Malware Detection using DNS Records and Domain Name Features

Conference Paper · June 2018

DOI: 10.1145/3231053.3231082

CITATION
1

READS
199

5 authors, including:

Monther Aldwairi
Jordan University of Science and Technology
67 PUBLICATIONS **332** CITATIONS
SEE PROFILE

Fatna Belqasmi
Zayed University
64 PUBLICATIONS **395** CITATIONS
SEE PROFILE

Malware Detection using DNS Records and Domain Name Features

Khulood Al Messabi
College of Technological Innovation,
Zayed University
Abu Dhabi, UAE
m80007533@zu.ac.ae

Monther Aldwairi*
College of Technological Innovation,
Zayed University
Abu Dhabi, UAE
monther.aldwairi@zu.ac.ae

Ayesha Al Yousif
College of Technological Innovation,
Zayed University
Abu Dhabi, UAE
m80007520@zu.ac.ae

Anoud Thoban
College of Technological Innovation,
Zayed University
Abu Dhabi, UAE
m80007525@zu.ac.ae

Fatna Belqasmi
College of Technological Innovation,
Zayed University
Abu Dhabi, UAE
fatna.belqasmi@zu.ac.ae@zu.ac.ae

ABSTRACT

As billions of people depend on Internet application to perform day to day tasks, the prevalent of malwares and online attacks cause a huge loss to global Internet economy prevalent. Domain name system is one of the core components of the Internet, which allows users to type in website names and resolves them to Internet addresses. Several studies proposed using DNS for malware detection, because it is the first step before visiting a specific website. Unfortunately, majority focused on malicious URLs back listing, botnets, top-level-domain, DNS and resolvers. This paper proposes a system to detect malicious domain names, by using eight unique features that accurately identify malicious websites before being visited. We implemented our approach of malicious domain names detection using Python, and experimented with five weeks of real-world data using Weka.The experimental results reports a 77.5% and low false positive rates 22.4%. That is very promising considering the approach detect website based on feature calculated based on URL and without downloading the file.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability;

KEYWORDS

DNS, domain name, malware detection, malicious domains

*Department of Network Engineering and Security
Jordan University of Science and Technology
PO. BOX. 3030, Irbid, 22110, Jordan

ICFNDS'18, June 26–27, 2018, Amman, Jordan

ACM ISBN 978-1-4503-6428-7/18/06…$15.00
https://doi.org/10.1145/3231053.3231082

ACM Reference Format:
Khulood Al Messabi, Monther Aldwairi, Ayesha Al Yousif, Anoud Thoban, and Fatna Belqasmi. 2018. Malware Detection using DNS Records and Domain Name Features. In *ICFNDS'18: International Conference on Future Networks and Distributed Systems, June 26–27, 2018, Amman, Jordan.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3231053.3231082

1 INTRODUCTION

Domain Name System (DNS) is a service that maps Internet websites and domain names to the corresponding IP addresses. DNS name resolution takes place before the user views a website and the attackers redirect or link users to other malicious website, with malware, drive-by-downloads, malicious scripts, malicious advertisements or other malicious content [17]. To combat that form of threats, the core network service providers, advertisers and search engines try as much as possible to check domain reputation before redirecting the users. Several, web of trust and reputation companies [15] specialize in web metrics and domain metrics analysis [21] and provide their services to governments and companies to block or blacklist those malicious domains. While blacklisting is still widely used, it is not enough and cannot keep up with new domains springing up daily such as botnets, fast flux networks [1], drive-by-downloads [6], phishing [22], spam [5] and malicious advertisements [17]. Blacklisting is the first line of defense and further more accurate security measures must be put in place to ensure more intelligence is used to protect the users

The main challenge for blacklisting, is the rate at which newer malicious domains manifest themselves and keep changing domain names and hosting services. It poses a serious challenge to develop a comprehensive accurate and up to date reputation lists of the tens of thousands of domains that are registered on a daily basis. Besides, the existing DNS reputation systems are characteristics of DNS lookups that originates from resolvers and are responsible for look up domains to differentiate malicious and legitimate domains. However, these kinds of DNS services before it could determine the reputation for a domain that can only occur upon compromise process [8].
To facilitate detection of an attack of a malicious domain before it occurs, we should study the previous DNS activities associated

with each domain and relate how the physical behavior for each defective domain is dissimilar to the legitimate domain [7]. It can be deduced that the registration of domain and establishment of record resources happens before the attacks and about 55 percent of spam messages may be sent everyday upon registration of domain as a spam message [5]. Most importantly, DNS infrastructure for the defective domain is centered in various address spaces and only a few IP address sections host services for a domain associated with malicious activity. Therefore, monitoring and analysis of DNS records are imperative when it comes to the detection and mitigation of malware.

The DNS is used for helping Internet users locate online services such as web servers and mail host. However, DNS may also be used for malicious purposes, one example is the use of DNS for the management of botnets, using command and control servers. This attack is commonly used for DoS attacks by stealing private information and sending a great deal of spam, to gain a financial profit[9]. Another example is phishing attacks, phishing websites usually appear as a legitimate website, with a domain name that sounds authentic [22].

Researchers proposed to detect malware by detecting patterns of activities similar to malware activities and different from normal activities baseline. Unfortunately, these methods have limitation in that they suffer fromh hight false positives and did not involve any dynamic DNS service [16]. On the other hand, Kopis, which uses a machine-learning algorithm, focuses on an automated malware analysis system. It identifies previously unknown malware-related domain names, before appearing in blacklists [8]. The Infloblox DNS Firewall and the malware protection system was introduced by FireEye Inc [14] and helps in protecting organizations from Advanced Persistent Threats (APTs). The Infloblox DNS firewall blocks connections by denying DNS requests, which results in users not accessing the malicious host that contains or controls the malware. It also blocks the malware from transmitting stolen information from the intended user. Nonetheless, it does not block all APTs, but it can identify, if not block, many infections in teh initial stages. Other studies propose the arrangement of domain names to exclude suspicious domain names [20], but these approaches need a database of common malicious domains.

In this paper, we present a unique approach to detect DNS malware by using several behavioral features to recognize the malicious domains from the legitimate domains before teh user opens them. The technique uses a combination of several existing methods of malicious domain names detection. However, this technique uses and collects the most prominent DNS-based feature used in previous work and combines them to achieve the best outcomes. The goal of DNS malware detection is to mitigate and lessen online attacks, by detecting the attach early one using selected DNS features of suspicious domains. First, the list of malicious domains and legitimate domains are stored into two separate datasets. Then, a python script extracts and classifies malicious domain names' features. Subsequently, nslookup is used to acquire the IP addresses of the malicious domains. This does not provide a final solution to detect malicious activities, but it gives a more accurate detection. The approach is predicted to detect malicious domain names by merely observing the obvious features of suspicious domains and combine the features with some of the DNS based attributes.

The rest of this paper is organized as follows. The following section describes the related studies relevant to DNS malicious domains detection. Section 3 discusses the proposed approach, presents a detailed explanation of the selected DNS and domain names based features. Section 4 presents the details of constructing the detection module, database collection, classifier details, training and testing stages and experiments. Section 5, presents the validation findings from the experiment and the results are thoroughly analyzed. Finally, the proposed future work will be discussed accompanying the conclusion.

2 RELATED WORK

Attackers have increasingly exploited the Domain Name System, to maintain and manage their malicious infrastructures. Thus, many researches have proposed number of approaches to detect malicious domains, however, most of the experiments have failed to deliver an ideal system that does not have shortcomings. Mockapetris initiates a reputation mechanism that dynamically assigns a scorecard that judges a new domain whether it is malicious or legitimate based on a new unknown domain score [18]. Another study used behaviors on a primary domain, that falls below DNS resolvers, to detect and identify those domains that generate defective activities. These domains are used for adult websites, fast flux networks [1], phishing malware [22], and spam emails [26]. Passive DNS analysis was used to recognize domains automatically. Another previous study also considers detecting malicious domains though supervising DNS migrations from local servers. However, Mockapetris introduce new traffic feature as well as a new vantage point that is concerned with leveraging the global physical outlook by supervising traffic at the top of the hierarchy on DNS [18]. This new vantage can adequately recognize malware domains by checking the certainties of the global DNS queries.

Moreover, one important area of botnet detection research proposed an approach that takes into consideration the subsequent and the arrangement of the domain names [20]. This approach is seen to isolate dangerous domain collection from the temporal similar DNS databases. However, this method requires a known malicious domains as anchors. Also, they studied different approaches for detecting malware activities such as monitoring and analyzing DNS traffic. As other methods focus on finding botnets, which use defective services, others consider the analysis of recursive DNS traffic from various networks. They created a system that when placed at the edge of a network have the capability to find out and correct botnet attacks on a connection by establishing defective flux domain, originating from malware in APT attacks, do not require Domain Generation Algorithm (DGA) domains.

Furthermore, two systems were proposed with the aim to detect malicious domains using passive DNS analysis. One recent work, studied DNS lookup behavior within a local domain below the DNS resolvers to build the domains reputation [10]. They introduced new system called EXPOSURE that uses large-scale and passive

DNS analysis techniques with aim to detect the malicious domains. The main approach of this technique is to detect domains that are used in malicious activities on the Internet with less training time and data. The analysis technique presented in this paper is novel and it is used for the malicious domains detection that is based on the request of the passive DNS analysis. The technique is distinct from previous techniques that target only the botnet operations that used the fast flux domains [5], and does not depend on previous knowledge about the service kind such as phishing, fast flux services and spam, that the malicious domain provides [1]. Also, the study identifies 15 behavioral features that are indicative of malicious behavior, and 9 of them are novel and have not been suggested in previous studies. A large volumes of DNS data were used and they were collected by the researches, within period of two and a half month, as offline data set for EXPOSURE. After that, 100 billion DNS queries were recorded, they resolved 4.8 million different domain names. Moreover, Exposure was deployed in real-life setting on an ISP network and used it in order to control the 30,000 clients DNS traffic. The findings of the performed experiments showed that the proposed technique is scalable, and is able to accurately and automatically recognize the legitimate domains from the malicious domains (botnet command and control servers, phishing sites, and scam hosts) with a low false positive rate. The limitations of the Exposure system is that, it depends on the monitoring of the DNS queries from a restrictive number of RDNS servers. Also, the visibility of DNS queries linked to huge DNS zones, is partial.

Exposure has been prepared to be used by several DNS operators, as it depend on sharing of data through different networks in order to gain a visibility, that has meaningful level, into DNS traffic [8]. In contrast to Exposure that used by RDNS-based system, Kopis detected malware domains by the information attained from the upper DNS hierarchy. The features of Kopis, enabled it to discover malicious domains in accurate manner without relying on the IP reputation information, needed for Exposure. The automated malware analysis system, Kopis, operates at the upper DNS hierarchy for malware detection based on resolution of DNS patterns [8]. It is considered to be the first system to operate at the top-level domain. It also provides early detection of malware to prevent damages before they occur. A six month real world data from two popular DNS authorities was used to test the system. It detected unclassified malware-related domains a couple of weeks before listing them in the blacklist or security forums. In addition, to that, this system was used in China to identify the DDoS botnet formation, which helped in removing botnets and minimizing the emergence of new ones. It is true that Kopis is more efficient in e new botnets detection, unfortunately, we cannot be installed or used by one local network [10].

One research that examines the TLD servers to cluster newly registered domains based on registration information and lookups explored by group of researchers [13]. They proposed the early DNS behavioral properties of attack domains, that enabled by two perspectives. First, the DNS infrastructure linked to the domain and it is noticeable from the resource records. Second, the DNS lookup patterns across the networks that initially search for the .com and .net domains. Their experiment findings help to detect the malicious domains in an early stage. Their findings from monitoring the malicious domains infrastructure are as follows.

- Attackers register the domain and establish the resource record before they place the attacks.
- Malicious domains DNS infrastructure, is placed in various address space regions and autonomous systems, different than the legitimate domains infrastructure.
- The newly registered malicious domains early lookup patterns are significantly different than the ones for the benign domains.

It is true that the observed features can be used to as the basis for an early DNS-based system alerting for attacks. However, one limitation is that the focused approach is only in fast-flux features.

Based on the intrusion detection related work survey, an anomaly-based approach was proposed to detect botnet that relies on periodic supervision and analysis of the DNS traffic [23]. This approach works by unrevealing the group activities in DNS database that are concurrently sent by bots, which are distributed in a way. In the same note, various features can be used to distinguish DNS traffic that comes from benign or botnets clients. However, the study only focuses on the collection of activities related to botnets and the features incorporated in the study are not be able to detect APT malware.

Signature-based detection and anomaly-based detections are other systems relevent to detecting malware attacks but not directly related to DNS. Signature-based detection is a technology that involves detecting malware infections based on the existing signature's database. Additionally, signature-based detection technology is capable of identifying malware in communication traffic through exact pattern matching. However, this technology is surrounded by a fatal disadvantage; it is incapable of detecting new infections, that is if the signature of the new malware does not reside in the already established signature database [4]. Snort is a well-known signature-based detection system. It has many rules in the VRT that aids in the detection of malicious code and suspicious network activities. Moreover, there exist several low false positives only when the attacks are defined in prior [26]; however, it is difficult to detect new or strange attacks using Snort.

On the other hand, tons of literature have tried to improve anomaly based intrusion detection methodology. This is a technique for detecting abnormal behavior that differs from the normal behavior baseline. It requires that the normal behavior of the network to be determined and studied first before identifying the abnormal behavior. This approach of anomaly-based intrusion is important because it has the capability to detect the unknown or new attacks [23]. This technique has a unique advantage over other similar technologies of same functionality, because it is capable of detecting new or unknown abnormal behaviors. However, anomaly-based intrusion detection can easily cause false positives due to the complicated networks and application behaviors. Most importantly, several legitimate applications produce similar behavior profile to that of abnormal behavior [19].

Wang and Shirley correlated the patterns of malicious activities in domains with the detection of DNS malwares. In basic forms, after registering domain names with a registrar, there must be a pattern of the amount of time that takes domains to resolve actively on the

Internet. Firstly, the pattern is described by comparing data from registries for numerous domains in the top-level and DNS data source. After that, the pattern was used to baseline and compare the pattern of malicious activities in domains. It turned out that malicious activities have different patterns than that of benign ones. This can be effectively used to detect malicious domains before they cause damage. Comprehensive DNS data was collected and observed from the Security Information Exchange (SIE) for two weeks. Information about domain names correlated to malicious code was collected. Malware attempts were included in the results and they seem to have different patterns. The experiment in this paper was biased as it excluded dynamic DNS services, which means that the latency can only be calculated for malware that look for domains in the top-level domains [24].

Infoblox DNS Firewall developed by Infoblox Inc., is a powerful DNS firewall that basically provides early detection of malwares. This DNS firewall basically detects and protects against malwares by using a subscription service of threat information and updated continuously on malware. It basically blocks the connection by denying the request for DNS communication. This prevents malwares relying on DNS from reaching its host. The system also provides a feature of determining infected devices in a short amount of time [14]. one of the disadvantages of this firewall is that the set-up process of DNS firewall is complicated.
Kara et al. proposed a system that detects malware in DNS traffic. The system basically observes all the activities performed in the DNS zone. This is done by collecting resource record type of access counts and defining distribution channels of payload. The system consists of one main module, which is the payload distribution detection module. The experiment showed that the proposed system can effectively detect malicious payload channels. It was also noticed from the experiment that DNS malware instances, which are based on payload distribution of user strong pattern to retrieve the payloads of attacks. This paper also provides an analysis of the distribution of malicious payload channels, which was accomplished with the help of malware dataset covering one year [16]. The strength of this paper is that it includes statistics of 2707 detected domains. On the other hand, the focus was mainly on DNS resolvers, which means that the DNS traffic that does not travel through DNS resolvers is not considered.

3 METHODOLOGY

This section provides a description of the approach to detect DNS malware, dataset assortment process, and features used for the detection of malicious domain names. The main goal of this study is to label the domains as benign, or malicious before the users visit them. The technique used and combined several existing methods of malicious domain names detection. We study and pick the most prominent DNS-based features and domain name based features from previous work. The goal of DNS malware detection is to mitigate and lessen online attacks, by detecting the he most important features of suspicious domains.

3.1 Features Selection

To determine the most important DNS-based features and domain names based features that are indicative of malicious domains, the analysis and measuring procedures for the data sets were performed. The nine features were chosen from previous work [24], [9] and from the our observation of malicious domains, to characterize the malicious domains. Table 1 illustrates the feature sets and the identified features. The selected features are described and explained in the following subsections.

3.1.1 Domain Name-based Features. Four main feature ar identified: basic feature, character indicator variables, top-level domains based features, and tokens.

- Number of characters: the mean number of characters in the domain name, which is usually 12-13 characters.
- Number of dots: the number of dots may indicate the level of suspiciousness in domain names. It was found for example, if dots recurred more than three times, it demonstrates a very high potential of the domain being malicious.
- Number of hyphens: the number of hyphens in the domain name may be indicative of maliciousness to some extent. Because most legitimate domains do not contain more than a hyphen or two.
- Number of numerical digits: the number of numerical digits in the domain name is an indicator of maliciousness.

3.1.2 Character Indicator Variables. Another feature that was assessed is the frequency of characters from a to z. The more repetitive the characters are the greater the probability of the domain name to fall under the malicious category.

3.1.3 Top-level Domains-based Features. An aspect of our domain names malicious features is including the calculation of most abused top-level domains used for malicious activity. Malicious domains TLDs were observed and taken into consideration. The result of the observations has shown that the most used TLDs in the malicious domains are listed in the Table 2.

3.1.4 Tokens. An additional source of features in domain names is keywords we refer to as tokens. First, an inappropriate words were filtered after the study of the malicious domains was done. we mean by inappropriate is obscene words that are usually filtered out in emails. Second, specific transfer-based words were observed in the malicious domains such as direct, redirect and transfer words. Attackers use these words in phishing URLs to redirect users to their websites in order to steal the personal information [3].

3.2 DNS Answer-Based Features

NS lookup command line was used to extract the IP addresses for domain names. We perform an experiments to determine whether the IP address add value to the detection accuracy. We use WEKA to evaluate the relevance of IP addresses, using a sample of 1000 IP addresses . However, the value of the IP address was deemed not useful and therefore was not used as part of our feature set

Table 1: Domain Names Based Features

Feature set	Number of features	Feature name
Basic features	1	Number of characters
	2	Number of dots
	3	Number of hyphens
	4	Number of numerical digits
Char indicator variables	5	Frequency of characters from a to z
Top-level domains	6	Most abused top-level domains
Tokens	7	Inappropriate words
	8	Transfer-based words

Table 2: Suspicious Top Level Domains

.zip	.date
.us	.click
.men	.study
.top	.stream
.webcam	.party
.country	.review
.kim	.cricket
.work	.link
.gq	.tq
.cn	.ru
.In	.tr
.bid	.download
.trade	.loan

4 BUILDING DETECTION MODEL

The following subsection explain the training dataset construction and classifiers testing and training. We experiment with numerous features combinations to measure detection accuracy.

4.1 Training Set Construction

The database for malicious domains were collected from malware domains website [12] , and the benign domains database were collected from Alexa [15] and Google search engine. Fundamentally, comparing these data sets, it is apparent that malicious domains are not as easily decipherable.

The next step was writing a python script to read URL, calculate and extract the aforementioned features. Those features will later be used in the training and teasing process.

4.2 Training and Testing

The python prepared a Weka inout file to validate the outcomes and rank features in terms of importance. We add the feature one by one and used information gain to assess the most important features and the least important ones. Based on the attributes' ranks, any feature that has a low rank will be removed. The feature of suspicious TLDs ranked the highest with a rank of (0.386). The remnant features were dots with a rank of (0.234), numerical character repetition (0.213), token (0.191), character length (0.187), and hyphens (0.126). The least important attribute rank derived were transfer words and inappropriate words with a percentage of (0.013) and (0.005), respectively. Because of these low findings in the words features, the attributes were filtered to have a more accurate result.

4.3 Classification

The attributes were then fed to the carious Weka classifiers.The classifier takes the data sets and uses it to build the detection model and then evaluates the test data to output matching statistics. Various classifiers were tested to compare the precision and based on that, several classifiers were chosen to be presented in the experiments section. Based on best performance, we select J48, which is a C4.5 decision tree. To classify a new items, J48 first creates a decision tree based on the attribute values of the available training data. J48 predicts value items that are missing, based on what is recognized about the attribute values for the other data records, mainly using rules that are constructed from the training sample [25]. The decision tree classifier was chosen because of the efficiency and accuracy of the results that these algorithms produce. Furthermore, the tree is built during the training phase in the decision tree classifier, so dividing malicious domains from the legitimate ones can be based on the best features [9].

5 EXPERIMENTAL EVALUATION

The following experiment are carried out to determine the accuracy of malicious domain detection, based on various distinct the feature sets.

5.1 Classifier Evaluation

The evaluation of the accuracy of the J48 decision tree classifier using 10-folds cross-validation. In 10-folds the dataset is partitioned onto 10 sets, 9 used to training and one for testing. The 10 sets are rotated in order for each one of them to be used as a testing set. Weka output the confusion matrix, which shows the accuracy of the classifier and gives information about the actual and predicted classifications done by the system. The matrix is composed of terms includes: True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN). Then we use the following equations to calculate the main metrics of precision or positive predictive value (PPV), recall or True Positive Rate (TPR), specificity or True Negative Rate (TNR), F-1 measure, and Matthews correlation coefficient (MCC).

$$Precision = PPV = \frac{TP}{TP + FP}$$

$$Recall = TPR = \frac{TP}{TP + FN}$$

$$specificity = FPR = \frac{FP}{FP + TN}$$

$$F - measure = \frac{2TP}{2TP + FP + FN}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FN \times (TN + FN)}}$$

In the experiment, the TP of malicious domains detected resulted in 3495 out of 5000 predicted malicious domains, which leaves behind an FP of 1505. For the legitimate domains, the TP outcome was 3792 out of 4,400 predicted legitimate domains, with a FP of 608. Table 3 demonstrates the results received from the J48 classifier. The highest classification detection rate was 77.52% percent. Out of a total of 9,400 malicious and legitimate domains: 7287 domains were classified correctly, and 2113 were incorrectly classified.
The precision is the fraction of retrieved instances that are relevant, whereas recall is the fraction of relevant instances that are retrieved. Therefore, they are based on an understanding measure of relevance; precision is the measurement of exactness and recall the measurement of completeness [11]. Precision for malicious domains= 3495/(3495 + 608) = 0.851 precision for legitimate domains=3792/(3792 + 1505) = 0.715
The ROC curve of the classifiers output is plotted with FPR on the X-axis, and TPR on the Y-axis. The line leaning towards the Y-axis, indicates how valid the classifier is. The Area Under ROC Curve (AUC) was is 80 percent, which is reasonable. The PRC area indicates that the classifier id better for detecting malicious domains than benign ones.

Table 3: The Accuracy of J48 classifier

	Malicious	Legitimate	Weighted Avg
TP Rate	0.699	0.862	0.775
FP Rate	0.138	0.301	0.214
Precision	0.852	0.716	0.788
Recall	0.699	0.862	0.775
F-Measure	0.768	0.782	0.775
MCC	0.564	0.564	0.564
ROC Area	0.805	0.805	0.805
PRC Area	0.821	0.730	0.778

5.2 Detection Rate

The highest average detection rate of our classifier was %77.5 percent by using the cross-validation evaluation on the training set. The approach is used is not comprehensive enough to detect all malware domains on the Internet. On the other hand, the goal of the approach is focusing on detecting the newly unknown malicious domains. In addition, all of the features can be calculated without downloading or accessing the domain website, that is there not risk to the user. This makes such a low detection rate acceptable, if we take into consideration that the other techniques use content based features. That is the domain must be visited or website must be downloaded to compute the features, which already defies the purpose.

6 CONCLUSIONS AND FUTURE WORK

Due to the expanding use of Internet, and the increasing number of malicious attacks, it is of great importance for Internet users and domain names registrars to be able to identify malicious domain names, by observation, without going through the process of accessing the website. The features that have been examined and analyzed, adds a notable predictive power to find out whether the domain is malicious or not. These features could progress over time as attackers might change their methods, which leads us to propose future works that could be done to monitor and detect developed domain names using a more advanced system.
The results obtained depend on the collected malicious domains, data that was acquired from a database containing DNS malware. Due to the time and resources limitation, there have been some shortcomings to this research, which could hopefully be carried out in the future researches. One of these limitations, considering a domain that is solely used for the purpose of advertising banners and links, to make profit. Internet users usually end up on parked domains because of typo squatting, which indicates a sense of maliciousness towards the existence of parked domains. It is optimistically anticipated that the features of parked domain names could be extracted similarly to what has been done in the extraction of other malicious domains features. Other prospect feature that could be evaluated are word segmentations in domain names and the geolocation of domain names. By adding into consideration, the use of word segmentation, it might give benefit to an additional attribute feature, which could result in a higher accuracy of the experiment. Conceivably, the geolocation of the IP addresses of the domain names, can filter out where the highest percentage malicious domain names originate. Finally, the quantitative improvement of sentiment analysis in Opinions Sandbox will be very useful in malware detection [2].

ACKNOWLEDGMENTS

This work was supported by Zayed University Research Office, Research Cluster Award #17079.

REFERENCES

[1] B. Al-Duwairi, A. Al-Hammouri, M. Aldwairi, and V. Paxson. 2015. GFlux: A google-based system for Fast Flux detection. In *2015 IEEE Conference on Communications and Network Security (CNS)*. 755–756. https://doi.org/10.1109/CNS.2015.7346920

[2] Feras Al-Obeidat, Eleanna Kafeza, and Bruce Spencer. 2018. Opinions Sandbox: Turning Emotions on Topics into Actionable Analytics. In *Emerging Technologies for Developing Countries*, Fatna Belqasmi, Hamid Harroud, Max Agueh, Rachida Dssouli, and Faouzi Kamoun (Eds.). Springer International Publishing, Cham, 110–119.

[3] Monther Aldwairi and Rami Al-Salman. 2011. MALURLs: Malicious URLs Classification System (The best paper award) *(Annual International Conference on Information Theory and Applications Canning)*. GSTF Digital Library (GSTF-DL). The best paper award.

[4] Monther Aldwairi and Niveen Ekailan. 2011. Hybrid Pattern Matching Algorithm for Intrusion Detection Systems. *Journal of Information Assurance and Security* 6, 6 (2011), 512–521.

[5] M. Aldwairi and Y. Flaifel. 2012. Baeza-Yates and Navarro approximate string matching for spam filtering. In *Second International Conference on the Innovative Computing Technology (INTECH 2012)*. 16–20. https://doi.org/10.1109/INTECH.2012.6457802

[6] Monther Aldwairi, Musaab Hasan, and Zayed Balbahaith. 2017. Detection of Drive-by Download Attacks Using Machine Learning Approach. *Int. J. Inf. Sec. Priv.* 11, 4 (Oct. 2017), 16–28. https://doi.org/10.4018/IJISP.2017100102

[7] Khalifa AlRoum, Abdulhakim Alolama, Rami Kamel, May El Barachi, and Monther" Aldwairi. 2018. Detecting Malware Domains: A Cyber-Threat Alarm System. In *Emerging Technologies for Developing Countries*, Fatna Belqasmi, Maxand Dssouli Rachida Harroud, Hamidand Agueh, and Faouzi Kamoun (Eds.). Springer International Publishing, Cham, 181–191.

[8] Manos Antonakakis, Roberto Perdisci, Wenke Lee, Nikolaos Vasiloglou, II, and David Dagon. 2011. Detecting Malware Domains at the Upper DNS Hierarchy. In *Proceedings of the 20th USENIX Conference on Security (SEC'11)*. USENIX Association, Berkeley, CA, USA, 27–27. http://dl.acm.org/citation.cfm?id=2028067.2028094

[9] Leyla Bilge, Engin Kirda, Christopher Kruegel, and Marco Balduzzi. 2011. EXPOSURE : Finding malicious domains using passive DNS analysis. In *NDSS 2011,*

18th Annual Network and Distributed System Security Symposium, 6-9 February 2011, San Diego, CA, USA. San Diego, UNITED STATES. http://www.eurecom.fr/publication/3281

[10] Leyla Bilge, Sevil Sen, Davide Balzarotti, Engin Kirda, and Christopher Kruegel. 2014. Exposure: A Passive DNS Analysis Service to Detect and Report Malicious Domains. *ACM Trans. Inf. Syst. Secur.* 16, 4, Article 14 (April 2014), 28 pages. https://doi.org/10.1145/2584679

[11] Lakshmi Devasena C. 2015. Comparative Analysis of Random Forest, REP Tree and J48 Classifiers for Credit Risk Prediction. *International Journal of Computer Applications* 3 (2015), 31–36.

[12] Malware Domains. [n. d.]. Malware Prevention through DNS Redirection (Black Hole DNS Sinkhole). http://mirror1.malwaredomains.com/

[13] Shuang Hao, Nick Feamster, and Ramakant Pandrangi. 2011. Monitoring the Initial DNS Behavior of Malicious Domains. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference (IMC '11)*. ACM, New York, NY, USA, 269–278. https://doi.org/10.1145/2068816.2068842

[14] Infoblox. [n. d.]. Infoblox Introduces a New Defense Against Advanced Persistent Threats. https://www.infoblox.com/company/news-events/press-releases/infoblox-introduces-new-defense-advanced-persistent-threats/

[15] Alexa Internet. 1996. Alexa an Amazon Company. https://www.alexa.com/

[16] A. M. Kara, H. Binsalleeh, M. Mannan, A. Youssef, and M. Debbabi. 2014. Detection of malicious payload distribution channels in DNS. In *2014 IEEE International Conference on Communications (ICC)*. 853–858. https://doi.org/10.1109/ICC.2014.6883426

[17] R. Masri and M. Aldwairi. 2017. Automated malicious advertisement detection using VirusTotal, URLVoid, and TrendMicro. In *2017 8th International Conference on Information and Communication Systems (ICICS)*. 336–341. https://doi.org/10.1109/IACS.2017.7921994

[18] Paul Mockapetris. 1987. *RFC 1035-Domain Names- Implementation and Specifications*. RFC 1035. RFC Editor. 1–55 pages. https://www.ietf.org/rfc/rfc1035.txt

[19] Khaldoon Mhaidat Monther Aldwairi, Yahya Flaifel. 2017. Efficient Wu-Manber Pattern Matching Hardware for Intrusion and Malware Detection *(International Conference on Electrical, Electronics, Computers, Communication, Mechanical and Computing (EECCMC))*. IEEE.

[20] J. A. Morales, A. Al-Bataineh, Shouhuai Xu, and R. Sandhu. 2009. Analyzing DNS activities of bot processes. In *2009 4th International Conference on Malicious and Unwanted Software (MALWARE)*. 98–103. https://doi.org/10.1109/MALWARE.2009.5403014

[21] myWOTC. [n. d.]. myWOT Web of Trust. https://www.mywot.com/

[22] M. A. Qbeitah and M. Aldwairi. 2018. Dynamic malware analysis of phishing emails. In *2018 9th International Conference on Information and Communication Systems (ICICS)*. 18–24. https://doi.org/10.1109/IACS.2018.8355435

[23] E. Stalmans and B. Irwin. 2011. A framework for DNS based detection and mitigation of malware infections on a network. In *2011 Information Security for South Africa*. 1–8. https://doi.org/10.1109/ISSA.2011.6027531

[24] Wei Wang and Kenneth E. Shirley. 2015. Breaking Bad: Detecting malicious domains using word segmentation. *CoRR* abs/1506.04111 (2015). arXiv:1506.04111 http://arxiv.org/abs/1506.04111

[25] Ian H. Witten and Eibe Frank. 2005. *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[26] G. Zhao, K. Xu, L. Xu, and B. Wu. 2015. Detecting APT Malware Infections Based on Malicious DNS and Traffic Analysis. *IEEE Access* 3 (2015), 1132–1142.