Journal of Information Processing Vol.23 No.5 623–632 (Sep. 2015)

[DOI: 10.2197/ipsjjip.23.623]

Regular Paper

Detecting Malicious Domains and Authoritative Name Servers Based on Their Distinct Mappings to IP Addresses

Yin Minn Pa Pa[1,a)] Katsunari Yoshioka[1] Tsutomu Matsumoto[1]

Received: December 5, 2014, Accepted: June 5, 2015

Abstract: As Domain Name System (DNS) provides flexibility and robustness in communications of hosts on Internet, not only legitimate users but also attackers often take advantages of it. If we know how attackers are managing their malicious domains with authoritative name servers, there is a possibility to detect not only malicious domains but also malicious authoritative name servers. In this study, we present a novel method for detecting malicious "domains" (noted as d) and malicious "authoritative name servers" (noted as ns-d) based on their distinct mappings to "IP addresses" (noted as IP). Namely, we present three features to detect them; 1) Single ns-d is mapped to many IP, 2) Single IP is mapped to many ns-d, and 3) Single IP is mapped to both ns-d and d. We evaluate proposed method in terms of accuracy and coverage in detection of malicious d and ns-d. The evaluation shows that our detection method can achieve significantly low false positive rate in detecting both malicious d and ns-d without relying on any previous knowledge, such as blacklists or whitelists.

Keywords: Malicious DNS

1. Introduction

As domain name system (DNS) is a very efficient, robust and low-cost communication channel, domains are widely used for malicious online activities, such as connecting a large number of compromised hosts and attacker's command and control (C&C) servers, phishing, etc. Attackers manage these malicious domains at authoritative name server, for example, changing corresponding IP address of malicious domain over time to hide IP addresses of C&C servers. There can be different cases in which attackers obtain control of authoritative name server. For example, the authoritative name server that attackers are abusing can be a server setup by DNS hosting service or attackers themselves. However, how attackers are abusing authoritative name servers to manage their malicious domains is not well studied. If we know this, there is a possibility to detect not only malicious domains but also malicious authoritative name servers.

In this study, we present a novel method for detecting malicious "domains" (noted as d) and malicious "authoritative name servers" (noted as ns-d) based on their distinct mappings to "IP addresses" (noted as IP). Namely, we present three distinct features to detect them; 1) Single ns-d is mapped to many IP, 2) Single IP is mapped to many ns-d, and 3) Single IP is mapped to both ns-d and d. We evaluate the proposed method in terms of accuracy and coverage in detection of malicious d and ns-d. The evaluation shows that our detection method can achieve significantly low false positive rate in detecting both malicious d and ns-d without relying on any previous knowledge, such as blacklists or whitelists.

There are previous studies partly focusing on some mappings of d, ns-d and respective IP. Even though Features 1 and 3 are proposed by paper [4], no evaluation is shown to explain how effective these features are. So we consider these existing features with new additional feature, namely Feature 2, and evaluate them in all possible combinations.

The rest of this paper is organized as follows: Section 2 describes related works. In Section 3, features used in the proposed method are presented. In Section 4, the proposed method is explained. In Section 5, evaluation of the proposed method is explained. Finally, in Sections 6, 7 and 8, evaluation on IP, discussion on monitoring period and conclusions are presented respectively.

2. Related Works

There are previous research efforts in finding malicious domains using passive DNS data, zone files or DNS whois database. In contrast with previous studies, we are not just focusing on finding malicious domains. We take a further step into understanding of how attackers are abusing authoritative name servers to manage their malicious domains. Based on this understanding, we try to detect not only malicious domains but also malicious authoritative name servers. We call domains and authoritative name servers that are relating to malicious online activities as malicious domains and malicious authoritative name servers, respectively. There may be variety of cases how authoritative name servers are prepared by attackers, such as setting up a dedicated server as malicious authoritative name server or abusing a legitimate server for malicious purposes, however, we do not differentiate them and consider both cases malicious in this study.

[1] Graduate School of Enrironment and Information Sciences and Institude of Advanced Sciences, Yokohama National University, Yokohama, Kanagawa 240–8501, Japan

[a)] yinminpapa@gmail.com

Antonakakis et al. developed a reputation based classification system called Notos [1] in which domains were reputed based on network based, zone based and evidence based. Bilge et al. designed EXPOSURE [2] in which behaviors of domains were analyzed focusing mainly on time series of domains being queried together with other features such as DNS answers based, TTL value based and domain name pattern based features. In both studies, only the mapping between d and respective IP was considered and ns-d was not considered.

Hao et al. [3] studied behavior of spam domains combining with active DNS behavior and registration information. Although they found that IP spaces used by spam domains were small, how d, ns-d and IP were related was not studied.

Hu et al. [4] studied active detection of fast-flux domain in which IP usage of fast flux domains were analyzed. They found IP overlap between fast flux domain and their authoritative name server. This finding is similar to feature three of our method although we are not focusing on detection of fast flux domains only.

The contribution of the proposed method is two-fold: (1) it can detect unknown malicious domains, name servers' domains, and their corresponding IP addresses that are not in existing blacklists, (2) it uses data that is publicly accessible and easy to obtain by a single DNS resolver while the existing methods rely on additional data that is available for certain entities such as a long period of historical data of domains and IPs [1], DNS traffic captured at large networks such as ISP [2], and DNS responses obtained by a large number of resolvers in different locations (continents) [3].

2.1 Comparison with Related Works

We understand that previous work [3] investigates on initial DNS behaviors (e.g. registration of domains) of malicious domains and reports some interesting characteristics but does not propose a method to detect malicious domains. Therefore, we excluded paper [3] from our comparison. We then compare our detection method with related works [1], [2], [4]. Comparing with [1], [2], [4], the main advantage of proposed method is that it can be realized with low requirements for necessary data. Namely, our method uses data that is publicly accessible and easy to obtain by a single DNS resolver while the other existing methods rely on additional data that are available for only certain privileged entities such as a long period of historical data of domains and IPs [1], DNS traffic captured at large networks such as ISP [2], and DNS responses obtained by large number of resolvers in different locations (continents) [3].

With the limitation of the data used for detection, our choices of features are limited. So we focus on the domains of name servers and their relationship with other domains and corresponding IP addresses and three features for detection. Even though Features 1 and 3 are proposed by paper [4], no evaluation is shown to explain how effective these features are. So we consider these existing features with a new additional feature, namely Feature 2, and evaluate them in their all possible combinations.

Regarding the accuracy and coverage, it is difficult for us to make a quantitative comparison with the existing methods because all three existing methods [1], [2], [4]. use additional data to which we have no access. From qualitative point of view, our method has many false negatives (i.e., low coverage) because we focus on only certain type of relationships between domains, name servers' domains, and their corresponding IP addresses and indeed there would be many malicious domains that do not have that relationship while all other existing methods use a variety of features with richer additional data. However, our method can detect some unknown malicious domains, name servers' domains, and their corresponding IP addresses that are not included in the blacklists we used for the evaluation. Moreover, we manually confirmed that some of the domains detected by our method are included in Alexa top rank domains (higher than Alexa top 1,000) that are conventionally considered legitimate.

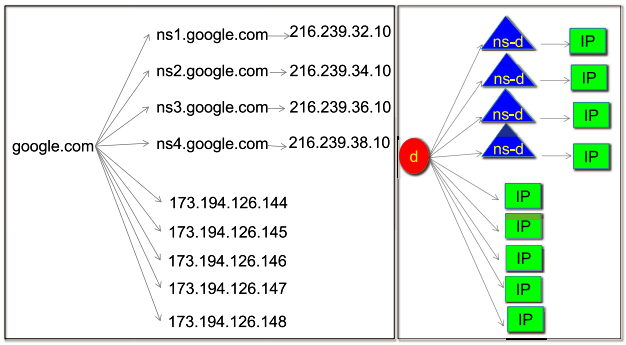


Fig. 1 Mappings of d, ns-d and respective IP.

3. Features

3.1 Mappings of d, ns-d and Respective IP

We first explain mappings of d, ns-d and their respective IP with real data example of "google.com" domain. In **Fig. 1**, google.com is d and ns1.google.com, ns2.google.com, etc., are ns-d. Both google.com and ns1.google.com have respective IP.

In the same way, for a particular d, it may have one or more corresponding ns-d. Both ns-d and d will have corresponding IP. In more detail, IP of ns-d is the IP address of a server running authoritative DNS service and IP of d may be the IP address of the server running other Internet service such as web.

3.2 Feature One: Single ns-d is Mapped to Many IP

As authoritative name server needs reliability for proper zone operation, IP of ns-d should not be changed frequently. On the other hand, attackers try to hide their authoritative name server by changing IP of ns-d. IP fluxing with IP of ns-d is a sign that ns-d is suspicious. Thus if a single ns-d is mapped to more than *Th1* IP addresses, we consider the mappings as a malicious case. The comparison between normal case and malicious case is shown in **Fig. 2**.

3.3 Feature Two: Single IP is Mapped to Many ns-d

Normally, different ns-d resolves to separate IP. For example, ns1.example.com and ns2.example.com resolve to separate IP. If many different ns-d resolve to single IP we consider the mappings as malicious case. Attacker with limited IP resources can take advantage in controlling his malicious domains with this feature. He can also hide his malicious authoritative name server by setting separate ns-d for each malicious domain. For exam-

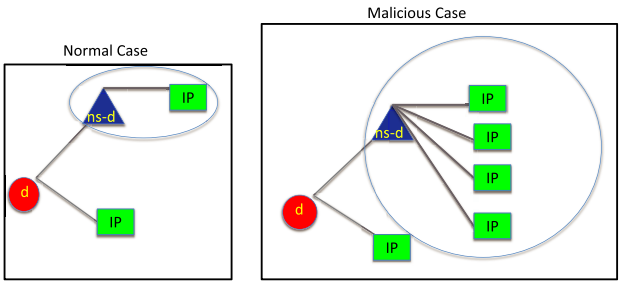


Fig. 2 Feature one.

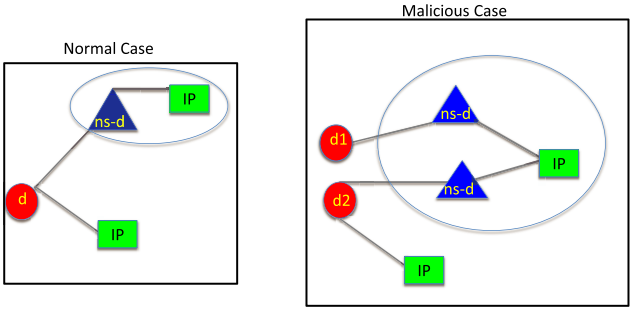


Fig. 3 Feature two

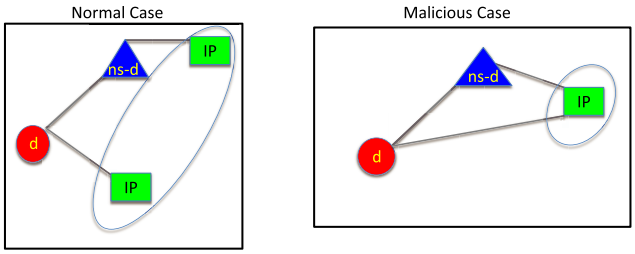


Fig. 4 Feature three.

ple, in registering malicious domains, attacker can setup to resolve malicious-1.com, malicious-2.com and malicious-3.com to ns.malicious-1.com, ns.malicious-2.com and ns.malicious-3.com respectively rather than resolving all malicious domains to a particular ns-d. In this way, if one hundred malicious d are registered, there will be one hundred different ns-d. All these ns-d are again setup to resolve to a single IP managed by the attacker so that he can manage all his ns-d with a single IP or a set of IP. That is why, in feature two, if single IP is mapped to more than Th_2 ns-d, we consider the mappings as malicious case. The comparison between normal case and malicious case is shown in **Fig. 3**.

3.4 Feature Three: Single IP is Mapped to Both ns-d and d

This feature is based on our finding that ns-d and d share the same IP. That is, DNS services and other malicious services, run in the same server. In the case of virtual hosting, one IP may be shared by many web sites. But it is practically very rare to share one IP with both DNS service and other service such as web.

As it is technically possible to run both web service and DNS service in the same server, a benign user of small business may install both services in the same server. In such case, the number of ns-d and d sharing the same IP should not be high. Therefore, if the total number of ns-d and d sharing the same IP is more than Th_3, we consider this as a malicious case. The comparison between a normal case and a malicious case is shown in **Fig. 4**.

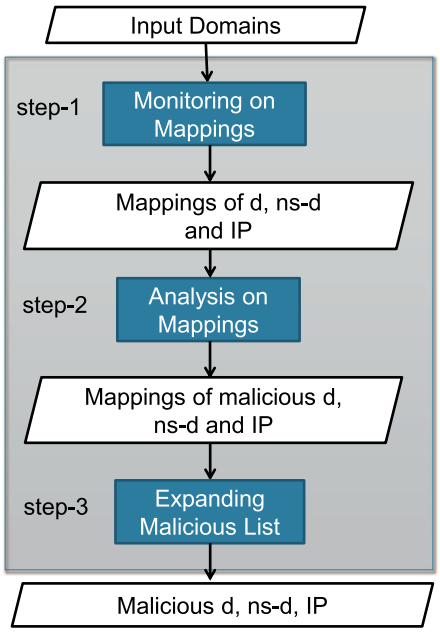


Fig. 5 Overview of proposed method.

4. Approach

4.1 The Proposed Method

We propose a method for detecting malicious d and ns-d based on their distinct mappings to IP addresses. Namely, we present three distinct features to detect them; 1) Single ns-d is mapped to many IP, 2) Single IP is mapped to many ns-d, and 3) Single IP is mapped to both ns-d and d. An overview of the proposed method is shown in **Fig. 5**.

The proposed method consists of three main steps: monitoring on mappings, analysis on mappings and expanding the malicious list. The input is a set of domains that are not known to be benign or malicious. Step one is monitoring on mappings of d, ns-d and IP. Step two is an important part in which we extract distinct mappings of malicious d, ns-d, and IP using all three features we proposed. In step three, we expand the malicious list and receive a list of malicious d, ns-d and IP as final output. Detail explanations of the three steps are described in the following sections. Analysis procedures and outcomes in each step of the proposed method are shown in **Fig. 6**.

4.2 Step One: Monitoring on Mappings

For every input d, we find 1) ns-d of d, 2) IP of ns-d, and 3) IP of d. **Figure 7** shows the process of finding mappings between d and ns-d, ns-d and IP and, d and IP.

In order to find ns-d of d, we simply query NS RR (Name Server Resource Record) of d. For example, we query NS RR of google.com so that we can get reply as “ns1.google.com” which is ns-d of google.com.

To look for IP of ns-d, we query A RR (IPv4 Address Resource Record) of ns-d. For example, we query A RR of ns1.google.com so that we can get reply “216.239.32.10” which is IP of ns1.google.com. After knowing ns-d and IP of ns-d, we check whether ns-d is really authoritative name server of d or not. For this, we query SOA RR (Start Of Authority Resource Record)

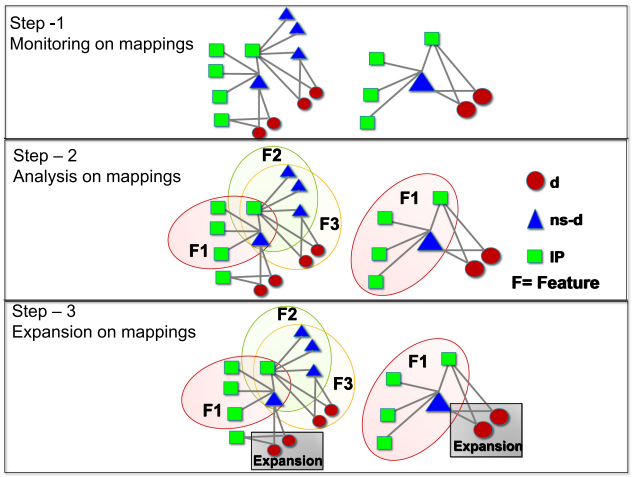


Fig. 6 Analysis procedure in each step of the proposed method.

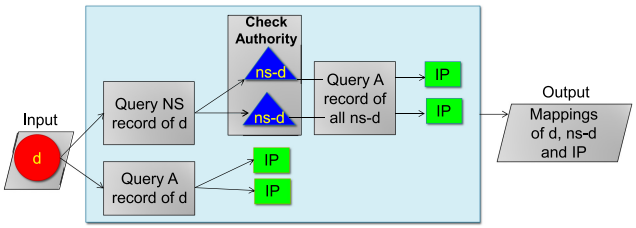


Fig. 7 Finding the mappings.

of d at ns-d and check reply packet whether aa (authoritative answer) bit is set or not. Only if aa bit is set in reply packet from ns-d, we assume that ns-d as authoritative name server of d.

Finally, to find the corresponding IP of d, we make A RR query of d. For example, we query A RR of google.com so that we can get a reply as "173.194.126.144" which can be one of the web servers of google.com domain.

For all queries, we use our recursive DNS server that query recursively to different levels of name servers in the DNS hierarchy till it reaches a final authoritative name server. For example, while querying A RR of d, our recursive DNS server talks directly to different levels of referral name servers in the DNS hierarchy starting from root servers till it reaches a final authoritative name server in which the corresponding IP of the queried domain is recorded in its zone file. We also set UDP (User Datagram Protocol) time out of queries to 1 second so that our resolver cannot be highly loaded. After finding all mappings of d, we obtain mappings between d and ns-d, ns-d and IP and, d and IP.

Step one is supposed to be continued for some period in order to obtain mappings of d, ns-d, and IP to be examined. In the experiment, we use the mappings obtained from the monitoring period of 214 days.

4.3 Step Two: Analysis on Mappings

Mappings obtained by step one are analyzed based on the following three features:

1. Single ns-d is mapped to many IP
2. Single IP is mapped to many ns-d
3. Single IP is mapped to both ns-d and d

The details of features are explained in Section 4. We depict the typical structure of features in **Fig. 8**.

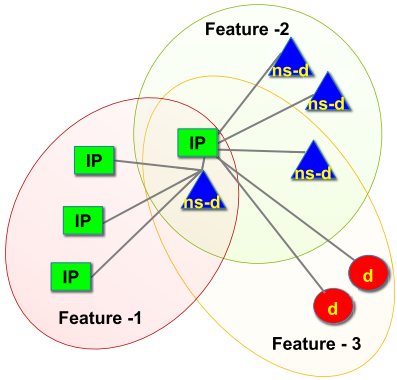


Fig. 8 Typical structure of all three features.

Table 1 Different combinations of features.

Combining three features	Combining two features	Separate Features
F1 ∧ F2 ∧ F3	F1 ∨ F2	F1 only
(F1 ∧ F2) ∨ F3	F1 ∨ F3	F2 only
F1 ∧ (F2 ∨ F3)	F2 ∨ F3	F3 only
(F1 ∨ F2) ∧ F3	F1 ∧ F2	
F1 ∨ (F2 ∧ F3)	F1 ∧ F3	
F1 ∨ F2 ∨ F3	F2 ∧ F3	
(F1 ∧ F3) ∨ F2		
(F1 ∨ F3) ∧ F2		

Firstly, we check the obtained mappings to see whether any of the three features is met. All three features have separate threshold values (noted as Th_1, Th_2, Th_3). Mappings exceeding threshold values will be considered malicious.

Indeed, in order to increase the accuracy of detection, we consider features in combined manners as shown in column 1 and 2 of **Table 1**. For example, for F1∧F2 combination, we look for mappings between ns-d and IP, that meet both feature one and two.

In general, feature one and two are mappings between ns-d and IP and only feature three is mappings of d and ns-d to IP. That is why only some combinations that has OR operation with feature three will consist of d in the result. For example, the result of "F1∧F2∧F3" combination will contain only ns-d and IP while the result of "F1∨F2∨F3" combination will include not only ns-d and IP but also d.

Output of step two will be the mappings of d, ns-d and IP that meet the combined features in Table 1. We consider all these d, ns-d, and IP of output as malicious.

4.4 Step Three: Expanding Malicious List

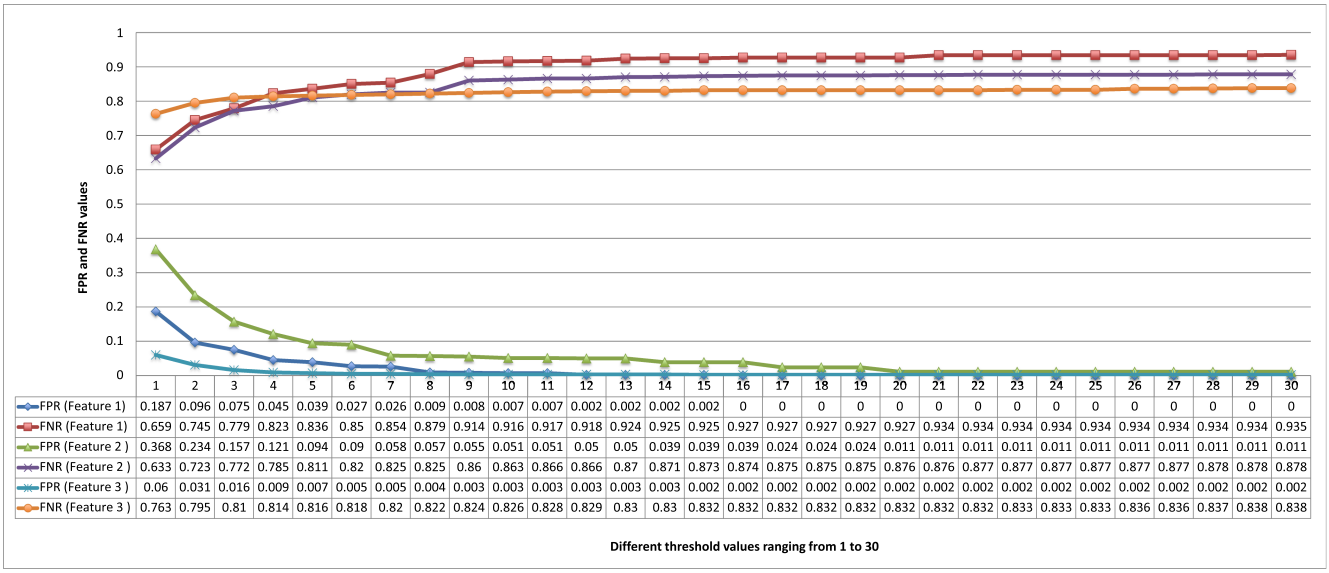
In step three, for each combination of features, we expand malicious d, respectively. Namely, we consider malicious for all d that are mapped to any of the ns-d or IP that construct malicious mappings identified in step two. Finally, we obtain lists of malicious d, ns-d and IP for each combination of the features.

5. Evaluation

5.1 Experiment and Results

5.1.1 Input Data Set

We collect and combine existing blacklist and whitelist to use it as input to the proposed method. Firstly, as known blacklist, we use malicious domains from DNS-BH project (malwaredomains.com) [5]. The total number of malicious domains we



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
FPR (Feature 1)	0.187	0.096	0.075	0.045	0.039	0.027	0.026	0.009	0.008	0.007	0.007	0.002	0.002	0.002	0.002	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FNR (Feature 1)	0.659	0.745	0.779	0.823	0.836	0.85	0.854	0.879	0.914	0.916	0.917	0.918	0.924	0.925	0.925	0.927	0.927	0.927	0.927	0.927	0.934	0.934	0.934	0.934	0.934	0.934	0.934	0.934	0.934	0.935
FPR (Feature 2)	0.368	0.234	0.157	0.121	0.094	0.09	0.058	0.057	0.055	0.051	0.051	0.05	0.05	0.039	0.039	0.039	0.024	0.024	0.024	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
FNR (Feature 2)	0.633	0.723	0.772	0.785	0.811	0.82	0.825	0.825	0.86	0.863	0.866	0.866	0.87	0.871	0.873	0.874	0.875	0.875	0.875	0.876	0.876	0.877	0.877	0.877	0.877	0.877	0.877	0.878	0.878	0.878
FPR (Feature 3)	0.06	0.031	0.016	0.009	0.007	0.005	0.005	0.004	0.003	0.003	0.003	0.003	0.003	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
FNR (Feature 3)	0.763	0.795	0.81	0.814	0.816	0.818	0.82	0.822	0.824	0.826	0.828	0.829	0.83	0.83	0.832	0.832	0.832	0.832	0.832	0.832	0.832	0.832	0.833	0.833	0.833	0.836	0.836	0.837	0.838	0.838

Fig. 9 FPR and FNR values of different threshold values.

Table 2 Numbers of d, ns-d and respective IP.

	d		ns-d		IP of d		IP of ns-d	
	Benign	Malicious	Benign	Malicious	Benign	Malicious	Benign	Malicious
	15,101	22,735	18,384	16,543	31,041	25,736	17,721	11,162
Unique total	37,836		32,280		54,754		25,657	

could collect within the whole analyzing period is 34,849 domains. Secondly, as known whitelist, we use top 10,000 domains from Alexa domains list [6]. The total number of benign domain we could collect within the whole analyzing period is 15,181 domains. In total, there are 50,030 domains as an input to the proposed method.

5.1.2 Step One: Monitoring on Mappings

The monitoring period is from April 1, 2014 to October 28, 2014. Within the whole period, we keep on monitoring all mappings between "d and ns-d," "ns-d and IP" and "d and IP," **Table 2** shows number of d, ns-d and respective IP we are able to find in step one.

We could only find mappings of 75% of input d. The main problem is because of NXDomain (Non Existence of domain). It is because of the short lifetime of malicious domains. Out of all input d, 17% of d becomes NXDomain in time of query. The rest 8% encounters errors such as ServFail (Server Fail), NoError (No Error), Refused (Query Refuse) and UDP query time out error. ServFail can be because of some failure in DNS service of authoritative name server. Although NoError literally means no error, we did not get any answer back for the query. It is because the RR type of d we are querying is not implemented although other RR type of d exist. For example, in querying NS RR of www.example.com, NS RR type of www.example.com does not exist although A RR type of www.example.com and NS RR of example.com both exist. In such case we receive NoError reply with no answer. Refuse error simply means that our query is refused. UDP timeout error is because of DNS query exceeding UDP timeout time.

5.1.3 Step Two: Analysis on Mappings

In this step, all mappings that meet any of the proposed three features are extracted as distinct mappings of a malicious case. We set value of Th_1, Th_2 *and* Th_3 to "three" as constant threshold value for all features because we would like to compare the strength of each feature and we think that 3 should be the smallest threshold value for detecting malicious domains and authoritative name servers according to many initial studies on malicious and benign domains.

An additional experiment on many different threshold values is conducted. By comparing the FPR and FNR values of different threshold values ranging from 1 to 30 as shown in **Fig. 9**, we would like to recommend 8 as the best threshold value for all features while FPR is as low as 0.004 and FNR is less than 0.9.

In our current experiment, to analyze data by feature one, we check all mappings between ns-d and IP. Then, we extract distinct mappings of a malicious case according to Th_1. As a result, in all mappings that meet feature one, there are 5,340 ns-d and 3,081 IP. In an extreme case, we found ns-d named "ns2.alfacoma.ru" (colored yellow in Fig. 10) that has 200 corresponding IP. We believe that these 200 IP can be IP of compromised hosts. All mappings extracted by feature one are visualized using force-directed graph drawing algorithm [7], [8]. **Figure 10** shows one example of mappings with 181 ns-d and 1,479 IP.

In order to analyze data by feature two, again, we check all mappings between ns-d and IP. But, this time, the analysis is focused on IP. For example, according to Th_2, if an IP has more than three corresponding mappings to ns-d, we think of it as a malicious case. As a result, there are 1,908 IP and 9,088 ns-d in all mappings that meet feature two. In an extreme case, to our surprise, we find a single IP related to 2,925 ns-d that are quite similar to each other such as ns1.com-fn41.net, ns1.com-fn62.net, ns1.com-fo30.net, etc. Some of the mappings that meet feature two exhibit similar structure when these are visualized. **Fig. 11** shows two mappings, both of which consist of exactly 7 IP and 560 ns-d. Although their relational structure is very similar, their actual ns-d and IP are different.

This may be an indication of the usage of the same administrative tool for these d and ns-d although a deeper investigation is

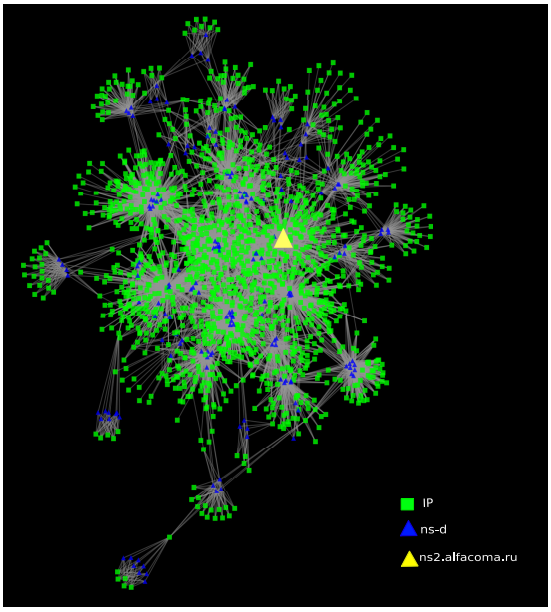


Fig. 10 Example of mappings that meet feature one.

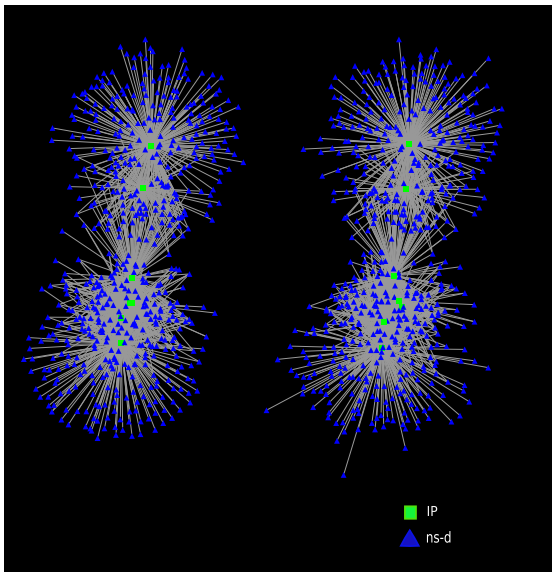


Fig. 11 Two examples of mappings with a similar structurethat meet feature two.

necessary.

To find mappings that meet feature three, we extract all mappings in which one IP is shared by both ns-d and IP. Then, for each detected mapping, it is checked whether the number of ns-d and d exceeds the threshold Th_3. As a result, there are 3,438 d, 5,477 ns-d, and 522 IP in all mappings that meet feature three. In an extreme case, we notice a single IP shared by 2,892 ns-d and 651 d.

An example of mappings that meet feature three is shown in **Fig. 12** consisting of 1,444 d, 1,420 ns-d and 70 IP. According to Fig. 12, we think that attackers are controlling a large number of d and ns-d with a limited number of IP resource.

After receiving all distinct mappings of a malicious case that meets features separately, we analyze features in a combined manner. The number of d, ns-d and respective IP obtained by different combinations of features are shown in **Fig. 13**.

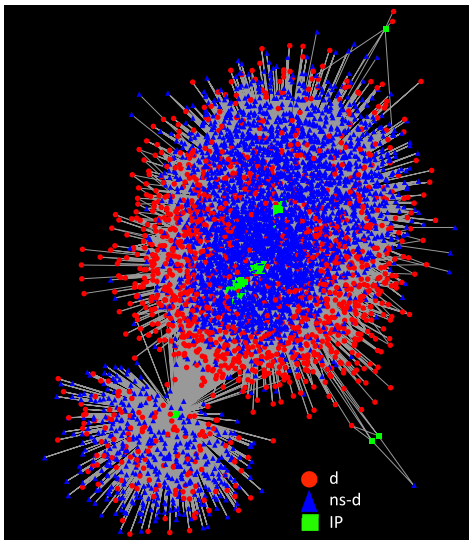


Fig. 12 Example of a mapping that meets feature three.

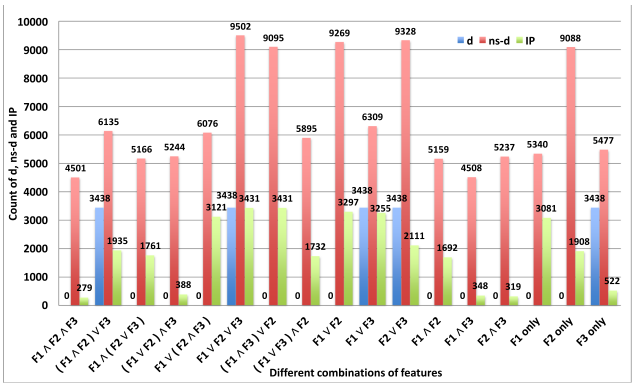


Fig. 13 Number of d, ns-d and respective IP obtained by step two.

Table 3 Benign and malicious instances from output of step two.

	d		ns-d		IP of d		IP of ns-d		Total(Unique)		
	Benign	Malicious	Benign	Malicious	Benign	Malicious	Benign	Malicious	d	ns-d	IP
Feature 1	0	0	621	4,719	0	0	2,544	1,258	0	5,340	3,081
Feature 2	0	0	1,837	7,251	0	0	1,533	1,123	0	9,088	1,980
Feature 3	157	3,283	597	4,880	180	448	411	244	3,438	5,477	522

Numbers of malicious and legitimate instances from output of step two are shown separately in **Table 3**.

5.1.4 Step Three: Expanding Malicious List

In this step, d mapping to malicious ns-d and IP obtained by step two are also treated as malicious d in order to expand the malicious domain list. By this way, we obtain d in all combinations of features. **Fig. 14** shows the output of step three.

5.2 Evaluation Methods and Results

The output of the proposed method is a list of malicious d, ns-d and IP obtained by different combinations of features. We evaluate our method by focusing on d and ns-d.

5.2.1 Evaluation of d

Firstly, the proposed method is evaluated in terms of accuracy and coverage in detecting malicious d. As ground truths, we consider all d in the input blacklist as malicious domains. As there are 323 domains that are in Alexa top 10,000 list and also detected as malicious domains by VirusTotal, we exclude these by utilizing Virus Total database from whitelist and then use the rest

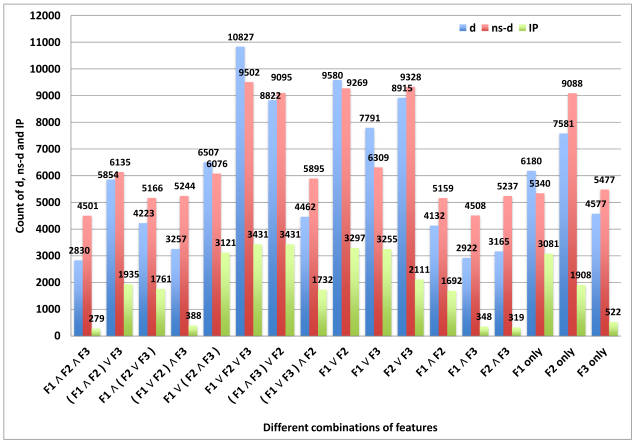


Fig. 14 Number of d, ns-d and respective IP obtained by step three.

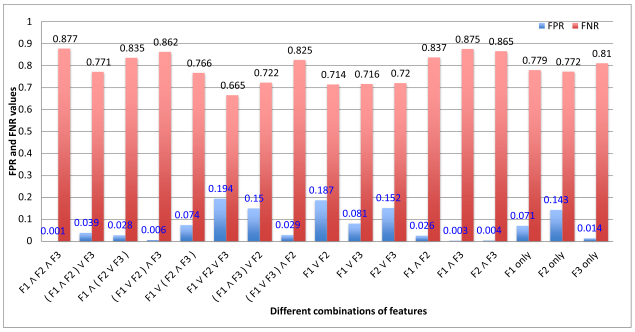


Fig. 15 FPR and FNR of d.

of domains in whitelist for evaluation. We determine accuracy by FPR (False Positive Rate). If FPR is low, it means the proposed method detects malicious d accurately. FPR is calculated by FP/N in which FP is the number of false positives d and N is the number of truly benign d. Coverage in detecting malicious d is determined by FNR (False Negative Rate). If FNR is high, it means the proposed method misses to detect a lot of malicious d. FNR is calculated by FN/P where FN is the number of false negatives d and P is the number of truly malicious d.

Figure 15 shows FPR and FNR of the proposed method for each combination of the three features.

By Fig. 15, low FPR values show that the proposed method is good in accuracy of detecting malicious d. On the other hand, a high FNR indicates that there are many malicious d we miss to detect. In Fig. 15, most FNR is more than 0.7. It is because the proposed method can detect only malicious domains that meet the features we are looking for and not all malicious d are based on features we used. That is why, in practice, we recommend to use our method in parallel with another method. By comparing results in Fig. 15, "(F1∨F2)∧F3" is acceptable while FPR is low and FNR is not the highest. When we see features separately, F3 is best for detecting malicious d accurately.

From the point of view of accuracy, the most strict case, namely "F1∧F2∧F3" combination, shows the lowest FPR of 0.1%.

Our method has a high false negative rate and therefore we should mention that it is not to be used in a single-handed manner. It is indeed to be used on top of an existing detection mechanism. In that sense, we believe that we need to show that what we detect by our method is indeed malicious (i.e., low false positive rate) and different from known malicious domains and IP addresses such as those included in the existing blacklists. We show this in **Fig. 16**.

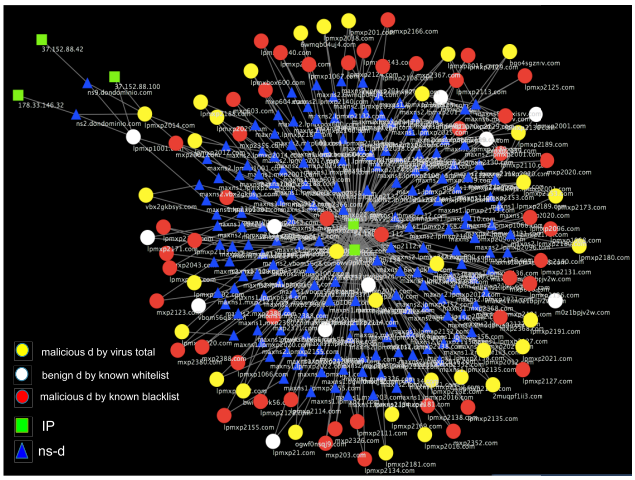


Fig. 16 Example of some evaluation results on d.

Table 4 Keywords and type of malicious activities.

Group Number	Keyword	Number of ns-d	Type
1	*.com-*.net	3923	Phishing Sites
2	maxns*.*.com	182	Malvertising
3	ns*.allfiles*.com	68	Drive-by-download
4	ns*.*arcinnia*.ru	108	Drive-by-download
5	ns*.*cloudbox*.com	146	Drive-by-download
6	*.cloudsvr*.com	100	Drive-by-download
7	*.*health*.ru	554	Rogue Online Pharmacies
8	*.*pharmacy*.ru		Rogue Online Pharmacies
9	*.*pill*.ru		Rogue Online Pharmacies
10	*.*tablet*.ru		Rogue Online Pharmacies
11	*.*drug*.ru		Rogue Online Pharmacies
12	ns51.*.*	357	Malware Site
13	ns52.*.*	354	Malware Site
14	ns53.*.*	335	Malware Site
15	ns54.*.*	331	Malware Site
16	*.*.orderbox-dns.com	184	Malware Site
	Total	6642	

5.2.2 Evaluation of ns-d

A challenge in evaluating ns-d is that there is no public benign or malicious ns-d list to the best of our knowledge. Thus, we cannot get ground truth for evaluation easily. To face this challenge, we make a malicious ns-d list and a benign ns-d list that we will be using as ground truths. For making a malicious ns-d list, we use two methods. The first one is manually searching ns-d on online web security reports and the second one is programmatically querying ns-d to VirusTotal database.

To search ns-d manually on online web security reports, we group all ns-d according to similarity of names. For example, ns-d such as "ns2.com-zy59.net," "ns3.com-fr26.net" and "ns3.com-gc22.net" are grouped according to their common name, "*.com-.*." Using a common name of each group as keyword, we search web reports and carefully read the reports in order to make sure that at least one ns-d of each group is related to malicious online activity. Then, we label each group according to malicious online activities described in the web report [9], [10], [11]. With this way, we can group 20% (6,460 ns-d) of all ns-d (32,218 ns-d) into 16 groups and we are able to label their relating malicious activities such as phishing, malicious advertising, drive-by-download, rouge online pharmacies and malware sites. **Table 4** shows keywords and malicious activities described in web reports.

In the second method, we query all ns-d (both malicious and

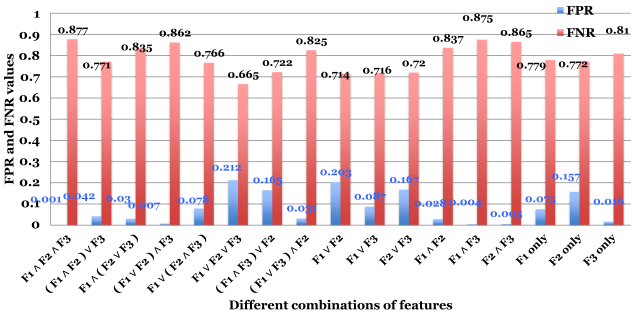


Fig. 17 FPR and FNR of ns-d.

benign ns-d) to VirusTotal and check whether any of them are known as malicious by antivirus products in VirusTotal. From this experiment, 18.4% (5,397 ns-d) of all ns-d are known as malicious.

Finally, we combine both results of two methods to get malicious ns-d list that we will be using as ground truth. As a result of two methods, we get 25.5% (8,247 ns-d) of all ns-d as malicious ns-d list. Then, the rest of the ns-d not included in our malicious ns-d list will be treated as benign ns-d. Finally, we receive a malicious ns-d list that includes 8,247 ns-d and a benign ns-d list of 24,034 ns-d. These two lists are used as ground truths in evaluation.

Using the ground truth data we prepared, the proposed method is evaluated in terms of accuracy and coverage in detection of malicious ns-d. We determine accuracy by FPR (False Positive Rate). If FPR is low, it means the proposed method detects malicious ns-d accurately. FPR is calculated by FP/N in which FP is the number of false positive ns-d and N is the number of truly benign ns-d.

Coverage in detecting malicious ns-d is determined by FNR (False Negative Rate). If FNR is high, it means the proposed method misses to detect a lot of ns-d. FNR is calculated by FN/P where FN is the number of false negatives ns-d and P is the number of truly malicious ns-d.

Malicious ns-d received by different combinations of features are evaluated in terms of FPR and FNR. The results are shown in **Fig. 17**. We also show FPR and FNR of each feature separately in order to compare features.

According to low FPR values in Fig. 17, it shows that the proposed method can detect malicious ns-d accurately. Moreover, FNR values are also not so high. All cases have FNR of less than 0.5 meaning the proposed method can detect more than 50% of malicious authoritative name servers. When we compare, FPR and FNR values of all combinations, we found that combinations that have AND operation with F2 can achieve significantly low FNR. Thus we think F2 is better to detect wide coverage of malicious ns-d comparing with F1 and F3. From the perspective of accuracy, the performance of F1 and F3 is better than F2. From aspect of false positive, in the most strict case, "F1∧F2∧F3" combination, FPR is 0.8%.

Finally, evaluation of ns-d shows that we can detect malicious ns-d with low FPR and FNR. That is why, we consider that the proposed method is strong enough in practice for detecting malicious authoritative name servers. But, we also need to notice that FPR and FNR are totally depending on the quality of ground truth data we prepared.

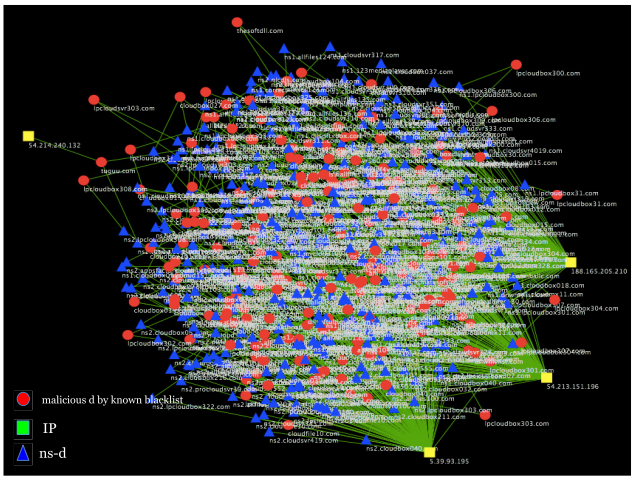


Fig. 18 Some malicious IP.

6. Evaluation on IP

We downloaded 575,147 blacklist IP addresses from public IP blacklists [18], [19], [20], [21], [22], [23], [24]. We match these IP blacklist with IP addresses output by the proposed method. The total number of output IP for all features by the proposed method is 3,431 IP addresses. As result, only 39 IP addresses (out of all 3,341 IP) match with a public blacklist. According to matching results, only 1% of our output IP addresses match with a public IP blacklist. We think that it is because output IP addresses by our proposed method are those of authoritative name servers and the blacklists we downloaded from Internet are not. Although most output IP addresses of the proposed method are not in public blacklist, we think that these IP are really malicious because of their very distinct mappings to ns-d. Some examples of mappings of IP that do not match with a public IP blacklist are shown in **Fig. 18**.

According to Fig. 18, there are only 4 IP addresses that involve with that much domains and authoritative name servers. We think that these IP must be really malicious. But, none of these four IP addresses are matched with a publicly known blacklist. In the same way, five IP addresses involving a lot with many different d and ns-d are not matched with a public IP blacklist although we think it as malicious.

7. Discussion on Monitoring Period

We analyze how detection results change according to the length of the monitoring period. The monitoring period for all possible mappings of a particular domain is difficult to determine because it depends on how a domain is managed by the owner. Of course, DNS records of benign domains are more stable than malicious domains. By experiment results of **Figs. 19** and **20**, FPR and FNR values do not have that much difference among results. That is why we think that one month is enough to monitor the change in DNS records of domains thoroughly.

We also analyze data of less than one month such as one day, one week, two weeks and so on. **Figure 21** shows how numbers of detected malicious domains are changing in monitoring period of one day, one week, two weeks and one month.

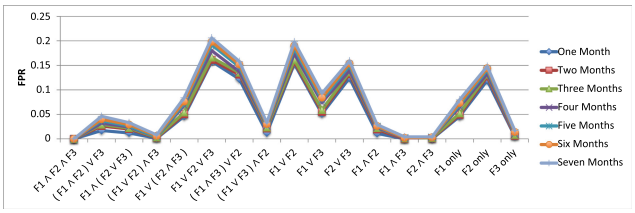


Fig. 19 FPR by each monitoring period (one month, two months, three months, etc...).

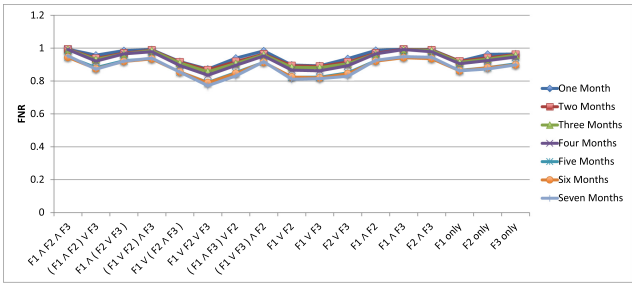


Fig. 20 FNR by each monitoring period (one month, two months, three months, etc...).

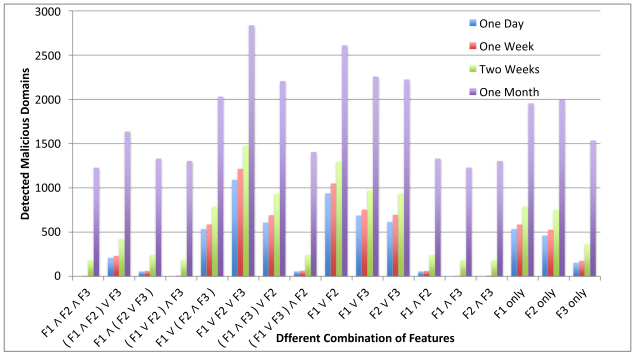


Fig. 21 Number of detected malicious domains in each monitoring period.

8. Conclusion

We proposed a method for detecting malicious d and ns-d based on their mappings to IP addresses. In the proposed method, we use three distinct features; 1) Single ns-d is mapped to many IP, 2) Single IP is mapped to many ns-d, and 3) Single IP is mapped to both ns-d and d. Detecting malicious d and ns-d includes three steps: 1) Monitoring on mappings 2) Analyzing mappings based on three features and 3) Expanding d according to malicious ns-d and IP found in step two. Finally, we evaluate the proposed method in terms of accuracy and coverage in detecting malicious d and ns-d. The evaluation shows that the proposed method can detect malicious d and ns-d with a high accuracy.

Lastly, we note that our method purely focuses on the mapping of d and ns-d to IP and does not rely at all on any previous knowledge, such as blacklists or whitelists in the detection method.

Acknowledgments A part of this was conducted under the auspices of the MEXT Program for Promoting Reform of National Universities and supported by PRACTICE (Proactive Response Against Cyber-attacks Through International Collaborative Exchange) project by the Ministry of Internal Affairs and Communications, Japan.

References

[1] Antonakakis, M., Perdisci, R., Dagon, D., Lee, W. and Feamster, N.: Building a Dynamic Reputation System for DNS, *USENIX security symposium*, pp.273–290 (2010).
[2] Kirda, L.B.E., Kruegel, C. and Balduzzi, M.: Exposure: Finding malicious domain names using passive dns analysis, *Proc. NDSS* (2011).
[3] Hao, S., Feamster, N. and Pandrangi, R.: Monitoring the initial DNS behavior of malicious domain names, *Proc. 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, pp.269–278 (2011).
[4] Hu, X., Knysz, M. and Shin, K.G.: Measurement and analysis of global IP ADDRESS-usage patterns of fast-flux botnets, *INFOCOM, 2011 Proceedings IEEE*, pp.2633–2641 (2011),
[5] DNS-BH (online), available from ⟨http://www.malwaredomains.com⟩.
[6] Alexa List (online), available from ⟨http://alexa.com⟩.
[7] Force Directed Algorithm (online), available from ⟨http://en.wikipedia.org/wiki/Force-directed_graph_drawing⟩.
[8] Cytoscape (online), available from ⟨http://cytoscape.org⟩.
[9] Norton Save Web (online), available from ⟨http://safeweb.norton.com⟩.
[10] Spam404 (online), available from ⟨http://www.spam404.com/⟩.
[11] VirusTotal (online), available from ⟨https://www.virustotal.com/⟩.
[12] Metcalf, B. and Spring, M.: Passive Detection of Misbehaving Name Servers, Technical report of CMU/SEI-2013-TR-010 (2013).
[13] RFC 1035 (online), available from ⟨http://www.ietf.org/rfc/ rfc1035.txt⟩.
[14] Pa, Y.M.P., Yoshioka, K. and Matsumoto, T.: Finding malicious authoritative DNS servers, *Information and Communication System Security* (*ICSS-2013*), Yokohama, Japan (2013).
[15] He, Y., Zhong, Z., Krasser, S. and Tang, Y.: Mining DNS for Malicious Domain name Registrations, *2010 6th International Conference on Collaborative Computing: Networking, Applications and Worksharing*, pp.1–6 (2010).
[16] Holz, T., Gorecki, C., Rieck, K. and Freiling, F.C.: Measuring and Detecting Fast-Flux Service Networks, *Proc. 16th Annual Network and Distributed System Security Symposium* (*NDSS*) (2008).
[17] Konte, M., Feamster, N. and Jung, J.: Dynamics of Online Scam Hosting Infrastructure, *Proc. Passive and Active Measurement* (*PAM*), pp.219–228 (2009).
[18] OpenBL (online), available from ⟨https://www.openbl.org/⟩.
[19] (online), available from ⟨https://myip.ms/browse/blacklist/Blacklist_IP_Blacklist_IP_Addresses_Live_Database_Real-time⟩.
[20] (online), available from ⟨http://malc0de.com/bl/IP_Blacklist.txt⟩.
[21] Malware Domain List (online), available from ⟨http://www.malwaredomainlist.com/hostslist/ip.txt⟩.
[22] (online), available from ⟨http://www.unsubscore.com/blacklist.txt⟩.
[23] (online), available from ⟨http://perishablepress.com/blacklist/ip.txt⟩.
[24] (online), available from ⟨https://zeustracker.abuse.ch⟩.

Yin Minn Pa Pa received her B.E. in Information Technology in 2006 from Mandalay Technological University, Myanmar and M. Phil in Infrastructure Management in 2013 from Yokohama National University, Japan. She is currently a Ph.D. candidate of Information Media and Environment Science Course of Graduate School of Environment and Information Sciences, Yokohama National University. Her research interest is network security.

Katsunari Yoshioka received his B.E., M.E. and Ph.D. degrees in Computer Engineering from Yokohama National University in 2000, 2002, 2005, respectively. From 2005 to 2007, he was a researcher at the National Institute of Information and Communications Technology, Japan. Currently, he is an associate professor for Division of Social Environment and Informatics, Graduate School of Environment and Information Sciences, Yokohama National University. His research interest covers a wide range of information security, including malware analysis, network monitoring, intrusion detection, and information hiding. He was awarded 2009 Prizes for Science and Technology by The Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology.

Tsutomu Matsumoto is a professor of the Graduate School of Environment and Information Sciences, Yokohama National University and directing the Research Unit for Information and Physical Security at the Institute of Advanced Sciences. He received Doctor of Engineering from the University of Tokyo in 1986. Starting from Cryptography in the early 80's, he has opened up the field of security measuring for logical and physical security mechanisms. Currently he is interested in research and education of Embedded Security Systems such as Smartcards, Network Appliances, Mobile Terminals, In-vehicle Networks, Biometrics, and Artifact-metrics. He is serving as a program officer of the JSPS Research Center for Science Systems, the chair of Japanese National Body for ISO/TC68 (Financial Services), and a core member of the Cryptography Research and Evaluation Committees (CRYPTREC). He was a director of the International Association for Cryptologic Research (IACR) and the chair of the IEICE Technical Committee on Information Security and served as an associate member of the Science Council of Japan (SCJ). He received the IEICE Achievement Award, the DoCoMo Mobile Science Award, the Culture of Information Security Award, the MEXT Prize for Science and Technology, and the Fuji Sankei Business Eye Award.