WE KNOW IT BEFORE YOU DO: PREDICTING MALICIOUS DOMAINS

Wei Xu, Kyle Sanders & Yanxin Zhang
Palo Alto Networks, Inc., USA

Email {wei.xu, ksanders, yzhang}@paloaltonetworks.com

ABSTRACT

Malicious domains play an important role in many attack schemes. From distributing malware to hosting command and control (C&C) servers and traffic distribution, malicious domains are essential to the success of nearly all popular attack vectors. The detection of malicious domains has always been a hot topic in the security research community. Much effort has been put into building reputation-based malicious domain blacklists (MDBLs). However, in order to evade detection and blocking by the domain reputation systems, many malicious domains are now only used for a very short period of time. In other words, a malicious domain has already served most of its purpose by the time its content is detected and the domain is blocked. In this paper, we propose a system for predicting the domains that are most likely to be used (or are about to be used) as malicious domains. Our approach leverages the knowledge of the life cycle of malicious domains, as well as the observation of resource re-use across different attacks. The proposed system is built on top of various data feeds from the real world and our evaluation demonstrates the effectiveness of the predicted malicious domains.

1. INTRODUCTION

From distributing malware to hosting command and control (C&C) servers and traffic distribution, malicious domains are essential to the success of nearly all popular attack vectors. To detect and block the malicious domains, most modern domain reputation systems are designed to search for evidence of malicious activities exhibited by the domains. Evidence includes: malicious content such as malware, web pages with exploit code, web pages with driveby downloads, etc.; and malicious behaviour such as communicating with infected hosts to collect private information, to launch attacks, etc.

However, nowadays many malicious domains are only used for a very short period of time. The reasons for the short 'live' time are twofold. The first reason is to evade detection. In some cases, the malicious content/behaviour is only present/exhibited for a short time, thus making it very easy for detection systems to miss the evidence. In other cases, malicious domains have served most of their purpose by the time the malicious content is detected, hence the blocking of the domains is little help in defending against the attacks. The second reason is that the cost of registering a domain and setting up a server to host its content has become very low. For example, registering a common '.info' domain with a domain name registrar such as GoDaddy.com costs only ten dollars per year[1].

To solve this problem, we propose a system that predicts the domain names which are most likely (or about) to be used for malicious purposes. In this way, the predicted malicious domains can be blocked before or at the beginning of their being used for malicious purposes. Our approach is based on our knowledge of the life cycle of malicious domains and our research into the connections and patterns exhibited by various detected malicious domains. More specifically, we discovered that before a malicious domain can be used, attackers have to complete multiple actions in order to activate the domain. These actions cannot be bypassed. More importantly, these actions will leave traces in different types of publicly available data feeds. By identifying the traces related to the preparation or initial use of malicious domains, we can predict or provide early warning of the malicious domains and apply effective blocking.

In this work, we have done the following:

- We proposed a novel system to predict the domains that will be used by attackers for different malicious purposes.
- We discovered the re-use of malicious domains and identified patterns in the re-use.
- We designed a system to leverage Domain Generation Algorithms (DGAs) to automatically predict future malicious domain names.
- We discovered different connections between malicious domains that were used by attackers at different times.
- We discovered temporal patterns in DNS queries of the malicious domains before their use.
- We applied the system to predict malicious domains and we evaluated the effectiveness of the system.

It should be noted that in this work, we only focus on newly registered malicious domains. In other words, we only focus on the domains that are created specifically for a malicious purpose. We do not discuss benign domains being hacked and used for malicious purposes. This is because, with most benign domains, we cannot take any action until malicious activities are actually observed coming from the domains. Otherwise, the functionality of the benign domains would be interrupted. Under this assumption, we believe that prediction of benign domains being used for malicious purposes has less value.

2. BACKGROUND

This work is built on top of two important observations: the life cycle of malicious domains and the re-use of valuable resources among different malicious domains.

2.1 Life cycle of malicious domains

Knowledge of the life cycle of malicious domains presents opportunities to predict a malicious domain before it is used. Figure 1 shows the life cycle of a malicious domain. In general,

[1] Privacy protection costs another 10 dollars per year, but since many malicious domains are registered with stolen credit card information, privacy is not a real concern for attackers.

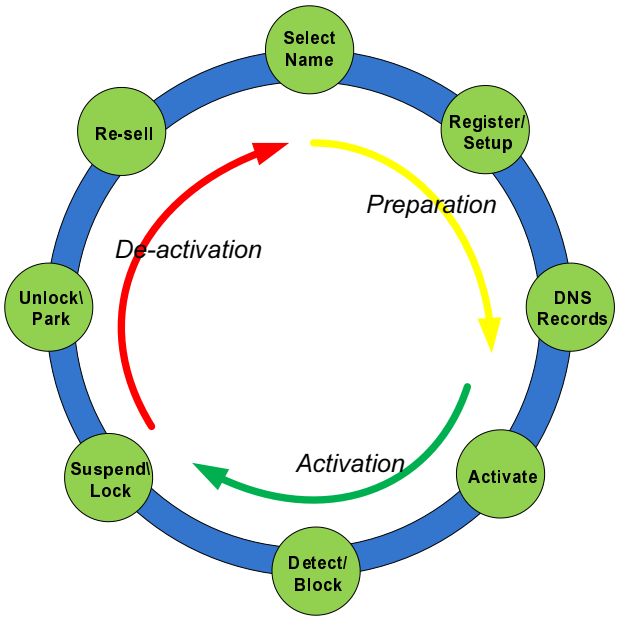

Figure 1: Life cycle of a malicious domain.

the life cycle consists of three phases: 1) preparation phase; 2) activation phase; and 3) de-activation phase.

In the preparation and activation phases, for a malicious domain to be usable (i.e. publicly accessible and able to serve content), attackers need to perform the following actions: selection of a domain name, registration of the domain name (if it has not already been registered), creation of DNS records so that the domain name can be resolved to an IP address controlled by attackers, and the setting up of a server using the IP address to serve malicious content (e.g. exploit pages, driveby downloads, C&C server, etc.). These actions follow certain sequences:

- The selection of domain names happens before registering the domain names.
- The registration of a domain name happens before the creation of DNS records for that domain name.
- Obtaining an IP address and using that address to set up a server also happens before DNS records are created, since the IP address is part of the DNS records.
- In a successful attack, the aforementioned actions all happen before the malicious domain is activated.

The de-activation phase happens when a malicious domain is detected and/or blocked. The domain name will be added into various malicious domain black lists (MDBLs). Meanwhile, the domain name will be suspended by the domain registrar. After a certain blocking period, the domain name may be resold to a new owner and will be disabled/removed from the MDBLs because listing a 'dead' malicious domain has no value.

The sequence of actions involved in activating a malicious domain, as well as the time interval between these actions, makes the prediction of malicious domains possible. However, it should be noted that the time interval between actions is not necessary for all of the prediction approaches presented in the remainder of this paper.

2.2 Re-use of valuable resources

The success of most popular attacks, such as DDoS, spamming, phishing and botnets, is dependent on many resources. These are often purchased. For example, domain names are registered or transferred for a price; bullet-proof servers are available for rent; large numbers of infected hosts are also available for rent. Some of the purchases are made through legitimate processes; others are made via illegal channels such as black markets, underground forums, etc. Many types of resources are made to be re-usable so that they can be resold multiple times to maximize financial gain. In fact, the re-use of valuable resources will become more and more prevalent given the fundamental underlying economic principles.

The re-use of resources across different attacks also presents opportunities for us to find connections between malicious domains. Using our knowledge of these connections, we can identify domains that are setting up and/or being prepared to be used for malicious purposes. We will dive into the details of how we exploit these opportunities in the following sections.

3. PREDICTING MALICIOUS DOMAINS

This section discusses how to predict malicious domains with the knowledge of their life cycle and re-use of resources. Based on the type of information our approaches leverage, the approaches can be categorized as follows:

- **Re-use of domain names**. As we have mentioned before, some malicious domains are created, used, abandoned/ suspended and then re-used. We propose an approach to find domain names that are most likely to be re-used.
- **Domain names**. Some domain names are generated by algorithms instead of humans, e.g. DGA domains. To predict these domains, we propose a system to automatically detect DGA malware and pre-generated domain names by feeding future inputs into DGA algorithms.
- **DNS queries**. DNS queries are often made after the registration and creation of DNS records, but before the domain is used. We have discovered several patterns in the DNS queries and we use these patterns to find malicious domains that are about to be used.
- **Connections between malicious domains**. By identifying connections between different malicious domains, especially domains used at different times, we can discover domains that have not yet been or are about to be used for malicious purposes. The connections we identified include: shared hosting IP addresses, shared DNS resolution infrastructure and shared domain registration information.

3.1 Re-use of domain names

We discovered the re-use of previously detected/suspended malicious domains in our malicious domain evidence monitor. The first characteristic is that the time interval between the

Domain name	Registered	First use	Registration changes	Most recent use
storagenl.info	2012-05-26	2012-09-24	2014-04-07	2014-05-13
markdownloads.info	2012-08-29	2012-10-31	2014-05-20	2014-05-04
installerlaunch-pz1.com	2012-04-06	2012-06-14	2014-03-27	2014-05-23
ncappworld.info	2013-02-15	2013-05-05	2014-03-26	2014-04-12

Table 1: Examples of malicious domain name re-use.

previous use and the current use is, on average, over a year. Since these domains have previously been detected, then suspended, this time interval ensures that two requirements are met: the domain names are back on the domain name transferring market, and they have been cleared from DNBLs.

The second characteristic is that most of the domains are still resolvable in the form of domain parking. If we take these two characteristics alone, many previously malicious domain names can be found, but the majority of the domains found will *not* be re-used for malicious purposes.

To find the domain names that are most likely to be re-used, we calculate a probability score of the domain being re-used.

This score is based on the domain name, TLD, changes in IP addresses and the price for domain transfer. When the score is higher than a certain threshold, we consider the domain is likely to be re-used for malicious purposes.

Table 1 lists some examples of domain re-use with the dates of the first time each domain was used for malicious purposes and the last time each domain was used for malicious purposes. The time spans between the first use and most recent use are all over one year. During the inactive time, the registration records of these domains have been changed. For example, the registration of 'storagenl.info' has been changed nine times over the inactive period, and the registration of 'installerlaunch-pz1.com' has changed seven times. Moreover, the malicious content of these domains has also changed between the first time they were used and the current use, suggesting they serve different malicious purposes.

3.2 DNS queries

Patterns of DNS queries for malicious domains can be used in prediction. We discovered several patterns in the DNS queries of a domain before the domain was used and/or detected as malicious. The patterns indicate different activities related to malicious domains, including preparing/testing the domain for malicious purposes.

In general, the patterns we discovered fall into three categories:

- **Change in DNS records**: The change in DNS records often happens when preparing a domain, e.g. setting up servers and modifying DNS records, testing DNS resolution, etc.
- **Temporal patterns of DNS queries**: DNS queries to different domains exhibit different patterns. We use two types of patterns. The first type detects a DNS query that exhibits a sudden outbreak in terms of volume; the second type detects a DNS query that exhibits periodic patterns.
- **Patterns in QNAME**: We are currently looking for three patterns. First, we are looking for a QNAME in a DNS query that looks like a DGA domain. Second, we are looking for a QNAME that is similar to known/detected malicious domains. Third, we are also looking for a QNAME that is similar to popular, legitimate domains, but which has no relationship to those domains.

It should be noted that not all of the domains found via DNS patterns are regarded as malicious. The domains have to be tested using two testing modules: a JS/HTML malicious content detector and a classifier based on features extracted from DNS records, AS, IP addresses, etc. If either tester returns a positive result, a domain is considered malicious. We acknowledge that prediction is a time-sensitive operation and that this extra layer of testing might lengthen the process, but we believe it is necessary to ensure the quality of predicted malicious domains.

3.3 Connections between malicious domains

Connections exist between malicious domains that have been used in different attack campaigns. We identified the following types of connections: 1) same name servers; 2) same IP addresses; 3) same registrant information. The existence of these connections is largely attributed to the re-use of resources. There are other reasons for the connections, e.g. embedding a pseudo identity of an attacker in the domain WHOIS information, multiple domains being registered using the same stolen credit card information, etc. For each type of connection, we first analyse the rationale behind the connection before we make use of it in our prediction.

Same name server

One name server can provide the DNS records for a large number of malicious domains. Given the important role of a name server, it is often found to be hosted on bullet-proof servers [1]. We detect and collect a list of name servers that provide DNS records for a number of domains, most if not all of which are malicious. Table 2 lists some examples of malicious name servers and the number of malicious domains that use the same name server[2]. To ensure the quality of the predicted domains, the collection of malicious name servers needs to filter out: 1) benign name servers (name servers of domain registrants, CDN providers, web hosting providers, etc.); 2) 'nonce' name servers (e.g. name servers that serve only one malicious domain). Also, the IP addresses of malicious name servers can be used to find other suspicious name servers hosted

[2] All of the malicious domains are registered/used at different times.

Name server	Number of malicious domains
ns*.starmiddle.com	6
ns*.01fonofni.ru	128
ns*.ganr.pl	16
ns*.erci.pl	7
ns*.qx9.pl	4

Table 2: List of known malicious name servers.

on the same IP addresses. However, we need to verify the maliciousness of the suspicious name servers before we can use them in our predictions.

To leverage the connection of using the same name server, we collect and maintain a known malicious name server list. We keep searching the Passive DNS (PDNS) data feed for domains that have NS records pointing to one of the name servers on the list. Any domain meeting this condition will be considered to be malicious.

Same IP address

The IP addresses used in prediction come from two sources. The first is the IP addresses used by known malicious domains; the second source is sinkhole-identified infected IP addresses. For IP addresses that come from the first source, we need to filter out those that are supposed to be used by different domains, e.g. floating IPs in CDN/web hosting. For IP addresses that come from the second source, we need to filter out 1) portal IP addresses; 2) commercial IP addresses. Our current IP-filtering scheme is mainly based on IP location/assignment information and existing knowledge of domains in CDN/web hosting/other enterprises. For example, to filter our CDN/web hosting IP addresses, we will look for IP addresses that have a owner that is a known CDN or web hosting provider.

The process of using the connection of use of the same IP addresses to predict malicious domains is similar to that of the same name servers. The challenging part is in accurately identifying the nature/functionality of the IP addresses so that we do not make false inferences.

Same registrant information

The WHOIS information for a malicious domain sometimes includes certain pseudo-identity information, e.g. the same/similar fake registrant name, the same registrant email, same registrant address, etc. Table 3 lists the field names that can be used to find connections.

WHOIS field	Match type
Registrant name	Fuzzy match
Registrant address	Exact match
Registrant phone	Exact match
Registrant email	Exact match

Table 3: WHOIS information used for discovering connections.

For registrant name, we apply a fuzzy match because we have noticed some very similar registrant names being used in different domains in the same attack campaign. For other information, although we have not observed cases in which attackers provide similar but not the same information to different domain names that he/she registers, we did not exclude the possibility of this happening.

For each type of connection, we create an individual module that follows the connections from known/detected malicious domains to unknown/unused malicious domains.

3.4 Pre-generating domain names

DGA is used to evade domain reputation systems [2], but once the DGA algorithm is reverse-engineered, it can be used to generate domain names that will be used in the future. Existing works [3, 4] on DGA-based malware detection assume that DGAs are used dynamically to generate a list of potential rendezvous points. Based on this assumption, researchers have proposed various approaches to identify the DGA-generated DNS queries in network traffic. However, based on our analysis, we found that DGAs are embedded in malware for other purposes. For example, we noticed that some malware use DGAs to generate traffic 'noise' so that the real rendezvous points (e.g. C&C servers) or communication channels can be obfuscated.

Therefore, in order to use DGA in prediction, several problems have to be resolved. The first problem is how to find DGA malware. The second is how to filter out DGA algorithms that are used to generate noise so that the real malicious domains can be hidden. We discovered that not all the DGA-generated domains are used for communication. The third problem is how to leverage the DGA algorithm to generate future domains without reverse engineering (the limitation of manual process). The fourth problem is that DGA algorithms normally generate hundreds to thousands of domain names per day (with the majority of the names being NX domains), making the discovery of domains that will actually be used more difficult. For the sake of scalability, the solution to these problems has to be an automatic approach.

To solve these problems, we designed and implemented a system that can automatically detect DGA malware using sandbox-based malware detection techniques. We leverage the idea of black-box testing and virtual machine execution acceleration to 1) differentiate the purpose of applying DGA algorithms in different malware; 2) generate a list of DGA domains that will be used in the future. We also track the registration and the DNS query to find the domains that will be used.

4. DATA SETS AND EVALUATION

4.1 Data sets

In order to predict malicious domains, we need to use and collect large volumes of different types of data. We will describe in general how we obtain the data, and the statistics of the data.

- **Passive DNS**: We have our own PDNS data feed as well as the data feed provided by SIE/Farsight [5]. On average, we collect 1.5 billion DNS responses per day.

- **Sinkhole**: We have built a sinkhole infrastructure that can automatically search sinkhole candidate domains, passively collecting connections to sinkhole domains, and store/process sinkhole log data. We have collected over 1.2 billion incoming connections to our sinkhole domains and over 28 million unique IP addresses of potential infected hosts.
- **WHOIS**: We have collected over 280,000 WHOIS records of malicious domains.
- **IP location database**: We purchased an IP location database from an external provider. The database contains the records of geolocation and assignment information of approximately 8.4 million IP prefixes.

4.2 Evaluation

Our evaluation process has three parts. The first part is to check *VirusTotal* for the detection rate across multiple anti-virus vendors. In this part, we log the time the domain is detected as malicious on *VirusTotal* and the detection rate.

In the second part, we check the PDNS data feed to look for the time difference between when we predict the domains and the time the domain names are queried.

In the third part, the predicted domain names are released in the form of DNS signatures that are used on a firewall to block matched DNS queries. After that, we check the log to find out which DNS signatures have been triggered.

We have analysed the results from over a month's data. During this time, we predicted 2,172 domain names, of which 1,793 (83%) have a *VirusTotal* detection rate of more than one vendor. The results of searching for the domain names in the PDNS feed show that, on average, our predictions are eight hours earlier than the first time the DNS queries to domains show up in the PDNS feed. Checking the signature-triggering log shows that 1,145 domains are actually triggered after they are released as DNS signatures. (It should be noted that the signature-triggering logs are collected from only five firewall devices.)

5. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed several approaches for predicting malicious domains. Our approaches are based on the observation that re-use of valuable resources occurs in setting up malicious domains, as well as the knowledge of the life cycle of malicious domains. We evaluated our approaches using a large set of data of different types, and the results suggest that these methods can predict malicious domains and could be used to effectively prevent attacks. In our future work, we will continue to look for more connections and evidence that suggest a domain name will very likely to be used for a malicious purposes.

6. REFERENCES

[1] Xu, W.; Wang, X.; Xie, H. New trends in fastflux networks. 2012. https://media.blackhat.com/us-13/US-13-Xu-New-Trends-in-FastFlux-Networks-WP.pdf.

[2] Wang, D. Y.; Savage, S.; Voelker, G. M. Juice: a longitudinal study of an SEO campaign. In Proceedings of the NDSS Symposium, 2013.

[3] Antonakakis, M.; Perdisci, R.; Nadji, Y.; Vasiloglou, N.; Abu-Nimeh, S.; Lee, W.; Dagon, S. From throw-away traffic to bots: detecting the rise of DGA-based malware. In Proceedings of the 21st USENIX security symposium, 2012.

[4] Antonakakis, M.; Demar, J.; Elisan, C.; Jerrim, J. Dgas and cyber criminals: A case study, 2011. https://www.damballa.com/downloads/r_pubs/RN_DGAs-and-Cyber-Criminals-A-Case-Study.pdf.

[5] The security information exchange (SIE), 2014. https://www.farsightsecurity.com/Services.