Journal of Information Hiding and Multimedia Signal Processing ©2014 ISSN 2073-4212
Ubiquitous International Volume **5**, Number **2**, April **2014**

Detecting Hybrid Botnets with Web Command and Control Servers or Fast Flux Domain

Chia-Mei Chen, Ming-Zong Huang, Ya-Hui Ou

Department of Information Management
National Sun Yat-sen University
70 Lienhai Road, Kaohsiung, 80404, Taiwan
cchen@mail.nsysu.edu.tw

Received February, 2013; revised March, 2013

Abstract. *Botnet consists of one or more command and control servers (C&C Servers) and infected computers (bots), where the communication between the two parties often goes through a commonly used network protocol, such as HTTP. Web-based botnet attacks become more serious and popular recently, as hacker takes advantage of the HTTP connections hiding the malicious transmissions in a vast amount of normal traffic that is not easily detectable. In addition, integrating with fast-flux domain technology, botnet attack may use a web server to issue attack command and fast-flux to extend the lifespan of the malicious website. Fast-flux domain also achieves stealth by preventing users from making direct contact with the malicious websites. Thus, fast-flux domain technology is a cloak technology preferred by hackers, as it is often able to circumvent the detection based on filter or blacklist. Therefore, this study not only attempts to conduct anomalous flow analysis on web botnets, but also explores fast-flux domain on botnets. The proposed detection examines flow traffic and web domains to identify a botnet either using HTTP as communication channel or using fast-flux domain for cloaking. The experimental results prove that the proposed method can effectively identify these botnets.*
Keywords: Fast-flux domain, web-based botnet, malware

1. **Introduction.** The Internet has brought great convenience to the masses, making social networking, information delivery, and even online shopping indispensable to the lives of most people. In contrast, the pervasiveness of the Internet also makes some hackers envious of the economic effects brought together by the status quo. This has led to network security problems that are more severe and diffcult to prevent. Botnet is not a new technique for attack. Rather, it is a combination of various forms of malware, such as Trojan horse, virus, worm, and spyware. Hackers often modify bot program code to quickly develop a new botnet attack, creating a serious threat that is diffcult to prevent. Traditional firewalls, anti-virus software, and IDS (intrusion detection systems) are all ineffective against a botnet attack. Infections can occur through system vulnerabilities or through social engineering that can induce a user to click on a malicious image or website, which can automatically trigger the execution of a pre-set function. This could then result in the system being remotely controlled by the pre-defined set of commands. Once partial control of the bot machine is achieved, spam or other attacks are then launched to the detriment of the user. Methods for detecting botnets are mostly based on analysis of the behavioral characteristics of the attack. The methodology used in the attack must first be fully understood before any attempt at prevention can be made.

By using an HTTP connection as a communication channel, a web-based botnet attack can avoid detection by a firewall and increase the threat of the attack. One of the attack characteristics is its small traffic signature, which also fits perfectly well within the normal traffic flow. Since most firewalls do not filter HTTP traffic, it is therefore not easy to detect any abnormal behavior. According to security-related information provided by various research institutions, web-based botnets not only present serious threats to today's network security, they are also found to be using fast-flux domain technology for seeking routes of transmission. From the hacker's perspective, the advantage of fast-flux domain is that one domain name can have multiple IP addresses that can be used for rapid switching, which can keep the malicious website from being detected, and thereby extend its lifespan. This technique can not only block blacklist detection methods, but also benefits from load balance. Fast-flux domain technique allows a fully qualified domain name (FQDN) pointing to multiple IP addresses. These IP addresses can be scattered all over the world, making a malicious domain diffcult to be tracked and analyzed. Hackers can make a fast-flux domain constantly associate with various IP addresses. The machines with the associated IP addresses are usually victim hosts, so the malicious domain hiding behind these machines can circumvent detection. Attacks based on web-based botnet and fast-flux domain offer the benefit of nearly undetectability and rapid spreading. These attacks raise attack capability but also increase the diffculty of anti-hacking defense. In this study, a malicious web-based botnet detection mechanism is proposed with the ability of fast-flux domain detection.

2. **Related Work.** Since most malicious attacks use the fast flux domain technique to evade detection, many studies have focused on analyzing the characteristics of fast flux server domain (FFSN) to formulate a method for determining the presence of fast flux domains. Along with the ever increasing threat of web-based botnets, a plethora of information security issues have arisen from web-based botnets and FFSN. The purpose of this study is to detect web-based botnets while considering fast flux domain technology, by exploring the changing relationship between the two; thereby helping users achieve a more accurate detection of this type of attack. A botnet usually takes advantage of standard network protocols such as IRC or HTTP to remotely control victim terminals for spreading malware. The main reason for choosing HTTP is so that hackers can write control commands directly into the web program, which can easily allow a web-based botnet to be hidden inside the normal traffic flow so that it can remain undetected until the actual attack is launched [1,2]. Lee et al. [3]discover the abnormal behavior of a botnet and found the characteristic of degree of periodic repeatability (DPR) which a hacker using BlackEnergy as attack tool. They further used DPR to judge whether the host linked to the web server is a web-based bot. BlackEnergy is a web-based bot written by a hacker from Russia, and is designed to attack multiple IP or domain servers simultaneously. The packet content that it sends out is usually encrypted to avoid feature code detection by anti-virus software. Once an activation command is received, the bots carry out DDoS attacks against the target websites by following what is in the packet content. In determining the attributes of a malware, Lakhina and Crovella [4] proposed using a method called sample entropy to find the traffic flow distribution characteristics of source IPs, source ports, destination IPs, and destinations. The resulting flow diagram generated by using such a method shows that it can more easily detect a variety of attacks, such as DDoS and port number scanning. Although this method can be used more widely in a variety of botnet detections, its major disadvantage lies in that it really only works to reveal attacks that are in progress, but is unable to detect any botnet traffic immediately prior to these attacks. The method is not designed to block the connection in time to stop

any command and control servers (C & C server) from communicating with the zombie host. Wang [5] proposed that a pattern table be used to record the number of occurrences of various patterns, such as port, packets, and bytes, and then use the characteristics of these pattern table recordings for detection purposes. But these proposed patterns were made too broadly, generating too high a rate of false alarms. Fast-flux domain refers to the setting of shorter TTLs (time to life) for the DNS resource record (RR), and which uses a loop approach to replace a group of IPs that possess traffic directing functionality. After a TTL has expired, if the resulting IP information of the DNS is found to be different than before, then it is reasonable to suspect that the system may be using fast flux techniques. The characteristics of a detected fast-flux are very similar to Round-robin DNS (RRDNS) or Content Distributed Networks (CDN) in that they are all single server multiple IPs. However, the biggest difference with RRDNS and CDN is in the attempt to enhance its ability for recovery by modifying the DNS. RRDNS and CDN also use flux agents and redirected functionality. Some sites with heavier traffic, such as Google and Myspace, quickly map their domain names to various physical machines to achieve network load balance. Hackers aim to use fast-flux domains to extend the duration of the shielding of malicious websites. If it is not connected to malicious websites, a fast-flux domain does not pose a threat to the user. In terms of security issues, most studies aim to explore the characteristics generated by malicious websites using this technology; such as looking up IPs that match to multiple countries or that belong to multiple ASNs, and have a TTL of less than 600 s. The authentication mechanism proposed by Holz et al. [6] for detecting a fast-flux domain in a network is the use of double DNS queries to obtain the following three features of DNS response: non-duplicated IP addresses, the number of name servers (NS) and the number of autonomous system numbers (ASN). Holz et al. [6]did not use the TTL feature, because CDNs also exhibit the characteristic of possessing a short term TTL. Such a characteristic cannot distinguish clearly between a fast-flux domain and a CDN. Based on the above detection method, Zhou et al. [7] proposed two additional methods to improve detection. One method involves the use of simultaneous queries to check multiple DNS hosts to observe the number of non-duplicated IPs to reduce the time required to detect a fast-flux domain. The other is through cross-comparison of fast-flux domain detection results to accelerate query performance. Because many fast-flux domains share the same IPs, if the query results from a FQDN to be tested are similar to those of the known fast-flux domains, then the probability that the FQDN to be tested is a fast-flux domain is high. Although the detection methods proposed by these studies already have a good detection rate, some web sites that legitimately use fast-flux domain technology-such as pool.ntp.org and database.clamav.net-are still classified as malicious websites, thereby increasing the rate of false alarm. Passerini et al. [8] used even more features divided into three categories, as shown in Figure 1.

Passerini et al.8 believed that TTL is important in determining a fast-flux domain, in that a short TTL can quickly match to various different IPs. The classi

cation of domain information by Passerini et al. is not the same as those of other studies. Because hackers often use the personal information of victims or randomly generated names to register malicious domains, such classification can find other malicious domains based on the personal registration information that has already been detected from a fast-flux domain. Fast-flux domains are not limited to web applications. Any applications using DNS can also use fast-flux domains. However, presently, most fast-flux domains facilitate web services. Regardless of which service a fast-flux domain uses, its detection method and characteristics are identical [9]. Yu et al. [9] developed a system and proposed two detection methods to validate the effectiveness of a service. They are average online rate (AOR) and minimum availability rate (MAR). Once the existence of a fast-flux

Category	#	Description
Domain name	F_1	Domain age
	F_2	Domain registrar
Availability of the network	F_3	Number of distinct DNS records of type "A"
	F_4	Time-to live of DNS resource records
Heterogeneity of the agents	F_5	Number of distinct networks
	F_6	Number of distinct autonomous systems
	F_7	Number of distinct resolved qualified domain names
	F_8	Number of distinct assigned network names
	F_9	Number of distinct organisations

FIGURE 1. Fast-flux domain feature classification

domain agent is discovered, its activities are monitored every hour using calculations based on AOR and MAR.In a legitimate fast-flux domain, these activities should be under complete control for around the clock service, and its AOR value should be close to 100%. If a fast-flux domain is malicious, then its AOR value is significantly less than that of a legitimate domain. Similarly, its MAR value is also smaller than that of a legitimate domain. Thus, finding MAR can help identify whether a fast-flux domain is malicious. Yu et al. also developed an agent monitoring system that builds on an IP database. Each time a new IP is found, the system immediately determines if the IP address is suspicious, and its findings are documented. These findings showed that a fast-flux domain service network has some fixed characteristics that can be used for detection. This study uses the number of ASNs and the time of registration as characteristics for measuring fast-flux domains. The number of ASNs can be used to determine whether a website is using a fast-flux domain. Because a malicious website using a fast-flux domain exists for a very short period, the registration time can be used to narrow down the selection range. Furthermore, through observation, this study discovered a correlation between the A records, the IP reverse lookup of its domain name, and the original FQDN as features of a malicious website. A hacker may be able to decide the FQDN, but their use of fast-flux technology is subject to the following restrictions: Unlike a CDN, which can choose its own hardware device and a specific IP, a fast-flux domain cannot guarantee service up- time. In addition, the domain name of a reverse IP lookup is determined by the network administrator of that IP address. Therefore, the FQDN of a malicious website is not usually relevant to the domain name of a reverse IP lookup.

3. **Proposed Detection.** A fast-flux domain requires numerous physical machines to switch IP addresses. A botnet can satisfy this requirement perfectly, increasing the ability of hackers to spread malware. This study demonstrates that a fast-flux domain not only shields malicious websites, but also masks the C & C server from being detected. Hackers can then use the fast-flux domain to send commands, taking control of the entire botnet. Additionally, a fast-flux domain agent is simply a relay station that possesses the traffic-redirection features of port 80 or port 53. Therefore, a web-based botnet using HTTP as communication channel can more easily redirect users to malicious websites. The architecture in this study aims to develop a detection system which can identify web-based botnets or malicious website using fast-flux domain technology to evade the detection. The data sources include URL traffic and spam archive. Observed from spam archives, the majority of the malicious activities relate to fast-flux domains. Therefore, spam is often used to discover fast-flux domains. Traffic data with URL information provides the website address of each HTTP connection for distinguishing visiting pattern

of a machine which is an important attribute for web-based botnet as well as fast-flux domain. The proposed system architecture is described in Figure 2. In the following sections, the botnet detection mechanism is explained first, followed by fast-flux domain detection.

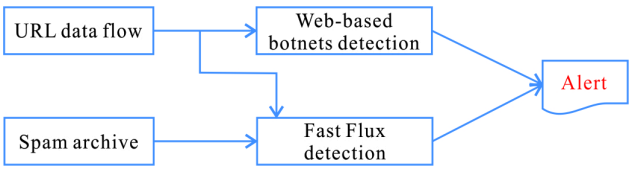

FIGURE 2. System architecture diagram

3.1. **Web-based Botnet Detection.** The web-based botnet attack model revealed that bot herder communicates with the server, the command & control server (C&C server), that issues commands and controls bots through HTTP connections. Based on the preliminary study conducted on Testbed TWISC, the HTTP connections exhibit the characteristics of periodic repeatability and identical webpage access. This study conducted preliminary experiments choosing web- based botnets from BlackEnergy and Zeus families. The results show that BlackEnergy does possess the characteristics of periodic repeatability and identical webpage accesses and can be identified through layer 4 flow traffic. The first sight of Zeus bots however seems no bot behaviors found, as its download time indicates a lack of regularity and the accessed web sites are not the same ones. When the URL data is taken into consideration, each web page (i.e., URL) visited by Zeus bots reveals regular browsing pattern, a characteristic of periodic repeatability. Based on the above preliminary experiments, the results demonstrate that using URL information as the basis for detection allows for more accurate web-based botnet detection. In addition, it can also improve the false alarms generated by a hybrid botnet, integration of web-based and fast-flux domain technology. As shown in Figure 3, initially the bot herder connects to the C&C server which facilitates fast-flux domain technology, turns it into a fast-flux domain agent, and establishes a connection between the two sides. Unfortunately, the detection based on IP address may fail to identify despite its fixed connection. It is also the case that since the server loses regularity of connectivity due to frequent changes of the IP addresses, which can cause the connectivity of the same original group to lose its grouping functionality as a result of dissimilar IPs. If URL information is used for detection based on groupings of FQDN in the accessed web pages, the detection can overcome the problem of dissimilar IPs and it can identify the connection regularity.

The proposed web-based botnet detection first screens out successful web connections which usually are issued by real clients, not bots. The connections to the same web server are grouped together for further examination. A normal web server often provides dynamic contents for users, while a C&C server's webpage remains the same for a period of time and bots repeatedly and periodically visit the same contents. Webpage attributes of each grouped traffic are extracted to discover the anomalous web connections. The attribute of webpage dynamicity indicates the degree of variance of a webpage and that of regularity rate represents the degree of periodic repeatability. A normal website usually provides rich and diversified contents for surfers. Therefore, the web links in a domain will be visited variously by different users. Due to the diversified contents and visitors, the first attribute, webpage similarity, is expressed by entropy, an information-theoretic statistic which measures the variability of the data. Let $f_1, f_2, \cdots, f_m$ be a group of flow

traffic on a domain X. $Sim(f_i, f_j)$ is the similarity of the ows with the selected fields which is suitable for measuring the first attribute.

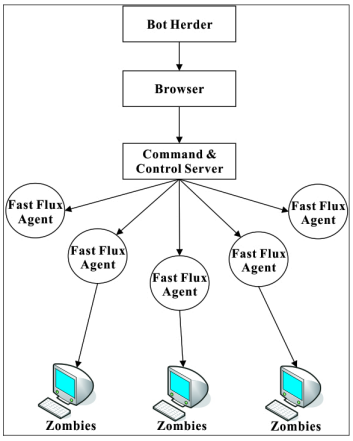

FIGURE 3. Web-based botnet and fast-flux domain integrated system architecture

The second attribute, regularity rate, is defined as $\frac{\sum d^x i}{\sum \max((d^x(i+1)-d^x i),1)}$, where $d^x i$ be the timespan of two consecutive flows,f_{i+1} and f_i. To avoid the two timespans are equal and the divider zero, a minimum value 1 is used in case they are the same. The reciprocal gives higher degree of regularity to small period discrepancy. Let G^X be a set of web connections accessing website $X, g^x i$ be the i th connection to website $X, d^x i$ be the timespan of two consecutive connections, $g^x i$ and $g^x(i+1)$, without loss of generosity, the i th connection is earlier than $i+1$ th's. Assume a flow traffic data contains m fields of information,$f_1, f_2, \cdots, f_m$. The flow data of connection $g^x i$ can be represented as ,$f_1(g^x i), f_2(g^x i), \cdots, f_m(g^x i)$. Let $f_t(g^x i)$ be the flow of connection $g^x i$ and $d^x i$ be the timespan of two consecutive connections, $g^x i$ and $g^x(i+1)$, where $p^x i = f_t(g^x(i+1)) - f_t(g^x i)$. The degree of periodic repeatability of a grouped traffic G^X , $R(G^X)$, is defined as $\frac{\sum p^x i}{\sum (p^x(i+1)-p^x i)}$. The attribute is normalized by its period and the reciprocal gives higher degree to small period discrepancy. The definition of webpage dynamicity is similar, but the rest of flow information is taken into consideration. The flow data remains the same, if the accessed webpage replies the same web contents.

3.2. **Fast Flux Domain Detection.** This study adopted command dig to get information about the associated IP and domain and to determine whether a domain name, FQDN, uses fast-flux. Furthermore, the A and NS records recorded in DNS are used to determine if the FQDN is legitimate or malicious. This study used dissimilar ANSs, the reverse lookup of the DNS, and the time of the domain registration as characteristics for distinguishing benign and illegal domains. Dissimilar ASNs (autonomous system number): Each IP has an ASN; IPs that are geographically close have the same ASN. The appearance of different ASNs indicates the presence of a potential fast-flux domain. As fast-flux domain agents, mostly victim's machines, spread all over the world, one IP may correspond to several dissimilar ASNs.

Though a legitimate site may also have multiple IPs, the regional distribution of their ASNs remains the same. The reason for having multiple IPs is to maintain a balanced load, thereby reducing traffic to the host. Most malicious websites use fast-flux domain to make multiple IPs appear in the same DNS resource records. However, this does not mean that all websites using fast-flux domain are malicious. In Figure 4, for example, www.ava**.com is a legitimate website. Its two groups of ASN numbers are 36351 and

21844, indicating that these host machines are divided geographically into two groups, and representing that fast-flux domain technology is used to place multiple IPs in the same DNS. Therefore, more features are required to determine malicious websites, rather than simply judging a website to be malicious based on the use of fast-flux domain.

www.ava**.com	2000	IN	A	67.228.*.*	(36351)
www.ava**.com	2000	IN	A	74.55.*.*	(21844)
www.ava**.com	2000	IN	A	74.55.*.*	(21844)
www.ava**.com	2000	IN	A	74.55.*.*	(21844)
www.ava**.com	2000	IN	A	74.55.*.*	(21844)
www.ava**.com	2000	IN	A	74.86.*.*	(36351)
www.ava**.com	2000	IN	A	174.36.*.*	(36351)
www.ava**.com	2000	IN	A	174.36.*.*	(36351)
www.ava**.com	2000	IN	A	174.37.*.*	(36351)
www.ava**.com	2000	IN	A	174.37.*.*	(36351)
www.ava**.com	2000	IN	A	174.123.*.*	(21844)

FIGURE 4. ASN in A record of www.ava**.com

Reverse lookup of DNS server: DNS server records the DNS resource records of the domain. The reverse lookup of a FQDN in that domain should match with that of the DNS server. The domain name can be used to determine if the reversed domain name has any correlation to the original domain name. If it is correlated, then it is judged to be a legitimate site. Conversely, illegal websites often have uncorrelated domain names, indicating that they do not belong to the same domain system. As shown in Figure 5, for example, the reverse IP lookup in the A record of a legitimate FQDN www.ava**.com generates the same domain name, www.ava**.com. Conversely, wave***.com, which is the reverse IP lookup in the A records of an illegal website www.tax.sta**-ca.net is fundamentally unrelated to the original domain name.

www.ava**.com	2000	IN	A	67.228.*.*	a***sl.ava***.com
www.tax.sta**-ca.net	180	IN	A	24.113.*.*	24.113.*.*.wave***.com

FIGURE 5. Reverse IP lookup corresponding to legitimate and malicious websites

Registration Time: The time for a fast-flux domain to stay active is 18.5 days on average. Domain names with a long survival time can usually be judged as legitimate sites. Figure 6, for example, shows that legitimate sites have a relatively long registration time, while most malicious websites have short registration times.

taipei****.com	benign	Creation Date: 06-dec-1998
ava****.com	benign	Creation Date: 06-oct-1997
jx2d****.com	malicious	Creation Date: 19-may-2012
lit****.com	malicious	Creation Date: 20-jul-2012
roko****.com	malicious	Creation Date: 12-jun-2012

FIGURE 6. Registration time comparison between legitimate and malicious websites

The above attributes are used for detection. The attribute values from benign and malicious websites were studied and evaluated to find good threshold for better detection rate. Therefore, Bayesian probability theory is chosen as the analysis method to explore the impact of parameters on the result predictions. Number of dissimilar ASNs: Based on our study, a legitimate site usually has only one ASN, but there are some legitimate sites with one to three dissimilar ASNs. However, illegal sites often have multiple dissimilar ASNs. As the discreteness of the attribute, the definition of ASN parameter,w_1, use a step-wise function: If a website has more than three dissimilar ASNs, it is judged to be

illegal. One with one to three ASNs is given a figure between 0 and 1. Based on our experiments, 0.2 exhibits a better performance.

$$w_1 = \left\{ \begin{array}{cc} 0, & \text{if } |ASN| = 1 \\ 0.2, & \text{if} 1 < |ASN| \leq 3 \\ 1, & \text{if } |ASN| > 4 \end{array} \right\}$$

Reverse IP lookup: This study applies the domain name obtained from reverse lookup for matching the domain name of FQDN to calculate the degree of similarity as the basis of web link similarity, where D_x denotes the domain name of FQDN X and D_y is its reverse domain. Each domain consists of a set of tokens (aka level domains) delimited by dots. The above two domains are compared in terms of the corresponding level domains and LCS (Longest Common Subsequence) is applied. Let the number of level domains be n, so the degree of domain name similarity is normalized by n. In summary the feature is defined as follows. $w_2 = LCS\frac{(D_x, D_x)}{n}$. Registration time: Malicious website typically has been registered for less than one month. Nevertheless, some malicious websites use fast-flux domain technology to evade detection and may have longer time. Based on our observation, one year is considered to be a trustful domain and less than one month is not. The registration time in between does not have significant difference and hence a step-wise function, w_3, is defined to represent the weight of the registration time to the legitimacy of a domain. The parameters of the step-wise function can be tuned based on training data and the range is between 0 and 1.

$$w_1 = \left\{ \begin{array}{c} a = \frac{registrarage \leq 1mouth}{maliciousnumber} \\ b = \frac{registrarage \leq 1year}{maliciousnumber} \\ c = \frac{registrarage > 1mouth}{maliciousnumber} \end{array} \right\}$$

The summation of the above three attribute values gives an anomaly score of the inspected domain name.

4. **Performance Analysis.** This study collected a list of legitimate websites from the 1,000 most renowned enterprises, and that of malicious websites derived from spam for performance evaluation. The legitimacy of the websites is inspected by McAfee [10] and other monitoring services, including a Blocklist Removal Center [11], a WOT [12], and Free PC Security [13], as each monitoring service checks for different security threats, spyware, phishing, malware, and other security threats. McAfee applied reputation analysis detecting the above mentioned threats actively. Its results are labeled as 'mark1' and other monitoring as 'mark2'.

4.1. **Web-based Botnet Detection and Experimental Analysis.** This study divided the experiments into close-ended and open-ended. A close-ended experiment used Zeus Bot as a subject of detection and open-ended used university campus website traffic for detection evaluation.

4.1.1. **Close-ended Experiment**. The close-ended experiment was carried on testbed environment, Testbed@TWISC [14], which simulated a real network environment and conducted attack scenarios there. The experimental network is illustrated in Figure 3, including three infected Zeus bot machines and a C&C server. The server acts as a web server communicating and controlling the bot machines and they connect to the server in a periodic way. In addition to sending out actual traffic from the bots, the communication is also mixed in with malicious traffic to test whether the system can detect the malicious

traffic sent by the bots. The results show that the proposed system can identify the malicious traffic efficiently, in a short period of time.

4.1.2. **Open-ended Experiment**. The open-ended experiment used network traffic in university dormitories and administrative units as experimental subjects, which include 13 dormitories and 19 administrative units that receive internal or external network traffic. The purpose of the experiment is to evaluate the applicability of the proposed method to the real network. The results show that the proposed system is able to identify the possible threats. For example, the study further analyzed the suspicious traffic, alerted by the proposed system, between the website http://*.*.*.*/download/echo.php and $140.X_9.Y_7.Z_6$ and verified the alert results. The reverse IP lookup of website http://*.*.*.*/download/echo.php matched the IP address $38.113.Y_1.Z_3$, which was determined to be the connection of http://$38.113.Y_1.Z_3$/download/echo.php. Malware URL detection results from other monitoring services showed that http://$38.113.Y_1.Z_3$/download/echo.php was a malicious website, which is consistent with our detection results. Another real attack incident identified by this study was the traffic between malicious website http://akakalat.com and $140.117.Y_3.Z_1$, which has lasted for 17 hours a day. Again, it was verified by McAfee forensic analysis as well.

4.2. **Fast Flux Domain Experiment and Analysis.** At this stage of the experimental analysis, the discussion focuses on the spam archive of malicious websites and legitimate campus websites as the basis for evaluating system performance. The monitoring report from McAfee website was considered as a reference and labeled as test result mark1, while that from Blocklist Removal Center, WOT, and Free PC Security as mark2. Spam archive: The following table was determined to contain malicious websites that had used fast-flux technology. In mark1 and mark2 of Figure 7, the forensics show mostly malicious websites, with similar system analysis results.

ID	URL	source	Mark1	Mark2
1	sigo****.com	spam	Malicious	Malicious
2	wvisit****osr29.com	spam	Malicious	Malicious
3	goph****good13.com	spam	Malicious	Malicious
4	hot****line10.com	spam	Malicious	Malicious
5	visit****sr23.com	spam	Malicious	Malicious
6	vph****.daa.tk	spam	Malicious	benign
7	ruy****.com	spam	Malicious	Malicious
8	medphar*****count9.com	spam	Malicious	Malicious
9	medphar*****count2.com	spam	Malicious	Malicious
10	bestgood****24.com	spam	Malicious	Malicious

FIGURE 7. System-determined malicious websites in the spam archive

4.3. **Fast Flux Domain Experiment and Analysis.** At this stage of the experimental analysis, the discussion focuses on the spam archive of malicious websites and legitimate campus websites as the basis for evaluating system performance. The monitoring report from McAfee website was considered as a reference and labeled as test result mark1, while that from Blocklist Removal Center, WOT, and Free PC Security as mark2. Spam archive: The following table was determined to contain malicious websites that had used fast-flux technology. In mark1 and mark2 of Figure 7, the forensics show mostly malicious websites, with similar system analysis results.

Spam archive: The following table was determined to contain malicious websites that had used fast-flux technology. In mark1 and mark2 of Table 2, the forensics show mostly malicious websites, with similar system analysis results.

Campus network: Figure 8 shows the websites in the campus network judged to be legitimate by the system. The McAfee website monitoring report and the forensic analysis of security validation results of Blocklist Removal Center, WOT, and Free PC Security also showed that the sites were legitimate.

ID	URL	source	Mark1	Mark2
1	sigo****.com	spam	Malicious	Malicious
2	wvisit****osr29.com	spam	Malicious	Malicious
3	goph****good13.com	spam	Malicious	Malicious
4	hot****line10.com	spam	Malicious	Malicious
5	visit****sr23.com	spam	Malicious	Malicious
6	vph****.daa.tk	spam	Malicious	benign
7	ruy****.com	spam	Malicious	Malicious
8	medphar*****count9.com	spam	Malicious	Malicious
9	medphar*****count2.com	spam	Malicious	Malicious
10	bestgood****24.com	spam	Malicious	Malicious

FIGURE 8. System-determined malicious websites in the spam archive

In addition, Figure 9 shows the data of the campus network judged to be malicious websites. In addition to the determination of bayjail.ru as a malicious website, other websites were judged to be malicious by mark1, but benign by mark2. The main reason was because the sites all belonged to underground gambling sites, and the mark2 authentication mechanism does not necessarily define websites that use fast-flux techniques as malicious.

ID	URL	source	Mark1	Mark2
6	ya***.com.tw	spam	benign	benign
5	tiny***.com	spam	benign	benign
4	pa***.com	spam	benign	benign
13	snigs****.file.word***.com	spam	benign	benign
16	mic****.com	spam	benign	benign
17	feed****.google.com	spam	benign	benign
18	rapid****.com	apam	benign	benign

FIGURE 9. System-determined legitimate websites in the university campus network

URL	source	Mark1	Mark2
www.15***.tk	brick	malicious	benign
www.ligh***.tk	brick	malicious	benign
www.sky***.tk	brick	malicious	benign
www.09***.tk	brick	malicious	benign
www.89***.com	brick	malicious	benign
bay***.ru	brick	malicious	malicious
www.qq***.tk	brick	malicious	benign

FIGURE 10. System-determined malicious websites in the university campus network

4.4. **Performance.** In analyzing performance, web-based botnet detection is designed to verify the malicious nature of suspected botnets for filtering actions. Fast-flux domain detection is based on the results of two verification mechanisms for comparison with the systematic analysis results of this study for performance assessment. In terms of the experimental results, precision refers to the ratio between websites determined to be malicious by the system, and websites actually verified as malicious. Data show that the performance of the campus network is lower than that of the spam archive. The cause of this misjudgment was as mentioned before: that the security authentication of the Blocklist Removal Center, the WOT, and Free PC Security had misjudged underground

gambling sites as legitimate sites. In terms of accuracy, the correct determination of legitimate and illegitimate websites attained an accuracy of up to 97 % or more, which ensured that under web-based botnets in conditions of fast-flux domain, the system could still achieve a very high filtering efficiency.

Assessment category	Spam archive	Campus network experiment	Overall system performance
Precision	97%	67%	88%
Recall	98%	81%	94%
Accuracy	97%	98%	98%
Miss rate	5%	1%	2%

FIGURE 11. Overall performance

5. **Conclusions.** Apart from using the bot program to identify web-based botnet characteristics, this study also used it to find malicious websites that use fast-flux domain technology by identifying its characteristics. In web-based botnet detection, this study used previous literature as the basis for modifying IP-based detection methods to help improve its accuracy. Apart from solving the botnet problem from a programming perspective, the role that botnet played in benefiting hackers also provided us with the direction for solving the botnet problem. Fast-flux domain requiring numerous IPs is a characteristic that makes it closely related to botnet, so that detection of fast-flux domain techniques can indirectly solve the botnet problem. Previous botnet detection methods mostly started from the bot program or malicious traffic, and did not solve the botnet problem from a different angle. Therefore, this study proposed to use connection regularity as the basis for web- based botnet detection, and combined it with fast-flux domain detection. In addition to enhancing the accuracy of detection, it can also detect different types of botnet. The longer-term goal of this study is for the system to integrate with IPS or Layer 7 firewalls. The output results of this system include IP and FQDN, which writes problem IPs into IPS (intrusion prevention system) rules, or writes FQDN into Layer 7 firewalls so that it can block the connection to suspected botnet. IPS or Layer 7 firewalls have also been widely used in local area networks. If they were integrated with the system of this study, the host in the local area network could be protected from botnet threats. Additionally, IPS or Layer 7 firewalls are scalable, which can be integrated with various malicious software detection methods to achieve the most rigorous protection.

REFERENCES

[1] G. Gu, J. Zhang, and W. Lee, BotSniffer: detecting botnet command and control channels in network traffic *Proc. of the 15th Annual Network and Distributed System Security Symposium*, 2008.

[2] M. Polychronakis, P. Mavrommatis, and N. Provos, Ghost turns zombie : exploring the life cycle of web-based malware, *Proc. of the 1st Usenix Workshop on Large-Scale Exploits and Emergent Threats*, 2008.

[3] J. S. Lee, H. C. Jeong, J. H. Park, M. Kim, and B. N. Noh, The activity analysis of malicious HTTP-based botnets using degree of periodic repeatability, *Proc. of International Conference on Security Technology*, pp. 83-86, 2008.

[4] A. Lakhina, M. Crovella, and C. Diot, Mining anomalies using traffic feature distributions, *Proc. of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*, pp. 217-228, 2005.

[5] Y. T. F. Chan, C. A. Shoniregun, and G. A. Akmayeva, A netflow based internet-worm detecting system in large network, *Proc. of The 3th International Conference on Digital Information Management*, pp. 581-586, 2008.

[6] T. Holz, C. Gorecki, F. Freiling, and K. Rieck, Measuring and detecting of fast-flux service networks, *Proc. of the 15th Annual Network and Distributed System Security Symposium*, 2008.

[7] C. A. Zhou, C. Leckie, and S. Karunasekera, Collaborative Detection of Fast Flux Phishing Domains, *Journal of Networks*, vol. 4, no. 1, pp. 75-84, 2009.
[8] E. Passerini, R. Paleari, L. Martignoni, and D. Bruschi, FluXOR: detecting and monitoring fast-flux service networks, *Proc. of the 5th international conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, LNCS 5137, pp. 186-206, 2008.
[9] S. Yu, S. J. Zhou, and S. Wang, Fast-flux attack network identification based on agent lifespan, *Proc. of the 2010 IEEE International Conference on Wireless Communications, Networking and Information Security* , pp. 658-662, 2010.
[10] McAfee, Available at: http://www.siteadvisor.com.
[11] SPAMHAUS, Available at: http://www.spamhaus.org/lookup.lasso.
[12] WOT, Available at: http://www.mywot.com/.
[13] Free PC Security, Available at: http://www.freepcsecurity.co.uk/.
[14] Testbed @ NCKU, Available at: https://testbed.ncku.edu.tw.