Predicting Zero-day Malicious IP Addresses

Amirreza Niakanlahiji
UNC Charlotte
aniakanl@uncc.edu

Mir Mehedi Pritom
UNC Charlotte
mpritom@uncc.edu

Bei-Tseng Chu
UNC Charlotte
billchu@uncc.edu

Ehab Al-Shaer
UNC Charlotte
ealshaer@uncc.edu

ABSTRACT

Blacklisting IP addresses is an important part of enterprise security today. Malware infections and Advanced Persistent Threats can be detected when blacklisted IP addresses are contacted. It can also thwart phishing attacks by blocking suspicious websites. An unknown binary file may be executed in a sandbox by a modern firewall. It is blocked if it attempts to contact a blacklisted IP address. However, today's providers of IP blacklists are based on observed malicious activities, collected from multiple sources around the world. Attackers can evade those reactive IP blacklist defense by using IP addresses that have not been recently engaged in malicious activities. In this paper, we report an approach that can predict IP addresses that are likely to be used in malicious activities in the near future. Our evaluation shows that this approach can detect 88% of zero-day malware instances missed by top five antivirus products. It can also block 68% of phishing websites before reported by Phishtank.

KEYWORDS

Zero Day Malware Prediction, Malicious IP Prediction

1 INTRODUCTION

Sharing Indicators of Compromise (IoCs), such as malicious IP addresses, malware hashes, and malicious URLs, is a key part of a modern cyber defense strategy. For example, most enterprises check an IP blacklist at the network perimeter to identify potentially malicious traffic. Such traffic is often blocked and, depending on policy, additional actions may be taken. For instance, a host sending information to blacklisted IPs may be investigated for zero-day infection. An obvious limitation with blacklists is that they only offer a rear mirror view of the threat landscape. Attackers can easily bypass an IP blacklist by using new IP addresses that have not been employed in malicious activities. Previous research works have attempted to predict IoCs that may be associated with new malicious activities, e.g., short DNS record TTL [1], a recently registered domain [3], and a misspelled domain name that are atypical to normal businesses [4]. In addition, others have shown that infrastructures that are used by attackers to launch their malicious activities tend to cluster in certain "neighborhood" (e.g. same hosting network) [2, 7].

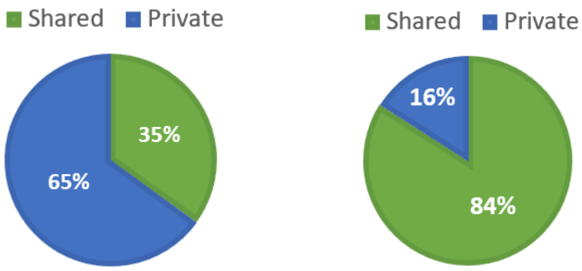
This paper presents an approach to predict IP addresses that are likely to be used for malicious activities based on Cyber Threat Intelligence (CTI) data sources. We start with the observation that attackers tend to find *soft targets* on the Internet to deploy the infrastructures such as drive by download, command and control (C&C), and web hosting, necessary for conducting their operations. All such infrastructure services require a public IP address, hence by blocking the IP addresses of *soft targets*, we can preemptively disrupt the attackers' operations without knowing about their attack plans. By *soft targets*, we mean that is least *costly* for the attacker. We define *cost* here more broadly to include both low price for purchasing hosting service or lax security measures in the following areas (1) low cost of exploiting existing resources such as a web server, or hijacking a domain name (2) low cost of renting new resources such as registering a domain name, (3) low risk of attrition due to lax verification of credentials, and (4) low risk of prosecution.

Based on our observation, we hypothesize that shared hosting services, where the services are shared by multiple independent entities, on the Internet are good candidates for soft targets and hence the probability of observing a shared hosting service involved in a malicious activity is considerably higher than a private hosting service. This hypothesis is based on the following reasons. First preliminary investigation suggests that the cost for shared hosting is significantly lower than private hosting. For example, we found a service provider advertising $99 hosting for life. Second, low-cost hosting providers typically offer little security service. Websites using such hosts may be easy targets for attackers so that attackers can acquire the IP address for *free*.

By resolving IP addresses for all .com and .net top level domains, we find that 35% of .com and .net domains run on shared hosts, where the same IP address is used for multiple unrelated domains. This is often provided by a hosting provider that adopts shared hosting as a business strategy to reduce cost. The rest, 65%, of .com and .net domains use private hosting services, where the IP address for the domain is not shared as shown in Figure 1a. In contrast, 84% of outbound malware traffic refers to shared hosts as shown in Figure 1b. For malware analysis, we used the GT Malware Passive DNS Daily Feed dataset (GT Malware dataset for short). GT Malware publishes a daily feed of DNS requests with about 250,000 malware instances.

Obviously, not all websites using shared hosting are malicious, and some shared hosting service providers offer effective security service. Another observation is that responsible businesses will

SafeConfig'17, November 3, 2017, Dallas, TX, USA
© 2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-5203-1/17/11…$15.00
https://doi.org/10.1145/3140368.3140369



(a) Among all service providers

(b) Among malicious service providers

Figure 1: Percentage of shared vs private hosting providers among (a) all service providers (b) malicious service providers on the Internet

choose online service providers with better security services. The networks operated by such providers exhibits less malicious activities.

In this paper, we proposed a new approach that is based on the previous two observations. The rest of the paper provides empirical evidence to support these observations. We show that our approach can detect 88% of zero-day malwares missed by the following AV software: Kaspersky, McAfee, AVG, Avast, and Symantec. It can also block 68% of phishing URLs before they are reported by Phishtank. We will also provide substantial evidence that our approach will not considerably impact the normal business needs of an enterprise.

The remainder of this paper is organized as follows: Section 2 describes the details of our prediction process for IP addresses that are likely to be engaged in malicious activities. Section 3 describes our evaluation results. We discuss and compare with related work in Section 4. Finally, section 5 concludes this work with discussion as well as directions for future research.

2 PREDICTING ZERO-DAY MALICIOUS IP ADDRESSES

The first task is to identify IP addresses that are engaged in the shared hosting behavior. We use Verisign's top level domain (TLD) zone files for .com and .net to identify shared hosting providers. We obtain daily Verisign TLD files and resolve IP addresses for .com and .net domains to identify IP addresses serving multiple domains owned by different organizations. According to [10], about 47 percent of all registered domains use .com as TLD. Therefore, our mapping is a good sample representation of all domain name to IP address mappings on the Internet.

Large enterprises often own multiple domains and point them to the same server. For example, both t.com and twitter.com are owned by Twitter. We treat these domains as aliases if they belong to the same organization. We use WHOIS registrant organization names for top domains on the Internet to identify such aliases.

The second task is gathering IP addresses associated with malicious activities from cyber threat intelligence data sources. In our prototype implementation, we collect data from GT Malware and Phishtank. GT Malware dataset contains DNS names requested by malware instances and the corresponding IP addresses. It contains approximately 250,000 new malware instances every day. Each malware instance is identified by its hash. On average there are 5,000 unique IP addresses reported as being associated with reported malware instances each day. We refer to this list of IP addresses as GT-IP-List. Phishtank is a community based website for sharing and validating phishing URLs. Users submit phishing URLs and other users check the URLs and vote to determine whether a URL is a valid phishing URL. Every day about 2,400 URLs are submitted by users on Phishtank.

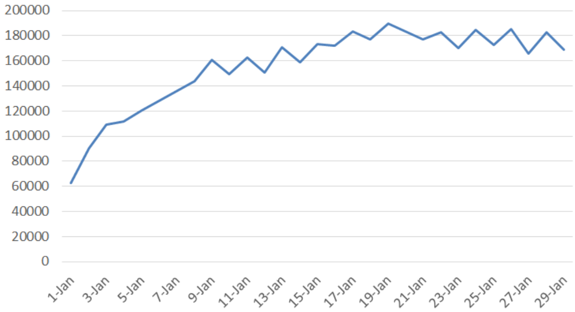


Figure 2: The size of IP Blacklist created based on GTMalware Dataset

Researchers have shown that an unreported IP address in a network that has many malicious IP addresses tends to be malicious than a network with a few such IP addresses[2, 7]. Our intuition is that shared hosting can play a vital role along with a /24 bit IP subnet block to predict IP addresses for future malicious activities. As a proof of concept implementation, we say IP address X to be likely engaged in malicious activities if it satisfies all the following conditions:

- X is hosting multiple websites operated by multiple entities (i.e. shared hosting)
- X has not previously known to be malicious
- X is in a /24 bit subnet that has as least one IP address in GT-IP-List over the past N days, where N is the time-window size.

In summary, to predict the list of potentially malicious IP addresses, we determine the /24 subnets that encompass reported malicious IP addresses and then for each of these subnets, we enumerate all the IP addresses that 1) were not appeared in the blacklists, and 2) is a shared host. We consider these IP addresses as potentially malicious IP addresses.

We also use a time window (seven days) to account for actions taken by service providers to "clean up the neighborhood". An IP address stays on the predicted list for only seven days if no new malicious activities are reported for rest of the IP addresses in the /24 subnet neighborhood. We have on an average 158,000 IP addresses each day in our blacklist after the first seven days. Figure 2 shows the graph for the number of Blacklisted IP addresses throughout January 2017.

3 EVALUATIONS

To evaluate our approach, we seek to answer the following research questions. First, how effective is this approach for preventing malicious activities? We choose to look at the detection of zero-day malware infections and blocking of phishing websites. In both cases, we benchmark our results against measures widely used by industry and show that we are better at detecting zero-day infections and blocking phishing websites. Second, how much impact would our predicted blacklist have on normal business functions? Third, what is the most effective time window for prediction?

3.1 Zero-day malware infections

We use GT Malware to evaluate the effectiveness of our approach in preventing zero-day malware instances. We use VirusTotal as an oracle to determine the maliciousness of hashes in GT Malware. VirusTotal is a public online file scanning service that determines whether a file is a malware. In addition to scanning a binary file, one can query the VirusTotal database by giving the hash of a binary file. VirusTotal provides results from more than 60 antivirus (AV) products. AV products can mistakenly identify a binary as malicious (false positive). As an oracle we use the the results of the following five high ranked antivirus products that are commonly used by today's businesses: `Kaspersky`, `McAfee`, `AVG`, `Avast`, and `Symantec`. To be more specific, a hash is labeled as malicious if it is regarded as malicious by at least one of the five AV vendors. To automate this process, we use VirusTotal Public API which was limited to 5,000 queries daily.

During Jan 2017, we randomly selected 5,000 unique hashes from GT Malware every day. We queried VirusTotal with each of those selected hashes immediately after GT Malware data becomes available. The line labeled "clean" in Figure 3 shows the daily number of hashes that were recognized as "benign" by all five AV products on that day. A significant subset of these "benign" hashes are predicted by our approach as malicious because they contacted IP addresses in our malicious IP prediction list. The number of daily predicted malicious hashes that evaded the five AV products are represented by the line "predicted" in Figure 3.

To evaluate the accuracy of our prediction, we asked VirusTotal to rescan the all "benign" hashes again in March 2017, two months after the initial query. In the intervening period, these AV products have changed the "verdict" for some of the hashes regarded as "benign" earlier. Hashes identified as malicious in the new scan by at least one of the five AV vendors were represented by the line "true positive" in Figure 3.

On average, our method predicted 88% of zero-day malware instances missed by all five AV vendors.

A practical application scenario for our approach might be as follows. A zero-day malware got past AV and infected a machine in an enterprise. As soon as the malware starts to generate DNS traffic, our predicted IP list will be able to detect this infection and timely quarantine the infected machine.

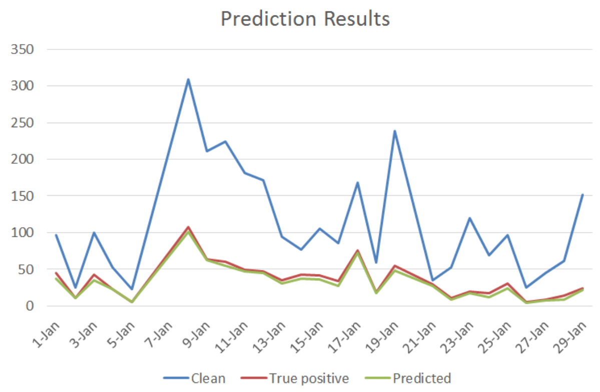


Figure 3: Zero-day malwares undetected by top 5 anti-viruses and predicted by our approach

Next, we evaluate the robustness of our approach. By robustness, we mean how many different malware families this approach is able to detect. One can imagine a situation that a specific malware family uses shared hosting as part of its infrastructure. It has many variants, and our approach may only be effective at detecting variants of this malware family.

For each malware instance we successfully predicted, we queried VirusTotal for its identity. VirusTotal would return multiple answers, each provided by a different AV vendor. For this evaluation, we used results from Symantec. Table 1 lists 28 malware family names for malware hashes detected by our approach during the period of evaluation (January 2017). Our approach appears to apply to a significant number of malware families.

Table 1: Diverse Variants of Malwares Detected

Trojan.ADH.2	PUA.Gen.2	Packed.NSISPacker!g4
Ransom.Cry	Downloader	PUA.Downloader
Trojan.Gen.2	Trojan.ADH	SecurityRisk.gen1
Infostealer	Trojan.Gen.8	Infostealer.Limitail
Trojan Horse	Trojan.Gen	Trojan.Gen.8!cloud
Trojan.Gen.6	Backdoor.Trojan	SecurityRisk.Downldr
PUA.DriverPack	PUA.InstallCore	Packed.Vmpbad!gen35
Ransom.Kovter	PUA.OpenCandy	Trojan.Zeroaccess!g3
PUA.Softonic	SMG.Heur!gen	PUA.ICLoader!g2
		ML.Attribute.High-Confidence

3.2 Zero-day phishing websites

We use Phishtank to evaluate the effectiveness of our approach in preventing zero-day phishing attacks. Phishtank is a community-based phishing dataset. It accepts reports of phishing URL. Phishtank allows users to vote to determine whether posted URLs are indeed phishing sites. This process is time-consuming and hence many published URLs in Phishtank are not verified as phishing URLs by the users.

We use VirusTotal as an oracle to determine the maliciousness of unverified URLs in Phishtank. If an unverified URL is identified as a phishing URL by at least two sources in VirusTotal, we will consider it as a phishing URL.

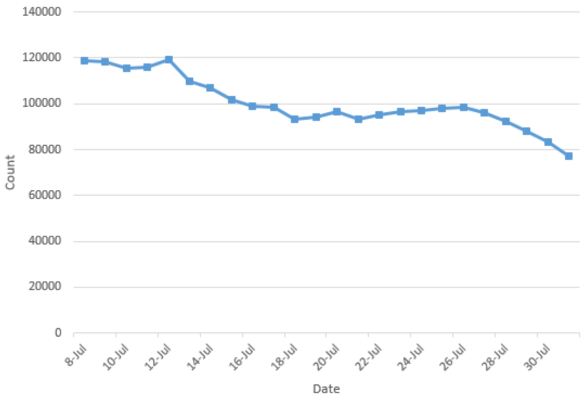


Figure 4: Number of zero-day IP addresses based on Phishtank dataset

We collected 70,953 phishing URLs published on Phishtank during July 2016. For user voting results, we collected voting results 30 days after each URL is first published on Phishtank. Based on user votes on Phishtank, 11,308 out of the published URLs were valid phishing URLs, and 319 URLs were not valid ones. For the remaining URLs, we queried VirusTotal and found that 54,724 URLs were reported as phishing URLs by at least two VirusTotal sources. We added these URLs to our dataset of valid phishing URLs

We applied our approach on the collected phishing dataset to determine the number of phishing URLs we could have predicted. In this experiment, instead of relying on GT-IP-LIST to mark /24 subnets on the Internet, we consider the IP addresses associated with valid reported phishing URLs. Figure 4 shows the number of predicted IP addresses on each day during July 2016. On average about 100K IP addresses will be added to the list of reported IP addresses on each day. Figure 5 shows the total number of reported phishing URLs and the number of URLs that we could have predicted on each day based on the resulted blacklist during July 2016. In our experiment, we could have blocked about 68 percent of phishing URLs before they are reported to Phishtank.

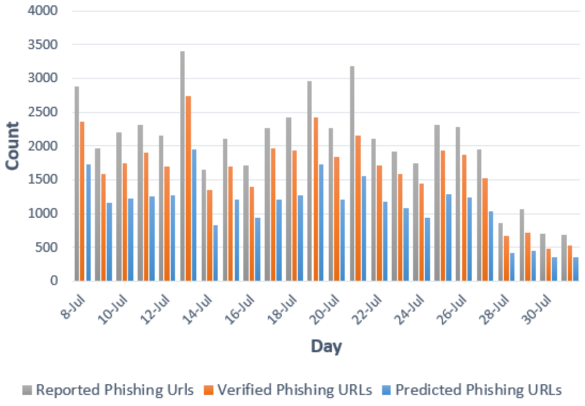


Figure 5: Number of predicted phishing URLs

We also checked whether the predicted phishing URLs belong to multiple phishing campaigns. To do so, we randomly selected a small number of phish links from predicted ones and manually examined them to determine which companies were the target of these phishing links. Based on the observation, the predicted URLs targeted different companies, which shows that our approach can block a broad range of phishing attacks.

3.3 Impact on normal business functions

In this section, we evaluate the impact of our approach on normal business function. Clearly not all IP addresses predicted are malicious. We start by evaluating our hypothesis that responsible businesses tend to have better cybersecurity and that attitude is reflected in the selection of hosting providers. We used Alexa top 1,000 websites as a proxy for responsible businesses. Over the period of January 2017, only the IP addresses of the following four of Alexa top 1,000 websites appeared in our predicted blacklists by our zero-day malware prediction approach: wordpress.com, wp.com, yandex.ua, 163.com, two of them were hosting WordPress contents. Note that WordPress sites are often blocked by large enterprises for poor security.

This evaluation suggests that vast majority of reputable businesses are not using service providers that may have higher security risks. Additionally, the average size of our predicted IP blacklist (160,000 IP addresses) is a very small fraction of the Internet (.004% of IPv4 space). One of our future research goals is to identify specific characteristics of a service provider we predict that is attractive to attackers.

To minimize the impact on normal business, one may distinguish human initiated browsing traffic vs. automated traffic. For example, visiting to a Wordpress site may be okay for human initiated browsing. Outbound traffic to likely malicious IPs that is not generated by human browsing may be blocked to minimize the risk of malware infection. Human browsing exceptions may be made utilizing commonly available safe browsing features in major browsers.

3.4 Prediction time window

Our approach predicts malicious IP addresses by considering observed malicious activities during a specific period of time. As we

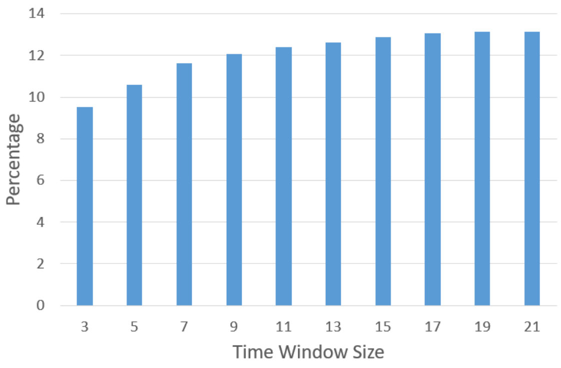


Figure 6: Avg. IP prediction percentages for different time-window size

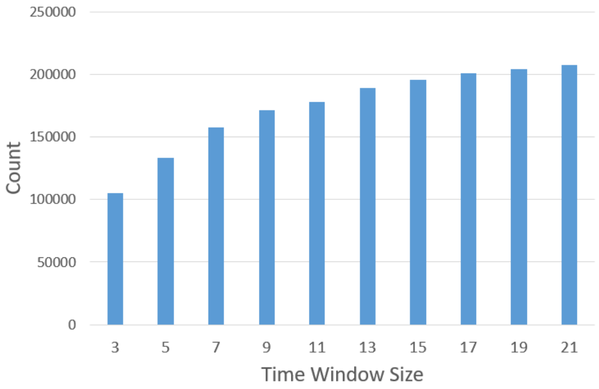


Figure 7: Change in Blacklist size over different time-window

increase this time window size, the number of predicted IP addresses increases as well. However, we should limit the size this time window because service providers often clean up malicious websites in response to reports. In this section, we evaluate the impact of the size of this time-window on the effectiveness of prediction using data from GTMalware.

Let $P_n(T_x)$ denote the percentage of malicious IPs (in GTMalware) predicted by our approach on day n for time-window T_x, where x is the size of time window. $P_n(T_x)$ is calculated according to Eq. 1 where $BlockedIP_n(T)$ is the number of IP addresses predicted by our approach and $NewIP_n$ is the number of new unique IPs in the GT Malware dataset on the n-th day.

$$P_n(T_x) = \frac{BlockedIP_n(T_x)}{NewIP_n} \tag{1}$$

Let $\overline{P(T_x, N)}$ denote the average malicious prediction rate for a given time window T_x over N days (N = 30 days, January 2017) as shown in Eq 2.

$$\overline{P(T_x, N)} = \frac{1}{N} * \sum_{i=1}^{i=N} P_i(T_x) \tag{2}$$

We calculated $\overline{P(T_x, N)}$ for a number of time windows ranging between 3 to 21 days as shown in Figure 6.

It is evident that IP predication rates increases with time window size. However, the rate of increase decreases quickly. We also calculated the impact of time-window on daily blacklist size. Figure 7 shows that the average blacklist sizes for different time window sizes over the same one month (January 2017) time period. From Figure 7, it is clear that blacklist size increases with the time-window, but the rate of increase is decreasing similar to the average prediction percentages. One possible explanation is that attackers constantly acquire new IP addresses in order to circumvent IP-blacklisting. As bad IP reports age, the chance that attackers still resides in the same subnet decreases. Moreover, as we mentioned earlier, service providers clean up the malicious domains once they are reported. Therefore the chance of predicting another IP addresses as malicious on the same subnet after the time-window will also decrease. As an evidence supporting this hypothesis, we observed in GTMalware that on average over 70% of the malicious IPs overlap in the same /24 subnet between two consecutive days. That overlap drops to 30% for two days apart.

Based on the results in Figure 6 and 7, selecting a time window between 7 to 21 days is reasonable depending on how much time one might want to give service providers to take down malicious activities.

4 RELATED WORKS

IP blacklisting is a well-established practice in the security community, and many companies are relying on blacklists to defend against attackers. Traditionally, IP blacklists are created by compiling cyber threat intelligence reports from different sources. Researchers have proposed ways of using blacklists to enable network firewalls to mitigate different types of attacks. Zhang et. al [11] proposed Highly Predictive Blacklisting (HPB), which is a PageRank-like algorithm to rank attack sources based on threat intelligence sources. Soledo et. al. [8] has an Implicit Recommendation System that extends HPB by considering temporal patterns of cyber attacks to prioritize attack sources.

Although compiling a blacklist from a set of threat data sources can be beneficial for cyber defense, such blacklists only offer a rear mirror view of the threat landscape. In recent years, many researchers have attempted to tackle this problem by identifying features that are shared among cyber threats that can be examined on incoming network traffic to determine whether they should be blocked.

Several research works, e.g. [2, 7, 9], have shown that malicious activities are not uniformly distributed over the Internet. In other words, malicious activities tend to cluster together and form high risk communities [7]. The goal in such works is to identify high risk networks that host such malicious activities. Collins et. al. [2] presented the idea of spatial and temporal uncleanliness in network to predict botnet IP addresses. Stone-Gros et. al. [9] presented FIRE, FInding Rogue nEtworks, to identify ISPs that are responsible for the most malicious activities. Moura et. al. [7] coined the term Internet Bad Neighborhood. They showed that spamming activities tend to be clustered in bad neighborhoods by analyzing spammer activities on the Internet. In such works, a network is considered as high risk if enough number of malicious activities (above some predefined threshold) reported by cyber threat intelligence sources are reside in that network.

Other research works such as [1, 5, 6] have suggested features that can be calculated on an incoming network request to determine whether it is maliciousness without requiring a collection of threat reports. McGrath et. al. [5] proposed several features such as number of IP addresses with a domain, number of ASs that these IP addresses reside in, and DNS record TTL that can be used to determine whether a domain name is using a fast flux technique that is commonly used by phishers. Moghimi and Varjani [6] proposed another set of features including the number of dots in URL, SSL certificate, URL length, blacklisted keywords to identify phishing URLs. Bilge et al. [1] proposed a system, EXPOSURE, to detect malicious domains. EXPOSURE consider four different sets of features: time-based features, DNS answered based features, TTL value-based features, and Domain name based features.

During the course of our research, we found that some of the proposed features are not effective in predicting zero-day IP addresses based on GT Malware data. For example, many research works such as [5] have reported a very short domain TTL is a good indicator for detecting malicious domains; however, we found that a significant number of reputable domains including Alexa top domains also have very short domain TTLs possibly due to the use of load balancers, or content delivery networks (CDNs). In this paper, we introduce a new salient feature, shared hosting, that is strongly correlated with malicious activities.

Ours is a hybrid approach in which cyber threat intelligence data sources as well as shared hosting are used to identify potentially high risk network neighborhoods. Our approach has a lower threshold for the number of observed malicious activities to identify high risk network neighborhoods as we are not solely rely on cyber threat intelligence sources to predict zero-day malicious IP addresses. Our approach is robust in that can be used to pro-actively identify a variety of infrastructures such as command and control servers, drive by download servers, and phishing web sites that are used by attackers to launch their attacks.

5 DISCUSSIONS AND FUTURE WORKS

In this paper, we propose a new approach to predict zero-day IP addresses that potentially will be used by the attackers in the near future based on recent cyber threat intelligence data reports. Through experimentation on two different cyber threat intelligence sources, we showed that with the presented approach, we can detect about 88% of unrecognized malwares by top five AV vendors and detect about 68% of phishing URLs.

Our results strongly support the idea that shared hosting services are being targeted by attackers and used for launching different types of attacks and hence is a good metric that can be used to detect zero-day attacks. Our preliminary investigation of such services suggests that these services are attractive for attackers mainly due to their low cost of renting as well as poor security. However, in the future, we want to investigate the reasons more deeply to understand the business model of the attackers and determine whether attackers can evade by changing their behaviors easily.

We plan to improve our algorithm of detecting shared hosting service providers such that we can distinguish their IP address from the ones that are used by content delivery networks (CDNs) and DNS parking servers and study each of these groups of IP addresses separately.

6 ACKNOWLEDGEMENT

This work is partly supported by a grant from the National Science Foundation under the award id 1266360.

REFERENCES

[1] 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 Transactions on Information and System Security (TISSEC)* 16, 4 (2014), 14.

[2] M Patrick Collins, Timothy J Shimeall, Sidney Faber, Jeff Janies, Rhiannon Weaver, Markus De Shon, and Joseph Kadane. 2007. Using uncleanliness to predict future botnet addresses. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement.* ACM, 93–104.

[3] Mark Felegyhazi, Christian Kreibich, and Vern Paxson. 2010. On the Potential of Proactive Domain Blacklisting. *LEET* 10 (2010), 6–6.

[4] Sujata Garera, Niels Provos, Monica Chew, and Aviel D Rubin. 2007. A framework for detection and measurement of phishing attacks. In *Proceedings of the 2007 ACM workshop on Recurring malcode.* ACM, 1–8.

[5] D Kevin McGrath, Andrew Kalafut, and Minaxi Gupta. 2009. Phishing infrastructure fluxes all the way. *IEEE Security & Privacy* 7, 5 (2009).

[6] Mahmood Moghimi and Ali Yazdian Varjani. 2016. New rule-based phishing detection method. *Expert systems with applications* 53 (2016), 231–242.

[7] Giovane CM Moura, Ramin Sadre, and Aiko Pras. 2011. Internet bad neighborhoods: the spam case. In *Network and Service Management (CNSM), 2011 7th International Conference on.* IEEE, 1–8.

[8] Fabio Soldo, Anh Le, and Athina Markopoulou. 2010. Predictive blacklisting as an implicit recommendation system. In *INFOCOM, 2010 Proceedings IEEE.* IEEE, 1–9.

[9] Brett Stone-Gross, Christopher Kruegel, Kevin Almeroth, Andreas Moser, and Engin Kirda. 2009. Fire: Finding rogue networks. In *Computer Security Applications Conference, 2009. ACSAC'09. Annual.* IEEE, 231–240.

[10] W3techs. 2017. Usage of top level domains for websites. https://w3techs.com/technologies/overview/top_level_domain/all. (2017).

[11] Jian Zhang, Phillip A Porras, and Johannes Ullrich. 2008. Highly Predictive Blacklisting.. In *USENIX Security Symposium.* 107–122.