IEICE TRANS. INF. & SYST., VOL.E106–D, NO.9 SEPTEMBER 2023

LETTER *Special Section on Log Data Usage Technology and Office Information Systems*

Malicious Domain Detection Based on Decision Tree

Thin Tharaphe THEIN[†a)], ***Nonmember***, **Yoshiaki SHIRAISHI**[†], ***Senior Member***, ***and*** **Masakatu MORII**[†], ***Fellow***

SUMMARY Different types of malicious attacks have been increasing simultaneously and have become a serious issue for cybersecurity. Most attacks leverage domain URLs as an attack communications medium and compromise users into a victim of phishing or spam. We take advantage of machine learning methods to detect the maliciousness of a domain automatically using three features: DNS-based, lexical, and semantic features. The proposed approach exhibits high performance even with a small training dataset. The experimental results demonstrate that the proposed scheme achieves an approximate accuracy of 0.927 when using a random forest classifier.

key words: *malicious domain detection, machine learning, domain name system (DNS)*

1. Introduction

With the advancements in information technology, the risk and complexity of cybersecurity threats are increasing at an alarming rate, and various malicious cyber-attacks emerge daily. In general, malicious domains are key components for attackers to run malicious activities over the Internet and to infect user devices. Attackers can cause users to be victims of spam, phishing, and drive-by-download. Subsequently, the attackers can compromise the user privacy, install malware, or cause financial losses. Therefore, it is critical to discover and block such malicious activities. Although the existing domain and IP blacklists can be used to block malicious domains, these blacklists cannot keep up with the continual increase in newly registered domains. Therefore, an effective approach for detecting malicious domains is desirable.

The Domain Name System (DNS) is a remarkable resource for detecting malicious domains, and extensive research has been conducted on the detection of malicious domains using DNS data. In this study, using the DNS-based, lexical, and semantic features, we built classification models with random forest, XGBoost, and AdaBoost to estimate the maliciousness of a domain.

2. Related Works

2.1 Approaches for Malicious Domain Detection

The detection methods for malicious domains proposed in previous studies can mainly be divided into two categories: the classification-based and graph-based approaches.

Classification-based Approach

This approach mainly uses machine learning algorithms with features that are manually extracted from domain names and DNS traffic data, which can differentiate between legitimate and malicious domains. Examples of such features are the domain length, number of characters, number of digits, and time-to-live (TTL). Bilge et al. proposed a system known as *EXPOSURE*, which can detect malicious domain names by using a decision tree algorithm with features extracted from passive DNS analysis [1]. Similarly, in [2], malware detection based on DNS records and domain name features to identify malicious domains by using decision tree classifiers was proposed. Chiba et al. proposed the *DomainProfiler* system that uses a random forest classifier to detect newly registered malicious domain names with time-series domain features [3].

Graph-based Approach

Previous studies relating to graph-based approaches used the association between the domains and IP addresses or clients to form a domain graph, and subsequently applied graph-based learning algorithms such as belief propagation, label propagation, and graph convolutional networks for the domain classification. Khalil et al. constructed a domain–IP bipartite graph from the association between the domains and IPs, and then used a path-based algorithm to discover potential malicious domains [4]. Kazato et al. proposed a graph convolutional network-based malicious domain detection method by building a domain relation graph [5]. This approach made use of the domain–IP relationship, domain owner information, and autonomous system number to construct the domain graph. In the graph-based domain classification approach, the association between the domains plays an important role in determining the classification accuracy.

2.2 DNS Traffic Analysis

In order to avoid domains from being blacklisted, the attackers keep moving their domain names across the DNS. There are two most commonly used techniques to get these behaviors, namely Fast-Flux and Domain-Flux (IP-Flux). The former means that each domain name is associated with multiple IP addresses that are continuously changing to avoid blacklisting attempts. The latter makes use of a popu-

Manuscript received November 22, 2022.
Manuscript revised March 23, 2023.
Manuscript publicized June 22, 2023.

[†]The authors are with Kobe University, Kobe-shi, 657–8501 Japan.
a) E-mail: thein.thin@gsuite.kobe-u.ac.jp

DOI: 10.1587/transinf.2022OFL0002

lar technique, known as the Domain Generation Algorithm, to dynamically create a large number of malware domain names, which are associated with only one or a few IP addresses. This makes it difficult to block the domain names since most of these domain names are short-lived. However, these techniques leave footprints within the DNS data. It is therefore important to analyze those traces in the DNS traffic to detect the malicious domain.

In general, DNS traffic data can be collected in two ways: active and passive DNS data. Active DNS data are obtained by intentionally sending DNS queries periodically by the data collector and then it recorded the respective responses for further analysis. Since the data collector issues each query, the active DNS data does not link with the behavior of actual users and thus easing the privacy concerns [6]. The active DNS data captures the DNS records of a given domain, such as the IP address (A), name server (NS), and mail exchange (MX) records. Active DNS data do not have privacy problems because they do not include information on the user query domains. Thales [14] is an example of privacy-preserving active DNS data collection system that actively queries and collects a large volume of active DNS data using domain names from various publicly accessible sources. Passive DNS data provide historical records of the domain and contain richer information than active DNS data. Passive DNS provides the fastest means of accessing historical data that may no longer exist in the DNS records. The collection method of passive DNS is more complex than that of active DNS, but there are paid services that provide access to passive DNS database. Passive DNS data is gathered by deploying sensors on multiple DNS servers and DNS server logs to obtain real DNS queries and response information, but there are certain limitations and privacy issues on collected data depending on the location of deployed sensors, especially if sensors are deployed between clients and resolvers. The authors of Ref. [14] provided the experiment on active DNS vs. passive DNS data, and it is shown that active DNS data has more DNS record types while passive DNS data provide a tighter connection graph. Based on this [6], active DNS data can be used to discover newly created and potentially malicious domains. In the proposed system, the DNS records of the domain from only active DNS data were used for domain classification.

3. Proposed Scheme

An overview of the proposed approach is depicted in Fig. 1. The DNS data and additional information relating to the domains are first collected by the data collector module. Thereafter, three groups of features (DNS-based, lexical, and semantic) are extracted for each domain name. The maliciousness estimation of the domains is subsequently performed by the ensemble classifiers.

3.1 Data Collector

The DNS traffic data relating to each domain are actively queried and the DNS server processes each query request. Thereafter, the DNS server responds with the corresponding data. Examples of response data of a domain include A records, NS records, and TTL. This DNS response dataset is further enriched by the domain WHOIS information, which includes the domain registration, expiration, and updated dates. The collected DNS data are used to estimate the maliciousness of the domains.

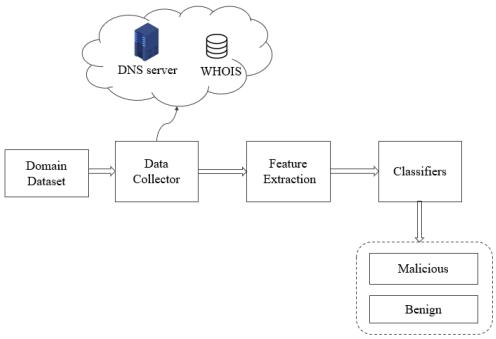

Fig. 1 Overview of proposed system.

Table 1 Domain feature set.

Type	Features
DNS-based features	Number of A records
	Number of NS records
	Number of MX records
	TTL
	Active time of domain
	Lifetime of domain
Lexical features	Number of consecutive characters
	Number of digits
	Length of domain
	Number of words
Semantic features	Domain reputation score

3.2 Feature Extraction

In this step, the previously collected data are processed to extract the features that can effectively distinguish between malicious and benign domains. Based on the observation and analysis of the large amount of DNS data obtained from the data collector, 11 features were extracted to build the classification model for malicious domain detection, as indicated in Table 1. The following section discusses how these features can be used to differentiate between benign and malicious domains.

DNS-based Features

The DNS response records of malicious domains are very different from those of benign domains. Malicious domains tend to have more A (address) records and lower TTL val-

ues. One of the reasons is due to the widespread use of the fast-flux domains [11]. The main idea behind the fast-flux is that each malicious domain is hosted on many different IP addresses which are changed quickly to avoid being blacklisted. Moreover, a more sophisticated type of fast-flux network, named double-flux, has an additional layer that makes it more difficult to track the malicious domains. The double-flux process is done by changing both the DNS A records and NS records frequently in a round-robin manner with a very short life span. This results in more A records and more distant NS records in DNS lookup. Moreover, fewer MX records are observed in malicious domains than in benign domains since domains associated with botnets attack usually have no or fewer MX records [15], [16].

Furthermore, the lifetime and active time of benign domains are typically much longer than those of malicious domains. The lifetime and active time of the domain were calculated using Eqs. (1) and (2). The lifetime of the domain is the interval between the expiration date and the registration date of the domain. Similarly, the active time of the domain is the interval between the updated date and the registration date of the domain. According to this information, the following characteristics were selected for the DNS-based features: number of A records, number of NS records, number of MX records, TTL, active time, and lifetime of the domain.

$$Lifetime = Date_{Expire} - Date_{Create} \tag{1}$$

$$Activetime = Date_{Update} - Date_{Create} \tag{2}$$

Lexical Features

In general, benign domain name strings are easily pronounceable and can be recognized with no trouble, whereas malicious domain names are mainly non-pronounceable by humans [12], [13]. The observation of a large number of malicious domains revealed that malicious domains contain more numbers, and the confusing mixture of numbers and words makes it difficult to pronounce malicious domains. Therefore, the following characteristics were selected for the lexical features of the domain: length of domain, number of digits, number of words, and number of consecutive characters.

Semantic Features

The conventional approaches for malicious domain detection include the use of DNS data and lexical features [1], [7], but in this research, in addition to DNS-based and lexical features, we also incorporated the semantic features of the domain. The previous study [17] introduced the detection of malicious domains using semantic features, whereby the domains with the highest accessed rates are selected as benign domains. Each domain name is then segmented by the N-gram method to construct whitelist domain name substrings, and those whitelist substrings are used to calculate the reputation (maliciousness) of a domain. This research followed a similar approach to the previous method for calculating the reputation value of a domain. First, as a ground truth to construct the whitelist domain substring, the top 100,000 domain names from Alexa Top Sites [8] were collected and segmented by the N-gram method by setting the lengths of N to 3, 4, 5, 6, and 7. A total of 344,503 domain name substrings were extracted from the top 100,000 domain names and used as the whitelist domain name substring. Thereafter, the reputation score of the testing domain was calculated according to Eq. (3):

Table 2 Domain reputation score.

Domain name	Reputation score	Label
duolingo.com	44.064	Benign
discord.com	62.8	Benign
dkdrlah12.0pe.kr	7.347	Malicious
dqy.qyuyu.com	0.567	Malicious
facebook.com	63.412	Benign
douate.com	20.185	Malicious

$$Reputation\ Score_{domain} = \sum_{i=1}^{k} \log_2 \left(\frac{S_N(k)}{N} \right), \tag{3}$$

where $S_N(k)$ is the total number of occurrences of the k-th domain name substrings in the whitelist domain name substrings and N is the length of the N-gram ($N = 3, 4, 5, 6, 7$). Several results of the domain reputation scores are presented in Table 2. It can be observed that the reputation score of the benign domain tended to be larger than that of the malicious domain because segmented benign domain substrings frequently occurred in the whitelist domain name substrings.

4. Evaluation

The performance of the proposed scheme was evaluated using three ensemble classifiers: random forest, Xgboost, and Adaboost. The benign and malicious domain names published on the Internet were collected, labeled, and used as ground truth data. The dataset was then divided into training and testing data to evaluate the effectiveness of the classification of each domain as benign or malicious.

4.1 Dataset

The dataset contained a total of 1,457 domain names, among which 680 domains were malicious and 777 were benign. The benign domains were collected from Alexa Top Sites [8], which is a website ranking system based on popularity. We considered that the top-ranked websites were legitimate domains. The malicious domain names were gathered from publicly published domain blacklist services. The resolvable malicious domains were randomly selected from malwaredomainlist.com [9] and a compromised domain list [10]. These domains were likely to be compromised by malware, and command and control communication and phishing.

Table 3 Experimental results.

Features	Classifiers	Accuracy	Precision	Recall
DNS	Random Forest	0.8973	0.8955	0.8824
	AdaBoost	0.9007	0.8741	0.9191
	XGBoost	0.9007	0.8794	0.9118
DNS+Lexical	Random Forest	0.9041	0.9033	0.9050
	AdaBoost	0.9096	0.9092	0.9113
	XGBoost	0.9068	0.9063	0.9084
DNS+Lexical+ Semantics	Random Forest	0.9270	0.9199	0.9219
	AdaBoost	0.9151	0.9146	0.9168
	XGBoost	0.9123	0.9115	0.9131

4.2 Evaluation Results

The domain classification was performed by three methods: random forest, AdaBoost, and Xgboost. A comparison of the experimental results is presented in Table 3. To show the effectiveness of the proposed model, we conducted the experiment based on three groups of features: (i) using only DNS, (ii) using DNS and lexical features, and (iii) using DNS+lexical+semantics features. All three classifiers (i.e., random forest, AdaBoost, and XGBoost) achieved above 89% accuracy in detecting the malicious domains in all feature modes, even though the domain name dataset was quite small. In the experiment, the dataset is divided into 75% training data and 25% testing data. All classifiers were trained and evaluated using 10-fold cross-validation. The experimental results demonstrated that using the combination of all groups of features performed better than using only DNS features or DNS+Lexical features. Random forest exhibited the best performance among the three methods in all evaluation metrics when all groups of features are used.

5. Conclusions

We have proposed an approach to classify a domain as malicious or benign by using active DNS traffic data and WHOIS information. Moreover, we incorporated semantic features in addition to the commonly used lexical and DNS-based features in an attempt to improve the malicious domain detection. The experimental results demonstrate that the proposed approach achieved an accuracy as high as 93% with the random forest classifier when using a small domain training dataset.

The current classification only recognizes domains as malicious or benign. We can further categorize the maliciousness of a domain as spam, phishing, command and control, or malware, making it a multiclass classification problem. Moreover, the use of a combination of passive DNS and active DNS data may enhance the ability to detect bad domains.

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant Number JP21H03444. This research results were partly obtained from the commissioned research under a contract of “Research and development on IoT malware removal/make it non-functional technologies for effective use of the radio spectrum” among “Research and Development for Expansion of Radio Wave Resources (JPJ000254)”, which was supported by the Ministry of Internal Affairs and Communications, Japan.

References

[1] L. Bilge, E. Kirda, C. Kruegel, and M. Balduzzi, “EXPOSURE: Finding malicious domains using passive DNS analysis,” NDSS 2011, pp.1–17, San Diego, CA, USA, Feb. 2011. DOI: 10.1145/2584679

[2] K.A. Messabi, M. Aldwairi, A.A. Yousif, A. Thoban, and F. Belqasmi, “Malware detection using DNS records and domain name features,” ICFNDS ’18, pp.1–7, Amman, Jordan, June 2018. DOI: 10.1145/3231053.3231082

[3] D. Chiba, T. Yagi, M. Akiyama, T. Shibahara, T. Yada, T. Mori, and S. Goto, “DomainProfiler: Discovering domain names abused in future,” Proc. 2016 46th Annu. IEEE/IFIP Int. Conf. DSN, pp.491–502, Toulouse, France, June 2016. DOI: 10.1109/DSN.2016.51

[4] I. Khalil, T. Yu, and B. Guan, “Discovering malicious domains through passive DNS data graph analysis,” ASIA CCS ’16, pp.663–674, Xi’an, China, May 2016. DOI: 10.1145/2897845.2897877

[5] Y. Kazato, Y. Nakagawa, and Y. Nakatani, “Improving maliciousness estimation of indicator of compromise using graph convolutional networks,” 2020 IEEE 17th Annu. CCNC, pp.1–7, Las Vegas, NV, USA, Jan. 2020. DOI: 10.1109/CCNC46108.2020.9045113

[6] Y. Zhauniarovich, I. Khalil, T. Yu, and M. Dacier, “A survey on malicious domains detection through DNS data analysis,” ACM Comput. Surv., vol.51, no.4, pp.1–36, 2018. DOI: 10.1145/3191329

[7] Z. Liu, Y. Zeng, P. Zhang, J. Xue, J. Zhang, and J. Liu, “An imbalanced malicious domains detection method based on passive DNS traffic analysis,” Secur. Commun. Netw., vol.2018, Article ID 6510381, 2018. DOI: 10.1155/2018/6510381

[8] “Alexa Top Sites,” https://www.alexa.com/topsites, Accessed June, 2020.

[9] “Malware Domain List,” https://www.malwaredomainlist.com, Accessed June, 2020.

[10] “Compromised Domain List,” https://zonefiles.io/compromised-domain-list, Accessed June, 2020.

[11] T. Holz, C. Gorecki, K. Rieck, and F.C. Freiling, “Measuring and

detecting fast-flux service networks," Ndss, 2008.

[12] S. Yadav, A.K.K Reddy, A.L.N. Reddy, and S. Ranjan, "Detecting algorithmically generated malicious domain names," Proc. 10th ACM SIGCOMM Conference on Internet Measurement, pp.48–61. 2010.

[13] S. Yadav, A.K.K Reddy, A.L.N. Reddy, and S. Ranjan, "Detecting algorithmically generated domain-flux attacks with DNS traffic analysis," IEEE/ACM Trans. Netw., vol.20, no.5, pp.1663–1677, 2012.

[14] A. Kountouras, P. Kintis, C. Lever, Y. Chen, Y. Nadji, D. Dagon, M. Antonakakis, and R. Joffe, "Enabling network security through active DNS datasets," International Symposium on Research in Attacks, Intrusions, and Defenses, pp.188–208, Springer, Cham, 2016.

[15] I. Prieto, E. Magaña, D. Morató, and M. Izal, "Botnet detection based on DNS records and active probing," Proc. International Conference on Security and Cryptography, pp.307–316, IEEE, 2011.

[16] S. Hao N. Feamster, and R. Pandrangi, "Monitoring the initial DNS behavior of malicious domains," Proc. 2011 ACM SIGCOMM Conference on Internet Measurement Conference, pp.269–278, 2011.

[17] R. Sharifnya and M Abadi, "A novel reputation system to detect DGA-based botnets," ICCKE 2013, pp.417–423, IEEE, 2013.