Bigdata platform based approach for defending against DDoS

Yoon Joo Chae, Nikitha Johnsirani Venkatesan, and Dong Ryeol Shin
{chaeyj, nikithajv, drshin}@skku.edu
Information & Communication Engineering, Sungkyunkwan University, Suwon, Gyeonggi Do, South Korea

Abstract - *Distributed denial-of-service (DDoS) is a rapidly growing problem. From the first known attack in 1999 to the highly publicized Operation Ababil, the DDoS attacks have a history of flooding the victim network with an enormous number of packets, hence exhausting the resources and preventing the legitimate users to access them. The variety of attacks are overwhelming now a days. However, the existing method for managing a number of computer with respect to an external invasion, such as hacking, is not strong. We look into the real time attacks of DDoS. Using a big data platform real-time processing capacity, we propose a way to manage a large number of individual computer log collecting, processing via the analysis at a public organization. We propose a way of former prediction by using system logs analysis. Therefore, to identify the malicious data, processing, analysis, are done using Logstash and Spark respectively.*

Keywords: DDoS, Spark, Logstash, Elastic Search and HUE

1 Introduction

DDoS attacks are one of the biggest challenge faced by security researchers on International scale. The losses caused by the security breach can cost up to several billions of dollars [10]. In a public sector, a hacker has managed to target multiple systems from a centralized computer. The existing defensive mechanisms for predicting the cyber-attacks are still in the basic level. Now a days hackers use sophisticated hacking system to invade the authorized systems of the users. The current security systems like even antivirus are failure in detecting the intrusion. Sometimes, outside intrusion such as downloading data from USB device is unavoidable. But, still using proper analysis of the data some of the malicious attacks can be overcome. Currently the attacks are of two types [11]. The first one is to send the malicious packet which is injected with virus called vulnerability attack as a running application. The second one is traditional method of draining the resources of the victim like input-output bandwidth, database bandwidth, CPU memory and the like.

Traditional defensive mechanisms try to rectify the virus after detecting it. For example, anti-virus which is installed in our system detect the virus after the virus got executed in the system. In this paper, using big data platform, we propose a way to detect the malicious virus before even entering the system. The framework manages the system logs from large number of computers thus eliminating the malicious attacks. The system logs are collected and processed and finally analyzed with the help of public organization. The real-time response measures in the cluster structure predict accidents in advance to allow flexible management about the incident.

The rest of the paper is organized as follows: Section 2 describes the DDoS architecture. Section 3 describes the related work done to defend the DDoS attacks and the real-time attacks happened in the organizations. We outlined the architecture and the experimental results in section 4. We also discussed the future work regarding this framework in section 5. Finally, section 6 concludes the paper.

2 DDoS Architecture:

A denial-of-service (DoS) is a type of attack in which the hackers prevent the authorized users from accessing the service [5]. A distributed denial-of-service (DDoS) is a kind of DoS attack. The difference between DoS and DDoS is that, DDoS involves multiple systems to target a single computer whereas DoS attack uses one system. The word "distributed" implies that an attack is focused within a team of disruptors who hack with a common goal of preventing the webservers from working normally. Normally, the source of attack is more than one, likely to be in thousands of unique IP address.

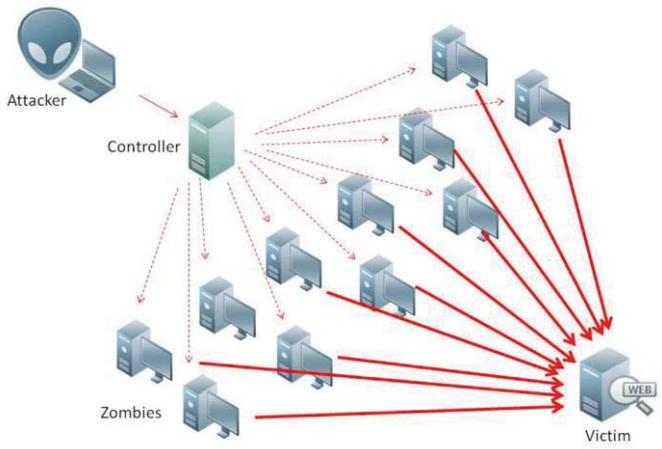

Figure 1. Overview of DDoS

The overview of the DDoS attack is depicted in figure 1. The attacker sets up the hierarchical attack architecture. Initially, an attacker selects one controller which is vulnerable to securities. After that, the zombies are selected following the same procedure as of the controller. But, the zombies are indirectly handled by the attacker through the controller. The zombies, otherwise called as selected agents perform the DDoS attacks by sending enormous amount of malicious traffic to the targeted system [7]. The controller and the zombies are commonly located in the external networks. Once the hacker successfully selected the controller and the zombies, he/she starts controlling the communication among the

controller, zombies and the targeted system. After completing the selection and the communication process, the attacker starts launching the DDoS attacks on the victim simultaneously. Normally, the communication between the controller, zombies and the victim will be encrypted for the safe information exchange.

There are many kinds of DDoS attacks [6]: Traffic attacks, Bandwidth attacks, Application attacks and the like. The DDoS attacks can be done by various techniques which are listed as follows [5]:

- Internet Control Message Protocol (ICMP) flood
- Teardrop attacks
- Peer-to-peer attacks
- Permanent denial-of-service attacks
- Application-layer floods.
- Nuke
- Slow Read attack
- Telephony denial-of-service(TDoS)

The above attacks are detailed in [5]. The main motivation of DDoS attacks are financial frauds, competitive rivalry, ideological hacktivism and extortion.
Virtually, all kind of resources that are connected to the Internet are vulnerable to DDoS attacks. Many existing systems are not capable of protecting the systems against these DDoS attacks. To protect the Internet, most of the organizations use security tools such as, Internet Service Providers (ISP), firewalls and secure web gateways. Although all these security tools act as first layer protection for the basic threats, they cannot give protection from advanced threats

3 Related works:

3.1 Intrusion practices in Institutions:

This section lists the known cyber-attacks on organizations which are made known to the public. The fields that are very often prone to DDoS attacks are listed below:

- Retail
- Communication
- Technology
- Health-care

We will discuss the real time scenarios of the cyber-attacks in the respective fields [8].

3.1.1 Retail:

In December 2013, 70 million individual's personal information was stolen along with their credit and debit card details from an organization named Target.
In October 2013, more than 9,000 credits cards were used fraudulently following the attack in Neiman Marcus. The employees who worked in the company were not even able to detect the attack for months because of the hacker's code.
In January 2014, 2.6 million customers in Michael's payment card were affected. The hackers targeted the POS system to gain access.
In May 2014, the contact and log-in information of 233 million customers was hacked in eBay along with their employee's details. Later, eBay requested all of its customers to change their password.
In September 2014, again 868,000 credit and debit card information was hacked from Goodwill Industries International. Malware infected the chain store through infected third party vendors.
In March 2012, the banks in South Korea named Jeju Bank, NongHyup, Shinhan Bank reported that Internet Banking servers were blocked temporarily. Some of the branches told that the computers were infected with virus and many important files had been erased [9].

3.1.2 Communication:

In January 2014, Yahoo mail reported that around 273 million accounts were hacked.
In April 2014, AT&T was hacked for two weeks completely from inside by someone who accessed all the users and social security information.
In June 2014, Feedly reported that 15 million users were temporarily affected by three DDoS attacks.
In September 2014, around 5 million usernames and passwords were hacked from Gmail users and they were released on Russian site.
In October 2014, the pictures of 200,000 users were hacked from snapsave. Snapsave is a third-party app for saving photos from Snapchat.

3.1.3 Technology

In June 2014, 100 million users from Evernote faced DDoS attacks.
In September 2014, Hackers used third party applications to access Apple user's online data storage. This attack leads to posting the celebrities personal pictures online.

3.1.4 Health care

In June 2014, credit and debit card information from Chang's restaurant was hacked and reported that they all sold online.
In August 2014, the personal data of 4.5 million patients were hacked from Community Health Services (CHS). CHS made a statement that all the patients who visited any of its branches might have their information hacked. They claim that malware used in the attack originated from China. The FBI warns eventhe other health care information might have been stolen.

4 System Architecture:

Figure 2 shows the real-time computer configuration Central Management System. We use the Logstash to collect system logs from Windows. The Logstash is used in Windows Management Instrumentation (WMI) to gather logs. Logstash is an open source platform which can process any data from

any source. It will centralize the data processing of all types and extend to custom log formats

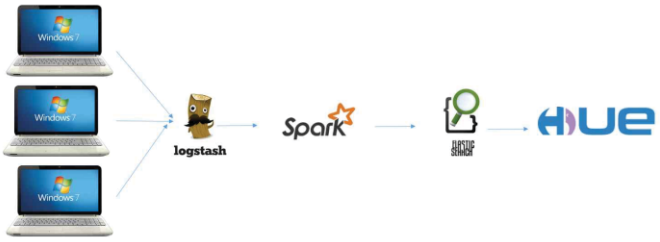

Figure 2. Architecture of the framework

WMI is a Microsoft implementation of the Web based enterprise management which gives the overall information. The data is collected and then sent to Logstash Server. We later use WMI in Apache Spark logs of collected system logs. Log data is transferred in real time to Spark for processing in memory. Now, the Apache Spark analyze the data for the malicious data. [3] Elastic Search stores the analysis data on the server Elastic Search. Finally, the analysis data is expressed in a common User Interface UI called HUE (Hadoop User Interface).

4.1 Experiments and Results:

We experimented the proposed framework in a server which contains 8 racks. The memory of the Rack is 16 core CPU * 8 (rack). Each rack has 32 GB memory. The server has 1 TB of storage size. The figure 3 depicts the results of the system logs from Logstash.

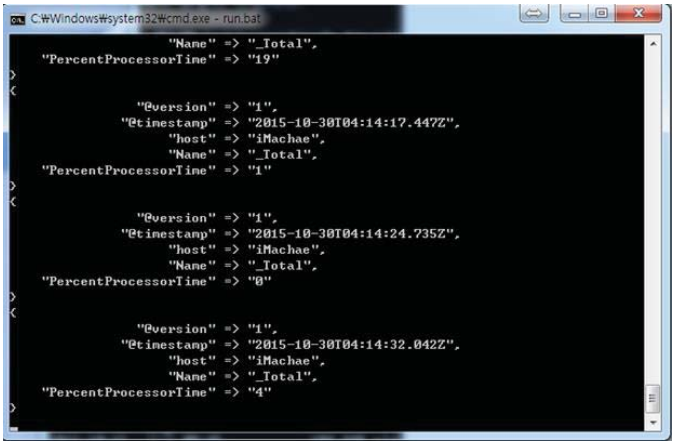

Figure 3. Logstash analysis

The analysis time took one hour in total and the log data size is 100 Mb. The interval between the analyses is 7 seconds.

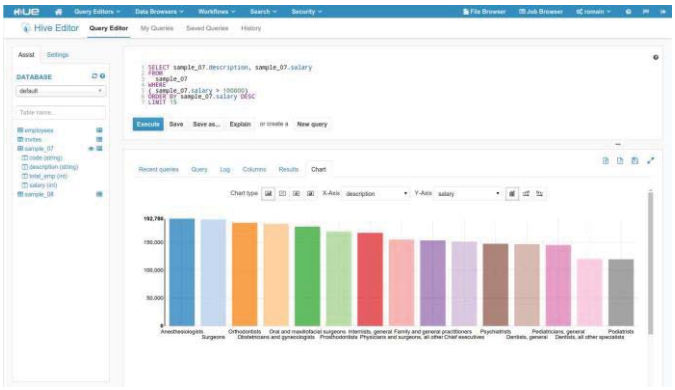

Figure 4. Results of the Analysis

The analysis results is done using HUE (Hadoop User Experience) and the results is given in figure 4.

5 Future work:

In the above work, since the analysis is done by spark in-memory cache, it can only process limited amount of data.
Real-time computer in the log file of the Windows WMI collected in real time by using the management system without passing has the categories of about 100. It collects logs for the individual machine (slave machine) to be used in public organizations in real time, it is expected to be with respect to the suspected role quickly diagnose and predict than the conventional method. In particular log , such as the (log of the operations access reservation using the Schedule service) and CPU (CPU management), network adapter management log contents of the job provided by WMI is after a certain period of time after infection , such as DDoS attacks respond in real time with respect to that work process and malicious , it can analyze the overall log management for individual computers, it is expected to be easier to manage than traditional methods of centralized management system.

6 Conclusion:

Conventional defensive method to DDOS attacks, usually response against such attacks in a scheduled treatment.
However, using the same platform, to examine the process, which are analyzed in advance with respect to a single computer that is managed by the public institution. In this paper, we have analyzed and checked the reservation process in advance before a hacking accident occurs. The framework can obtain information in advance without any timely schedule where we can block easily.

7 Reference:

[1] Zaharia, Matei, et al. "Spark: cluster computing with working sets." Proceedings of the 2nd USENIX conference on Hot topics in cloud computing. Vol. 10. 2010.

[2] Krishna, T. Lakshmi Siva Rama, T. Ragunathan, and Sudheer Kumar Battula. "Customized Web User Interface for Hadoop Distributed File System." Proceedings of the Second International Conference on Computer and Communication Technologies. Springer India, 2016.

[3] Kim Joo-hyuk, imjinsu., "The latest information security issues and encryption technology overseas study Trend", National Internet Development Agency of Korea, 2014 236 Proceedings of 2015

[4] Kim Ji Hoon., "DDoS Internet chaos", AhnLab Available at: http://www.ahnlab.com/kr/site/securityinfo/secunews/secuNewsView.do?menu_dist=3&seq=16241

[5] Available at: https://en.wikipedia.org/wiki/Denial-of-service_attack

[6] Peng, Tao, Christopher Leckie, and Kotagiri Ramamohanarao. "Detecting distributed denial of service attacks by sharing distributed beliefs." *Information Security and Privacy*. Springer Berlin Heidelberg, 2003.

[7] Lee, Keunsoo, et al. "DDoS attack detection method using cluster analysis. "*Expert Systems with Applications* 34.3 (2008): 1659-1665.

[8] Walters, Riley. "Cyber-attacks on us companies in 2014." *Heritage Foundation Issue Brief* 4289 (2014).

[9] Available at: https://en.wikipedia.org/wiki/2013_South_Korea_cyberattack

[10] Singh, Kamaldeep, et al. "Big data analytics framework for peer-to-peer botnet detection using random forests." *Information Sciences* 278 (2014): 488-497.

[11] Tripathi, Shweta, et al. "Hadoop based defense solution to handle distributed denial of service (DDoS) attacks." *Journal of Information Security* 4.3 (2013): 150.