Detection of Fast-Flux Domains

Chia-Mei Chen, Sheng-Tzong Cheng, and Ju-Hsien Chou

***Abstract*—Botnets create harmful attacks nowadays. Lawbreaker may implant malware into victim machines using botnets and, furthermore, he employs fast-flux domain technology to improve the lifetime and robustness of botnets. To circumvent the detection of command and control servers, a set of bots is selected to redirect malicious communication and hides the communication within normal traffic. As the dynamics of fast-flux domains, blacklist mechanism is not efficient to prevent fast-flux botnet attacks. It would be time consuming to examine the legitimacy of the domains of all the connections. Therefore, a lightweight detection of malicious fast-flux domains is desired. Based on the time-space behaviors of malicious fast-flux domains, the network behaviors of domains are formulized in this study to reduce the time complexity of modeling features. According to the experimental results, the malicious fast-flux domains collected from the real networks are identified efficiently and the proposed solution outperforms the blacklists.**

***Index Terms*—Botnet, fast-flux domain, malware, command and control server.**

I. Introduction

The topology of a 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. HTTP-based botnet attacks become more serious and popular recently. Lawbreaker (aka bot herder) builds a web server as the C&C server, hiding and blending malicious transmissions in a vast amount of normal web traffic. C&C server plays a vital role in a botnet, as it contains the information of the bot machines and controls the attacks. If the C&C server is controlled by law enforcement, the botnet can be taken down accordingly because the bot machines would report to the server regularly. To invade the detection of C&C servers, a bot herder adopts fast-flux [1]-[3] technique to extend the robustness of botnet. Based on the structure of fast-flux domain as shown in Fig. 1, fast-flux agent acts as a relay station between the website and the client site. Fast-flux domain, mapping the domain to one of the flux agents dynamically, achieves stealth by preventing users from making a direct contact with the malicious website and making the detection of C&C servers difficult. Thus, fast-flux domain technology is a cloak technology preferred by bot herder to circumvent detection.

Manuscript received December 9, 2012; revised March 11, 2013.

Chia-Mei Chen is with the Department of Information Management, National Sun Yat-sen University, Kaohsiung, Taiwan, R.O.C. (email: cchen@mail.nsysu.edu.tw)

Sheng-Tzong Cheng and Ju-Hsien Chou are with the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan, R.O.C (email:stcheng@mail.ncku.edu.tw, p7896127@mail.ncku.edu.tw)

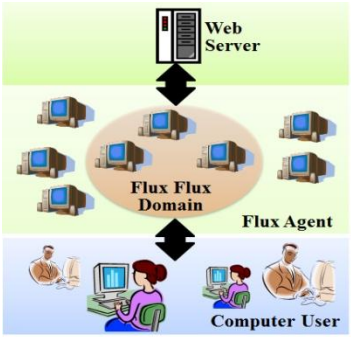

Fig. 1. Structure of a fast-flux domain

Some researches [2]-[5] identify the discrepancy of network behaviors between malicious and normal fast-flux domains. To balance traffic load, fast-flux domain distributes the traffic to a set of flux agents. Therefore, the record Time-to-Live (TTL) of a normal fast-flux domain query is short in order to direct the traffic efficiently. Famous websites, such as Google, Twitter, and Facebook, often employ such scheme to provide a better quality of service for the large amount of user requests, while bot herder deploys fast-flux domain for cloaking malicious traffic. Further analysis of the network behaviors should be examined to identify distinct attributes to classify the fast-flux domains.

As network generates large network connections each day, an efficient and light-resource detection method is needed to examine all the domains requested. To reduce resource demands and to improve detection time, the proposed approach models the network behaviors of domains into formulas by mathematical analysis of benign and malicious fast-flux domains.

This article focuses on the computation and time performance to develop a lightweight malicious fast-flux domain detection mechanism. The various features of network behaviors are digitized and formulated such as the relationship of flux agents, the distribution of IP addresses from flux agents, and the domain age in order to simplify the steps of the detection of domain. Finally the weight functions of each formula are given to determine the malicious fast-flux domain exactly.

II. Related Work

Nazario [2] discovered many particular features such as domain names, active lifetimes and relationships in malicious fast-flux domains. FluxOR [4] was to identify malicious fast-flux domains by three categories (domain name, the availability of network, and the heterogeneity of agents) and nine features of DNS/IP information. Huang [6] applied IP addresses to find the actual geographical coordinate distribution to determine the malicious fast-flux

DOI: 10.7763/JACN.2013.V1.30

domain. Hsu [7] considered the stability of transmission latency time as C&C flux agents may have variant network bandwidth and dynamic response time. Zhou [8] queries collected IP addresses among multiple DNS servers as well as to compare the results of DNS RR at the same time so that collaborative detection is more efficient.The network behaviors of malicious fast-flux domains are summarized as follow:

Time relativity of malicious fast-flux domain: The latency (sleep) period of malicious fast-flux domains [2] is the significant delay between registration and use in a fast-flux operation, which is about one week to months. Besides, the active lifetime of malicious fast-flux domain [1], [2], [4] is about 18.5 days to five weeks.

The number of flux agents: A malicious fast-flux domain links to different flux agents and changes them often to evade detection or blacklist blocking. The enough flux agents ([4], [8], [9], at last eight in [9]) are collected to determine the domain legitimacy. Comparing with the malicious, the benign only have a limited number of flux agents for load balancing.

The distribution of flux agents [2], [4]: An IP address implies valuable information such as autonomous system number (ASN), reverse lookup domain, and location. It is useful for distinguishing the malicious from the normal. A reputed domain uses multiple flux agents of the same ASN, while a malicious one randomly selects its agents from the compromised machines. In Table 1, the flux agents of a benign domain, Twitter, are from the same provider, ASNs and reverse lookup domains, while those of a malicious one have all different information.

TABLEI: THE FLUX AGENTS OF A REPUTED DOMAIN AND A MALICIOUS ONE

twitter.com				thesinglestoday.com			
IP Address	ASN	CC	Reverse Domain	IP Address	ASN	CC	Reverse Domain
199.59.148.10	13414	US	twttr.com.	85.*.221.*	13110	PL	icpnet.pl.
199.59.148.82	13414	US	twttr.com.	87.*.240.*	9116	IL	NULL
199.59.150.7	13414	US	twttr.com.	123.*.129.*	4134	CN	163data.com.cn.
199.59.150.39	13414	US	twttr.com.	123.*.229.*	4780	TW	tbcnet.net.tw.
199.59.149.230	13414	US	twittrr.com.	222.*.31.*	4766	KR	NULL

The computation resources needed by [8] are higher to query implement two mechanisms at the same time. The work [7] must maintain stable network quality to record the delay caused by domain and to prevent the inaccuracy of the delay time from flux agent queries.

Researches [6], [8], [10]-[12] inspected fast-flux domain based on the information collected and stored in a database. Referencing database required a certain amount of storage for processing and the information should be up-to-date or the detection rate might be affected.

Flux score [9] proposes an equation to judge the existence of fast-flux domain accurately.

$$f(x) = 1.32 \cdot n_A + 18.54 \cdot n_{ASN} + 0 \cdot n_{NS} \qquad (1)$$

n_A, n_{ASN} and n_{NS} represent the number of IP addresses, ASNs, and NS of domain *x*, respectively. Based on equation (1), a domain with $f(x)$ higher than 142.38 is classified as malicious. It needs to have at least 8 IP addresses located in 8 different ASNs to over the threshold. Our preliminary experiment indicates that a malicious fast-flux domain has average 4.9 IP addresses and 4.2 ASNs in one NS query. We also found that Flux score required the process time of 3 minutes to 6 hours, average 38 minutes, to classify one domain, which might not be practical for a real network.

This study focuses on the computation and time performance to develop a lightweight malicious fast-flux domain detection mechanism.

III. DETECTION FUNCTION

A. Finding Network Behaviors

Commands, dig and whois, are applied for collecting the network behavior of a domain. Command dig is to get not only domain information, such as IP address, NS and TTL, but IP information including reverse lookup domain, ASN, and country. Command whois returns the information of domain, administrator, and creation date. The query results are shown in Fig. 2. To reduce the processing time, the proposed detection system only issues the query commands 5 times in a short time frame s to collect enough IP addresses from the target domain, where s is set to 5 second in our evaluation experiments.

To analyze the discrepancy between the benign and the malicious, top 500 websites from Alexa[13] represent benign domains and 108 malicious from ALTAS [14] are considered in different network attributes including the number of IP addresses, ASNs, mapped countries, reverse domains and domain age. Asshown in Fig. 3, 70% of the benign use a single IP address, while the malicious have at least 4 IPs and over 60% have 5. 90% benign have single ASN and benign with multiple IP addresses mostly affiliate to one ASN, just like Twitter in Table 1. Contrarily, malicious have at least 4 ASNs. 98% of benign are set in single country even though some have multiple IPs, while all malicious have more than one. The distribution of the reverse lookup domains of benign, shown in Fig. 3(d), is similar to that of ASN, shown in Fig. 3(b). The age of all the benign is more than one year old, while all malicious are less than four months. There are significant discrepancies between benign and malicious. The results indicate that the decision model does not need many attributes to improve the detection efficiency because of the significance of the attributes examined.

B. Formulaic Features

Massive network flows throughout networks and a large volume of connection requests should be examined for malicious fast-flux domains every day. Thus an efficient and light-resource detection method is desired to examine all requests in real-time. To reduce resource demands and process time, the proposed approach adopts only three attributes and models the network behaviors of domains into formulas by mathematical analysis of attributes described above. For our best knowledge, this study is the first attempt to adopt formulaic attributes for malicious fast-flux domain detection.

To examine a large scale of network traffic, this study identify few features which have most significance on distinguishing malicious and benign fast-flux domains so that the detection consumes less computing resource and becomes responsive. Based on our observation and literature

review, a malicious fast-flux domain has multiple flux agents located at different geographic locations, as they are bots controlled by the bot master and it may have been dormant for a short period of time before it operates online. The above key behaviors will be the basis for formulating our novel features. Our preliminary study concluded that the malicious fast-flux domains all have more than 3 IPs (flux agents). To reduce the process time, only the domains with more than 3 IPs ($N_{IP} > 3$) will be examined in the system. The proposed three formulated features are described below.

```
$ dig +short google.com
173.194.72.139173.194.72.102173.194.72.100
173.194.72.113173.194.72.101173.194.72.138
$ dig +short 113.31.125.74.origin.asn.cymru.com
TXT
"15169 | 74.125.31.0/24 | US | arin | 2007-03-13"
$ dig +short -x 173.194.72.113
tf-in-f113.1e100.net.
$ whois google.com | grep -E 'reat'
   Creation Date: 15-sep-1997
```

Fig. 2. An illustration of query results of dig and whois

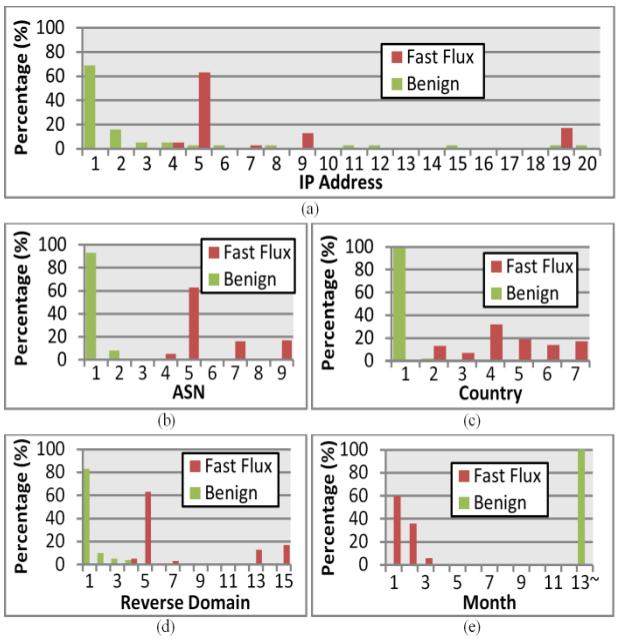

Fig. 3. Network behaviors of benign and malicious fast-flux domains

w_1: The number of countries

The flux agents of a fast-flux domain are distinct IP addresses. Even if the number of flux agents of a benign domain is as many as that of a malicious one, the agents of the benign are located in the same country. Those of malicious are not as they are victims worldwide compromised by attacker. Thus the number of countries is considered as a significant feature. Our preliminary study shown in Fig. 3 also verified the phenomenon. Feature w_1 is one, if the agents of the domain are located at more than one country; otherwise, it is zero. Feature w_1 is expressed as follow:

$$w_1 = \begin{cases} 0, & C_{country} = 1 \\ 1, & C_{country} \geq 2 \end{cases} \quad (2)$$

w_2: The relationship of ASNs to the flux agents

The preliminary analysis shows that the probability that a domain with $N_{IP} < 4$ is malicious is very small and that a benign domain often returns the same and few flux agents in Fig. 3(a). To identify fast-flux domains efficiently, feature w_2 is calculated only if the flux agents collected within a short time frame s is more than 3.

Some benign domains have multiple flux agents, whose IPs are affiliated to the same or few network organizations (i.e. one or few ASNs). Contrarily, the flux agents of the malicious have many ASNs because ofthe bots located worldwide and selected randomly. An extreme, but often occurred, situation for a malicious fast-flux domain is that every of its flux agents is belonged to a different ASN; one to one mapping from the flux agents to the ASNs. Even if the malicious mimic the benign and response only few flux agents, the mapping still shows its anomaly. This study discovered that there is an inverse proportional relationship the likelihood of a malicious domain and its number of agents associated with its ASNs. Thus the reciprocal of the number of flux agents in the same ASN is adopted to express the likelihood of a domain being malicious.

In real situations, some organization's networks are administrated by or subscribed from ISP. Contrarily, the malicious fast-flux domains have most flux agents come from different ASNs and original network owners. Therefore, this study combines the multiple ASNs with the same reverse lookup domain to one ASN.

This study discovered that the chance that a domain is malicious reduces if its ASN is associated with multiple agents and developed feature w_2 based on the above findings. As the number of the flux agents obtained from queries within a fix short time period may vary, the novel feature w_2 expresses the average likelihood with respect to all the ASNs from the query results to leverage the malicious mimicking the benign behavior as well as the benign with multiple ASNs.

Let N_{asn} denote the number of the distinct ASNs of the flux agents of a domain and A_i be the number of the flux agents (IP addresses) in i-th ASN of a domain. An ASN associated with more agents is less likely malicious. The likelihoods are averaged for a domain with multiple ASNs. Feature w_2 is as follows.

$$w_2 = \frac{1}{N_{asnR}} \times \sum_{i=1}^{N_{asnR}} \frac{1}{A_i} \quad (3)$$

TABLEII: THREE AVERAGE PROBABILITY OF THE MALICIOUS FAST-FLUX DOMAINS

	w_1	w_2	w_3
Fast-flux	1	0.8916	0.8376

w_3: Age of domain

A benign domain usually owns its domain for long and its domain age will be long when it is examined. Based on the statistics shown in Fig. 3, most malicious domains are newly registered. The researches [2], [4] show that a malicious domain may operate malicious actions, including spam, and phishing web site, and is often dormant a period of time before online operating. They also concluded that the dormant period of a malicious fast-flux domain is about than one month and the active lifetime is 5 days to weeks, no more than 18.5 days.

The domain age in this study is defined as the duration from the day that the domain is registered to that it is

examined by the system. The domain age of a malicious fast-flux domain is no more than the dormant time plus its active time, as it would be blocked once it is detected.

Based on our sample malicious, the average domain age of malicious fast-flux domains is 1.55 months. In Fig. 3(e), 60% have the domain age less than 1 month and the number of the malicious domains decreases progressively as the domain age get higher. Thus Poisson distribution is adopted to formulate the likelihood that a domain is malicious based on its domain age. *k* denote as the domain age of the domain to be examined in terms of months, where *k* equals the current time minus its registration time. Feature w_3 is expressed below in Poisson distribution with the expected value equals to the average domain age of malicious fast-flux domains, 1.55.

$$w_3 = \frac{e^{-1.55} \times 1.55^k}{k!} \times 3.04 \qquad (4)$$

Each of the three features contributes differently in the detection, so a weight function has to be calculated. The weight function is trained with 108 malicious fast-flux domains from ATLASto compute the probability that a domain is malicious fast-flux in terms of each proposed feature and the average probabilities are shown in Table2. The ratio of the three parameters is 1:0.8916:0.8376, normalized to 0.3664:0.3276:0.3069, and is used as the weights of the features. The detection function *P*(*Malicious Fast Flux Domain*) is expressed as follows.

$$P(Malicious\ Fast\ Flux\ Domain) = 0.3664 \times w_1 + 0.3276 \times w_2 + 0.3069 \times w_3 \qquad (5)$$

The proposed features is ulated by dig and whois, where dig is used for calculating w_1 and w_2, and whois is for w_3.

IV. System Evaluation

The detection function *P*(*Malicious Fast Flux Domain*) of a domain in equation (5) is computed accordingly based on the proposed three features. A threshold probability is required for determining the legitimacy of the domain to be examined. 500 benign websites from Alexa and 108 malicious fast-flux domains from ALTAS were applied to determine the threshold *h*. The selected threshold should be able to distinguish them from the malicious. Fig. 4 shows the distribution of each feature and it concludes that the threshold is 0.6.

This work proposes three features based on probability and does not require database or heavy computation resources. The features are formulated in functions and the proposed detection is suitable for large scale and real-time detection. The proposed detection limits the number of queries to avoid long delay and still maintains the high detection rate.

After calculating all domains, all malicious fast-flux domains have multiple country codes because worldwide bots are always selected to be the flux agents randomly without management. Even though the number of country codes is almost the unique condition of detecting malicious fast-flux domains, the specific gravity of the ASN, reverse lookup domain and domain age must still be calculated. It is in order to prevent that flux agents are selected withsome rules such as in the same country, ASN, or network segment rather than selected randomly.

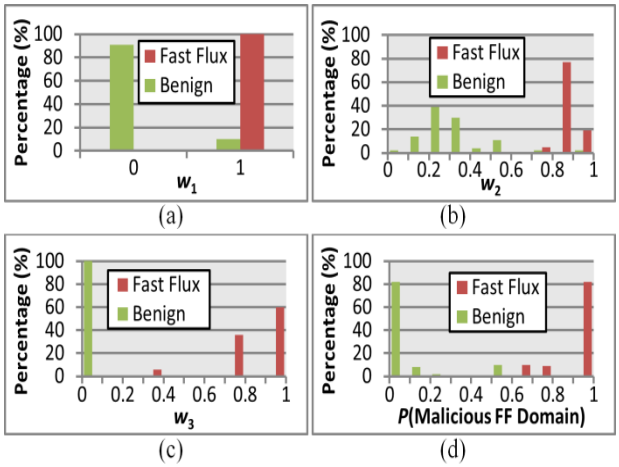

Fig. 4. Probability parameters of benign and malicious fast-flux domains with respect to each feature.

V. Conclusion

This study proposes three efficient formulaic features and the proposed detection can achieve the same detection rate as that with more information while maintaining much less process time. It is applicable for real-time and large scale detection as it requires only 15 seconds for examining a domain and can still achieve high detection rate.

During tracking and evaluating the malicious fast-flux domains, about 10% of the domains cease their operation. The average lifetime of the malicious is 2.61 months. Such short lifetime decreases the applicability of blacklist.

During the analysis of the malicious, some malicious fast-flux domains have the same set of flux agents. In Fig. 3(a), the set of 19 flux agents is shared by 18 malicious. It can be seen that bot herder may register many domains to evade blacklist detection and to extend the lifetime of the botnet. The analysis on the sleep time of the malicious verifies multiple malicious domains are owned by the same organization or bot herder as their sleep patterns are similar. The findings are helpful for crime investigation and defense.

References

[1] D. Kevin McGrath, Andrew Kalafut, and Minaxi Gupta, "Phishing Infrastructure Fluxes All the Way," *IEEE Security and Privacy Magazine special issue on Securing the Domian Name System*, pp. 21-28, September 2009.

[2] J. Nazario and T. Holz, "As the Net Churns: Fast-Flux Botnet Observations," in *Proc. the 3rd International Conference on Malicious and Unwanted Software*, pp. 24-31, Oct. 2008.

[3] J. Y. Wu, L.W. Zhang, J. Liang, S. Qu, and Z. Q. Ni, "A Comparative Study for Fast-Flux Sevice Networks Detection," in *Proc. Networked Computing and Advanced Information Management (NCM) 2010 Sixth International Conference*, pp. 346-350, 2010.

[4] E. Passerini, R. Paleari, L. Martignoni, and D. Bruschi, "FluXOR: detecting and monitoring fast- flux service networks," *Springer: Lecture Notes in Computer Science*, vol. 5137, pp. 186-206, 2008.

[5] M. Konte, N. Feamster, and J. Jung, "Fast-flux Service Networks: Dynamics and Roles in Online Scam Hosting Infrastructure," Technical Report GT-CS-08-07, September 2008.

[6] S. Y. Huang, C. H. Mao, and H. M. Lee, "Fast-Flux Service Networks Detection Based on Spatial Snapshot Mechanism for Delay-Free Detection," in *Proc. the 5th ACM Symposium on Information, Computer and Communications Security (ASIACSS 2010)*, Apr. 2010.
[7] C.H. Hsu, C.Y. Huang, and K.T. Chen, "Fast-Flux Bot Detection in Real Time," in *Proc. the 13th international conference on Recent Advances in Intrusion Detection*, Sep. 2010.
[8] C. V. Zhou, C. Leckie, and S. Karunasekera, "Collaborative Detection of Fast-flux Phishing Domains," *Journal of Networks*, vol. 4, no. 1, pp.75-84, 2009.
[9] T. Holz, C. Gorecki, F. Freiling, and K. Rieck, "Measuring and Detecting of Fast-Flux Service Networks," in *Proc. 15th Annual Network& Distributed System Security Symposium*, 2008.
[10] A. Caglayan, M. Toothaker, D. Drapaeau, D. Burke and G. Eaton, "Behavioral Patterns of Fast-flux Service Networks," in *Proc. the 43rd Hawaii International Conference on System Sciences* (HICSS), Jan. 2010.
[11] R. Lua and K.C. Yow, "Mitigating DDoS Attacks with Transparent and Intelligent Fast-Flux Swarm Network," *IEEE Network*, vol. 25, no. 4, pp. 28-33, July-August, 2011.
[12] S.Yu, S.Zhou, and S. Wang, "Fast-flux Attack Network Identification Based on Agent Lifespan," in *Proc. IEEE International Conference on Wireless Communications, Networking and Information Security (WCNIS) 2010*, pp.658-662, 2010.
[13] Alexa. [Online]. Available: http://www.aleza.com
[14] ATLAS. [Online]. Available: http://atlas.arbor.net/summary/fastflux

Chia-Mei Chen has joined in the Department of Information Management, National Sun Yat-Sen University since 1996. She was Section Chef of Network Division and Deputy Director, Office of Library and Information Services in 2009-2011. She serves as a coordinator of TWCERT/CC (Taiwan Computer Emergency Response Team/Coordination Center) since 1998 and continues working for the network security society. Her current research interests include mobile networks, multimedia systems, and network security.