International Journal of Innovative Research in Information Security (IJIRIS) ISSN: 2349-7017(O)
Issue 1, Volume 2 (January 2015) ISSN: 2349-7009(P)
www.ijiris.com

Intelligent Network Surveillance Technology for APT Attack Detections

Byungik Kim
Korea Internet & Security Agency
Seoul, Korea

Hyeisun Cho
Korea Internet & Security Agency
Seoul, Korea

Taijin Lee
Korea Internet & Security Agency
Seoul, Korea

Abstract— Recently, long-term, advanced cyber-attacks targeting a specific enterprise or organization have been occurring again. These attacks occur over a long period and bypass detection by security systems unlike the existing attack pattern. For such reason, they create problems such as delayed real-time response and detection after damages have already been incurred. This paper introduces the design of technology that applies real-time network traffic monitoring to detect unknown functional cyber-attack on the network. Specifically, the algorithm was verified and evaluated in terms of performance in an actual commercial environment. Cyber-attack detection performance is expected to be improved by enhancing the algorithm and processing large volumes of traffic.

Keywords— APT Attack, Network Surveillance, Malwares, Command and Control Server, URLs

I. INTRODUCTION

This document is a template. An electronic copy can be downloaded from the Journal website. For questions on paper guidelines, please contact the journal publications committee as indicated on the journal website. Information about final paper submission is available from the conference website.

In the past, major cyber-attacks were mostly isolated intrusion cases such as simple system hacking, web server attacks, or distribution of malicious code. Over time, however, more attacks have been targeting a specific country, organization, or enterprise. Such cyber-attacks are launched for the purpose of attaining economic gain or creating social confusion, aiming at a clear target. Called advanced persistent threat (APT), the attack occurs quietly for a long period of time; the existing security systems have limitations in detecting it [1],[2],[3],[4],[9].

An APT attack is divided into four steps (see Figure 1): “Preparation and Intrusion”; “Command Control”; “Inspection of Internal Vulnerability and Gathering of Information,” and; “Leak of Information and Infliction of Damage.” Such steps are carried out over a long period [5],[11].

The first step, “Preparation and Intrusion,” involves securing the path to the target organization or enterprise. Here, the APT attacker inspects the social networking service (SNS) postings of the employees of the target organization or enterprise to find out the interests or personal information of the employees [5]. It then sends a fabricated e-mail or web link related to the interest or hobby of the SNS user. The e-mail or web distributes the malicious code to enable the APT attacker to access the internal network of the target organization. When an employee of the target organization accesses the e-mail or web link, the PC automatically downloads the malicious code and gets infected. The downloaded malicious code uses the latest technique to dodge detection by the vaccine program of the PC [5].

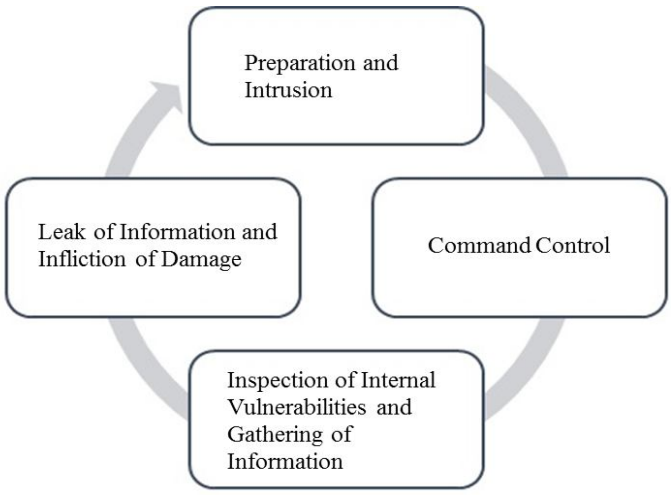

Figure 1. APT Attack Phases

The next step is “Command Control” wherein the infected PCs inside the target organization convey to the attacker the fact that they are infected and wait for the command. Here, the attacker will know that zombie PCs were created in the internal network of the target organization.

© 2015, IJIRIS- All Rights Reserved

The attacker checks if a zombie PC is the PC actually used by a user or if it is utilized to analyze the malicious code and sends additional malicious code for the next step – "Inspection of Internal Vulnerability and Gathering of Information" – if it is the actual user PC.

The additionally downloaded malicious code inspects the internal network of the target organization. Specifically, it inspects the vulnerability of various servers and PCs, and then sends the information to the attacker. For its part, the attacker selects the servers and PCs to attack based on the received information. The attacker obtains access privilege using the vulnerability of the servers or PCs to be attacked and collects the main information of the devices.

Lastly, the attacker sends the collected data to an external server and causes economic or social loss to the organization. Moreover, the attacker destroys the devices connected to the internal network of the target organization or disables the services of major systems after the information is collected [4],[5].

TABLE 1.-- APT ATTACK VS. OTHER CYBER-ATTACKS

	APT	**BotNet**	**Malware**
Target	Targeted	Unspecific	Unspecific
Frequency	Long-term	One-time	One-time
Instrument	New Malware	URL Downloaded Bot	Malicious program

Since the APT attacks occur over a long period and use unknown malicious code, they are difficult to detect with existing technologies such as IPS/IDS, which can detect known attacks only. Moreover, because the attacks occur over a long period, it is difficult to analyze all data from the beginning to the end of the attack. Table 1 shows the difference between the existing attacks and APT attacks.

Korea is particularly known to generate the second biggest number of APT attacks in the world [13]. The analysis of DDoS attack on March 4, 2011 and cyber-terror on March 20, 2013 suggests that the attack occurred over a long period of 7 months or longer but took place so quietly that the attack was detected only after the damage was done.

This paper describes a technique of increasing the accuracy of detecting such APT attacks. The technique involves collecting all network data in real time to extract important data and detecting the anomalous behavior penetrating within. The detected anomalous behavior is additionally analyzed to detect "Preparation and Intrusion" and "Command Control," which are the first steps of an APT attack. Such prevents the "Inspection of Internal Vulnerability and Gathering of Information" and "Leak of Information and Infliction of Damage" steps.

The rest of this paper is organized as follows: Chapter II describes the existing technologies to detect the attack penetrating within enterprises; Chapter III suggests the algorithm detecting the initial step of APT attacks based on key network data; Chapter IV verifies the suggested algorithm in an actual system; Chapter V presents the conclusion.

II. RELATED RESEARCH

Chapter II introduces the existing technologies for detecting and preventing intrusion attacks. Most of the intrusion attacks start with infecting the user PC or server with malicious code. The major technologies for detecting such malicious code can be divided into the vaccine-based intrusion malicious code detection and prevention technology and network traffic monitoring-based malicious behavior-detecting technology.

A. Detection of Vaccine-Based Intrusion Attack

The vaccine-based intrusion attack detection method inspects the files flowing into or running in a PC or a server to check if they contain malicious code. The vaccine detects malicious code using the specific unique value of the already analyzed malicious code. To create the unique characteristics of the malicious code, the malicious behavior file must be identifiable using the result of the analysis by a skilled malicious code analyst. The unique value of the malicious behavior file is then extracted and made into a signature to be distributed to each user PC. Figure 2 shows the detection process and malicious code signature creation method of a vaccine [6],[7],[12].

Vaccine-based intrusion attack detection can inspect all known malicious code in a short period. Moreover, it consumes only a small amount of system resources for detection. Note, however, that it can detect malicious code only when the malicious code used for the attack is already known. It will not detect cases such as APT attack, which uses unknown malicious code or attack pattern.

© 2015, IJIRIS- All Rights Reserved

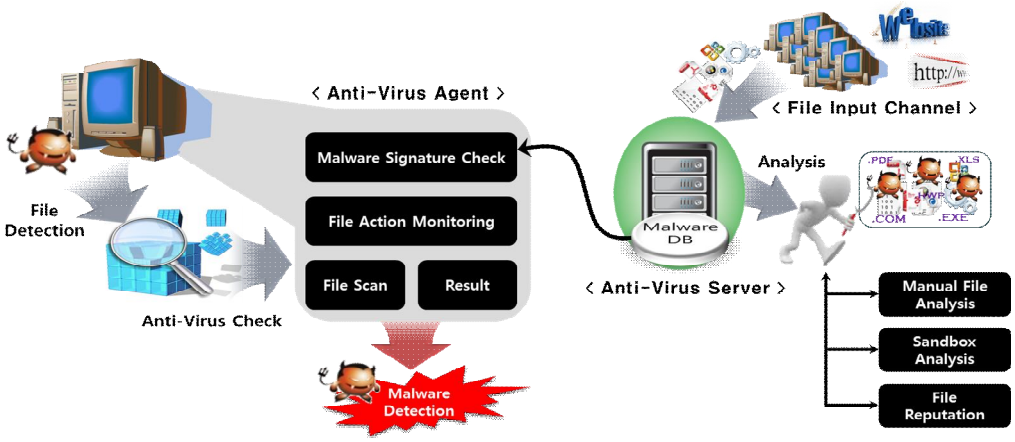

Figure 2. Operating Steps of Anti-Virus Products

B. Prevention of Intruding Malicious Network Traffic

This section discusses the network traffic-based intrusion detection/prevention technology, which is similar to the method suggested in this paper. The network-based prevention of intrusion technology can allow or prevent data similar to the specific condition. It includes IPS/IDS and firewall, which can analyze and prevent the data penetrating into the network in real time [14],[15].

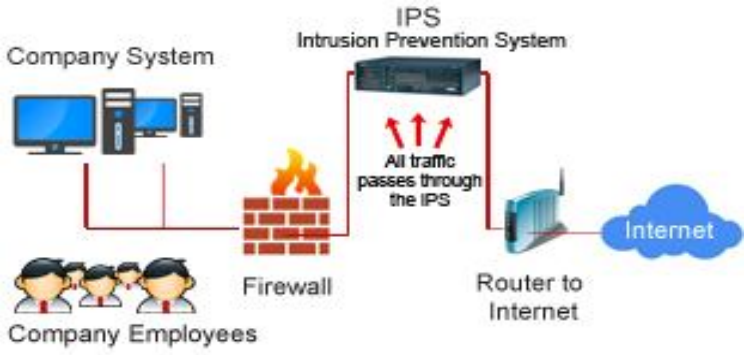

Figure 3. IPS System Concept

Using the specific data of network traffic, intrusion attacks can be detected and prevented. Records of outside IP, information-leaking server IP, and anomalously connecting IP used in past intrusion attacks can aid in preventing the recurrence of the same attack. Moreover, it can allow or prevent intrusion of the specific service of the traffic to prevent unintended access. By monitoring traffic in and out for a specific period, it can detect anomalously created traffic [8],[10].

Like the detection of vaccine-based intrusion attack, however, its limitation is that it can perform prevention based only on known attack or analyzed data. Although it can detect unknown anomalous traffic, it can detect traffic during a short period of 1-2 days only, but not attacks such as APT, which occurs over a long period.

This paper suggests an algorithm that overcomes the technical limitation of existing technology to detect and prevent APT attacks.

III. MEASURES FOR THE PREEMPTIVE DETECTION OF APT ATTACKS

As described in Chapter I, an APT attack targets a specific organization and occurs quietly for a long period; thus, it is very difficult to detect with existing IPS/IDS or vaccine program. To detect such attacks, data over a long period must be analyzed. Note, however, that such APT attacks must intrude the malicious code into the specific attack target space for "Preparation and Intrusion" and "Command Control." Chapter III suggests an algorithm for detecting the attacks based on network data using such characteristics of ATP attacks.

A. Total Traffic Collection and Monitoring

The biggest issue for an attacker is securing the zombie PC to execute the attacker's command within the target organization. For that, it tries to transmit the malicious code in various ways. The suggested algorithm monitors the entire network traffic and extracts the key data. It then installs a system to collect network traffic at the point of contact, which connects the inside of the target organization and its internal network. The installed system collects the entire in/out TCP/UDP traffic and gathers the necessary data, particularly the source IP/port and destination IP/port.

The entire traffic is inputted to the network collection system through tapping or mirroring. The inputted traffic is separated by the network interface card (NIC) of the network and recorded in the system memory in real time.

Figure 4 illustrates the network-collecting position where the suggested system is applied.

© 2015, IJIRIS- All Rights Reserved

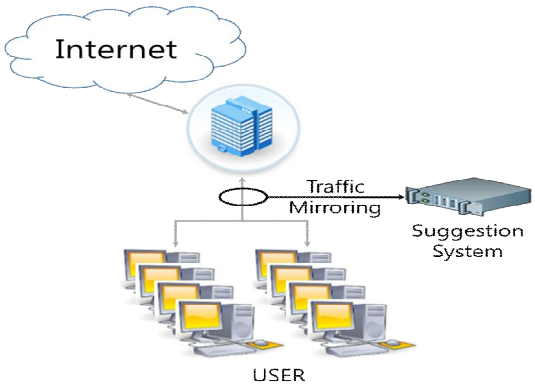

Figure 4. Network-Collecting Position

B. Extraction of Key Network Traffic Data

To detect an attack intruding into an organization, the key data in the traffic must be detected. The detected key data include the internal/external IP/port pair and traffic transfer between them as well as the file penetrated through the payload inside the traffic. The IP/port data become the key information to check the communication between a zombie PC inside and the outside attacker or command and control (C&C) server. The file inside the traffic payload is used by the attacker to distribute malicious code to secure the zombie PCs.

To check and manage such key data easily, the IP/port pair of the traffic and other data are recorded in the memory. The data recorded in the memory are used to extract the malicious code and detect the malicious code-downloading URL, distributing URL, and anomalous traffic.

The data recorded in the memory are saved as a text file every minute for use in detecting unknown attacks. Figure 5 shows an example of the generated text file.

```
2014-02-18 09:17:42.636529, 2014-02-18 09:17:42.678938, 074.125.235.183, 80, 211.231.150.125, 55735, 5, 1163
2014-02-18 09:17:42.637134, 2014-02-18 09:17:42.637529, 211.231.150.125, 55735, 074.125.235.183, 80, 3, 180
2014-02-18 09:17:43.637975, 2014-02-18 09:17:43.275486, 211.231.150.125, 55736, 074.125.235.183, 443, 191, 13184
2014-02-18 09:17:42.653466, 2014-02-18 09:17:42.772755, 185.015.247.245, 55690, 211.231.150.126, 3389, 17, 1030
2014-02-18 09:17:42.653990, 2014-02-18 09:17:42.772936, 211.231.150.126, 3389, 185.015.247.245, 55690, 33, 35084
2014-02-18 09:17:43.679280, 2014-02-18 09:17:43.316862, 074.125.235.183, 443, 211.231.150.125, 55736, 375, 504548
```

Figure 5. Network Traffic Log

C. Inspection of Intruding File

To detect the files penetrating into an organization, the traffic payload data recorded in the memory must be analyzed. It detects the PE file whose file header begins with MZ in the payload data and extracts the size of the file. After checking the file size, it extracts the data in the memory for the amount of file size and creates a separate file. The created file is sent to the vaccine inspection system and vaccine agent to check for existing malicious code.

Since most APT attacks use unknown malicious code, they are not detected by the vaccine inspection system. To detect such unknown malicious code, the file is sent to the system to analyze the file automatically. The suggested system sends to the “malicious code automatic analysis system” the files detected as normal by the vaccine. The “malicious code automatic analysis system” automatically performs static analysis and behavioral analysis of the transferred file. Moreover, it additionally analyzes the API data used by the file to check for anomalous behavior. The inspection result is sent to the system suggested by this paper.

D. Selection of Inspection Target URLs and In-depth Analysis

If the analysis of the PE file reveals an anomaly, it additionally analyzes the traffic creating the file. It checks the key data of the traffic created in Sections III.A and III.B and analyzes the path of access from outside to inside at the time of malicious code penetration.

For that, it collects the IP/port pair in the intruding traffic and URL data in the traffic header. The URL type contains the destination URL of the traffic and origin URL (referrer URL) creating the traffic. Figure 6 presents the header data and referrer URL data in an actual traffic.

```
GET /google-evil-plan HTTP/1.1
Host: www.blindfiveyearold.com
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10.7; rv:19.0) Gecko/20100101 Firefox/19.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
Referer: http://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&docid=Dlvw4-9W67f7WM
Cookie: wp-settings-2=m0%3Dc%26m1%3Do%26m2%3Dc%26m3%3Dc%26m4%3Dc%26m5%3Do%26m6%3Dc%
Connection: keep-alive
```

Figure 6. TCP Traffic Header and Referrer URL

© 2015, IJIRIS- All Rights Reserved

Using the data in the traffic, the URL from which the malicious code was downloaded is backtracked to identify the URL actually accessed by the internal user. Note, however, that the URL distributing the malicious code uses new sessions or separate traffic only to download the file to evade backtracking. To analyze such evasion of malicious code download detection, the suggested technology extracts all URLs accessed by the IP downloading the malicious code for the past three minutes. The destination URL and referrer URL in the extracted URLs are used to restore the order of traffic movement. From the restored URLs, the intersection URLs (destination URL == referrer URL) are extracted to be set as subjects of inspection by visiting. Figure 7 depicts the algorithm for extracting the subjects of inspection by visiting.

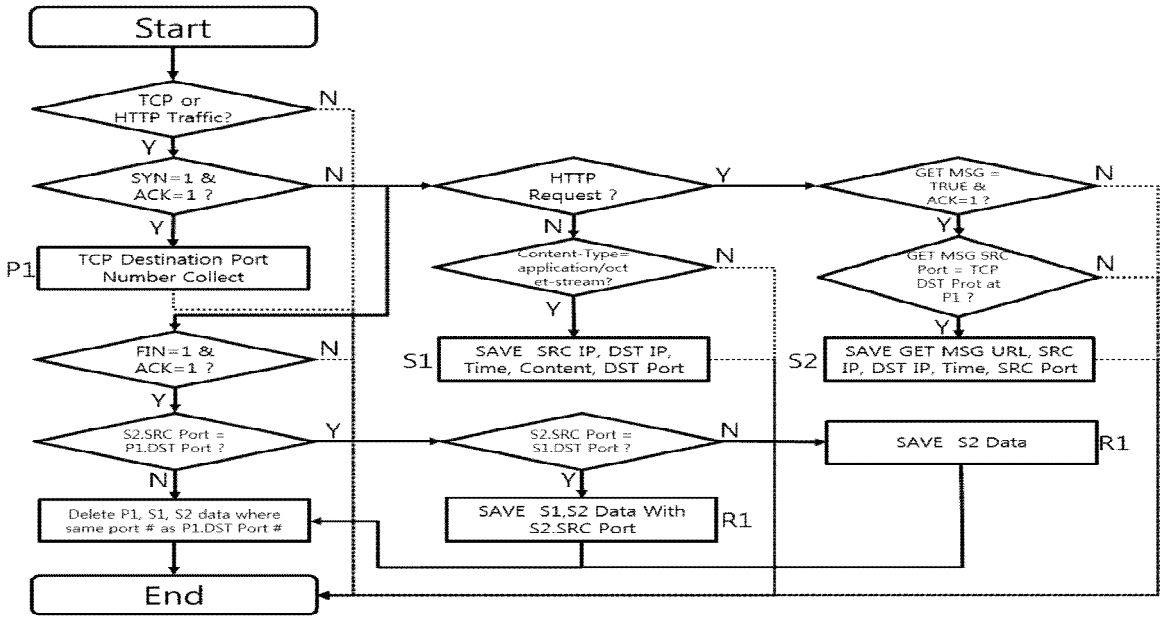

Figure 7. Proposed Algorithm

A URL set to be the subjects of inspection by visiting is sent to the "automatic malicious code collection system" to be checked whether it actually downloaded the malicious code. If the inspection indicates that the collected file is the malicious code, it sets the visited URL to be the URL that distributes and passes the malicious code. The result is sent to the system suggested in this paper.

E. Detection of Intrusion by APT Attack

The data detected in Chapter III Section B, C, and D are combined so that the attack intruding the target organization can be analyzed. The outside attacker can be detected by analyzing the URLs that distribute or pass the new or existing malicious code.

The reason detection of APT attack is so difficult is that it requires inspection of the entire intruding traffic and corresponding manpower to analyze the malicious code. The technology suggested in this paper provides only the data needed for analysis; thus, it can reduce the analysis time and automatically detect the anomalous attack of the network. It detects the internal IP downloading the new malicious code as well as the external attack command and control (C&C) server. The internal IPs accessing the detected C&C are automatically monitored to detect the zombie PCs inside the organization as well as to detect and prevent additional downloading of malicious code. Moreover, it can prevent the creation of additional internal zombie PCs based on the URL data distributing the malicious code.

IV. VERIFICATION OF THE SUGGESTED ALGORITHM

The technology and algorithm suggested in Chapter III were verified in an actual commercial environment. The verification tested the functionality and performance of "key network data extraction," "identification of URLs to be inspected," and "detection of intrusion attacks."

A. Key Network Data Extraction

Key data of traffic in/out of the network include the source IP/port, destination IP/port, detected time, and accessing URL data. They are information essential to detect cyber-attacks; the system processing performance can be compared for each network bandwidth. In an actual commercial environment, the network bandwidth dynamically changes; hence the difficulty of measuring "key network data extraction" performance. For that reason, the Avalanche system, which can generate network traffic, was used to test the performance. The performance test was conducted by varying the network bandwidth, session connection, etc., for a week beginning February 17, 2014. Figure 8 illustrates the detection performance at different network bandwidths. It processed traffic of around 700Mbps without any problem and up to 1Gbps assuming about 5% traffic loss was acceptable.

© 2015, IJIRIS- All Rights Reserved

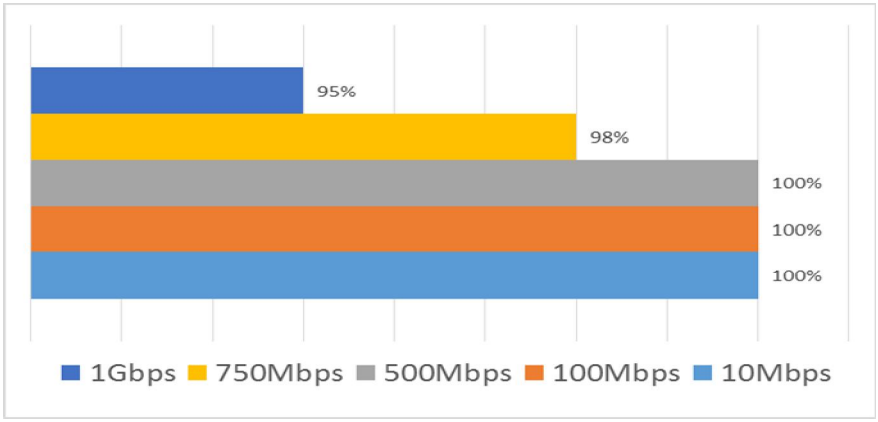

Figure 8. Detection Performance of Different Network Bandwidths

Table 2 shows the session connections per second that the system can handle.

TABLE 2. - DETECTION PERFORMANCE OF DIFFERENT **SESSION** CONNECTIONS

	# of Session Connections Created (by Sec)	# of Session Detections (by Sec)	Session Detection Rate
Test1	2,357	2,357	100%
Test2	3,541	3,541	100%
Test3	4,223	4,223	100%
Test4	4,821	4,821	100%
Test5	6,172	6,172	100%
Test6	6,721	6,721	100%
Test7	7,153	7,153	100%
Test8	7,781	7,781	100%
Test9	8,071	5,981	74%

The test shows that it can detect around 7,800 session connections per second and extract the data. Note, however, that the number of processed sessions rapidly decreased when there were more than 8,000 session connections per second.

B. Identification of URLs Subject to Inspection

For the month of February 2014, a business site using KT's Internet in Korea was tested for the performance of this algorithm. Daily average traffic volume was around 600Mbps, and 231,258 URLs to be inspected were extracted per minute. The server used for algorithm verification was a mid-size server. Figure 9 shows the graph of bandwidth inspected for the month of February, with Figure 10 presenting the average number of URLs collected per minute.

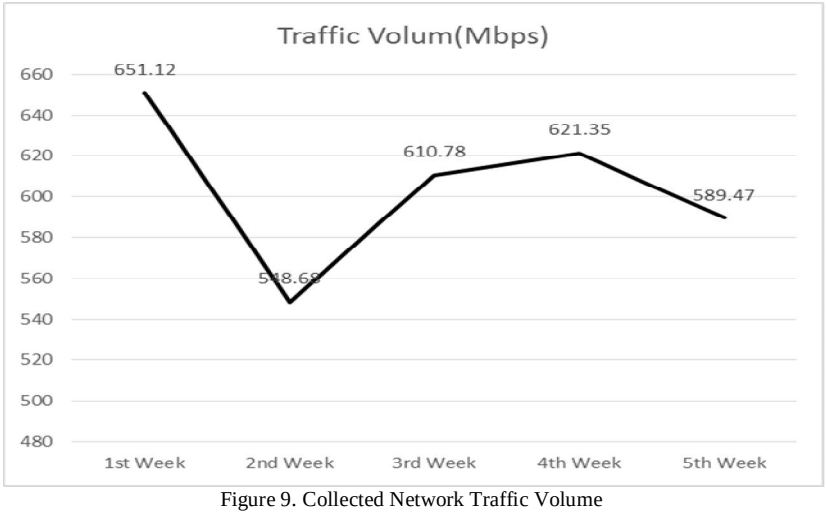

Figure 9. Collected Network Traffic Volume

© 2015, IJIRIS- All Rights Reserved

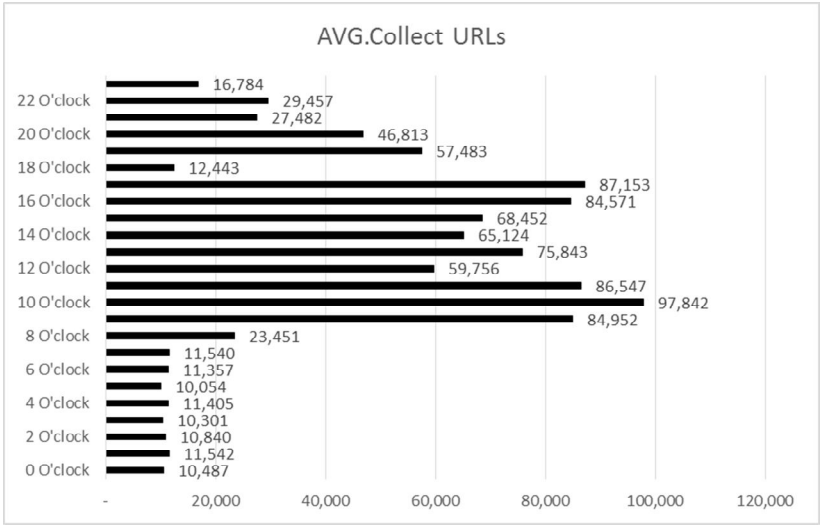

Figure 10. Average Number of Collected URLs

When the algorithm for the "identification of URLs to be inspected" was applied to the collected URLs, 38,657 visit inspection subject URL sets, or 1/6 of the total URLs, were extracted; this is equivalent to 5.996 times' detection performance efficiency improvement. Moreover, 147 URLs related to file creation were detected among the referrer URLs extracted by the suggested algorithm. Among them, 2 URLs were found to be related to the actual penetration of malicious code.

Figure 11 shows an example of the log file of the detected URL data during the test period.

2014-02-18 09:19:00.196, 183.098.222.181, 1345, 210.112.003.023, 80, www.bmw.co.kr/_common/html/css/generic_seo.css, http://www.bmw.co.kr/kr/ko/
2014-02-18 09:19:00.196, 183.098.222.181, 1345, 210.112.003.023, 80, www.bmw.co.kr/_common/html/css/generic_seo.css, http://www.bmw.co.kr/kr/ko/
2014-02-18 09:19:00.199, 183.098.222.181, 4034, 210.112.003.023, 80, www.bmw.co.kr/kr/ko/_common/html/css/generic_seo.css, http://www.bmw.co.kr/kr/ko/
2014-02-18 09:19:00.199, 183.098.222.181, 4034, 210.112.003.023, 80, www.bmw.co.kr/kr/ko/_common/html/css/generic_seo.css, http://www.bmw.co.kr/kr/ko/
2014-02-18 09:19:00.209, 183.098.222.181, 4203, 210.112.003.023, 80, www.bmw.co.kr/_common/html/js/bmw_scriptlib.js, http://www.bmw.co.kr/kr/ko/
2014-02-18 09:19:00.209, 183.098.222.181, 4203, 210.112.003.023, 80, www.bmw.co.kr/_common/html/js/bmw_scriptlib.js, http://www.bmw.co.kr/kr/ko/
2014-02-18 09:19:00.209, 183.098.222.181, 4498, 221.157.125.141, 80, image.buxmovie.com/screenshot/images/20486_2.jpg, http://www.buxmovie.com/
2014-02-18 09:19:00.209, 183.098.222.181, 4498, 221.157.125.141, 80, image.buxmovie.com/screenshot/images/20486_2.jpg, http://www.buxmovie.com/
Figure 11. Detected URL Log Sample

C. Detection of Intrusion Attacks

Lastly, the developed technology was tested to check unknown attack detection performance in a commercial environment. To verify performance at a commercial environment, two colleges and one small company in Korea were used for network dump, which was used because there may be problems of system stability and service availability when applying the developed system in an actual environment. Therefore, the dump file containing the actual network traffic without processing was analyzed. The dumped file was recreated using Avalanche. The traffic generated by the tested organizations for around two weeks from December 20, 2013 to January 3, 2014 were tested.

TABLE 3. -- MALICIOUS BEHAVIOR DETECTION RESULTS

	IP/URL	**Port**
Zombie PC	192.168.41.14	24157
	192.168.41.198	21145
	192.168.10.27	24154
	192.168.35.211	22141
	192.168.29.33	19875
	…	…
Malware Distribution URL or IP	http://zsd.hXXXX.or.kr	81
	http://upload.teXXX.com/jungsan	80
	…	…
C&C	64.149.84.XXX	80
	…	…

Files extracted from the traffic were analyzed to detect 12 known malicious code. The analysis of distributing URLs that downloaded the detected files confirmed 24 malicious code distribution data. At least 47 internal PCs downloaded the intruding malicious code, and 7 IPs accessed by these PCs to leak the information or receive the command were detected.

© 2015, IJIRIS- All Rights Reserved

V. CONCLUSION AND FUTURE RESEARCH

This paper has introduced an algorithm that analyzes the network traffic and collected data to detect an APT attack against a specific organization. The key functions of the introduced algorithm were verified in a commercial environment.

The system developed using the suggested functions and algorithm is expected to be applicable at public agencies, small companies, and hospitals using around 700Mbps network bandwidth. It can also prevent unknown attacks against the organizations using conventional IPS/IDS systems.

Note, however, that most companies and agencies process 1Gbps or higher traffic; thus, the suggested algorithm needs to be lighter to be applied in those organizations. Further research would be needed to process and analyze all data in a single system by overcoming the dependence on other systems for the inspection of collected files and URLs subject to inspection.

ACKNOWLEDGMENT

This work was supported by the ICT R&D program of MSIP/IITP. [10044938, The Development of Cyber Attacks Detection Technology based on Mass Security Events Analysing and Malicious code Profiling]

REFERENCES

[1] Ajay K. Sood, "Modern Malware and APT: What You May be Missing and Why'" AtlSecCon, March 2012.
[2] Giura.P, Wei Wang, "A Context-Based Detection Framework for Advanced Persistent Threats," Cyber Security 2012 International Conference, pp. 69-74, 2012.
[3] Mandiant, the Advanced Persistent Threat, M.Trends, 2010.
[4] Michael K. Daly, "The Advanced Persistent Threat," LISA `09.
[5] Ashit Dalal, "Advanced Persistent Threat(APT)-A Buzzword or an Imminent Threat?," ISACA `12.
[6] Wei Yan, Erik Wu, "Toward Automatic Discovery of Malware Signature for Anti-virus Cloud Computing," Complex Sciences, 2009, pp. 724-728.
[7] Peter Mell, Karen Kent, Joseph Nusbaum, "Guide to Malware Incident Prevention and Handing," Computer Security, NIST, 2005.
[8] Radoslav Bodo, Michal Kostenec, "Experiences with IDS and Honeypots-Best Practice Document," GEANT, 2012.
[9] Martin Lee, Daren Lewis, "Clustering Disparate Attacks: Mapping the Activities of the Advanced Persistent Threat,"Virus Bulletin Conference October 2011.
[10] Suchita Patil, Pallavi Kulkarni, Pradnya Rane, B.B.Meshram, "IDS vs IPS," IRACST, 2012, pp.86-90.
[11] Counter Threat Unit research, "Lifecycle of an Advanced Persistent Threat," Dell SecureWorks, 2012.
[12] Feng Xue, "Attacking the Antivirus," Black Hat Europe Conference, 2008.
[13] http://www.ajunews.com/view/20140820120217970
[14] Suchita Patil, Pallavi Kulkarni, Pradnya Rane, B.B.Meshram, "IDS vs IPS," IRACST, 2012, pp.86-90.
[15] Robert Drum, "ISD and IPS placement for network protection," CISSP, 2006.

© 2015, IJIRIS- All Rights Reserved