning, Tadayoshi Kohno, and Adam Shostack, Control-Alt-Hack: A Card Game for Computer Security Outreach, Education, and Fun, Technical Report UW-CSE-12-07-01, University of Washington, 2012.
- [2] Mark Gondree and Zachary N.J. Peterson, Valuing Security by Getting [d0x3d!]: Experiences with a Network Security Board Game, Proceedings of the 6th Workshop on Cyber Security Experimentation and Test (CSET), 2013.
- [3] Zagal, J. P., Rick, J., and Hsi, I. Collaborative games: Lessons learned from board games. *Simulation and Gaming* 37, 1 (March, 2006).
- [4] Cyber Security Defense Competition, <http://www.nationalccdc.org/>, accessed 6/10/14

# A Behavior-Based Covert Channel in a MMO

Brian Rowe and Daryl Johnson  
 Department of Computing Security  
 Rochester Institute of Technology  
 Rochester, New York USA  
 bxr9458@rit.edu, daryl.johnson@rit.edu

**Abstract**—This article proposes a behavior-based covert communication channel in World of Warcraft (WoW), a popular Massively Multiplayer Online Role Playing Game (MMORPG), to transfer data between 2 clients. This is done by modulating binary information onto the senders cast routine and capturing this routine on the receiver. This routine, while demonstrated in WoW, could also be applicable to any other MMORPG in existence. WoW has the advantage of a large user base, so common actions will draw very little attention.

## I. INTRODUCTION

Lampson first described covert channels in *A Note on the Confinement Problem* in 1973 [6]. He noted the difficulty of confining a program, so that it cannot communicate with any program besides its caller. Whether by shared resources, storage, or piggybacking on legitimate information, it is extremely difficult to prevent all inter-process communication. He called these illegitimate channels covert channels. Since then, these covert channels have been broken into two main groups with a proposed third: storage, timing, and behavioral. A storage channel is one in which a sending process communicates by directly or indirectly writing to a storage location where the receiving program can directly or indirectly read it, whereas a timing channel sends information by modulating system resource utilization so that the receiving program can observe this and derive information [5]. A third type, a behavioral channel, is one in which behavior patterns are modified to communicate a message between parties [4]. This last channel is more difficult to detect than a storage or timing one because an individual would have to know and understand a units normal behavior to be able to detect the irregularity, or modulation, resulting in a plethora of potential arenas in which to secretly and inconspicuously leak information. This article explores using an MMORPG as just such a carrier.

The idea of using a game as a covert channel is not a new one. In 2008, Zander, Armitage, and Branch proposed the idea of covert channels in first person shooters [9]. They described communication by encoding pitch, yaw, x, y, and z coordinates of a character to send encoded messages to a receiver which would then have to record and decode this information. An even simpler example, a state game called Magnetron created in 2009, allowed players to send, save, and receive messages by passing a predetermined authentication pattern into the game [4]. This subsequently allowed the following moves to either send or receive a message between other clients. Later in this article, the authors brought up the concept of

covert channels in an MMORPG, mentioning that it would be difficult to identify the clients and even harder to then examine them [4]. An MMORPG is a game environment in which large numbers of people can interact both with each other and with the game environment. These worlds create a different type of carrier to transfer information by using seemingly normal actions. The most popular MMORPG at this time is World of Warcraft. In an official press release from Activision Blizzards Earnings Call (a quarterly report detailing company finances) for the period ending December 31, 2012, this game had approximately 9.6 million paying subscriptions [1]. According to their website, WoW “is an online game where players from around the world assume the roles of heroic fantasy characters and explore a virtual world full of mystery, magic, and endless adventure [2].” This world containing millions of human players creates many normal environmental interactions that could be modulated to transfer information between clients. One example is explained below.

## II. METHOD

World of Warcraft has many different ways of interacting with the environment that could be modulated to covertly relay information between a sender and receiver. This paper focuses on using the combat system and a targeting/training dummy as the communication channel within the legitimate channel of the game itself.

World of Warcraft has training dummies in every major in-game city (see figure 1). A training dummy allows a player to perform spells or melee attacks (depending on the class of the character as chosen by the user) against this opponent to test attack sequences or rotation for Player vs. Environment (PVE) encounters (also known as player versus computer). Within the game there are many kinds of spells that can be cast. Two of these, the Frostbolt and Ice Lance spells, have been chosen to represent the binary digits 0 and 1 to encode messages. They were chosen because they are “no cool down” spells which means they can be used repeatedly without a waiting period between the casts. This is the medium that will be modulated for the covert channel.

First, a Perl script called conv.pl, written by Ivo on cool-commands.com [3], is used to convert the input message into binary code. Then, AutoHotKey [7], a free open source automation, hotkey, and scripting language, is used to compile .ahk script files into .exe executable files. The script 1ahk.ahk (see figure 2) modulates a binary 0 to the press of the 1 key,



Fig. 1. World of Warcraft Training Dummy

the script 2ahk.ahk (see figure 3) modulates a binary 1 to the press of the 2 key. The WoW.ahk script (see figure 4) activates a running WoW process bringing its window to the forefront. Then a third spell, the 3 key, is used to modulate the beginning and the end of the sequence.

```
#NoEnv
; Recommended for performance and compat
; with future AutoHotkey releases.
; #Warn
; Enable warnings to assist with detecting
; common errors.
SendMode Input
; Recommended for new scripts due to its
; superior speed and reliability.
; Ensures a consistent starting directory.
SetWorkingDir %A_ScriptDir%
Send 1
```

Fig. 2. 1ahk.ahk AutoHotKey script to modulate a 0

```
#NoEnv
; Recommended for performance and compat
; with future AutoHotkey releases.
; #Warn
; Enable warnings to assist with detecting
; common errors.
SendMode Input
; Recommended for new scripts due to its
; superior speed and reliability.
; Ensures a consistent starting directory.
SetWorkingDir %A_ScriptDir%

Send 2
```

Fig. 3. 2ahk.ahk AutoHotKey script to modulate a 1

To begin the sender and receiver must be logged into their respective characters in the World of Warcraft game. The receiver begins by issuing the command /combatlog in the chat box to enable WoW's built-in combat log to begin recording the actions and events in the environment around him/her to Program Files\World of Warcraft\Logs\WoWCombatLog.txt (see figure 5). The environment that is recorded by the combat

```
#NoEnv
; Recommended for performance and compat
; with future AutoHotkey releases.
; #Warn
; Enable warnings to assist with detecting
; common errors.
SendMode Input
; Recommended for new scripts due to its
; superior speed and reliability.
; Ensures a consistent starting directory.
SetWorkingDir %A_ScriptDir%

IfWinExist World of Warcraft
{
    WinActivate
}
```

Fig. 4. WoW.ahk AutoHotKey script to activate a WoW process

log is limited to a 200 yard range surrounding the player. This is a hard coded limit and cannot be changed by the player. The receiver needs to be within the 200 yard proximity - at the same time and on the same server - as the sender for the log to capture the senders actions. The sender then invokes the sender.pl script (see figure 6) which uses the conv.pl script to convert the ASCII message into a binary string. Sender.pl then calls upon 1.ahk.exe and/or 2.ahk.exe to send the binary 0's and 1's in the converted string by generating the corresponding key presses. The receiver logs out of the game when the animation ceases indicating that the message is complete. The combat log file is save locally. The receiver then executes the receiver.pl script (see figure 7) against the combat log file. This converts the Frostbolt spell back into a binary 0 and the Ice Lance spell back into a binary 1. The receiver.pl then call upon conv.pl to convert the binary sequence back into an ASCII string.

### III. LIMITATIONS

Limitations still exist in this system and the biggest in this example is bandwidth. A general formula for throughput is (60 seconds) divided by (cast time) equals casts per minute (bits). Depending on whether a 0 (in-game Frostbolt) or 1 (in-game Ice Lance) is sent for this demonstration, the cast

```

3/25 09:22:08.444 SPELL_CAST_START,0x018000003AC841B,"wowName",0x511,0x0,0x0000000000000000,nil,0x80000000,0x80000000,116,"Frostbolt",0x10,0x0180
3/25 09:22:10.364 SPELL_CAST_SUCCESS,0x018000003AC841B,"wowName",0x511,0x0,0xF130B637000011C9,"Training Dummy",0x10a28,0x0,116,"Frostbolt",0x10,0
3/25 09:22:10.654 SPELL_AURA_APPLIED,0x018000003AC841B,"wowName",0x511,0x0,0xF130B637000011C9,"Training Dummy",0x10a28,0x0,116,"Frostbolt",0x10,0
3/25 09:22:10.654 SPELL_DAMAGE,0x018000003AC841B,"wowName",0x511,0x0,0xF130B637000011C9,"Training Dummy",0x10a28,0x0,116,"Frostbolt",0x10,0xF130B
3/25 09:22:10.814 SPELL_CAST_START,0x018000003AC841B,"wowName",0x511,0x0,0x0000000000000000,nil,0x80000000,0x80000000,116,"Frostbolt",0x10,0x0180
3/25 09:22:12.654 SPELL_CAST_SUCCESS,0x018000003AC841B,"wowName",0x511,0x0,0xF130B637000011C9,"Training Dummy",0x10a28,0x0,116,"Frostbolt",0x10,0
3/25 09:22:13.094 SPELL_AURA_APPLIED_DOSE,0x018000003AC841B,"wowName",0x511,0x0,0xF130B637000011C9,"Training Dummy",0x10a28,0x0,116,"Frostbolt",0
3/25 09:22:13.094 SPELL_DAMAGE,0x018000003AC841B,"wowName",0x511,0x0,0xF130B637000011C9,"Training Dummy",0x10a28,0x0,116,"Frostbolt",0x10,0xF130B
3/25 09:22:13.454 SPELL_CAST_START,0x018000003AC841B,"wowName",0x511,0x0,0x0000000000000000,nil,0x80000000,0x80000000,44614,"Frostfire Bolt",0x14
3/25 09:22:16.054 SPELL_CAST_SUCCESS,0x018000003AC841B,"wowName",0x511,0x0,0xF130B637000011C9,"Training Dummy",0x10a28,0x0,44614,"Frostfire Bolt"
3/25 09:22:16.484 SPELL_DAMAGE,0x018000003AC841B,"wowName",0x511,0x0,0xF130B637000011C9,"Training Dummy",0x10a28,0x0,44614,"Frostfire Bolt",0x14
3/25 09:22:17.024 SPELL_CAST_START,0x018000003AC841B,"wowName",0x511,0x0,0x0000000000000000,nil,0x80000000,0x80000000,116,"Frostbolt",0x10,0x0180
3/25 09:22:19.164 SPELL_AURA_APPLIED,0x018000003AC841B,"wowName",0x511,0x0,0x018000003AC841B,"wowName",0x511,0x0,44544,"Fingers of Frost",0x10,0
3/25 09:22:19.164 SPELL_CAST_SUCCESS,0x018000003AC841B,"wowName",0x511,0x0,0xF130B637000011C9,"Training Dummy",0x10a28,0x0,116,"Frostbolt",0x10,0
3/25 09:22:19.304 SPELL_AURA_APPLIED_DOSE,0x018000003AC841B,"wowName",0x511,0x0,0xF130B637000011C9,"Training Dummy",0x10a28,0x0,116,"Frostbolt",0
3/25 09:22:19.304 SPELL_DAMAGE,0x018000003AC841B,"wowName",0x511,0x0,0xF130B637000011C9,"Training Dummy",0x10a28,0x0,116,"Frostbolt",0x10,0xF130B

```

Fig. 5. World of Warcraft Combat Log

```

#!/usr/bin/perl

#getting input from user
#converting into binary string
print "Enter phrase to be transmitted\n";

my $ui = <STDIN>;
chomp($ui);
my $string = `perl conv.pl -b \"$ui`";
print "The binary string is " . $string;
chomp($string);
@array = split(/,/, $string);

#call wow.exe which brings WoW to the foreground
system("WoW.exe");

#foreach in array, if the binary translation is a 0 press 1
# and if a 1 press 2 waiting for in-game cooldown appropriately
foreach (@array)
{
    if($_ == 0) {
        print "A 0\n";
        system("1ahk.exe");
        select(undef,undef,undef,1.9);
    }
    else {
        print "A 1\n";
        system("2ahk.exe");
        select(undef,undef,undef,1.5);}
}

```

Fig. 6. sender.pl script

time for 1 bit of information is either 1.9 or 1.5 seconds, respectively. This limits the throughput to a best case of 40 bits per minute (60 seconds per minute divided by 1.5 seconds per cast), a worst case scenario of 31 bits per minute (60 seconds per minute divided by 1.9 seconds per cast), with a median of 35 bits per minute. The National Institute of Standards and Technology [5] labels this as a low bandwidth channel because it transmits less than 100 bits per second.

There are three classes of covert channel bandwidths: high, medium, and low. While there are no commonly agreed upon numbers for these ranges, their uses are clear. Document and image exfil/infiltration typically will require a high bandwidth channel. If the channel can be used over an extended period of time, a medium bandwidth channel would suffice. Messaging and botnet command and control work best with at least a medium bandwidth channel. Low bandwidth channels can be used for signaling if prolonged channel usage is possible. If the communication is highly encoded that would reduce

the bandwidth required. For example, using a dictionary of 255 words would only use 8 bits per word. Low bandwidth channels modify the environment the least and consequently offer the potential for the greatest stealth.

While this channel is considered low bandwidth, it is still not insignificant as coordinates and encryption keys are still only a few bits to a few hundred bits in length. Logically, sensitive information could still be leaked very easily through this channel as important information can be worth the wait. This channel would also be a candidate for a botnet command and control channel.

A second limitation is that both clients have to pre-negotiate spell translations, character names, and a training dummy meet-up location to facilitate the communication. A scheduled meet-up time is also beneficial to prevent drawing unwanted attention, even though it isn't necessary for communication as the receiver can capture logs indefinitely until he runs out of local storage space.

```
#!/usr/bin/perl

#open input wow combat log
$binstring='';
open ( INPUT, "<C:\\Program Files (x86)\\World of Warcraft\\Logs\\WoWCombatLog.txt" )
    or die "File does not exist!\n";
print "The printed text is:\n";

#change to match your sending character's name
$character="SecretSquirrel";

#foreach line of input from the combat log, if regex matches output the resulting 0 or 1
#then once the string size is 8, convert back to ascii

foreach (<INPUT>)
{
    chomp($_);

    if ($_ =~ m/SPELL_CAST_SUCCESS/ && $_ =~ m/$character/)
    {
        if ($_ =~ m/Frostbolt/) { $binstring .= '0'; }
        elsif ($_ =~ m/Ice/) { $binstring .= '1'; }
    }

    if(length($binstring) == 8)
    {
        $ascstring = `perl conv.pl -t \"$binstring\"`; chomp($ascstring);
        $binstring=''; print $ascstring;
    }
}
print "\n";
```

Fig. 7. receiver.pl script

A final limitation is that both clients have to have WoW. Fortunately, WoW now allows players free play up until level 20 with no credit card necessary. It only requires basic information and an email address that can be made up readily, so this isn't much of a drawback.

#### IV. DETECTION AND PREVENTION

As stated by multiple papers above [4], [9], it is extremely difficult to detect behavior-based covert channels. Monitoring the network traffic or the players actions via an automated system would not raise any flags as the actions are seemingly normal. Every action being carried out by sender and receiver is common in the game and would not draw any unwanted attention. There are two additional factors that can make this channel even more inconspicuous: the server population and the time of day. On a high population WoW server (having a large number of human players) at night or on the weekends, there are almost always players attacking these dummies. Monitoring a few high population servers at night (between 6pm and midnight server time) and on the weekends, there was on average less than one minute per hour during which it did not have at least one player attacking it. Therefore, these actions are extremely common and would go unnoticed.

To detect this communication channel the casting on every targeting dummy across all realms would have to be monitored. First, a baseline of the normal usage would have to be taken. Then, a person (or program) that understands casting

rotations and mechanics would have to monitor the cast cycle and decipher the sender's casts from all other combat events occurring simultaneously. During this examination the monitor would have to pick up on the irregularities of the casting routine, realizing that it is not normal and is potentially a modulated signal, and have to begin monitoring the sequence to figure out the encoding scheme and finally decode the transmission. This would be not only very difficult, but borderline impossible to actually implement as all preceding factors would have to fall perfectly into place.

Even if all of the above events were to happen, detection of the receiver would be even more difficult than the sender as the only requirement for the actions to be logged is that the sender and receiver be on the same server and within 200 yards. The player does not have to target, or even be within line of sight of the caster for these events to be logged. The only feasible way to track down a receiver would be to first identify the sender, then monitor all players within 200 yards and hope for a repetition of proximity. Like a One-Time Pass, this identification could be easily thrown off by simply creating new free accounts every time the individuals want to send information. They could also create a new character on a different server, thereby not making a repetitive pattern that could be followed and traced. Simply stated, if tracking down the sender is already almost impossible then tracking the receiver would be unthinkable.

## V. IMPROVEMENTS

There are several improvements that could be made to improve upon this channel. For starters, the communication could be encrypted via a shared key between the sender and receiver which would provide an added layer of security in the event that transmission was properly intercepted.

A second improvement would be to find a class that has more than three no cool-down casts (not requiring a mandated wait time between casting the same spell again), like the Mage used in this example, to modulate more than 1 bit per cast. If a class could be found with 4, 6, or 8 moves, it could modulate 2, 3, or 4 bits per cast, respectively. This would drastically increase the bandwidth while still keeping the channel discrete.

Modulation of binary files rather than only ASCII characters onto the cast stream would be yet another enhancement. Then the only bandwidth limitation would be the server restart that happens once a week on Tuesday, inherently interrupting communication. In theory, 35 (bits per minute) times 10080 (minutes in a week), or 352,800 bits could be sent between server restarts.

Another idea, proposed by Benjamin Wollak [8], had a programming implementation that would allow players to communicate via emotes, a sound effect or movement used to express emotion, which are available in almost every MMORPG and have no cast time, making them almost instantaneous. Unfortunately, the game does not log these events and this idea was never completed as it would require programming a recording interface to interpret these visual events on the receiver.

Finally, fault tolerance and a checksum could also be introduced to verify the integrity of the transmission. This would provide the receiver with a guaranteed correct message rather than just a one-time chance. (WoW does perform network-based fault tolerance so this should not be a problem.)

## VI. CONCLUSION

As the number of devices in our homes, cars, and on our persons grows, the challenge of securing these devices also grows. Firewalls are often employed to control and filter traffic protecting them from malicious activities. Covert channels like the one discussed in this paper can thwart typical measures employed for security. For example, this covert channel could be used by malware installed on a system inside a firewalled network to provide a command and control channel for the malware. It could also enable network scanning and reconnaissance behind the firewall. Both of these activities could be accomplished with a low bandwidth covert channel and would be invisible to typical protection mechanisms.

This covert channel illustrates both the simplicity of creating and the difficulty in detecting a behavior-based covert channel. To detect these channels, a fundamental idea of the normal behavior of the events would have to be understood and a baseline created. Only then might one be able to detect potentially modulated behavior sequences. The implications for covert communication are not restricted to this game.

The major attraction in an MMORPG is the ever evolving human and environmental interaction that will continue to

create new items and sequences that can be modulated with information. Nobody knows what the next big game will be, but one can bet that people will look for new ways to use it as a medium for covert channels.

## REFERENCES

- [1] Inc Activision Blizzard, *Activision blizzard announces better-than-expected fourth quarter and calendar year 2012 results*, "http://files.shareholder.com/downloads/ACTI/2310472046x0x634081/fbb6a75d-f965-442b-8d6a-bde335918118/Q4\_2012\_atvi\_press\_release.pdf".
- [2] Inc Blizzard Entertainment, *World of warcraft game guide*, <http://us.battle.net/wow/en/game/guide/>.
- [3] Inc. CoolCommands, *Coolcommands conv.pl script*, "http://coolcommands.com/".
- [4] Bo Yuan Daryl Johnson and Peter Lutz, *Behavior-based covert channel in cyberspace*, 2009.
- [5] *Nist dod trusted computer system evaluation criteria*, <http://csrc.nist.gov/CSC-STD-001-83>, dtd 15 Aug 83.
- [6] Butler W. Lampson, *A note on the confinement problem*, *Commun. ACM* **16** (1973), no. 10, 613–615.
- [7] Chris Mallet, *Autohotkey*, <http://www.autohotkey.com/>.
- [8] Benjamin Wollak, *Behavior based covert channel in mmorpgs*, Unpublished.
- [9] S. Zander, G. Armitage, and P. Branch, *Covert channels in multiplayer first person shooter online games*, *Local Computer Networks*, 2008. LCN 2008. 33rd IEEE Conference on, 2008, pp. 215–222.

# Building a Virtual Cybersecurity Collaborative Learning Laboratory (VCCLL)

A. Julien Murphy<sup>1</sup>, B. Edward Sihler<sup>2</sup>, C. Maureen Ebben<sup>3</sup>, D. Lynn Lovewell<sup>4</sup> and E. Glenn Wilson<sup>5</sup>

<sup>1</sup>Philosophy, University of Southern Maine, Portland, Maine, USA

<sup>2</sup>MCSC, University of Southern Maine, Portland, Maine, USA

<sup>3</sup>Communication and Media Studies, University of Southern Maine, Portland, Maine USA

<sup>4</sup>MCSC, University of Southern Maine, Portland, Maine, USA

<sup>5</sup>Technology, University of Southern Maine, Gorham, Maine, USA

**Abstract** - In fall 2013, the Maine Cybersecurity Cluster (MCSC), was invited to assist the United States Coast Guard with cybersecurity training. MCSC conducted training activities that created the conditions under which Coast Guard personnel could experience and respond to cyber attacks first-hand. A major result of this endeavor was the recognition of two critical needs: 1) the necessity for a flexible, learning laboratory to address the increased security requirements presented by the Internet of Things (IoT), and 2) the need for applied education and training for students going into information assurance professions. To fill these gaps, MCSC designed plans for the creation of a Virtual Cybersecurity Collaborative Learning Lab (VCCLL). The lab would operate inter-institutionally and offer innovative, hands-on, collaborative learning experiences aimed at preventing and mitigating cyber attacks in real time. This paper delineates the background, design, and benefits of the VCCLL.

## Keywords

1. Cybersecurity, 2. Virtualization, 3. Education, 4. Training, 5. Laboratory, 6. Collaboration

## 1 Introduction

The chief objectives of Maine Cybersecurity Cluster (MCSC) are twofold: 1) to address network vulnerabilities across a spectrum of technologies in public and private sector organizations, and 2) to develop student education and skills around information assurance for workforce development. Central to cyber security education is the skill to detect network vulnerabilities. This skill is best acquired in an applied, dynamic, virtual laboratory, one that allows for students to uncover, understand, and resolve a variety of documented cyber security exploits in a practical manner [1]. The MCSC Virtual Cybersecurity Collaborative Learning Laboratory (VCCLL) will offer

the opportunity for students to work in a collaborative culture and to engage in solving challenging cybersecurity problems. Matching student skills with the MCSC objectives, the VCCLL draws on a team of faculty and other expert practitioners to work with undergraduate and graduate students to study and resolve security issues. Students in this program will go on to careers in security work or pursue other IT professions, or, at the very least, will become more aware network users. The latter is vital to our increasingly wired and interconnected society--the Internet of Things (IoT)--that is rapidly becoming the norm.

## 2 VCCLL Background

The VCCLL concept arose from an earlier pilot project created by MCSC. MCSC built a small cyber range for the United States Coast Guard, Sector Northern New England, to provide training about security issues related to data vulnerabilities under shared network conditions. This training is necessary because there is no way to ensure absolute separation between an individual's online presence inside and outside of a work environment [2]. Constant vigilance is necessary and must include awareness and training beyond the typical worship of the complex password and avoidance of nefarious sites. Temptations for security breaches through the use of public and other outside networks arise, for instance, when employees travel and use networks at airports and hotels. Similarly, vulnerabilities exist in everyday life when individuals visit coffee shops and jump onto open networks. The VCCLL training is aimed at increasing participants' awareness and skills about data security, and is encapsulated in a set of exercises called. "Evil at the Coffee Shop" (ECS). The aim of these exercises is to sensitize participants about myriad cybersecurity exploits that can occur in routine and informal settings.

### 3 “Evil at the Coffee Shop” became the inspiration for the VCCLL

In order to understand how the “Evil at the Coffee Shop” (ECS) training inspired the creation of the VCCLL, it is helpful to describe its initial design in context. The request from the United States Coast Guard was to target non-IT personnel and provide them with a brief introduction that would make the cyber threat “real.” An additional goal was to both supplement and reinforce mandated annual cybersecurity training. The ECS simulation was created to meet these goals. ECS was first deployed at an active US Coast Guard Base, Sector Northern New England. The simulation entailed disabling the Coast Guard network during a staged “crisis.” In addition, the simulation was planned to occur at the same time as a disaster drill that included a simulated extreme weather event, an epidemic, and a hypothetical terrorist attack at a harbor or port.

The cyber range developed for the simulation was comprised of two laptops for end users, a wireless router, and two laptops acting as control with various virtual machines (VMs) to handle spoofed web pages and Domain Name System (DNS) changes. During this activity, three scenarios were experienced by participants: 1) control of DNS which sent the participants to a set of spoofed web pages, 2) a Denial of Service (DOS) attack, and 3) a phishing simulation. The experiential activity was followed up with discussion and critique. If time permitted or the participants’ questions made it appropriate, a packet capture technique was also demonstrated. The expectation was that two Coast Guard personnel would participate at a time. However, the simulation attracted a group at least twice that size and higher, with the largest group numbering more than thirty Coast Guard personnel. When this consistently occurred, we knew we were on to something.

#### Synopsis of “Evil at the Coffee Shop”:

- Users surf the web
- Activate the spoofed web pages and suggest the Coast Guard personnel surf to one of them (i.e., CNN, Fox News, or Yahoo) and discuss how the Domain Name System (DNS) changes impacted the “look and feel” of the page
- Have the group surf to a page that requires a login (e.g., Facebook, LinkedIn, etc.); show participants that we could capture their password
- Discussion of https versus http. Demonstration of packet sniffing
- The question of how to get users onto our “bad” or “poisoned” network would always be raised and thus a DOS would be demonstrated and explained
- Phishing techniques were demonstrated. For

example, how to read the headers and links for indications of phishing with a fake Amazon email

Several lessons were learned from the “Evil at the Coffee Shop” (ECS) pilot simulation, and these are taken into account for the design of the VCCLL. Lessons include:

- Reviews and observations of the exercise indicated that each participant needs to have his/her own laptop, simply watching others is not nearly as engaging and effective
- Setup time is about 20 minutes with two researchers/instructors helping
- Giving participants about 10 minutes to surf and become comfortable produces a bigger “a-ha” moment when directed to a spoofed page
- Participants almost instantly recognize that the DNS exploit could be executed in any number of public venues
- Exactly what phishers were trying to do is very clear to participants from the earlier DNS exploit
- The logical progression of “this is what can happen in a public place” to “this is what can happen via an email” is why an expired and self-signed certificate are a cause for alarm

### 4 VCCLL Design

While the “Evil at the Coffee Shop” simulation worked well at a small scale, the plan is to build out this concept to afford more sophisticated training for IT students, as well as offer basic level training for non-IT persons. The VCCLL is designed to be flexible to serve both university students as well as the larger community and organizations. Located within the currently existing MCSC Cybersecurity Research lab, the VCCLL will use remote nodes made up of virtual machines to run different simulations. The virtual machines would be configured to simulate a real-time complex network environment. For on-site nodes, Linux will serve as the base operating system, with virtual machines installed as the user operating system(s). Nodes will be linked via a Virtual Private Network (VPN) to merge the nodes into a single private lab.

The objective is to develop and evaluate the feasibility of an inter-institutional, virtual cybersecurity collaborative learning laboratory to foster teamwork among undergraduate and graduate students across distances [3]. The VCCLL would link three virtual laboratory nodes: one at York County Community College, one at the University of Maine at Fort Kent, and one at University of Southern Maine (head node). Students from remote areas of Maine would be able to work with students from Maine’s economic and population center in Portland. This design meets the five criteria for a virtual cybersecurity laboratory: 1) increase advanced, hands-on learning in networking and security courses; 2) reduce cost and the need for specialized computer labs; 3) provide an agile

and secure computer environment for information assurance (IA) education; 4) foster collaboration and teamwork among students in distant locations; 5) enable inter-institutional collaboration for shared resources in cybersecurity education.

The VCCLL is designed to accommodate fifteen to forty-five students depending on the simulation scenario and required roles. To participate in the simulations, students are required to have basic knowledge in networking and related information security. At the outset, students will be given an outline of the goals and objectives of the project and information explaining how these are integral to the goals and objectives of their courses, including the ways in which professional ethics and strategic communication play a role in information security systems and practices.

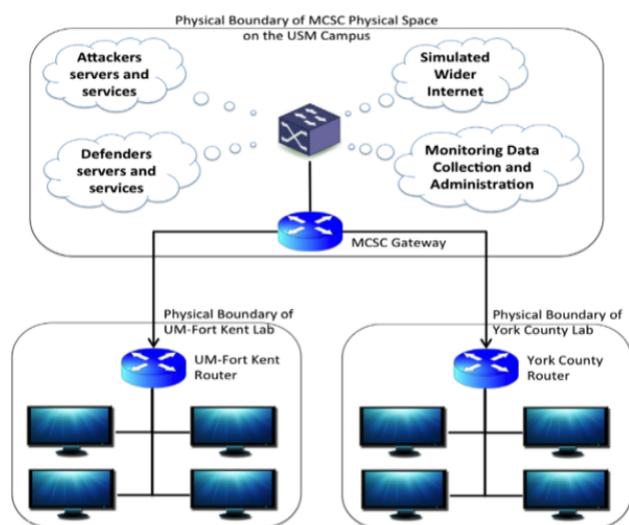


Figure 1

Maine Virtual Cybersecurity Collaborative Learning Laboratory

Notes: The elements denoted by clouds within the USM MCSC space are part of the base Cybersecurity Research Laboratory. The links between the sites will take place over VPNs to ensure that the hosting institutions are protected from activities within the lab. The virtual lab does not have any connections to either the hosting institutions networks or the wider Internet.

Figure 1 illustrates the flow of activities in the VCCLL. The expectation is that the VCCLL concept will surpass existing models by developing challenging exploit scenarios that require participants to: 1) learn and exercise highly developed interpersonal, collaborative skills; 2) generate specialized exploit mitigation skills; and 3) blend both of these skill sets while working in a dynamic virtual environment [4]. A unique aspect of the VCCLL is the affordance of achieving the foregoing with students from disparate geographical and cultural regions in Maine through randomizing exploit scenarios and team membership. This closely resembles the real-world conditions which ensue upon the discovery of exploits. The aim is to prepare students for the escalating complexities that emerge before, during, and after exploits, and for students to become accustomed to

working in realms where rules and roles change, and lives and major assets are at risk. Further, the VCCLL concept employs a three (3) node network that integrates two scenarios that exemplify common and frequent exploits into cybersecurity classes. These cybersecurity events are integral to the courses, and planned for prior to the event, with role and protocol assignments. They conclude with a debriefing and evaluation session—a post event “hot wash.” All students have the opportunity to partner both locally and remotely across virtual distances and will develop techniques and procedures for effective communication and collaboration [5].

## 5 Creation of simulation scenarios for the VCCLL

The VCCLL design and pedagogy builds on the experiences and insights gleaned from the “Evil at the Coffee Shop” (ECS) Coast Guard training pilot. MCSC faculty, staff, and other experts comprise the “Coffee Shop Team” and are charged with the creation of innovative, collaborative, and resource-shared cybersecurity simulation scenarios. Criteria for the scenarios are that they reach across diverse rural and urban cultures to strengthen the foundation of computer offense and defense security knowledge. Each simulation must be designed to last approximately four hours to allow for participants to apply collaboratively offensive and defensive skills and techniques. Distributed Denial of Service (DDOS) and Malware Eradication will continue to be refined as exploit scenarios for the VCCLL. Other simulation scenarios include general and common exploits such as: understanding the impact of Domain Name System (DNS) spoofing; understanding the differences between http and https and why those differences matter; understanding how a Denial of Service (DOS) attack can be used to drive users to a malicious access point, and using basic tools for differentiating phishing emails. The Coffee Shop Team hopes that the training is infectious and spreads to friends, colleagues, and associates of end-users, in part, by the increased awareness of participants and the need to share this so all can act to secure the network. Further examples of VCCLL exercises include:

*DDOS “Hactivists”:* In this exercise, participants learn to identify the major components on the network (improve documentation); identify the nature of the attack; select and configure effective teams; request help from outside; deploy help; respond to the attack; and contain the attacks.

*Malware Eradication:* Participants identify the nature of the outbreak; build and maintain records that reflect spread, investigation, and eradication; select and establish effective teams; use external sources to categorize and identify malware; request external help; deploy help; contain the attack. In addition, participants learn to communicate effectively with management about the event including investigation, efforts to eradicate,

advocacy for contacting outside resources, and communicate effectively with end users about the event.

*Integration of Professional Ethics and Strategic Communication Skills:* Participants gain increased understanding about when, how, and with whom to communicate network vulnerabilities and security breaches. This aspect of security may be just as important as awareness. The Coffee Shop team for the VCCLL will include a professional ethicist and a communications expert. This interdisciplinary aspect of simulated security attacks will allow participants to explore underlying reasons for responsible use of the internet, the importance of security for the ethical values of privacy and confidentiality that allow for autonomy and maintain civil society.

## 6 VCCLL Benefits

The VCCLL offers benefits that are lacking in current virtual models. First, the laboratory provides an in-lab experience using real world breaches [6]. This mirrors the working environment found in medium size or larger organizations. The geographic diversity of the ad hoc teams reflect structures often found in government and industry security groups. Using well-established and relatively common exploits in the virtual laboratory, students will experience multi-dimensional / multi-way simultaneous attacks and will be trained to address, correct, and guard against such activities in an ad hoc collaborative setting.

Second, central to the virtual laboratory model, is that students understand and appreciate the value of working in effective ad hoc teams in a highly decentralized laboratory--where they may or may not have peers with their levels of technical expertise physically present. This means that students will have to be confident in their ability to communicate and collaborate using technologies that, as yet, cannot convey the complete subtleties of *in situ* human interaction. This scenario has particular relevance to SCADA systems that are often physically remote and require local personnel to act as the security teams eyes and hands. Students will have to communicate using remote systems, and they will have to make decisions collectively and execute them without the luxury of physical presence. In the traditional classroom setting, an intervention or repair may be difficult, and a team decision will have to be made, but there is comfort for students in that they can discuss the solution in real-time and in each other's presence (with all the non-verbal cues that humans rely upon) in a critical situation.

Third, students and their teams learn to communicate effectively on many levels and often at off hours or during extreme conditions. Such extreme conditions make regular communication difficult or strained. Adding the factors of massive outages of data or physical infrastructure, remote, long distance, and or virtual communications frequently

fail. These failures and their solutions are addressed in the VCCLL. Frequently, capacity building models address research and development on a particular topic. In this case, the research and development both targets the virtual environment and uses the virtual environment and its tools, techniques, and culture.

## 7 VCCLL Integration

Lastly, the VCCLL is fundamentally interdisciplinary integrating tactical knowledge in cybersecurity with fundamental principles in strategic communication and professional ethics. Strategic communication concepts and practices are enacted including proactive, pre-crisis planning, ongoing communication management across work activities and groups, and post-crisis strategic response deployments. Through these activities students gain appreciation of communication behaviors that may influence crisis prevention and outcomes such as patterns of interpersonal and small group communication and decision making, forms and methods of communication with stakeholders, and effective use of media to communicate information about the crisis. Professional ethics are considered in terms of personal privacy, confidentiality of financial and personal information, and the importance of the ethical value of trust, which is often at the heart of all cybersecurity undertakings. Trust is central to maintaining personal autonomy and to securing the integrity of social and virtual networks [7] [8]. Trust includes trusted systems, nodes, and identification, which are all subject to attack or subversion. The model also includes the trusting relationships among students (student to student individually or in teams) or students to machine(s). Without the interpersonal (and person to machine) trust developed transactionally through solid communication practices, collaboration will not take place or will dwindle rapidly.

## 8 Results

Over time, it is envisioned that the VCCLL concept could be scaled up and applied to training for the general public who are now consumers and users of products and services that inherently carry security risks in the world of the Internet of Things (IoT), including networked homes, schools, libraries, and offices. The chief outcome of any successful VCCLL is increased participant awareness and caution around security along with deeper understanding of the responsibilities of network users to others beyond specific networks. While the VCCLL could be reconfigured to support SCADA breaches, educating non-IT staff using and supporting these systems about cybersecurity would greatly enhance the security of these systems. Anticipated results suggest improvements in the logistical design and implementation in virtual network nodes and information will aid in the successful execution of exploits in a distributed virtual collaborative laboratory over distances.

VCCLL ideas and concepts have been further developed and submitted to the National Science Foundation as a proposal under the CyberCorps Scholarship for Service program.

## 9 Conclusions

From the work with the US Coast Guard and our students, it is clear that there is a proven and appreciated need for hands on and real-world activities and training on typical cyber security exploits. Such injects are both fascinating to the everyone, whether they are seasoned IT workers or undergraduate students. Moreover, both groups need continued and upwardly scaling (quantity and complexity) experiences across geographic regions, networks, systems, and scenarios. Therefore, the establishment of a highly scalable virtual cyber security collaborative cyber range is a logical next step based on the preliminary work done over the last year at the Maine Cyber Security Cluster. Further, the need for a collaborative interdisciplinary approach is essential to establish effective communication and ethical behavior combined with technical expertise in order to overcome cyber security exploits.

## 10 References

- [1] Nance, K., Hay, B., Dodge, R., Seazzu, A. and Burd, S. (2009). Virtual Laboratory Environments: Methodologies for Educating Cybersecurity Researchers. *Methodological Innovations Online*, 4(3) 3-14.
- [2] Kott, A. (2014). Towards Fundamental Science of Cybersecurity. *Network Science and Cybersecurity, Advances in Information Security*. Volume 55, pp 1-13.
- [3] Bonabeau, E. (2013). Cybersecurity: Human Behavior Matters. Icosystem Corporation. Retrieved from <http://www.icosystem.com/cyber-security-human-behavior-matters/> on February 13, 2014.
- [4] Viveros, M. and Jarvis, D. (2013). Cybersecurity education for the next generation: Advancing a collaborative approach. Center for Applied Insights. IBM Corporation.
- [5] Willems, C., Klingbeil, T., Radvilavicius, L., Cenys, A., and Meinel, C. (2011). A distributed virtual laboratory architecture for cybersecurity training. Published in 2011 International Conference for Internet Technology and Secured Transactions (ICITST). Pp 408-415.
- [6] Zlateva, T., Burstein, L., Temkin, A., MacNeil, A., and Chitkushev, L. (2008). Virtual Laboratories for Learning Real World Security. Proceedings of the 12th Colloquium for Information Systems Security Education. University of Texas, Dallas, TX June 2 – 4.
- [7] C. Ess and Thorseth, M (2011). Trust and Virtual Worlds. Peter Lang Press.
- [8] Vallor, S. (2012). Flourishing on Facebook: Virtue friendship and new social media. *Ethics and Information Technology*, 14 (3). 185-199.

# LearnFire: A Firewall Learning Tool for Undergraduate Cybersecurity Courses

Alicia Kirkland  
Lewis & Clark College  
Portland, OR 97219  
akirkland@lclark.edu

Jens Mache  
Lewis & Clark College  
Portland, OR 97219  
jmache@lclark.edu

**Abstract** - Cybersecurity is a fairly new topic in computer science. Firewalls are one of the most important elements in keeping a network secure. This paper describes the design, function, and goals of LearnFire, a collection of exercises for firewall education. LearnFire is designed to be used in the classroom as a hands-on learning tool, but can also be used by students independently. Each element of LearnFire aims to test varying levels of knowledge concerning firewalls. LearnFire is unique for three main reasons. First, it exists completely in the cloud, allowing students to access it inside and outside the classroom. Secondly, LearnFire tests the ability to build a firewall and to analyze an existing firewall for functionality and effectiveness. Lastly, and most importantly, LearnFire provides feedback for students to help further their learning and assess their progress.

**Key Words:** security, firewalls, education, exercises

## I. Introduction

With the rise in cyberterrorism and hacktivism, companies are seeking people with cybersecurity experience more than ever. An article from Reuters reports that some of the largest companies in the United States are hiring cybersecurity experts to serve on their executive boards, which indicates increased concern with the threat network attacks pose today [1]. The demand for people with cybersecurity experience is on the rise, meaning that the demand for students who have experience

building and analyzing firewalls is increasing. Firewalls are “network devices whose purpose is to enforce a security policy across its connections by allowing or denying traffic to pass into or out of the network” [2]. They play a huge role in cybersecurity. Therefore, it follows that in cybersecurity education, there should be tools aimed at teaching student how to properly build a new firewall and analyze an existing one.

Since class time is limited, exercises used in class must be efficient and useful. They should not take a considerable amount of time to set up and troubleshoot. Additionally, exercises must supplement the lesson, test students’ ability to produce content, and check that students truly understand the meaning of what they have learned. Simply stated, good learning tools test students’ ability to build and analyze, while providing feedback that students can use to assess their learning progress. LearnFire does exactly that while existing conveniently in the Amazon Web Services (AWS) cloud. LearnFire emerged as an alternative to two other firewall learning tools: DETERLab and FireSim [3, 4].

## II. Related Work

In the Cybersecurity course at Lewis & Clark College, we used two exercises to practice firewall skills: a DETERLab scenario and a firewall simulation game called FireSim.

## A. DETERLab

DETERLab is an environment in the cloud that provides virtual machines for students to perform experiments. Instructors must work with DETERLab to set their students up with accounts. After this, starting an experiment takes less than ten minutes on average. DETERLab provides reading material for students as an introduction to firewalls, specifically IPTables. This is important because the syntax for IPTables is complicated and can be confusing for students who have never worked with them.

The DETERLab scenario is a fairly simple lab exercise. The student acts as a security administrator for a company. The lab gives students a list of requirements for the firewall and the students create a firewall from scratch.

This scenario is a great first step in learning about firewall configuration. However, it lacks complexity. Students must simply build a firewall and submit it to their instructor. The lab instructions provide methods to test the firewall, but there is no feedback beyond that (besides instructor feedback). It does not test students' analysis skills. Setup time for this is minimal. Students must begin their experiment on DETERLab and log in to the virtual machines using an ssh client.

## B. FireSim

FireSim is a competitive learning tool that requires students to build a firewall as protection against attacks from other students. Professor Ken Williams of North Carolina A&T State University developed FireSim using a Java applet and XML files. To use FireSim, instructors must download a group of files on a computer that acts as a web server. There is an additional computer that acts as the administrative computer that the instructor uses to add tasks over time. Each student operates his or her own computer.

The scenario lists a series of requirements that the students must respond to by building a firewall that allows and denies traffic according to the requirements. For example, the initial configuration

of the firewall must allow domain name server access, access by the public to the student's website, and email from other email servers using SMTP. Students then attack each other's networks, gaining points when their attacks are successful and losing points when their network is attacked. Students update their firewall in response to successful attacks and new requirements designated by the administrator.

FireSim has an excellent concept. It provides a competitive environment for a classroom, which engages students differently than a lecture or lab. The goal of the game is easy to explain and understand. However, some of the tasks are a little tricky and vague. Some of the tasks do not require a rule, but students do not know or understand this due to the minimal feedback provided by the game. Additionally, FireSim itself is buggy. Beyond that, the game could give more feedback and does not engage the students with what's happening under the hood. Students will understand how to write a rule to block access, but they might not understand what the rule means or why it blocks access.

## III. LearnFire Scenarios

LearnFire is a collection of exercises for students to practice building and analyzing firewalls. It will be part of EDURange, a cloud based resource for hosting on-demand interactive cybersecurity scenarios [5]. LearnFire will have at least three scenarios, with more being developed over time and as new firewalls and methods emerge in the field. These scenarios test skills concerning various types of software and hardware firewalls. The initial scenarios focus on IPTables, Berkley Packet Filter (BPF) [6], and Palo Alto Networks.

Feedback is a very important feature in LearnFire. It allows students to further understand the topic at hand and gives them clear guidelines on where there are gaps in their knowledge and where they need to invest more time.

### A. Scenario 1

The first scenario tests students' ability to analyze an existing firewall and create firewall rules. The student has access to two or more nodes in the cloud, represented by A, B, and C in Figure 1. Each node has its own set of firewalls (using IPTables and/or BPF), represented by A1, B1, and C1. The student must complete a series of tasks such as pinging one node from the other, sending a file, using SSH, and more. Students will have to edit, add, or delete rules in order to complete their tasks. They must record what they did to complete each task. This scenario only requires a network connection and a command line where a student can sign in to connect to the virtual machine. In terms of feedback, the scenario acknowledges completed tasks and gives hints when prompted by the user.

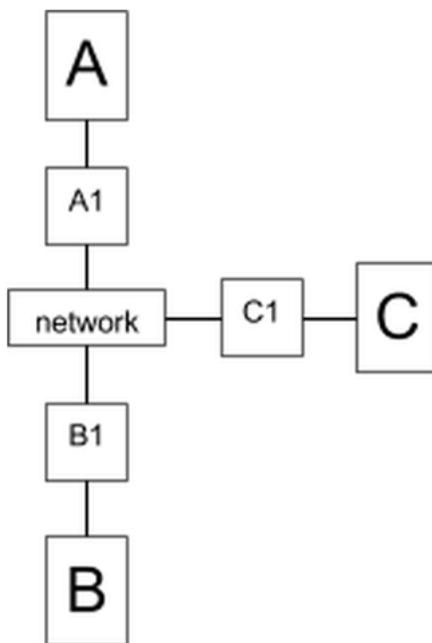


Figure 1: Conceptual Diagram of an example topology for Scenario 1

## B. Scenario 2

Similar to FireSim, this scenario will provide a more competitive platform for students, represented as Alice and Bob in Figure 2. Each student builds their own firewall, represented by A1 and B1, to

prevent their opponent from gaining access to their machine. The scoring agent will be live, constantly checking the command line and status of each virtual machine, and sending messages to each student to update them on the points they have won through penetrating their opponent's system and points they have lost due to their opponent succeeding in penetrating their system. The game can be timed or untimed, depending on the instructor or student preference. The scoring agent provides feedback by telling the students that they have gained or lost points and why that's happened.

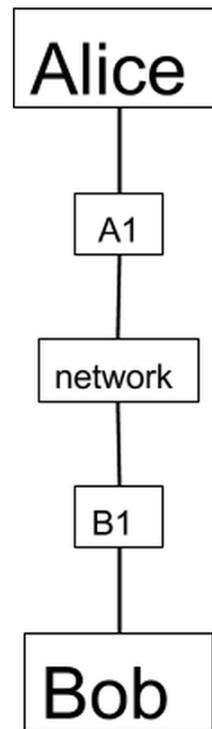


Figure 2: Conceptual Diagram of Scenario 2

## C. Scenario 3

This scenario shifts the focus from software firewalls to hardware firewalls. Palo Alto Networks (PAN) worked with us to configure a virtual machine that uses their user interface to build a firewall for a network. PAN is a next-generation firewall [7]. Rather than filtering traffic based on ports and IP addresses, PAN allows the user to build a firewall that filters traffic using user

identification, content identification, and application identification. This new technology is more similar to what someone would experience in the professional world as the security administrator for a network.

Students will have access to three virtual machines: a management console, a machine inside the network, and a machine outside the network. The students will be able to interact with the management console to access the traffic rules and use the other machines to generate traffic. A question posed by the scenario might ask, "Which rule limits Alice's ability to send a Facebook message?" or the scenario might call for the student to perform a task and check the rules if the task fails. Students must analyze the existing firewall and think critically about the meaning of each rule. Feedback would appear in the form of telling the student whether their answer is correct and why or why not.

#### IV. Future Work

As this is currently in development, it has not yet been tested with students. The scenarios will be tested using students at Lewis & Clark College as well as other student volunteer groups in the Pacific Northwest from participating institutions.

#### V. Conclusion

LearnFire creates a learning environment in the cloud for students to develop firewall skills. The minimal setup time makes it an effective exercise for in class work and its cloud availability allows

students to work from home. The feedback features engage students and allow them to build on their knowledge and push themselves to learn more. Overall, LearnFire is an excellent tool for educating students in order to fulfill the demand for cybersecurity experts.

#### VI. Acknowledgements

The National Science Foundation grant 1141314, the John S. Rogers Science Research Program of Lewis & Clark College, and the James F. and Marion L. Miller Foundation provided funding for this project. Special thanks goes to Richard Weiss.

#### VII. References

- [1] Nadio Mamouni (2014, May 30). *U.S. companies seek cyber experts for top jobs, board seats* [Online]. Available: <http://reuters.com>
- [2] Wm. A. Conklin and G. White. "Intrusion Detection Systems and Network Security" in *Principles of Computer Security: CompTIA Security+ and Beyond*, 3<sup>rd</sup> ed. Emeryville, CA: McGraw-Hill/Osborne, 2012, Ch. 13 pp. 334
- [3] K. Williams. Firewall Simulation [Online]. Available: <http://williams.comp.ncat.edu/FireSim/index.htm>
- [4] P. Peterson and P. Reiher. POSIX Permissions and Stateful Firewalls [Online]. Available: [https://education.deterlab.net/file.php/12/PermissionsFirewalls\\_UCLA/Exercise.html](https://education.deterlab.net/file.php/12/PermissionsFirewalls_UCLA/Exercise.html)
- [5] S. Boesen, R. Weiss, J. Sullivan, M. Locasto, J. Mache, E. Nilsen, "EDURange: Meeting the Pedagogical Challenges of Student Participation in Cybertraining Environments", CSET Workshop, USENIX Security Symposium, 2014
- [6] S. McCanne and V. Jacobson, "The BSD Packet Filter: A New Architecture for User-level Packet Capture," LBL, Berkeley, CA, Dec 1992
- [7] J. Snyder. *What is a next-generation firewall?* [Online]. Available: <http://networkworld.com>

# From Air Conditioner to Data Breach

G. Markowsky and L. Markowsky

School of Computing & Information Science, University of Maine, Orono, Maine, USA

**Abstract**—*This paper examines the 2013 Target Data Breach in detail with the intent of developing some lessons learned that can serve security educators. The Target Data Breach originated in the network of a trusted vendor and then spread to Target's network. The rush to put more objects on the Internet is introducing many vulnerabilities into networks, so Target's experience of being attacked from a "trusted" source is likely to be repeated from many new sources. This paper then discusses the concept of a "kill chain" and how it could be of use to defenders. Finally, it discusses the relevance of the cyber castle metaphor to the design of hybrid networks and some approaches to building secure hybrid networks.*

**Keywords:** Target Data Breach, Internet of Things, IoT, Cyber Castle, hybrid network

## 1. Introduction

On December 18, 2013, Brian Krebs posted [1] an item in his blog about Target investigating a data breach. On December 19, 2013, the giant retailer released a statement [2] confirming that they were indeed investigating a massive data breach. Target's statement included the following section.

Approximately 40 million credit and debit card accounts may have been impacted between Nov. 27 and Dec. 15, 2013. Target alerted authorities and financial institutions immediately after it was made aware of the unauthorized access, and is putting all appropriate resources behind these efforts. Among other actions, Target is partnering with a leading third-party forensics firm to conduct a thorough investigation of the incident.

The Target Data Breach inspired many news articles such as Goodin [3], Mick [4] and Krefit [5]. Mick [4] traces the source of the attack to an HVAC (heating, ventilation and air conditioning) company, Fazio Mechanical Systems, located in Sharpsburg, Pennsylvania. He notes that among Fazio's clients are Walmart, Costco, Exxon Mobil, and many other companies. Krefit [5] notes that the breach was caused by the loss of one of Fazio's employee's credentials and that Target gave Fazio access so they could remotely login and perform efficiency updates.

Yang and Jayakumar [6] report that in addition to the 40 million stolen credit cards, personal data for up to 70 million Target customers was also stolen, and that some customers

might be in both groups. Although the Target Data Breach was large, it is not the largest known breach [7], [8]. See [7] for an interactive visualization.

## 2. Details of the Target Data Breach

On March 26 a U. S. Senate Committee released a report [9] about the Target Data Breach. Figure 1 from that report shows many interesting details about this breach. First, the attack took place over almost three months beginning in September 2013 and ending on December 15, 2013. Thus the attack was not some spur of the moment event carried out by a teenage hacker. It shows a great deal of planning and patience. Ironically, the attack began about the same time that Target was certified as PCI-DSS [10] compliant. The attack began with the theft of credentials from one or more Fazio employees. As noted in [4], Fazio has a number of large retailers as clients, and we do not know whether the attackers were specifically interested in exploiting Target or just discovered that Target was an easier "target" than other retailers.

According to [9], the attackers first breached Target's network on November 12, 2013. They spent nearly two weeks (11/15-11/28) testing malware on Target's point-of-sale (POS) system. Interestingly more than two weeks passed before Symantec and FireEye software detected the intrusions and alerted Target. At this point, no damage had been done and no data had been stolen. So far no one has come up with an explanation as to why Target chose to ignore the warnings that it received from its own systems.

Riley [11] contains some additional information about how Target was compromised. Six months before the data breach, Target purchased a computer security system called FireEye for \$1.6 million. On multiple occasions FireEye warned Target about the presence of intruders in its networks and about some of their activities. These warnings were reviewed by Target's security staff and ignored. Finally, on December 12, 2013, the U. S. Department of Justice notified Target that its network had been breached and data stolen. It took Target another three days to remove the malware and attackers from its system.

Riley [11] contains many additional details about the malware and how it was installed on Target's network. It also includes a discussion of how the stolen credit card numbers were offered for sale and the fact that one of the websites that sold the stolen credit card numbers, Rescator.so, was broken into and the logins, passwords and payment information of carders were posted online.

# A Timeline of the Target Data Breach

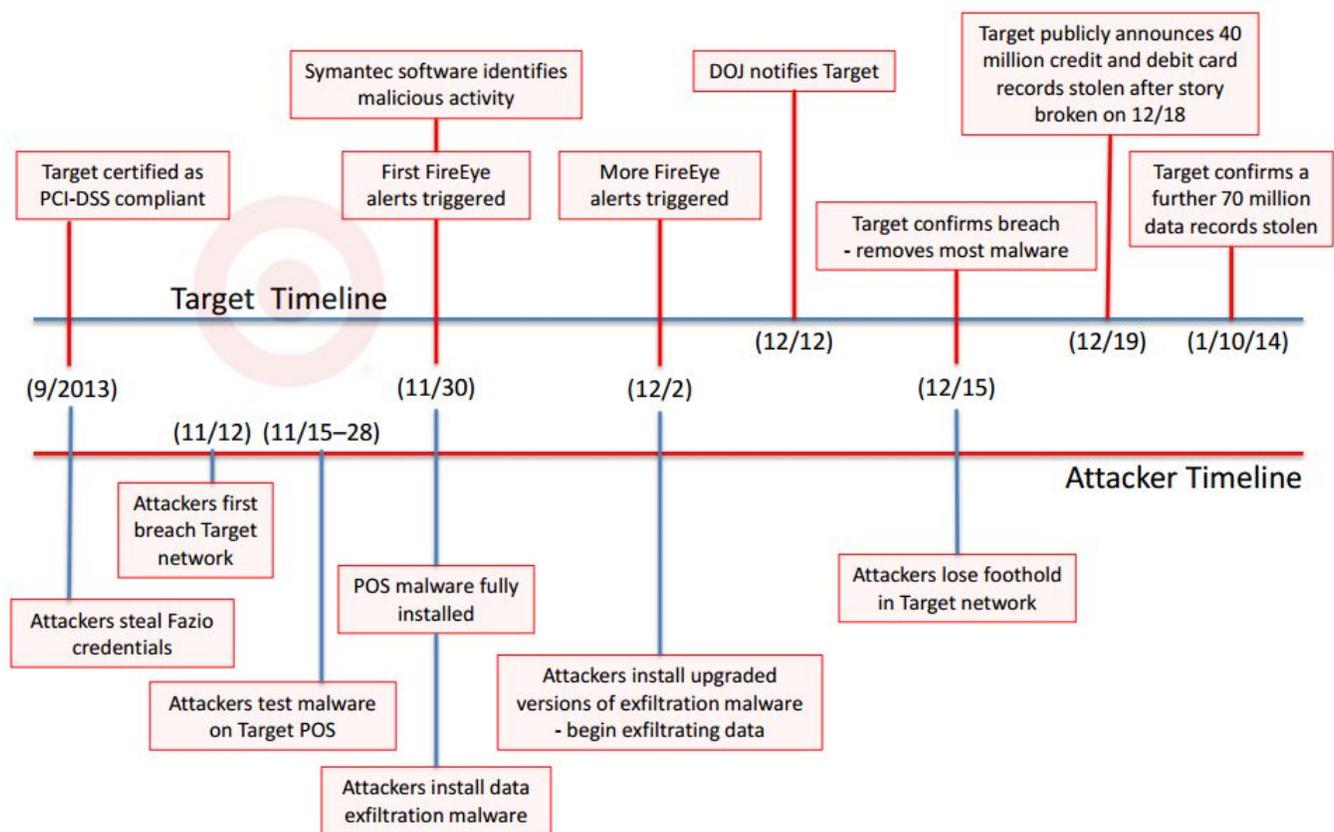


Fig. 1: Timeline from the Senate Report [9]

This data breach was very costly to Target and its staff. Target's profits fell 46% during the holiday season. In addition, several lawsuits were brought against Target, which will likely result in additional losses and legal fees. The data breach led to the resignation of Beth M. Jacob [12], its Chief Information Officer and Executive Vice President for Technology Services in March 2014. Ms. Jacob had no training in computer science or cybersecurity and it is unclear how much of a factor this was in the Target Data Breach. Her resignation was followed by the resignation of Target's CEO, Gregg Steinhafel, in May 2014 [13].

One consequence of the Target Data Breach is the acceleration in the adoption of chip-containing credit cards by Target and other retailers. For more details see [14].

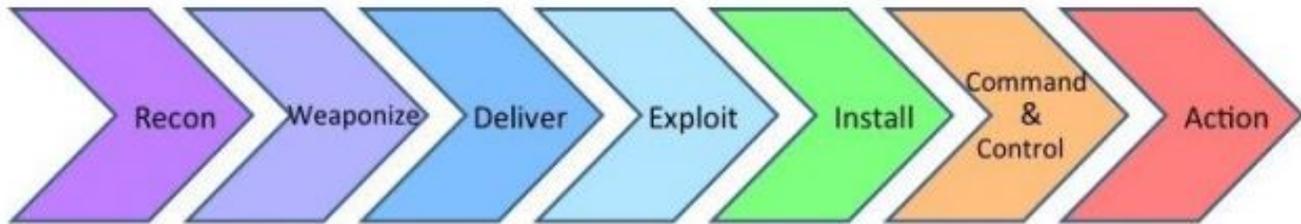
### 3. Defending Against Target-Type Data Breaches

One of the reasons for studying data breaches is to figure out ways to reduce the likelihood of future data breaches. The Senate Report [9] mentioned in the previous section

discusses the use of a "kill chain" in defending against Target-type data breaches. This concept was introduced by the Lockheed Martin Computer Incident Response Team in 2011 [15]. The goal of the kill chain approach is to redress the perceived imbalance between attackers and defenders. Typically, attackers need to only find one weak spot to proceed with exploitation, while defenders must protect all areas of a network. Users of kill chains try to mount an active defense and to model the attacker's steps by a kill chain of steps. The term kill chain comes from the fact that the attacker needs to carry out all the steps in the chain of steps to be successful, while the defender only needs to interrupt any one of the stages to prevent the attack. Kill chains are viewed as a defensive weapon against advanced persistent threats (APTs) such as the Target Data Breach.

The steps of the intrusion kill chain are shown in Figure 2. The following is a list of the steps and a brief explanation each.

- 1) *Reconnaissance*. This involves collecting as much information about the target as possible. This is done us-



Source: Lockheed Martin

Fig. 2: The Intrusion Kill Chain [9]

ing as many resources as possible. Amazing amounts of information can be collected from the Internet.

- 2) *Weaponization*. This involves putting together an exploitation package for the intended victim. This package is often built by combining a standard document such as a PDF file, a word processing document or a spreadsheet with some type of remote access trojan.
- 3) *Delivery*. This involves getting the payload to the intended victim. Three common methods for doing this are: email attachments, compromised websites and infected USB drives. Almost all these methods require some cooperation from the intended victim.
- 4) *Exploitation*. This involves getting the payload activated and getting a foothold on the target system. This step provides the first link between the attacker and the victim's system.
- 5) *Installation*. This involves expanding the bridgehead into a persistent presence on the victim's system.
- 6) *Command and Control (C2)*. This involves setting up full control of the system and the escape path for such things as stolen data.
- 7) *Actions on Objectives*. This involves the attacker accomplishing whatever were the original goals of the attack.

The steps in the kill chain are familiar to cyber defenders. The novelty of using the concept of a kill chain is that it provides a strategy for an active defense that has the ability to disrupt many APTs. [15] provides a detailed case study of the use of this technique. We will follow the lead of the Senate Report [9] and apply this kill chain method to the Target Data Breach with the idea of suggesting how an active defense can thwart such attacks. The concept of a kill chain means that the attacker can be stopped at any point along the chain. Some steps in the chain might be easier to disrupt than other steps, and it is best to focus on those steps.

- 1) *Reconnaissance*. It appears that the Target attackers carried out their reconnaissance through Internet searches and by using Target's supplier portal and facilities, which were active as of February 12, 2014

[16]. In this step, the attackers identified Target's third-party vendors. Some of Target's vendor information sites continue to be active as of August 30, 2014 [17], [18]. This publicly available information permitted the attacker to map Target's internal network prior to the breach. One good defensive action for most organizations might be to limit the amount of publicly available information about themselves. This is difficult to do since no one has full control of the information available about them. For example, one does not need to have a Facebook page to have a Facebook presence: it is enough to have friends with Facebook accounts who choose to mention you. While security researchers tend to disparage "security through obscurity," making it difficult for an adversary to learn about your systems might encourage the adversary to search for an easier target. There is a reason why carnivores typically pick less vigorous animals when there is a choice. Organizations can help their defense by encouraging their employees and collaborators to expose as little information as possible to the public. As countries have learned in wartime, "loose lips, sink ships."

- 2) *Weaponization*. It is speculated that the weapon used to initiate the Target Data Breach was most likely a modified PDF or Microsoft Office document that was emailed to a Fazio employee. At that time Fazio was using the free version of Malwarebytes's Anti-Malware software, which does not provide real-time protection and is not licensed for commercial use. In general, organizations should invest in protective software. Although this protective software is not foolproof, it does catch many instances of malware and helps raise cybersecurity awareness among users. In some sense, it is hard for a defender to disrupt the weaponization stage since it is totally in the hands of the attacker. At best one can prepare for different sorts of weapons once they get delivered.
- 3) *Delivery*. The weapons appear to be delivered to Fazio via a phishing email. Once Fazio was compromised,

it was relatively easy for the attackers to get into Target's network. The PCI-DSS standard requires two-factor authentication for network access from outside the network as shown in the following quote [10, p. 47].

8.3 Incorporate two-factor authentication for remote access (network-level access originating from outside the network) to the network by employees, administrators, and third parties.

Organizations can disrupt the delivery or an exploit by getting their employees to recognize the dangers of phishing emails. One effective method is to send phishing-type emails to your own staff and display a "Gotcha" type of message when people click on links they shouldn't click on. Phishing is surprisingly successful even at security companies so it is important to take it seriously and to take steps to make it less effective. Similarly, use two-factor authentication as much as possible. It is obviously less convenient, but the consequences of security breaches are becoming quite severe. It is especially important not to ignore calls for two-factor authentication when one is supposedly adhering to some standard such as PCI-DSS.

- 4) *Exploitation.* To defeat exploitation one must make one's systems as secure as possible. Ironically, Target's FireEye software system had a feature that would automatically eradicate malware, but that feature was disabled at the time of the attack. Being aware of which attacks are likely to be deployed can help defeat exploitation. For example, in 2013 Visa issued warnings in April [19] and August [20] describing exactly the sort of attack that was used against Target. Had the Target security staff been on the lookout for such attacks, they would have likely responded earlier and more effectively to the attack launched against them. Organizations should seek to learn as much as possible about current threats. This can be through reading as widely as possible and attending conferences and workshops. The cyber threat landscape is constantly changing and one needs to stay current. In general, defenders will get better results if they are active defenders rather than first responders once a disaster has occurred. Of course, if one has defensive systems and they issue warnings, it is imperative that the defender understand exactly what is triggering the warnings. It is a bad idea to routinely dismiss warnings as false alarms. If one is indeed troubled by false alarms from a system, then either the system needs to be better configured or replaced.
- 5) *Installation.* It is not clear how the installation step was carried out by the attackers. There is some speculation that the attackers might have exploited a default account in a BMC Software information technology

management system. In general, system security is increased by securing or removing all default accounts and making sure that all default passwords have been replaced with real passwords. This is a requirement of the PCI-DSS standard [10, p. 24].

- 6) *Command and Control.* Figure 1 shows that about a month passed between the time the Target network was first breached and the time that the Department of Justice notified Target that its systems had been breached. The details of how the attackers maintained their position in the network are not known; however it is known that the attackers seemed to be able to roam freely throughout Target's system. Target would have benefited from having strong firewalls between various systems. It would also have benefited from blocking or filtering Internet connections that are commonly used for command and control. Networks containing sensitive data should be very unfriendly landscapes for roaming by unauthorized users. There should be frequent barriers and challenges to all who traverse this landscape. We will revisit this point in our last section.
- 7) *Actions on Objectives.* The data stolen from Target's servers was exported by FTP in plain text to several servers, at least one of which was located in Russia. At a minimum, Target should have had network rules in place that prohibited connections to countries with which it had no business relations. This would have complicated the data exporting for the attackers. In general, it is important to watch outgoing traffic for suspicious activities. Many firewalls focus on filtering incoming traffic. While this is important, it is only part of the story. Sometimes outgoing traffic is easier to analyze for suspicious activities. In general it is good to have whitelists, graylists and blacklists to help interrupt malicious activities and to expedite benign activities.

Organizations need to create attack scenarios to give themselves an opportunity to critically review their own security posture. The analysis applied to the Target Data Breach in this section can be applied by organizations to their own systems.

## 4. Implications for The Internet of Things

The Internet of Things (IoT), sometimes referred to as the Cloud of Things (CoT) or even the Internet of Everything (IoE), is a term that refers to the growing interconnected ensemble of objects that use the Internet to provide connectivity. Some authors include computers and smartphones in the Internet of Things, while others exclude them.

The security threat introduced by the IoT, and the relevance of the Target Data Breach, is that now computers are

sharing the same cyberspace as thermostats, air conditioners and countless other “smart” devices. The growth of these smart Internet-connected devices promises to swamp the growth of computers, tablets and smartphones. If every appliance has some sort of connectivity, along with every TV, game box, burglar alarm, heating and ventilation system, fire alarm, etc., then it is easy to see that the average household might soon have more “things” devices connected to the Internet than traditional devices. As early as 2003 [21] many luxury cars had 100 or more processors. Even the run of the mill economy car in 2003 already had several dozen processors. Now with the development of the “connected car” [22], [23] all of these microprocessors will be vulnerable. It is not surprising that Gartner [24] estimates that there will be 26 billion devices in the IoT by 2020, and that ABI Research [25] estimates that there will be 30 billion devices in the IoT by 2020. Note that both estimates do not include computers and smartphones. Gartner [24] notes that “by 2020, component costs will have come down to the point that connectivity will become a standard feature, even for processors costing less than \$1.”

At least one think tank has declared the Internet of Things to be one of the major security threats for 2014 [26]. Studies by Norse [27] and Hewlett Packard [28], [29] identify many types of objects, including such things as printers, thermostats, and security systems, that can be and have been compromised.

One of the reasons that the Internet of Things is such a security threat is that security takes a backseat to innovation [30]. Two powerful forces driving the growth of the IoT are profit and convenience. Companies see the Internet of Things as a very lucrative market. The market for set top boxes grew to \$20 billion in 2013 [31]. The markets for many other devices are also expanding rapidly. In addition, businesses expect that the “Big Data” generated by the armies of sensors and intelligent devices will help them develop new products and increase their profits [32].

## 5. Conclusions and Recommendations: The Castle Metaphor Revisited

It is clear from the analysis of the Target Data Breach and the growth of the Internet of Things that networks in organizations are going to be hybrids. For this hybrid environment, the castle metaphor [33] both conveys the concepts of cyberdefense and complements the concept of a kill chain. Castles inspire many people from an early age and provide a physical model for security that some people might relate to better than just a purely virtual model.

Applying the castle metaphor to the Target Data Breach, we conclude with the following observations:

- 1) Real castles were always part of an overall defensive strategy and were often constructed first, before the surrounding cities were built. This was not always

possible in the case of older cities, but in many cases cities grew around castles that were able to provide local defense. Many computer networks grow in an arbitrary and unplanned manner, without a strategy to meet the organization’s objectives and needs. Clearly, the Target network would have benefited from a better design.

- 2) Castles were subdivided into a number of subareas that could be defended even if some of the defenses were breached. Organizations need to run through various scenarios on the assumption that their defenses will be breached. In particular they should focus on information that they do not want attackers to get and think about how to protect it better. It is clear that sensitive data in Target’s internal network was insufficiently protected.
- 3) Castle defenses were active and castle defenders thought hard about how to put as many obstacles in the path of attackers as possible. As noted earlier, kill chains are designed to work with an active defense.
- 4) Castle defenses had multiple walls constructed so that they supported each other. For example, some castles had two sets of walls. The inner walls were taller than the outer walls so that even if the enemy were to capture the outer walls, they would not be able to look down upon the defenders on the inner walls. This reinforces the idea that defenses need to be designed with proper separation and defense given to particular items.
- 5) Castles directed attackers in particular directions and made them work for every inch of territory. Since cyber crime has become a business, having defenses that require more time from an attacker to overcome will often encourage the attacker to go elsewhere. The FireEye system that Target installed forced the attackers to use its facilities and enabled it to spot the intrusion. Regretably, Target security personnel ignored the FireEye warnings.
- 6) Castles had removable bridges and narrow passages that made defense easier. The various restrictions proposed on FTP traffic function as narrow passages and removable bridges.
- 7) Castles used guile and deceit to redirect attackers and to confuse them. The FireEye system used by Target is an example of guile when used properly.
- 8) Castle defenders usually had a good idea of who would be likely to attack them and how. The Visa alerts [19] and [20] outlined exactly the sort of attack that Target might be subject to. Unfortunately, Target ignored these timely warnings.

Few doubt that providing secure cyber services is becoming more challenging. It will require all of us to devote more attention to cybersecurity in order to prevent future Target-like data breaches.

## References

- [1] Brian Krebs, "Sources: Target Investigating Data Breach," Krebs on Security, December 18, 2013, <http://krebsonsecurity.com/2013/12/sources-target-investigating-data-breach/>.
- [2] Target Press Release, "Target Confirms Unauthorized Access to Payment Card Data in U.S. Stores," Minneapolis, December 19, 2013, <http://pressroom.target.com/news/target-confirms-unauthorized-access-to-payment-card-data-in-u-s-stores>.
- [3] Dan Goodin, "Point-of-sale malware infecting Target found hiding in plain sight," *ars technica*, January 15, 2014, <http://arstechnica.com/security/2014/01/point-of-sale-malware-infecting-target-found-hiding-in-plain-sight/>.
- [4] Jason Mick, "HVAC Firm at Center of Target Data Breach Also Counts Wal-Mart, Costco as Customers," *Daily Tech*, February 5, 2014, <http://www.dailytech.com/HVAC+Firm+at+Center+of+Target+Data+Breach+Also+Counts+Walmart+Costco+as+Customers/article34278.htm>.
- [5] Elizabeth Kreft, "How One HVAC Worker May Have Led to the Entire Target Data Breach," *The Blaze*, February 6, 2014, <http://www.theblaze.com/stories/2014/02/06/how-one-hvac-worker-may-have-caused-the-entire-target-data-breach/>.
- [6] Jia Lynn Yang and Amrita Jayakumar, "Target says up to 70 million more customers were hit by December data breach," *The Washington Post*, January 10, 2014, [http://www.washingtonpost.com/business/economy/target-says-70-million-customers-were-hit-by-dec-data-breach-more-than-first-reported/2014/01/10/0ada1026-79fe-11e3-8963-b4b654bcc9b2\\_story.html](http://www.washingtonpost.com/business/economy/target-says-70-million-customers-were-hit-by-dec-data-breach-more-than-first-reported/2014/01/10/0ada1026-79fe-11e3-8963-b4b654bcc9b2_story.html).
- [7] Information is Beautiful Website, July 1, 2014, an interactive visual display of data breaches, <http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/>.
- [8] Information is Beautiful Dataset, July 1, 2014, a spreadsheet that lists many data breaches that occurred prior to the Target Data Breach, <https://docs.google.com/spreadsheets/cc?key=0Aqe2P9sYhZ2ndFpGb0pHeEdKVndwTHFYt3BHS0dLN1E#gid=1>
- [9] "A 'Kill Chain' Analysis of the 2013 Target Data Breach," *Majority Staff Report for Chairman Rockefeller*, March 26, 2014, [http://www.commerce.senate.gov/public/?a=Files.Serve&File\\_id=24d3c229-4f2f-405d-b8db-a3a67f183883](http://www.commerce.senate.gov/public/?a=Files.Serve&File_id=24d3c229-4f2f-405d-b8db-a3a67f183883).
- [10] Payment Card Industry (PCI) Data Security Standard, version 2.0, October 2010, [https://www.pcisecuritystandards.org/documents/pci\\_dss\\_v2.pdf](https://www.pcisecuritystandards.org/documents/pci_dss_v2.pdf).
- [11] Michael Riley, Ben Elgin, Dune Lawrence, and Carol Matlack, "Missed Alarms and 40 Million Stolen Credit Card Numbers: How Target Blew It," *Bloomberg Businessweek*, March 13, 2014, <http://www.businessweek.com/articles/2014-03-13/target-missed-alarms-in-epic-hack-of-credit-card-data>. A video presenting this information can be viewed at <http://www.cbsnews.com/news/target-ignored-systems-hacking-warnings-report-says/>.
- [12] Elizabeth A. Harris, "Target Executive Resigns After Breach," *New York Times*, March 5, 2014, [http://www.nytimes.com/2014/03/06/business/a-top-target-executive-resigns.html?\\_r=0](http://www.nytimes.com/2014/03/06/business/a-top-target-executive-resigns.html?_r=0).
- [13] Elizabeth A. Harris, "Faltering Target Parts Ways With Chief," *New York Times*, May 5, 2014, <http://www.nytimes.com/2014/05/06/business/target-chief-executive-resigns.html>.
- [14] Megan Geuss, "Chip-based credit cards are a decade old; why doesn't the US rely on them yet?," *Ars Technica*, August 2, 2014, <http://arstechnica.com/business/2014/08/chip-based-credit-cards-are-a-decade-old-why-doesnt-the-us-rely-on-them-yet/>
- [15] Eric M. Hutchins, Michael J. Cloppert, and Rohan M. Amin, "Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains," *Lockheed Martin White Paper*, 2011, <http://www.lockheedmartin.com/content/dam/lockheed/data/corporate/documents/LM-White-Paper-Intel-Driven-Defense.pdf>.
- [16] Brian Krebs, "Email Attack on Vendor Set Up Breach at Target," *Krebs on Security*, February 12, 2014, <http://krebsonsecurity.com/2014/02/email-attack-on-vendor-set-up-breach-at-target/>.
- [17] Target's Partners Online Website, [https://wamlogin.partnersonline.com/securitybrokerage/pub/login.htm?TYPE=33554433&REALMOID=06-0fb16762-e63c-4dfe-a532-445551a2cc51&GUID=&SMAUTHREASON=0&METHOD=GET&SMAGENTNAME=\\\$SM\\\$cc00Zase4DnYh8i1ouaTStD0m1cY5nFYdGTFhrO4YJmPrPT7LYI0X0\%2bdoEIocNJ&TARGET=\\\$SM\\\$https\%3a\%2f\%2fwww\%2epartnersonline\%2ecom\%2f](https://wamlogin.partnersonline.com/securitybrokerage/pub/login.htm?TYPE=33554433&REALMOID=06-0fb16762-e63c-4dfe-a532-445551a2cc51&GUID=&SMAUTHREASON=0&METHOD=GET&SMAGENTNAME=\$SM\$cc00Zase4DnYh8i1ouaTStD0m1cY5nFYdGTFhrO4YJmPrPT7LYI0X0\%2bdoEIocNJ&TARGET=\$SM\$https\%3a\%2f\%2fwww\%2epartnersonline\%2ecom\%2f)
- [18] Target's Property Development Website, [https://pdzone.target.com/portal-target/templates/html/login\\_new.jsp?TYPE=33554433&REALMOID=06-f980003e-9313-4487-9ac0-98f792cd3f2f&GUID=\&SMAUTHREASON=0&METHOD=GET&SMAGENTNAME=-SM-pE3oyZbm8QUgajXyF0U\%2bN90sEDwYJr1Uba9SdRrCZO15LpIIHPKsYwdcu8cz5\%2f&TARGET=-SM-HTTP\%3a\%2f\%2fpdzone\%2etarget\%2ecom\%2fportal--target\%2ftemplates\%2fhtml\%2fexternal\\_content\\_display\%2ejsp\%3fcontentid\%3dPRD02--035451](https://pdzone.target.com/portal-target/templates/html/login_new.jsp?TYPE=33554433&REALMOID=06-f980003e-9313-4487-9ac0-98f792cd3f2f&GUID=\&SMAUTHREASON=0&METHOD=GET&SMAGENTNAME=-SM-pE3oyZbm8QUgajXyF0U\%2bN90sEDwYJr1Uba9SdRrCZO15LpIIHPKsYwdcu8cz5\%2f&TARGET=-SM-HTTP\%3a\%2f\%2fpdzone\%2etarget\%2ecom\%2fportal--target\%2ftemplates\%2fhtml\%2fexternal_content_display\%2ejsp\%3fcontentid\%3dPRD02--035451).
- [19] Visa Data Security Alert, "Preventing Memory-Parsing Malware Attacks on Grocery Merchants," April 11, 2013, <http://usa.visa.com/download/merchants/alert-prevent-grocer-malware-attacks-04112013.pdf>.
- [20] Visa Data Security Alert, "Retail Merchants Targeted by Memory-Parsing Malware - UPDATE," August, 2013, [http://usa.visa.com/download/merchants/Bulletin\\_Memory\\_Parser\\_Update\\_082013.pdf](http://usa.visa.com/download/merchants/Bulletin_Memory_Parser_Update_082013.pdf).
- [21] Jim Turley, "Motoring with microprocessors," August 11, 2003, <http://www.embedded.com/electronics-blogs/significant-bits/4024611/Motoring-with-microprocessors>.
- [22] Wikipedia, "Connected car," August 15, 2014, [http://en.wikipedia.org/wiki/Connected\\_car](http://en.wikipedia.org/wiki/Connected_car).
- [23] Charlie Osborne, "Verizon on Internet of Things, the connected car: Location is key," *ZDnet*, July 22, 2014, <http://www.zdnet.com/verizon-on-internet-of-things-the-connected-car-location-is-key-7000031860/>
- [24] Gartner Press Release, "Gartner Says the Internet of Things Installed Base Will Grow to 26 Billion Units By 2020," December 12, 2013, <http://www.gartner.com/newsroom/id/2636073>.
- [25] ABI Research Press Release, "More Than 30 Billion Devices Will Wirelessly Connect to the Internet of Everything in 2020," May 9, 2013, <https://www.abiresearch.com/press/more-than-30-billion-devices-will-wirelessly-conne>.
- [26] Steve Durbin, "Security Think Tank: ISF's top security threats for 2014," *ComputerWeekly.com*, <http://www.computerweekly.com/opinion/Security-Think-Tank-ISFs-top-security-threats-for-2014>.
- [27] Norse Blog, "Threat Thursday: Compromised Internet Connected Devices on Your Network," December 12, 2013, <http://www.norsecorp.com/blog-thursday-devices-131212.html>.
- [28] HP Press Release, "HP Study Reveals 70 Percent of Internet of Things Devices Vulnerable to Attack," July 29, 2014, <http://h30499.www3.hp.com/t5/Fortify-Application-Security/HP-Study-Reveals-70-Percent-of-Internet-of-Things-Devices/ba-p/6556284#U-412vldXNk>.
- [29] HP Study, "Internet of Things Research Study," [http://fortifyprotect.com/HP\\_IoT\\_Research\\_Study.pdf](http://fortifyprotect.com/HP_IoT_Research_Study.pdf).
- [30] Mohana Ravindranath, "Analyst: In 'Internet of Things,' security often takes a backseat to innovation," *The Washington Post*, November 21, 2013, [http://www.washingtonpost.com/business/on-it/analyst-in-internet-of-things-security-often-takes-a-backseat-to-innovation/2013/11/21/b50db616-52ed-11e3-9e2c-e1d01116fd98\\_story.html](http://www.washingtonpost.com/business/on-it/analyst-in-internet-of-things-security-often-takes-a-backseat-to-innovation/2013/11/21/b50db616-52ed-11e3-9e2c-e1d01116fd98_story.html).
- [31] Riley Snyder, "Set-top box revenue grows to record \$20 billion," *Los Angeles Times*, July 16, 2014, <http://www.latimes.com/business/technology/la-fi-tn-set-top-box-sales-20140716-story.html>.
- [32] Kurt Marko, "How the Internet of Things Will Change Your Business," *Information Week*, December 31, 2013, [http://reports.informationweek.com/abstract/81/11996/Business-Intelligence-and-Information-Management/How-the-Internet-of-Things-Will-Change-Your-Business.html?cid=smartbox\\_techweb\\_analytics\\_7.300001221](http://reports.informationweek.com/abstract/81/11996/Business-Intelligence-and-Information-Management/How-the-Internet-of-Things-Will-Change-Your-Business.html?cid=smartbox_techweb_analytics_7.300001221).
- [33] 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*, July 18-21, 2011, Las Vegas, Nevada, USA, pp. 507-511, <http://worldcomp-proceedings.com/proc/p2011/SAM5059.pdf>.



## **SESSION**

# **LATE BREAKING PAPERS AND POSITION PAPERS: security systems and applications**

## **Chair(s)**

**Prof. Hamid R. Arabnia**



# An Optimized Iris Recognition System for Multi-level Security Applications

S. Soviany<sup>1</sup>, C. Soviany<sup>2</sup>, S.Puscoci<sup>1</sup>

<sup>1</sup>T.C.T. Department, National Communication Research Institute (I.N.S.C.C), Bucharest, Romania

<sup>2</sup>IDES Technologies, Bruxelles, Belgium

**Abstract** - *This paper proposes an optimized iris recognition system which is suitable for client applications with several security levels. The innovations are the low dimensionality of the iris feature space and the hierarchical classification-based identification method; in this approach the biometric samples are processed in 3 stages to provide identification final decisions. The proposed design solution provides scalability, flexibility and a low computational complexity in biometric data processing. These benefits are important for costs optimization in security systems design.*

**Keywords:** rejection, identification, multi-level security

## 1 Introduction

Iris is a very accurate biometric because it provides a lot of discriminating features. Most of the actual iris-based biometric systems are based on Daugman algorithm, which is considered to be very accurate for both verification and identification [1],[2]. Despite of these advances there are still significant challenges; this is especially true while addressing large-scale identification applications with several security levels. A huge searching space for 1-to-many matching requires a significant processing and time complexity for feature extraction, encoding and matching operations. Another challenge relates to the identification accuracy while using a reduced iris features set; there are applications requiring optimal trade-offs between the processing complexity and the achieved accuracy, but with various security levels.

The proposed iris recognition system is based on a regional and textural approach for feature extraction and a hierarchical design for data classification.

The feature extraction uses co-occurrence matrices, ensuring an easy way to adjust the features number. This low complexity approach is different from most of the actual solutions based on Wavelet transforms and Gabor filters, which are not always optimal for large-scale identification because they provide a big number of features, with a significant impact on the further processing in classification stage [3],[4].

The biometric data classification decisions are provided with a hierarchical algorithm; this classifier with 3 stages is designed according to the application-defined security levels.

The essential of our contribution is given by the classification approach, with a 3-level hierarchical model according to the end-users authorization degrees. Each of these levels defines a security level. Within this

hierarchical classifier, the first 2 stages decisions are given by detectors; a detector is an one-class classifier with target vs. non-target outputs. The third classification stage discriminates among the other enrolled identities and it is called only if the detection stages failed on their target identities. The detectors application in biometric recognition is reasoned by their design principle; a detector is only trained on a single target class and therefore provides some computational complexity saving because the model should not compute all of its basic parameters for all the classes (enrolled identities). This approach provides a significant advantage in processing time and is suitable for identification systems design if the goal is to achieve an optimized identification process for applications with several security levels; the applied classification method provides an accurate identification just for the most authorized users of the protected resource, which define the target identity for the detection stages of the biometric data hierarchical classifier.

The remainder of this paper is structured as follows. Section II presents some related works in this area. Section III presents the iris recognition system architecture. Section IV presents the whole identification method with its main stages. Section V presents and discusses some experimental results. Section VI concludes the paper and proposes future research directions to improve this hierarchical identification method for various security requirements and several biometrics.

## 2 Related Works

Many of the well-known approaches in iris recognition systems are derived from the Daugman algorithms, in which the pre-processing stage performs segmentation and normalization of the eye original images while applying Daugman's integro-differential operator, Hough Transform and Gabor filters to extract and optimally encode the iris features [1],[2]; this encoding generates the required biometric template for the classification stage which is based on Hamming distance evaluation for matching. Although the published works (including Wildes and Boles approaches in [5] and [6], respectively) show very accurate results in individual recognition, these results are achieved on high-quality datasets, and without independent testing activities. Another issue is that the generated features number is still too high in order to provide a reasonable complexity for mobile applications in which the complexity issue is still a challenge; this is because the end-users mobile terminals have limited resources in terms of processing capabilities

and requirements for the biometric sensor should be suitable.

Given the huge potential of iris recognition for various security applications, other researchers recently approached this biometric with other related methods in feature encoding and data classification. They started their research just from the same dimensionality issue of the Daugman-based iris recognition systems; the standard size of the feature vector (2048 bits in the Daugman-based approaches) is considered to be not very suitable for storage and processing [3],[4] especially while having real-time applications or mobile devices with a lot of processing-related constraints. Also the transmission bandwidth is another reason for the actual focus on feature space dimensionality reduction.

In [3], a wavelet transform-based algorithm for iris recognition was implemented using MATLAB wavelet toolbox. The authors achieved significant performance improvement for feature vector lengths in range of 120 to 480 bits, therefore for lower feature space dimensionalities. They applied an image segmentation method based on Hough Transform after the image conversion to 8-bit grey scale and size reduction to 225 x 300 pixels. However, the Hough Transform is computationally intensive and therefore it is not suitable for real-time or eventually mobile applications with resources constraints. The authors applied Hamming distance for iris templates matching, but this approach is not very suitable for multi-level security applications because the various security levels are typically defined according to particular requirements of the end-users, with various thresholds for individuals identification.

Another recent approach in iris recognition systems design and implementation is given in [4]. The authors proposed an algorithm for pupil detection and feature extraction algorithm in iris recognition. They focused on get some efficiency in terms of computational and execution performance by applying a scanning method for pupil detection; the feature extractor implementation is done with five level decomposition techniques, using haar wavelet; the feature vector size is 90 in their approach. This feature space size provides advantages in its computational simplicity and speed, especially while performing the iris templates matching with Hamming distance. However, the achieved results are given on CASIA database, with high quality images. Also the focus on FAR, FRR and EER as performance measures is sometimes not very convenient for applications in which the focus is towards the identification accuracy; this means that there are a lot of applications in which the designed system has to guess who a real person is, without any additional identifier.

### 3 The System Architecture

The biometric security system architecture is depicted in fig. 1.

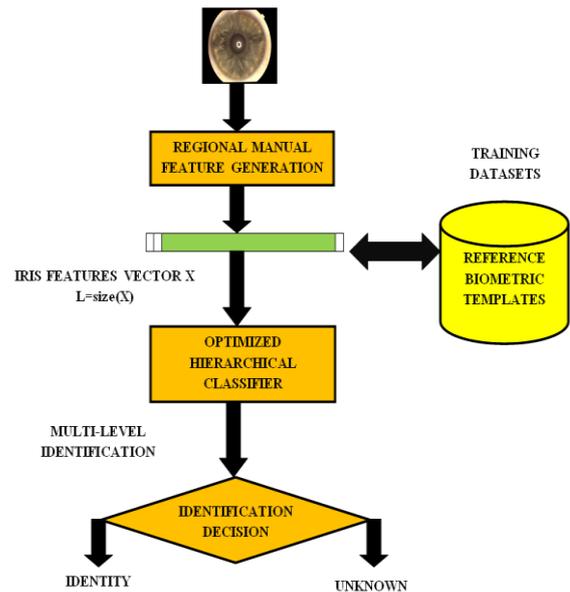


Figure 1: The iris –based biometric system architecture

The system functional architecture includes the modules performing the typical biometric recognition process: data introduction (either for enrollment and authentication), iris images pre-processing for feature generation, data classification and final identification decision provided to the end-user application. The system is used in identification mode, therefore it is trained to accurately guess who a real person is, without any additional identifier such as a username.

The system is designed for multi-level security application with an *optimal trade-off between the complexity and accuracy*. A *multi-level security application* defines several security layers depending on the end-users authorization degrees and the resources vulnerabilities. The reference database contains iris samples which are provided from 80 individuals. Each of them provided 5 iris images from the right eyes of the individuals.

### 4 The Iris Recognition Method

The individuals identification is performed in following stages:

- Pre-processing stage for feature generation;
- Data classification for the end-user recognition.

#### 4.1 Pre-processing stage for feature Generation

The feature generation stage is depicted in fig. 2.

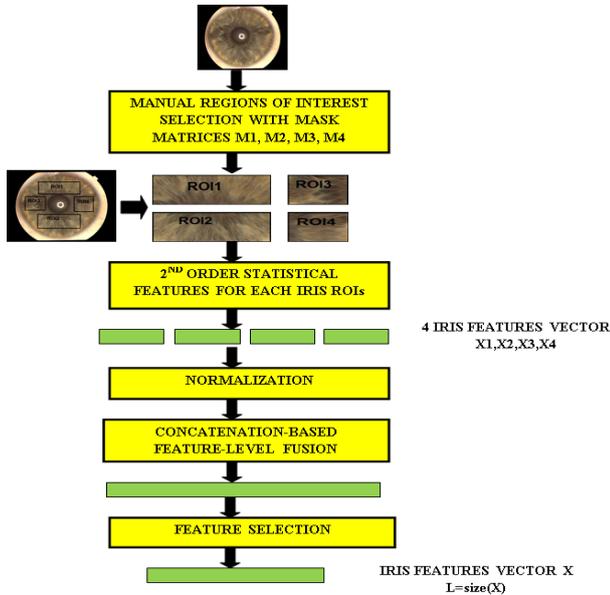


Figure 2: The feature generation stage

First we have to convert the input iris color images to black-and-white (gray-scale) images using the procedure given in [7]; this is useful just for iris texture analysis. Although there are other methods for this image conversion, the reason for this procedure applying is its simplicity; it is suitable especially for iris textural analysis, as stated in [7].

Then we perform the feature generation in the following sub steps, with the algorithm given in fig. 2:

- 1: *The manual selection of 4 regions of interests (ROIs)* within the input iris image;
- 2: *The 2<sup>nd</sup> order statistical features computation* from each ROI;
- 3: *The features normalization* in order to achieve the same numerical range for all the feature vectors components;
- 4: *The concatenation-based feature-level fusion* in order to get a single feature vectors;
- 5: *The feature selection* in order to provide the optimal feature set with reduced dimensionality.

This procedure is different from Daugman's method. As much as our target is to achieve a **small dimensional iris features space**, this computational complexity reduction allows to train the classifiers with small number of samples per class, according to [8].

In the 1<sup>st</sup> substep (*manual ROI selection*), we define 4 rectangular regions within the input iris image:  $ROI_k, k = \overline{1,4}$ . On each of the defined ROIs we apply the corresponding mask matrices.

Then we compute *the 2<sup>nd</sup> order statistical features* for each of the 4 ROIs. These features are derived from **the co-occurrence matrix** and provides information about the relative positions of the several gray levels within the input image [8]. For each iris ROI image the feature vector is derived by converting the gray-level co-occurrence matrix to the vector representation [9]. Each co-occurrence matrix element allows to evaluate the probability of a certain gray-level for one pixel within the

input image while another displaced pixel has another gray level [10],[11],[12]. This approach allows to capture the structure of the underlying texture in the input image [10]. It also allows to easily exploit the texture just on the selected regions of interest [11]; we use the iris texture to get the optimally reduced features sets. The co-occurrence matrix is actually a 2D histogram which encodes structural information supporting derivation of informative data representation useful for texture classification problems [12]. The **co-occurrence matrix-based feature extractor** has the following parameters [11],[12]:

- *the number of gray-level bins (GLB)*. We apply a smaller value of GLB ( $GLB = 7$ ), ensuring a resulted co-occurrence matrix with many significant values and less null values;
- *the displacement distance*, the number of pixels between the pixel pairs which are used to fill the 2D histogram [12]. For our application,  $DISPL-DIST = 2$ ; this parameter value should not exceed a certain limit otherwise the number of pixel pairs will decrease and the amount of useful information will also decrease.

The reason for using the co-occurrence matrices-based feature extractor is related by its simplicity while comparing with the traditional methods with segmentation, normalization and wavelet-based feature encoding. Also these automatic feature extractor parameters could be easily adjusted in order to find out the optimal feature space dimensionality for the designed application purposes. Actually the co-occurrence matrix-based feature vector size  $X$  is [12]

$$size(X) = (GLB)^2 \quad (1)$$

Besides of this achievement, the co-occurrence matrix allows to compute the following additional 2<sup>nd</sup> order statistical features [8],[10]: angular second moment, contrast, inverse difference moment and entropy. Therefore, the 4 iris feature vectors sizes are:

$$L_k = size(X_k) = (GLB_k)^2 + 4, k = \overline{1,4} \quad (2)$$

In the 3<sup>rd</sup> sub step the components of all the 4 feature vectors are **normalized** leading to homogeneous-ranged values for all the extracted features. We apply a sigmoid function-based normalization technique which provides a common range of [0, 1] for the resulted feature values:

$$f(X_k) = \frac{1}{1 + \exp(-\alpha_k \cdot X_k - \beta_k)}, k = \overline{1,4} \quad (3)$$

The scaling ( $\alpha_k$ ) and offset ( $\beta_k$ ) coefficients are resulting from our experimental data and their optimal values are in the following ranges:

- for  $k = 1(ROI1)$ :  $\alpha_1 \in [1,1.5]$  and  $\beta_1 \in [2,2.5]$ ;
- for  $k = 2(ROI2)$ :  $\alpha_2 \in [1.5,2]$  and  $\beta_2 \in [1.5,2]$ ;
- for  $k = 3(ROI3)$ :  $\alpha_3 \in [2.5,3]$  and  $\beta_3 \in [1.25,1.75]$ ;
- for  $k = 4(ROI4)$ :  $\alpha_4 \in [2.75,3.25]$  and  $\beta_4 \in [1,1.5]$

The 4 iris feature vectors are then *concatenated* resulting in a single vector with its size given by  $L_0 = size(X_0 = [X_1 | X_2 | X_3 | X_4]) = 4 \times L_k, k = \overline{1,4}$  (4)

This iris feature space size is still too high for our optimization purposes (complexity vs. performance trade-off). Therefore the final sub step in feature generation process should reduce this high dimensionality through a suitable *feature selection* procedure. On the available datasets we evaluate the following non-optimal and non-exhaustive feature selection methods: *forward-searching feature selection*, *backward-searching feature selection*, *individual ranking feature selection*, *random feature selection* and *floating-search feature selection*. The evaluation is done using *1-NN* (nearest-neighbor rule) *classification error rate* as a performance measure for features because this rule is asymptotically at most twice as bad as the Bayes rule [13]; actually it is a typically feature evaluation criterion in many pattern recognition systems. Finally we choose the *individual ranking for feature selection* because of its high speed, therefore providing a significant improvement in execution time; on the available experimental data this method provided the best time in optimal feature selection, which was almost 4 times higher than for the sequential searching methods (forward, backward and floating) and also 3.5 times higher than for random searching-based feature selection. The resulted **optimal iris textural feature number** was **12**, which is obviously better than the value resulted from the concatenation-based feature-level fusion.

### 4.2 Processing Stage for data classification

The identification process is handled as a multi-class problem in which all the iris samples belonging to one person are defining a separate class [14],[15]. The iris data classifier output supports decisions for individuals identification; this is why the classifier should be trained for every person recognition. On the other hand the target application has several security levels and therefore the classification algorithm should be hierarchical; in this approach the identification decision results after several processing stages. Each of these stages relates to a **security level**. Figure 3 depicts the classification stages in our hierarchical approach for the **multi-level security application**.

There are 3 stages in this process. The first 2 stages perform the **detection** for the target identities, and the 3<sup>rd</sup> stage performs the **discrimination** among the other  $N - 5$  enrolled identities ( $N = 80$ ); after the detection stages we already have decisions for 5 enrolled identities. A **biometric detector** is a classifier which is trained for only one identity recognition, providing decisions for target or non-target identification; the non-target identification decision relates to all the other enrolled identities.

*The 1<sup>st</sup> security level* accurately identifies the 2 most authorized users. The identification decisions result from 2 biometric detectors, each of them being trained for its own target identity. If both detectors fail on their targets,

then the process go to *the 2<sup>nd</sup> security level*, in which the target identities are provided from 3 users with a lower

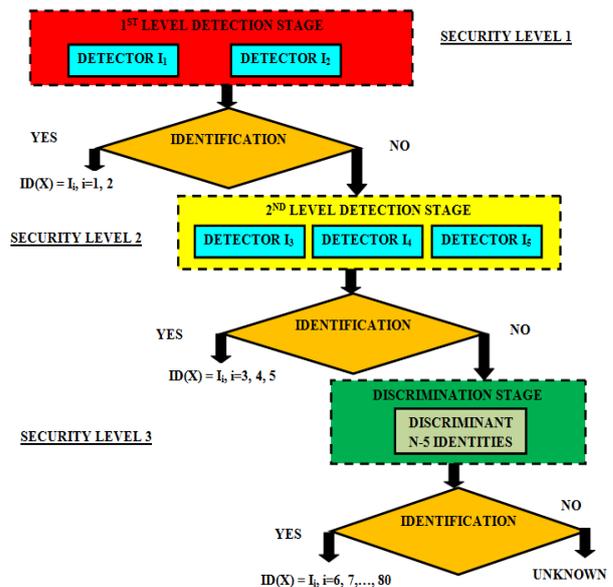


Figure 3: The classification stage

authorization degree; this classification stage performs detection for the 3 target identities. If the 2<sup>nd</sup> classification stage fails on the target identities recognition, the hierarchical classifier proceeds to the 3<sup>rd</sup> stage according to *the 3<sup>rd</sup> security level*; it should identify the end-users with the most restrictive access permissions. This stage performs the discrimination among the other previously non-recognized identities.

The optimal models for the **biometric detectors** are based on Gaussian mixtures, according to the experimental data. The differences are only related to the number of Gaussian components and iterations, respectively. So each biometric detector model is

$$p(X | I_i) = \sum_{j_{i1}=1}^{N_{I_i}} p(X | j_{i1}) \cdot P_{j_{i1}}, i = \overline{1,5} \quad (5)$$

in which:

$I_i$  is the target identity label for the designed biometric detector. The first 2 enrolled identities are the most authorized and therefore they should be recognized in the 1<sup>st</sup> security level; the other 3 individuals are identified in the 2<sup>nd</sup> detection stage;

$X$  is the iris pattern which is applied to the detector input. It is represented as a feature vector;

the coefficients  $P_{j_{i1}}$  are the mixture weights;

$p(X | j_{i1})$  are the Gaussian components.

$N_{I_i}$  is the mixture components number

The unknown parameters are computed using EM (Expectation Maximization) Algorithm [8], [13].

The output decisions for both detection stages are based on the following Bayesian function evaluation [16]:

$$g(X) = \frac{n_{Z,i}}{n_Z} \cdot \sum_{j_{i1}=1}^{N_{I_i}} p(X | j_{i1}) \cdot P_{j_{i1}}, i = \overline{1,5} \quad (6)$$

in which:

$n_{z,i}$  is the number of the training iris samples belonging to the target enrolled identity  $I_i$ ;

$n_z$  is the overall number of the training iris samples for all enrolled identities.

Therefore the identification decision rule is:

$$ID(X) = \begin{cases} I_i, & \text{if } g(X) \geq \theta_i \\ non - I_i, & \text{otherwise} \end{cases}, i = \overline{1,5} \quad (7)$$

where:

$ID(X)$  is the target or non-target detector decision for the most important end-users identification;

$\theta_i$  is the threshold for the identity  $I_i$  recognition.

In the 3<sup>rd</sup> stage the hierarchical classifier should only **discriminate** among the other  $N - 5$  enrolled persons, because the previous 2 detection stages already failed on their target identities recognition. In our application the resulted iris feature space (with only 12 features) allows to train the discriminant models with a small number of samples per identity and to select lower complexity classifiers. Another criterion for the classifiers selection is their learning curves smoothness; a smooth learning curve reveals a more predictable classifier behavior on the available data. Based on these criteria we select the following classifiers: Near Mean Classifier (NMC), Linear Discriminant Classifier (LDC), Fisher Classifier and Quadratic Discriminant Classifier (QDC). Figure 4 depicts their corresponding learning curves on the available iris samples dataset with reduced feature.

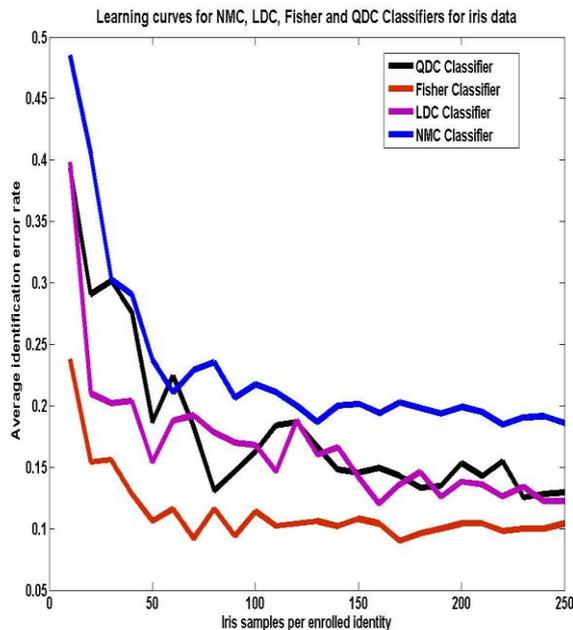


Figure 4: Learning curves for NMC, LDC, Fisher and QDC classifiers on iris data with reduced feature space size

The performance measure is the average per-identity classification error rate. Also the training subset size is ranging from 10 to 250 samples per identity.

One can see that Fisher and LDC classifiers have the best behavior especially for training dataset sizes not exceeding 50 samples per enrolled identity. Actually on the whole range of training set sizes the lowest average identification error rate is provided by Fisher classifier; this behavior is achieved on the available biometric data which we used for the iris recognition system design and training. We applied a 5-fold cross validation for the discriminants and the depicted results are averaged on 20 experiments.

According to this behavior which resulted from the learning curves analysis, we should use Fisher classifier for the discriminant stage of the designed iris recognition system.

## 5 Experimental Results

For the designed system performance evaluation we use the average **True Positive Rate** vs. **rejection threshold** representation. We apply several classification rejection thresholds to compare the achieved TPRs and to find out the optimal threshold. In biometric applications the rejection threshold allows to adjust the individuals recognition accuracy while considering the input biometric samples quality.

The experimental data are provided from 80 end-users of a medical database; the security application is designed for a telemedicine system. We perform 20 experiments; the performance measure (TPR) is averaged per enrolled identity and over all these experiments. This measure applies per class applied and therefore we should average the achieved values over all the enrolled individuals. The classifiers are only trained with 50 samples per identity, both in detection and discrimination stages.

Figure 5 shows the average TPRs vs. rejection threshold trade-offs while considering the following 3 cases:

- **A:** without detection stages, with only discrimination among 80 enrolled identities (one security level);
- **B:** with a single detection stage (for 5 target identities) and a discriminant stage (2 security levels);
- **C:** with 2 detection stages and a discriminant stage (3 security levels).

The achieved results for different rejection thresholds are given in table 1. One can see that for a significant range of rejection thresholds the best recognition performances are provided for the multi-stage hierarchical classifier with 2 detection initial stages (operational case C). However, in biometric applications the typical classification rejection thresholds should be around 5%, because this is the typical fraction of individuals who are

TABLE 1: THE IRIS RECOGNITION SYSTEM PERFORMANCES				
Operational Cases	Averaged TPRs for different rejection thresholds $\theta$			
	0%	5%	10%	15%
A	0.761	0.760	0.759	0.758
B	0.850	0.805	0.785	0.760
C	0.950	0.910	0.890	0.880

not able to provide optimal input

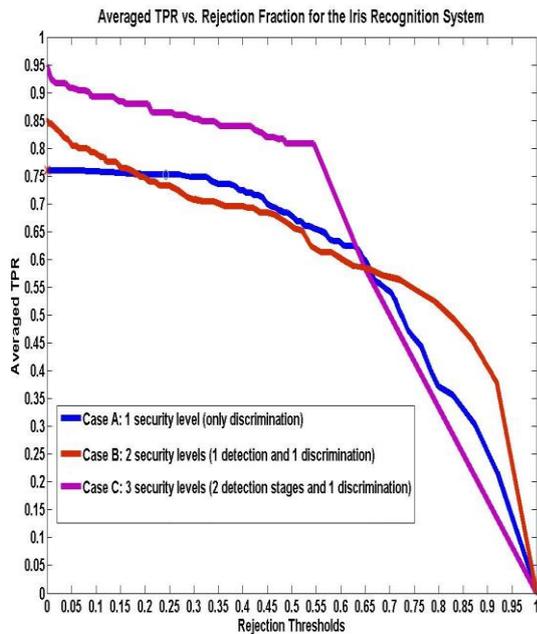


Figure 5: Averaged TPRs vs. rejection thresholds

biometric samples. On the other hand this threshold should not exceed a certain limit otherwise the system performance for individuals recognition will faster decrease.

The iris-based recognition system is optimized by fixing the classifiers operating points (given according to the end-user application specific thresholding) allowing to achieve a desired average TPR. Also this optimization allows to fix the suitable rejection fraction for different security levels, according to the performance requirements of the application. This is another advantage of the proposed approach in biometric systems design and optimization for multi-level security applications.

From our experiments one can see that for an application with 3 security levels (2 detections and one discrimination), the optimal rejection threshold is in range of [0; 5] %, while the identification performance decreasing is still limited; this range provides an average TPRs which still exceeds 0.925. For an application with only one security level (i.e. without detection and with

only discrimination), the allowed rejection range could be larger; however the individuals identification performance is obviously lower than for the multi-level cases, especially for rejection thresholds between 0 and 15%.

## 6 Conclusions

The security system design and optimization is still a big challenge especially for biometrics. This is because of the high complexity in biometric data processing and especially for applications with several security levels. Also the identification still remains a bottleneck because of the huge searching space for large-scale identification systems.

In our paper we designed an optimized iris-based recognition system for a multi-level security application. This system operates on a reduced feature space dimensionality; this is a significant improvement relative to the actual iris-based biometric systems. Another advantage is that the proposed approach allows to perform optimizations according to the security levels number; these security levels relate to the users authorization degrees and the protected resources vulnerabilities; the goal of these differentiated optimizations is to improve the performance/cost ratio in biometric security systems design.

Despite of this multi-level security approach, further theoretical and experimental research should be conducted on extended biometric datasets with various feature space sizes and several feature extraction methods, also to get a quantitative view concerning the influence of the classifier hierarchy depth on the execution time; this is especially important in order to optimize the large-scale identification systems.

## 7 References

- [1] Daugman J.: "How iris recognition works" (PDF). IEEE Transactions on Circuits and Systems for Video Technology 14 (1): 21–30, January 2004.
- [2] Verma P., Dubey M., Verma P., Basu S.: "Daugman's Algorithm Method for Iris Recognition-A Biometric Approach", International Journal of Emerging Technology and Advanced Engineering (IJETA), Vol. 2, Issue 6, June 2012
- [3] Panganiban A., Linsangan N., Caluyo F.: "Implementation of Wavelet Transform-Based Algorithm for Iris Recognition System", International Journal of Information and Electronics Engineering, Vol. 2, No. 3, May 2012
- [4] Roselin V., Waghmare L.: "Pupil detection and feature extraction algorithm for Iris recognition", AMO-Advanced Modeling and Optimization, Vol. 15, No. 2, 2013
- [5] R. Wildes, J. Asmuth, G. Green, S. Hsu, R. Kolczynski, J. Matey, S. McBride. A system for automated iris recognition. Proceedings IEEE Workshop

on Applications of Computer Vision, Sarasota, FL, pp. 121-128, 1994

[6] W. Boles, B. Boashash. A human identification technique using images of the iris and wavelet transform. IEEE Transactions on Signal Processing, Vol. 46, No. 4, 1998

[7] Bhattacharyya D., Das P., Bandyopadhyay S.K., Kim T.: "IRIS Texture Analysis and Feature Extraction for Biometric Pattern Recognition", International Journal of Database Theory and Application, vol. 1, nr. 1, pp. 53-60, decembrie 2008

[8] Theodoridis S., Koutroumbas K.: "Pattern Recognition" 4th edition, Academic Press Elsevier, 2009

[9] Eleyan A., Demirel H.: "Co-occurrence matrix and its statistical features as a new approach for face recognition", Turk J Elec Eng & Comp Sci, Vol.19, Nr.1, 2011

[10] Zucker S.W., Terzopoulos D.: "Finding Structure in Co-Occurrence Matrices for Texture Analysis", Computer Graphics and Image Processing nr. 12, 1980

[11] Bino S. V, A. Unnikrishnan and Kannan B.: "Gray level Co-Occurrence Matrices: Generalisation and some new features", International Journal of Computer Science, Engineering and Information Technology (IJCEIT), Vol.2, No.2, April 2012

[12] \*\*\* "PerClass Documentation: Feature extraction from images", 2014

[13] Devroye L., Györfy L., Lugosi G.: "A Probabilistic Theory of Pattern Recognition", Springer, 1997

[14] Soviany S., Puşcoci S., Jurian M.: "A Detector-Discriminant Model for Biometric Security Systems", International Conference on Information Technology and Computer Networks (ITCN 2012), Viena, 10-12 november 2012

[15] Soviany S., Puşcoci S., Soviany C.: „A Multi-Detector Security Architecture with Local Feature-level Fusion for Multimodal Biometrics”, Journal of Communication and Computer (JCC), no. 9/2013, David Publishing Company

[16] Zhang David, Song Fengxi, Xu Yong, Liang Zhizhen: "Advanced Pattern Recognition Technologies with Applications to Biometrics", Medical Information Science Reference, IGI Global, 2009

# A Distributable Hybrid Intrusion Detection System for Securing Wireless Networks

David Tahmoush<sup>1</sup>

<sup>1</sup>University of Maryland, University College, Maryland, USA

**Abstract** - We developed a hybrid design to a NIDS that enables the seamless insertion of a machine learning component into a signature NIDS system that significantly improves throughput as well as captures additional networking traffic that is similar to known attack traffic. The throughput improvement by incorporating a normalcy classifier is significant, estimated to be the inverse of the false alarm rate which can easily net a factor of 1000. However, this can be diminished by updates that can trigger a retraining of the normalcy classifier. The addition of a normalcy classifier front-end also makes the system more highly scalable and distributable than the signature-based NIDS. The new hybrid design also allows distributed updates and retraining of the normalcy classifier to stay up-to-date with current threats, and makes a number of important performance and quality guarantees. The distributable hybrid implementation is very useful for securing wireless networks with multiple access points.

This system design also has the capability to recognize new attacks that are similar to known attack signatures. The hybrid design also can provide significant information on new attack traffic. By finding the signature of suspicious traffic that is similar to the signature of a known attack, it can be isolated and analyzed as a potential variant of a known attack.

**Keywords:** IDS, hybrid

## 1 Introduction

Machine learning classifiers can be used to discover the patterns hidden within large data sets, and one of the largest datasets is the information being passed through a network every day. Many information technology applications have been proposed and also used to classify network traffic [1, 2, 3, 4, 5]. Intrusion detection systems (IDS) monitor the system or network events and detect violations or threats to computer security policies, acceptable use policies, or standard security practices [6], and are one of the most significant counter measures [7, 8, 9, 10] against security threats. Intrusions can be found using signature based detection of known threats, but there are also anomaly detectors. Signature based detectors look for specific log entries or a specific payload in a data packet known to be indicative of misuse.

The IDS monitors the network traffic from a system or through a network and looks for any abnormal behavior in the

network activity which indicates a possibility of unwanted and malicious network traffic and take appropriate action if such situation occurs. The IDS uses signature detection for specific known threats or anomaly detection for unknown threats to analyze the data. However, many unknown threats are merely updated versions of known threats. Since machine learning techniques can determine whether new threats are similar to known threats, there is the potential to combine anomaly detection with approximate signature detection. One of the most significant aspects of an IDS is the use of artificial intelligence [11] to train the IDS about possible threats. The Intrusion Detection can gather information about the various traffic patterns and rules can be formed based on these patterns, to distinguish between normal traffic and anomalous traffic in the network. Machine learning techniques have the ability to generalize from limited, noisy data that is not complete to broader categories on new data. This generalization capability provides the potential to recognize patterns similar to known patterns but not exactly matching. The IDS should ideally recognize not only previously observed attacks but also future attacks that have not yet been seen [12].

## 2 Machine Learning in Intrusion Detection Systems

Some significant contributions to IDS have been made using Fuzzy Logic. Fuzzy inference combined with artificial neural networks were used for real time traffic analysis by building a signature pattern database using protocol analysis and neuro-fuzzy learning techniques [13]. Fuzzy rule-based classifiers for IDS were modeled [14]. A fuzzy intrusion recognition engine (FIRE) used Fuzzy Logic and data mining techniques to produce fuzzy sets based on the input traffic data to detect security threats [15]. Association-based classification of normal and anomalous attacks was performed on the basis of a compatibility threshold [16]. Association rules along with data mining techniques and classification was used on suspicious events in real-time [17]. Fuzzy rules gave the best detection rate when compared to linear generic programming, decision trees, and support vector machines on the DARPA 1998 dataset [18]. Fuzzy logic with an expert system performed better than 91.5% detection rate over all attack types with a reduced complexity over traditional fuzzy number ranking techniques [19]. Fuzzy adaptive resonance theory have also been used to implement network IDS [20] as well as fuzzy rules [21, 22].

A lot of work has been done on IDS using genetic algorithms. Genetic algorithms using both temporal and spatial information of the network connection during the encoding phase were used to identify anomalous network behaviors [23]. Genetic algorithms were used to find the best possible fuzzy function and select the most significant network features [24]. Genetic programming was used to derive classification rules with traffic data on the network [25]. Multiple agent technology with genetic programming was used to detect anomalies in the network [26]. A combination of information theory to filter the traffic data with genetic algorithms was used to detect anomalous behaviors in the network with reduced complexity [27].

Artificial neural networks are a popular machine learning technique, and it has been applied to IDS. A hybrid neural network was proposed using a combination of Self-Organizing Map (SOM) and Resilient Back-Propagation Neural Network (BPNN) [28]. Another hybrid system using a BPNN and a C4.5 Decision Tree was built [29] which showed that the certain network attack types could not be detected without a hybrid system. A multi-layer artificial neural network was used to classify network activity [30]. A multi-classification IDS system was built that showed a higher detection rate in each classification category than when only a single class was used to classify all non-normal data [31].

Additional approaches have included graphlets [32], decision trees [33], clustering [34], and deviation from normal traffic [35].

### 3 Data

Many researchers have proposed IDS classification algorithms based on machine learning techniques, but they have used older datasets from DARPA and others to evaluate their approaches. This dataset used is a network packet dataset consisting of normal network activity as well as many network attack types. The dataset is based on the DARPA98 dataset from MIT Lincoln laboratory, which provides answer class (labeled data) for evaluation of intrusion detection [36]. This dataset was created in 1998 and lacks of many current attack types. This paper uses current signatures from an IDS as an oracle for machine learning to form a new, faster IDS with the generalization capabilities of a machine learning built in. This avoids the work of manually labeling a dataset and provides more current signature information, but the quality of the initial IDS information determines the baseline for the new artificial intelligence based IDS.

### 4 Real-Time Intrusion Detection

A system that can detect network intrusion while an attack is occurring is called a real-time detection system. There are very few real-time network IDS approaches. A real-time IDS using Self-Organizing Maps (SOM) to detect normal network activity and differentiate it from a DoS attack was proposed [37]. A Bayesian classification model for

anomaly detection was also built [38]. A real-time IDS was built using two unsupervised neural network algorithms with a detection rate over 97%, separating normal traffic data from network attacks [39]. A real-time network IDS using fuzzy association rules could separate the normal network activity from network attacks [40]. A high-speed intrusion detection model using TCP/IP header information was built to detect denial of service (DoS) attacks [41].

One of the most widely used and well-known IDS is called SNORT, and it has become a standard in IDS [42]. SNORT is a commercial tool that does not use machine learning, basing its detection on regular expressions that match to known signatures of network attacks. Its attack signature rules are available only to their registered customers. The signature rules or patches have to be frequently updated and installed in order to detect current attack types or variations in known attack types.

Although some researchers are investigating real-time IDS with machine learning techniques, most of the work is based on accurate learning without good real-time performance measures and without good generalization capabilities. This paper presents a real-time hybrid design that can guarantee improved real-time performance with equivalent false alarm rates.

## 5 Advantages of a Machine Learning System

There are multiple advantages to a machine-learning based system over a signature-based system. A signature-based system needs to store attack signatures and download new signatures when they are updated, while a machine-learning system merely updates the weights on its classifier. A signature system can be difficult to parallelize with a shared signature database, while a machine-learned system can run multiple instances due to its lightweight nature. The speed of a machine learned system can be faster, and that advantage only increases with the growth in the size of the signature database to search over. The machine-learned system will have slightly more false positives and will not give detailed information about the true positives, so a signature IDS can be run on the output from the machine-learned system for labeling as well as false-positive reduction.

The primary advantage of a machine learning system over a signature system is the ability to generalize to new but similar data. This was the dream of machine learning with an IDS, that the IDS should ideally recognize not only previously observed attacks but also future attacks that have not yet been seen [12]. There are some systems that can generalize their detection well from learned attack patterns to new attack patterns [47], especially on probing attacks [48]. A machine learning system also has some ability to generalize to patterns not seen in the training data, and this was seen anecdotally in this project.

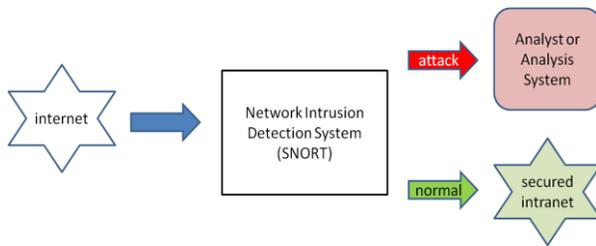


Figure 1. A standard setup for a network intrusion detection system.

## 6 Normalcy Classifier and a Hybrid Intrusion Detection System

A current network IDS setup is shown in Figure 1. Adding in a front-end with the capability to replicate the detections of a signature NIDS creates a hybrid system that can significantly improve the speed at equivalent false alarm rates but with a slightly higher false negative rate. For a hybrid system, the labeling and analysis of detection does not need to be implemented because a version of the signature

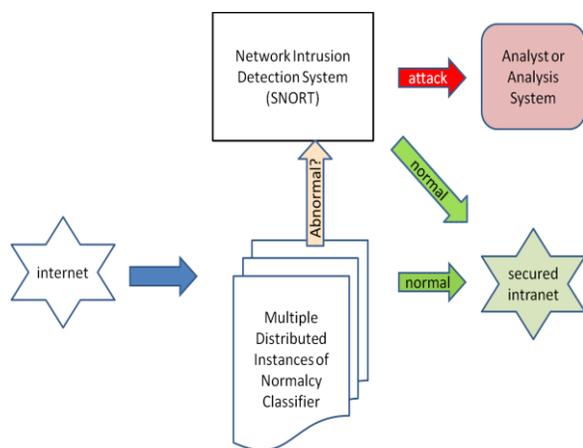


Figure 2. A hybrid setup for a network intrusion detection system with a distributed normalcy classifier. The network traffic considered abnormal by the normalcy classifier but normal by the signature system may contain new attack traffic that can also be analyzed.

NIDS should be run on the detections for labeling and analysis of the suspicious network traffic as shown in Figure 2. This hybrid system will outperform the signature NIDS as a standalone in speed since the high percentage of network traffic will be classified as normal and not sent to the labeler. It will also be more scalable, since additional normalcy classifiers can be run with significantly less overhead. The resulting system will produce the same level of labeling quality since the abnormal traffic would be routed through the signature component. The level of false alarms would not rise since the signature component would be run on the abnormal output from the normalcy classifier to reduce false alarms. The hybrid system would run faster, scale more easily, and

use far less resources than a series of signature NIDS instances. The cost is the slightly increased false negative rate caused by missed detections in the normalcy classifier. However, the abnormal output from the normalcy classifier may contain significant information about a new or unrecognized attack pattern. This output can be sent to an analyst or to an anomaly classifier.

To understand the improvement in speed and capabilities of a hybrid IDS, consider a signature based system with a number of packets that it can analyze per second. Adding in a normalcy classifier with a 2% false positive rate would improve the number of packets that could be classified by a factor of fifty. If you can achieve a 1% false positive rate, the improvement in packets per second jumps to a factor of 100. For a given false positive rate  $fp$ , the performance boost can be estimated at  $1/fp$ . This estimate neglects the overhead of running the normalcy classifier or classifiers, but that is typically negligible compared to the signature-based component and can be run in parallel. With a 0.1% false positive rate, the performance boost is a factor of one thousand. Unfortunately, no normalcy classifier can be perfect, so there will be a cost in the false negatives from the normalcy classifier that will be passed through into the network. Typically there is an inverse relationship between the false positives that would require signature processing and the false negatives which incurs risk to the network. This implies that an optimal performance can be achieved by varying the normalcy classifier to process all of the network data given the constraint provided by the signature section. However, this adds complexity to the normalcy classifier which would complicate the retraining.

Retraining is a significant issue in a hybrid system, since the normalcy classifier should be retrained every time the signature database is updated. The complexity of a deployable normalcy classifier can be limited by the allowable retraining time if there are not other workarounds while the normalcy classifier is retraining. In the case where the normalcy classifier is taken offline while retraining and the signature component runs without hybrid support, the performance gains can be eroded. Given a training downtime  $d$ , the expected performance boost drops by a similar factor of  $d$ , for example a downtime of 20% drops the performance boost by 20%. In selecting a normalcy classifier, the cost of this training downtime may be a significant consideration. However, updates are one large advantage of a hybrid system over developing a brand-new system. If the normalcy classifier is trained on the signature NIDS outputs, a new signature inserted into the system can trigger retraining of the classifier and redistribution of the training weights. This leaves the development of new signatures in the signature-based component of the hybrid system and then distributes the signatures to the normalcy classifier.

One large advantage of a hybrid system is the guarantee of no increase in false alarms. Since the positives of the normalcy classifier are analyzed by the signature-based component, the

output will be the same as if the signature-based component was run on all of the data except for the increase in false negatives.

Another advantage to a hybrid system is the consistent labeling when running the output through the signature component. This provides additional incentives for developing a hybrid system over building from scratch. Utilizing a signature-based approach to consistent labeling of any suspicious traffic enables the use of additional software that analyzes those labeled packets. By using a hybrid approach, the insertion of a machine learning component into the current system can be relatively seamless because the signature system is not changed or replaced, merely augmented and improved.

The use of a distributable network IDS system can be very useful in wireless networks, where the network could be infiltrated at almost any node. A normalcy classifier front-end provides a small distributable section that could be used to help provide network security on wireless networks.

## 7 Possible Implementation

Network packets are small collections of text. An N-gram can be used to break up the text into series of letters of a specified length to be used for classification [43]. This maps a network data packet into a high dimensional space where machine learning can be challenging. The high dimensional space can be hashed into a lower dimensional space without losing the ability to directly match the same packet [44, 45]. However, the hash is not a unique identifier and other similar packets may have the same hash. This approximation makes the system run faster, but the approximation can result in a large number of false positives if the dimension of the hash is too small. The tradeoffs for development of a hybrid IDS include the complexity of the algorithm, the required size and speed for the target platform, the training time for processing updates, the acceptable loss in detections. The runtime and retraining time can also be affected by the complexity of the algorithm. One possible implementation like this explored some of the performance tradeoffs [46] like size and speed, but this is an area for greater exploration with a larger and more relevant data set.

Though the abnormal output from the normalcy classifier has been shown to contain some information about new or unrecognized attack patterns, this has not been well characterized and represents a significant area for future research. The use of a normalcy classifier to capture relevant attack network traffic and a signature NIDS to remove known attacks leaves a much smaller set of network traffic that is similar to a known attack, or suspicious traffic. By matching the signature of the suspicious traffic to the signature of the known attack that it is similar to may provide additional insights to the suspicious traffic. The suspicious traffic that is similar to known active attack traffic can be isolated and analyzed as a potential variant of a known attack.

## 8 Hybrid Design Guarantees

Several guarantees can be made with this hybrid design pattern for improved NIDS performance. First, the resulting system will produce the same level of labeling quality as original NIDS. Second, the level of false alarms would not rise since original NIDS would be run on the abnormal output from the normalcy classifier. Third, the hybrid system would run faster, scale more easily, and use far less resources than a series of NIDS instances. Fourth, the false negative rate is going to increase slightly. Fifth, the cost of development and more significantly the cost of support and maintenance are significantly less than developing a new NIDS. These guarantees can help mitigate the development risk and can be used to understand the system tradeoffs when considering the overall design.

## 9 Conclusions

We developed a hybrid design to a NIDS that enables the seamless insertion of a machine learning component into a signature NIDS system that significantly improves throughput as well as captures additional networking traffic that is similar to known attack traffic. The throughput improvement by incorporating a normalcy classifier is significant, estimated to be the inverse of the false alarm rate which can easily net a factor of 1000. However, this can be diminished by updates that can trigger a retraining of the normalcy classifier. The addition of a normalcy classifier front-end also makes the system distributable across the network and more easily scalable.

The hybrid design also can provide significant information on new attack traffic. By finding the signature of suspicious traffic that is similar to the signature of a known attack, it can be isolated and analyzed as a potential variant of a known attack.

## 10 References

- [1] Maiolini, G., Baiocchi, A., Iacovazzi, A., & Rizzi, A. (2009). Real time identification of SSH encrypted application flows by using cluster analysis techniques. In NETWORKING 2009 (pp. 182-194). Springer Berlin Heidelberg.
- [2] Chen, R. C., Cheng, K. F., & Hsieh, C. F. (2010). Using rough set and support vector machine for network intrusion detection. arXiv preprint arXiv:1004.0567.
- [3] Este, A., Gringoli, F., & Salgarelli, L. (2009). Support Vector Machines for TCP traffic classification. *Computer Networks*, 53(14), 2476-2490.
- [4] Horng, S. J., Su, M. Y., Chen, Y. H., Kao, T. W., Chen, R. J., Lai, J. L., & Perkasa, C. D. (2011). A novel intrusion detection system based on hierarchical clustering and support

- vector machines. *Expert systems with Applications*, 38(1), 306-313.
- [5] Nguyen, T. T., & Armitage, G. (2008). A survey of techniques for internet traffic classification using machine learning. *Communications Surveys & Tutorials, IEEE*, 10(4), 56-76.
- [6] Scarfone, K., & Mell, P. (2007). *Guide to intrusion detection and prevention systems (idps)*. NIST special publication, 800(2007), 94.
- [7] Yao, J. T., Zhao, S. L., & Saxton, L. V. (2005, March). A study on fuzzy intrusion detection. In *Defense and Security* (pp. 23-30). International Society for Optics and Photonics.
- [8] Bace, R. G. (2000). *Intrusion detection*. Sams Publishing.
- [9] Dasarathy, B. V. (2003). *Intrusion detection*. *Information Fusion*, 4(4), 243-245.
- [10] Allen, J., Christie, A., Fithen, W., McHugh, J., & Pickel, J. (2000). State of the practice of intrusion detection technologies (No. CMU/SEI-99-TR-028).
- [11] Bobor, V. (2006). *Efficient Intrusion Detection System Architecture Based on Neural Networks and Genetic Algorithms*. Department of Computer and Systems Sciences, Stockholm University/Royal Institute of Technology, KTH/DSV.
- [12] Han, S. J., & Cho, S. B. (2005). Evolutionary neural networks for anomaly detection based on the behavior of a program. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 36(3), 559-570.
- [13] Chavan, S., Shah, K., Dave, N., Mukherjee, S., Abraham, A., & Sanyal, S. (2004, April). Adaptive neuro-fuzzy intrusion detection systems. In *Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on* (Vol. 1, pp. 70-74). IEEE.
- [14] Abraham, A., Jain, R., Thomas, J., & Han, S. Y. (2007). D-SCIDS: Distributed soft computing intrusion detection system. *Journal of Network and Computer Applications*, 30(1), 81-98.
- [15] Dickerson, J. E., & Dickerson, J. A. (2000). Fuzzy network profiling for intrusion detection. In *Fuzzy Information Processing Society, 2000. NAFIPS. 19th International Conference of the North American* (pp. 301-306). IEEE.
- [16] Tajbakhsh, A., Rahmati, M., & Mirzaei, A. (2009). Intrusion detection using fuzzy association rules. *Applied Soft Computing*, 9(2), 462-469.
- [17] Barbará, D., Couto, J., Jajodia, S., & Wu, N. (2001). ADAM: a testbed for exploring the use of data mining in intrusion detection. *ACM Sigmod Record*, 30(4), 15-24.
- [18] Abraham, A., & Jain, R. (2005). Soft computing models for network intrusion detection systems. In *Classification and clustering for knowledge discovery* (pp. 191-207). Springer Berlin Heidelberg.
- [19] Liao, N., Tian, S., & Wang, T. (2009). Network forensics based on fuzzy logic and expert system. *Computer Communications*, 32(17), 1881-1892.
- [20] Ngamwittayanon, N., Wattanapongsakorn, N., & Coit, D. W. (2009). Investigation of fuzzy adaptive resonance theory in network anomaly intrusion detection. In *Advances in Neural Networks-ISNN 2009* (pp. 208-217). Springer Berlin Heidelberg.
- [21] Toosi, A. N., & Kahani, M. (2007). A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers. *Computer communications*, 30(10), 2201-2212.
- [22] Tsang, C. H., Kwong, S., & Wang, H. (2007). Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection. *Pattern Recognition*, 40(9), 2373-2391.
- [23] Li, W. (2004). Using genetic algorithm for network intrusion detection. *Proceedings of the United States Department of Energy Cyber Security Group*, 1-8.
- [24] Bridges, S. M., & Vaughn, R. B. (2000, October). Fuzzy data mining and genetic algorithms applied to intrusion detection. In *Proceedings twenty third National Information Security Conference*.
- [25] Lu, W., & Traore, I. (2004). Detecting new forms of network intrusion using genetic programming. *Computational Intelligence*, 20(3), 475-494.
- [26] Crosbie, M., & Spafford, G. (1995, November). Applying genetic programming to intrusion detection. In *Working Notes for the AAAI Symposium on Genetic Programming* (pp. 1-8). MIT, Cambridge, MA, USA: AAAI.
- [27] Xia, T., Qu, G., Hariri, S., & Yousif, M. (2005, April). An efficient network intrusion detection method based on information theory and genetic algorithm. In *Performance, Computing, and Communications Conference, 2005. IPCCC 2005. 24th IEEE International* (pp. 11-17). IEEE.
- [28] Jirapummin, C., Wattanapongsakorn, N., & Kanthamanon, P. (2002, July). Hybrid neural networks for intrusion detection system. In *Proceedings of International Conference on Circuits, Computers and Communications* (pp. 928-931).

- [29] Pan, Z. S., Chen, S. C., Hu, G. B., & Zhang, D. Q. (2003, November). Hybrid neural network and C4. 5 for misuse detection. In *Machine Learning and Cybernetics, 2003 International Conference on* (Vol. 4, pp. 2463-2467). IEEE.
- [30] Moradi, M., & Zulkernine, M. (2004, November). A neural network based system for intrusion detection and classification of attacks. In *Proceedings of the 2004 IEEE international conference on advances in intelligent systems-theory and applications*.
- [31] Ngamwitthayanon, N., Wattanapongsakorn, N., Charnsripinyo, C., & Coit, D. W. (2008). Multi-stage network-based intrusion detection system using back propagation neural networks. In *Asian International Workshop on Advanced Reliability Modeling (AIWARM), Taiwan* (pp. 609-619).
- [32] Pukkawanna, S., Visoottiviseth, V., & Pongpaibool, P. (2007, November). Lightweight detection of DoS attacks. In *Networks, 2007. ICON 2007. 15th IEEE International Conference on* (pp. 77-82). IEEE.
- [33] Lee, J. H., Lee, J. H., Sohn, S. G., Ryu, J. H., & Chung, T. M. (2008, February). Effective value of decision tree with KDD 99 intrusion detection datasets for intrusion detection system. In *Advanced Communication Technology, 2008. ICACT 2008. 10th International Conference on* (Vol. 2, pp. 1170-1175). IEEE.
- [34] Katos, V. (2007). Network intrusion detection: Evaluating cluster, discriminant, and logit analysis. *Information Sciences*, 177(15), 3060-3073.
- [35] Chen, C. M., Chen, Y. L., & Lin, H. C. (2010). An efficient network intrusion detection. *Computer Communications*, 33(4), 477-484.
- [36] Lee, W., Stolfo, S. J., & Mok, K. W. (1999, August). Mining in a data-flow environment: Experience in network intrusion detection. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 114-124). ACM.
- [37] Labib, K., & Vemuri, R. (2002). NSOM: A real-time network-based intrusion detection system using self-organizing maps. *Networks and Security*, 1-6.
- [38] Puttini, R. S., Marrakchi, Z., & Mé, L. (2003, March). A Bayesian classification model for real-time intrusion detection. In *AIP Conference Proceedings* (pp. 150-162).
- [39] Amini, M., Jalili, R., & Shahriari, H. R. (2006). RT-UNNID: A practical solution to real-time network-based intrusion detection using unsupervised neural networks. *Computers & Security*, 25(6), 459-468.
- [40] Su, M. Y., Yu, G. J., & Lin, C. Y. (2009). A real-time network intrusion detection system for large-scale attacks based on an incremental mining approach. *Computers & security*, 28(5), 301-309.
- [41] Li, Z., Gao, Y., & Chen, Y. (2010). HiFIND: A high-speed flow-level intrusion detection approach with DoS resiliency. *Computer Networks*, 54(8), 1282-1299.
- [42] Chakrabarti, S., Chakraborty, M., & Mukhopadhyay, I. (2010, February). Study of snort-based IDS. In *Proceedings of the International Conference and Workshop on Emerging Trends in Technology* (pp. 43-47). ACM.
- [43] Brown, P. F., Desouza, P. V., Mercer, R. L., Pietra, V. J. D., & Lai, J. C. (1992). Class-based n-gram models of natural language. *Computational linguistics*, 18(4), 467-479.
- [44] Shi, Q., Petterson, J., Dror, G., Langford, J., Strehl, A. L., Smola, A. J., & Vishwanathan, S. V. N. (2009). Hash kernels. In *International Conference on Artificial Intelligence and Statistics* (pp. 496-503).
- [45] Shi, Q., Petterson, J., Dror, G., Langford, J., Smola, A., & Vishwanathan, S. V. N. (2009). Hash kernels for structured data. *The Journal of Machine Learning Research*, 10, 2615-2637.
- [46] Chang, R. J., Harang, R. E., & Payer, G. S. (2013). Extremely Lightweight Intrusion Detection (ELIDe), ARL-CR-0730.
- [47] Kumar, G., Kumar, K., & Sachdeva, M. (2010). The use of artificial intelligence based techniques for intrusion detection: a review. *Artificial Intelligence Review*, 34(4), 369-387.
- [48] Hwang, T. S., Lee, T. J., & Lee, Y. J. (2007, June). A three-tier IDS via data mining approach. In *Proceedings of the 3rd annual ACM workshop on Mining network data* (pp. 1-6). ACM.

