International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 2 (2019), pp. 264-268

Visualization of the Relationship between Malicious Code Distributing and Landing Sites

Jun Yeong Jeong
Department of Cyber Security Management,
Sangmyung University, Seoul, Republic of Korea.

Jinho Yoo[1]
Department of Business Administration,
Sangmyung University, Seoul, Republic of Korea.
Orcid: 0000-0003-4359-8009

Abstract

In recent years, we have been looking for ways to identify websites that first require fundamental countermeasures in order to prevent malicious code distributing. In this paper, we will analyze the relationship with infected sites that were not performed in conventional visualization analysis. The infected sites are divided into distributing and landing sites. Through visualization analysis, we will define the relationship between the distributing and landing sites and look closely at the characteristics of the infected sites to find the intentions of the attackers.

Keywords: Distributing sites, Landing sites, Malicious code, Visualization

I. INTRODUCTION

The number and range of new malicious codes detected every month is a good indicator of the trends of cybercriminals who create and distribute malicious code. Kaspersky Lab, a global cyber security company, said that the number of new codes detected per day by its technology was 70,000 in 2011, but increased to 360,000 in 2017, five times that. In addition, the number of new malicious codes detected in the first 10 months of 2018 reached 346,000 a day[1]. As such, the threat to malicious code infection is becoming very high for internet users.

Damages from malicious codes are becoming expanded in various forms such as Ransomware, Advanced Persistent Threat(APT). According to a report released by the Ransomware Computer Emergency Response Team Coordination Center in Korea, the number of Ransomware infection happened in 2018 was 1,290 in the first quarter, 1,037 in the second quarter, 906 in the third quarter, and 1,050 in the fourth quarter expectedly. In the case of Ransomware, the number of the cases infected in 2018 is 4,283, the total number of victims is 285,000 users, and the total damage amount is estimated to be 1.5 trillion won. Ransomware infections paths is accounted for 80% of the Internet, followed by 17% of e-mails. Looking at the infection status of Ransomware by industry, 43% of SMEs, 25% of small business owners and 22% of individuals were analyzed[2].

In recent years, we have been looking for ways to select sites that require fundamental countermeasures in order to solve this problem, as websites with diversified malicious code distribution channels and poor management tend to be abused as distributing and landing sites. There is not much research that has focused on analyzing the characteristics of malicious code distributing sites in a situation where the risk of malicious code distributing is increasing significantly. Therefore, this paper uses visualization method to analyze the characteristics of malicious code distributing sites and to find out the implications through relationship analysis of the distributing and landing sites that caused malicious code infection

II. PREVIOUS RESEARCH

Distributing site is a website that distributes and installs malicious code on the user's PC. Landing site is a website that leads to malicious code distributing sites[13]. It includes redirect codes that users cannot recognize or codes for preparation and execution to download malicious codes.

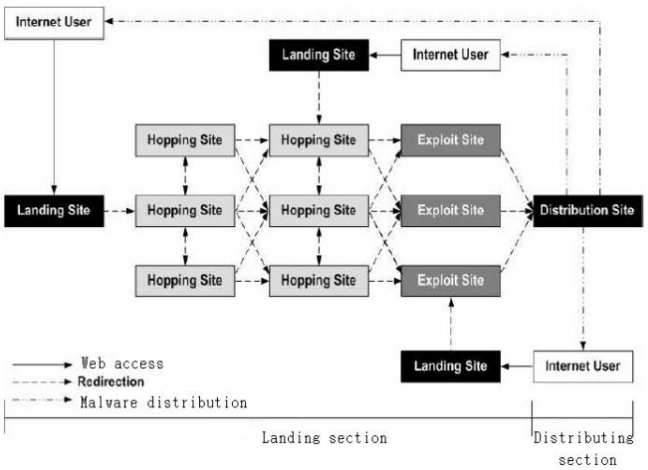

Fig.1. Distributing and landing routes of malicious codes

According to the analysis report of the malicious code distributing route of KISA(Korea Internet Development Agency), the distributing and landing routes are expressed as

[1] *Corresponding Author, E-mail: jhyoo@smu.ac.kr*

Figure 1. The malicious codes that started from the distributing site are connected to the landing sites, and when the user visits the landing sites, the malicious codes are downloaded to the user's PC from the distributing site. Attackers exploit and use distributing sites for expanding malicious codes. They continuously use distributing sites connected to a large number of landing sites and causes damage to user. Attackers may also work to use an infected landing site as a new distributing site. If the landing site infected with malicious code does not continue to take countermeasures against insufficient security measures or site vulnerabilities, user visiting the site can be infected and their assets can be stolen.

Ying Pan et al.(2006) suggested a way to detect phishing sites. Based on various criteria defined by characteristics of malicious code detected sites, they used anomaly detection method to judge the malignancy of other sites[7].

According to studies by Niels Provos et al.(2008), about 40% of malicious codes detected sites were accessed without a different route in the middle, that is, without going through a landing site, directly to the malicious code distributing site URL, and the remaining 60% reached the malicious code distributing sites through the landing sites[8].

To detect malicious code distributing sites, Dongwon Seo et al. (2018) introduced website relationship graphs, parallel coordination, and Amazon Web Service Systems, and sought to identify the pros and cons of each technique and the characteristics of malicious code distributing sites. The feature is that the top-level main page of the link does not exist or is often very simplely decorated, and the URL link is often an IP address rather than a domain name[6].

Sung-hyun Kim and Jinho Yoo (2017) predicted the likelihood of occurrence of malicious code infection by applying the change of malicious code infection to the Markov chain. They conducted a study to use the frequency of infection with malicious code as a basic data to estimate the probability[10].

Dong young Lee (2018) calculated the risk of the landing sites, and he developed new variables that can identify the infection characteristics and risk of the website to quantify the risk. Each variable analyzed how long the malicious code infection lasted and how much fast countermeasures were performed[11].

In Soo Song et al.(2010) quantified malicious code analysis and classification techniques and used parallel coordinates method to analyze and reclassify malicious codes. They analyzed the similarity between malicious codes and described changes and trends over time. They collected each attribute such as the identification value, type, infection route, installation route, and detected date of malicious code, and implemented network visualization. They identified patterns and correlations of malicious codes through multi-dimensional analysis[12].

Until now, previous researches has analyzed and classified the attributes of the malicious code itself, and presented malicious code pattern analysis with a visualization method. In this paper, we will analyze the relationship with infected sites that were not performed in conventional visualization analysis. The infected sites are divided into distributing and landing sites. Through visualization analysis, we define the relationship between the distributing and landing sites and look closely at the characteristics of the infected sites to find the intentions of the attackers.

III. RESEARCH METHODS

The following is the flow of this study: 1) Raw data collection, 2) URL parsing of the distributing and landing sites, 3) visualization analysis, 4) analysis of the characteristics of malicious code distributing sites. Detected data from the MCF(Malicious Code Finder) system of KISA(Korea Internet & Security Agency) were used for this research. The MCF system detects and analyzes the hidden malicious codes of distributing and landing sites. It also collects and stores URLs and distribution patterns on websites suspected of distributing malicious code from the source of the web page using the web crawling function. We used data of six months of malicious code distributing and landing sites detected from January 1, 2015 to June 30, 2015. During that period, the total number of malicious code distributing and landing sites was 30,134.

To express visualization, we conducted data parsing on raw data. Raw data was parsed after de-identification about domain names of sites. For parsing large amounts of raw data, Python (Windows 64bit Python 3.7 Version) was used.

Detected Date	Landing site	Distributing site
2015-06-30	http://e****.****.****	http://n****.****.****.*****
2015-06-30	http://l****.****.****	http://e****.****.****
2015-06-30	http://e****.****.****	http://n****.****.****
2015-06-30	http://m****.****.****	http://1****.****.****
2015-06-30	http://o****.****.****	http://b****.****.****
2015-06-30	http://s****.****.****	http://b****.****.****
2015-06-30	http://c****.****.****	http://s****.****.****
2015-06-30	http://m****.****.****	http://121.****.****.****
2015-06-30	http://m****.****.****	http://121.****.****.****
2015-06-30	http://ma****.****.****	http://121.****.****.****
2015-06-30	http://i****.****.****	http://121.****.****.****
2015-06-30	http://h****.****.****	http://121.****.****.****
2015-06-30	http://h****.****.****	http://121.****.****.****
2015-06-29	http://f****.****.****	http://3****.****.****
2015-06-29	http://s****.****.****	http://d****.****.****
2015-06-28	http://j****.****.****	http://s****.****.****
2015-06-28	http://e****.****.****	http://f****.****.****
2015-06-28	http://b****.****.****	http://1****.****.****
2015-06-27	http://s****.****.****	http://s****.****.****
2015-06-27	http://s****.****.****	http://e****.****.****
2015-06-27	http://b****.****.****	http://www.t****.****.****
2015-06-27	http://w****.****.****	http://h****.****.****
2015-06-27	http://b****.****.****	http://www.t****.****.****

Fig. 2. Example of Parsed Data Form

In this paper, the focus is on a detailed analysis of malicious code distributing sites. In the relationship between malicious code distributing and landing points, the central distributing sites are connected to many nearby landing sites. The more the landing sites are connected to one distributing site, the greater the influence and impact of the distributing site are. We used APIs of Intelje and D3 (https://d3js.org/) to analyze and visualize the parsed data.

IV. ANALYSIS AND RESULTS

The visualization of the relation from source to target is represented as shown in Figure 3. It means that source is the distributing site and target is the landing site. At the direction of the arrow, distributing sites are connected to landing sites. It means that the malicious codes transmitted from distributing sites to landing sites can be transmitted again to another secondary landing sites. H and G serve only as distributing site, but in the case of D, they serve as a landing site and another distributing site. This distributing and landing sites have their own characteristics and roles, and if they are not taken blocking action, the spread rate of malicious code would be higher. Since the spread of malicious code infections is repeated in various ways, rapid countermeasures to the distributing and landing sites are very important.

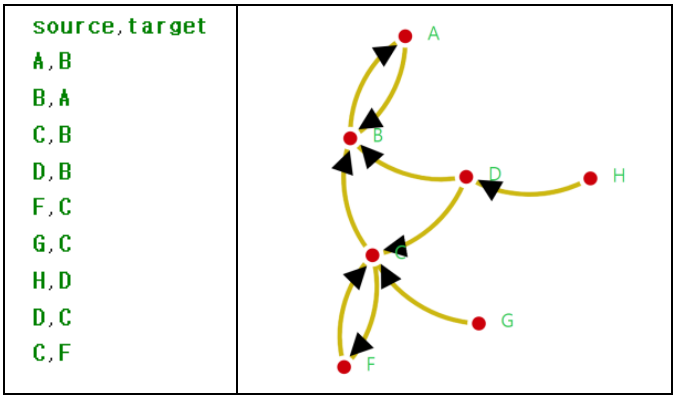

Fig. 3. An Example of visualization of the relationship from source to target

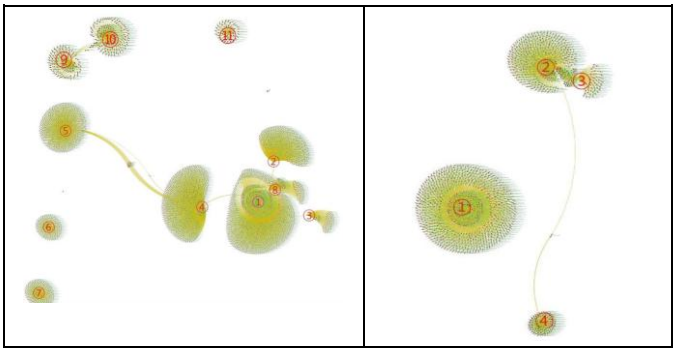

Fig. 4. Visualization of the relationship between malware distributing & landing sites

The result of visualization of distributing sites linked to more than 100 landing sites is shown in Figure 4. In January 2015, there were 11 distributing sites linked to more than 100 landing sites. It means that Attackers used 11 distributing sites as a source to disseminate malicious codes through connecting to more than 100 landing sites. Although relationship between mass distributing sites has not been shown significantly, some of them can be seen spreading to the same landing site. In February, the number of distributing sites linked to more than 100 landing sites was decreased to four, which means that some landing sites were detected and taken blocking action, but it means that the blocking action was not taken for distributing sites.

Monthly mass distributing sites is shown in Figure 5. In particular, site A(http://***. **cs/x.htm) was connected to 4,531 landing sites in January, and 1,613 landing sites in February, 376 landing sites in March and 207 landing sites in April. Despite the massive distributing of malicious code, it can be seen that countermeasures for the distributing site are not fully taken and are left infected. In May, it was somewhat hesitant, but it was not completely treated because it was distributing again to 119 landing sites in June. We can see that the fact that we are not doing well on very dangerous sites that are spreading malicious code in such a large amount is the reality that we are actually facing. Therefore, it is necessary to prevent the spread of malicious codes by treating these distributing sites first and more thoroughly and reducing the possibility of recurrence again.

Monthly change of malicious code distributing sites can be divided into six categories as Table 1. Six categories are as follows: ① over 100(connected to 100 and more landing sites), ② 20 to 99(connected to 20~99 landing sites), ③ 10 to 19(connected to 10~19 landing sites), ④ 4 to 9(connected to 4~9 landing sites), ⑤ 2 to 4(connected to 2~4 landing sites), ⑥ only 1(connected to only one landing site). ① ② ③ categories are shown in Table 1.

As shown in the table, there are new distributing sites that appear every month, and there are a lot of distributing sites that repeat next month and repeat again in the 3rd month. Repeating/repeating again for 2nd to 3rd month means that it is not taken action for infection treatment and is neglected. The monthly repeating of malicious code distributing is very dangerous, but it can be considered even more dangerous because the distributing sites are left infected and they can infect to user's PCs. The continued occurrence of the same domain every month means that the site is vulnerable and attackers are constantly using it as a malicious code distributing site. Therefore, it is necessary to take action and respond to these repeating sites first.

January		February		March		April		May		June	
Domain Name	count	Domain Name	count	Domain Name	count	Domain Name	count	Domain Name	count	Domain Name	count
http://*.***css/x.htm**	**4531**	**http://***.***css/x.htm**	**1613**	://*** ***.spamfree/index.htmlx.l	3339	http://***.***dd/1.js	708	http://***.***com.js	539	http://***.***product/index.html	483
http://***.***/index.htm	2253	p://***.***.***/view.jsoard/view	574	http://***.***.clinic/index.html	637	http://***.***2010/index.htmlml	376	http://***.***cc/index.html.html	447	http://***.***board/top.js	241
http://***.***/index.htm	757	ttp://***.***.***/view.jss/view.js	115	http://***.***/f/ck.gif	388	***.***function/index.html/index	349	http://***.***diary/index.html	410	http://***.***event/index.htmlhtm	147
http://***.***.***/top.js	595	http://***.***.***/index.html	100	**http://***.***css/x.htm**	**376**	http://***.***cms/index.htmlml	323	http://***.***board/top.js	406	**http://***.***css/x.htm**	119
http://***.***.***/s.gif	241	http://st***.***.***	54	http://***.***xe/index.html	344	ttp://***.***include/index.htmltm	241	***.***php/TTuwmmfC.phpmf(	358	http://***.***210/index.html	104
http://***.***.***/sty.htm	235	http://uh***.***.***	49	http://***.***conf/i/w.gifgif	251	**http://***.***css/x.htm**	207	ttp://***.***event/index.htmlhtm	273	http://21***.***.***	45
http://***.***.***/right.jsht.js	198	http://cs***.***.***	40	http://oh***.***.***	77	http://***.***keby.m.php	148	ttp://***.***include/index.htmltm	257	http://ww***.***.***	21
http://***.***.***/jp.js	191	http://ju***.***.***	38	http://ww***.***.***	71	http://***.***board/top.js	138	http://***.***product/index.html	248	http://hy***.***.***	19
http://***.***.***/index.html	178	http://fi***.***.***	34	http://19***.***.***	63	http://yg***.***.***	81	http://***.***kcp/ak.gifk.gif	200	http://21***.***.***	15
http://***.***.***/ck.gif	153	http://ww***.***.***	34	http://i-***.***.***	57	http://61***.***.***	73	http://***.***com/jp.js	168	http://ju***.***.***	15
p://***.***.***/view.html/view.h	131	http://i-***.***.***	32	http://s.***.***.***	51	http://my***.***.***	44	http://***.***2010/index.html	145	http://se***.***.***	12

Fig. 5. Monthly mass distributing sites

Table 1. Monthly change of malicious code distributing sites

Category	January	February	March	April	May	June
① over 100	New:11	New:2	New:5	New:7	New:10	New:4
						Repeat:0
					Repeat:3	Repeat:1
				Repeat:1	Repeat:1	Repeat:0
			Repeat:0	Repeat:0	Repeat:0	Repeat:0
		Repeat:2	Repeat again:1	Repeat again:0	Repeat again:0	Repeat again:0
② 20 to 99	New:36	New:8	New:24	New:17	New:14	New:7
						Repeat:0
					Repeat:6	Repeat:0
				Repeat:0	Repeat:0	Repeat:0
			Repeat:2	Repeat:0	Repeat:0	Repeat:0
		Repeat:5	Repeat again:2	Repeat again:0	Repeat again:0	Repeat again:0
③ 10 to 19	New:20	New:11	New:13	New:11	New:8	New:7
						Repeat:0
					Repeat:1	Repeat:0
				Repeat:1	Repeat:1	Repeat:0
			Repeat:0	Repeat:0	Repeat:0	Repeat:0
		Repeat:3	Repeat again:1	Repeat again:0	Repeat again:0	Repeat again:0

V. CONCLUSION

In this paper, we identified the characteristics of malicious code distributing sites through visualization analysis of the linkage between malicious code distributing and landing sites to find implications. The implications of malicious code visualization analysis are as follows. Monthly malicious code distributing sites are newly created, and they repeat and repeat again. If the new malicious code distributing sites are appropriately not taken action, attackers use the distributing sites as an easy and continuous tool to infect other sites and eventually infect the user's PCs visiting the infected landing sites to take away the user's assets or sensitive personal information.

The most widely distributing site A(http://***. **cs/x.htm) was consistently shown every month. It was identified as the high-risk distributing site that connected to many landing sites over six months. It is very important to manage distributing and landing sites because they will be used as a malicious code distributing source in cyberspace. Therefore, it is necessary to establish a response system for security through organic cooperation with government agencies and business organizations, and to block the spread of malicious code through prompt response after detection. If we do not quickly take blocking action on the distributing sites connected to a large number of landing sites, which can be used as a source to infect many users' PCs.

In this paper, there is a limit that only data detected through a specific MCF of KISA. Therefore, in the future, we'll have a plan to conduct research on the characteristics of the distributing sites with data acquired through various detection systems and find more meaningful implication.

Acknowledgements

This research was supported by a 2018 Research Grant from Sangmyung University.

REFERENCES

[1] https://www.dailysecu.com/?mod=news&act=articleView&idxno=42969

[2] https://www.boannews.com/media/view.asp?idx=74441

[3] http://www.zdnet.co.kr/view/?no=20180709113121

[4] https://byline.network/2018/05/3-13/

[5] https://www.samsungsds.com/global/ko/support/insights/111517_Org_Security1.html

[6] Dongwon Seo, Arindam Khan, Heejo Lee(2018), "A Study on Detecting Malcodes Distribution Sites," The Korea Information Processing Society Fall Conference, Volume 15, Volume 2, 2018.11

[7] Ying Pan , Xuhua Ding(2006), "Anomaly Based Web Phishing Page Detection," Computer Security Applications Conference, 2006. ACSAC '06. 22nd Annual, Dec.

[8] Niels Provos, Panayiotis Mavrommatis, Moheeb AbuRajab, Fabian Monrose(2008), Google technical report provos-2008a: All Your iFRAMEs Point to Us, Feb.

[9] Niels Provos, Dean McNamee, Panayiotis Mavrommatis, Ke Wang and Nagendra Modadugu(2007), The Ghost In The Browser: Analysis of Web-based Malware, Workshop on Hot Topics in Understanding Botnets (HotBots), Apr.

[10] Sung-hyun Kim, Jinho Yoo(2017), "A Study on Predicting Malware Infection in Websites Using Markov Chain", Journal of Security Engineering, Vol.14, No.1, pp.9-20

[11] Dong young Lee (2018), "A Study on Web Site Risk Classification by Malicious Code Infection Characteristics," Graduate School, Sangmyung University

[12] In Soo Song, Dong Hui Lee, Kui Nam Kim(2010), "A Study on Malicious Codes Grouping and Analysis Using Visualization", Journal of information and security, 10(3), pp.51 – 60

[13] KISA, Analysis Report on the Path of the Distribution of Malignant Codes, 2014

[14] KISA, Analysis of Large-scale malicious code distribution trend, 2014