Monitoring the Initial DNS Behavior of Malicious Domains

Shuang Hao
Georgia Tech
Atlanta, GA, USA
shao@cc.gatech.edu

Nick Feamster
Georgia Tech
Atlanta, GA, USA
feamster@cc.gatech.edu

Ramakant Pandrangi
Verisign, Inc.
Dulles, VA, USA
rpandrangi@verisign.com

ABSTRACT

Attackers often use URLs to advertise scams or propagate malware. Because the reputation of a domain can be used to identify malicious behavior, miscreants often register these domains "just in time" before an attack. This paper explores the DNS behavior of attack domains, as identified by appearance in a spam trap, shortly after the domains were registered. We explore the behavioral properties of these domains from two perspectives: (1) the *DNS infrastructure* associated with the domain, as is observable from the resource records; and (2) the *DNS lookup patterns* from networks who are looking up the domains initially. Our analysis yields many findings that may ultimately be useful for early detection of malicious domains. By monitoring the infrastructure for these malicious domains, we find that about 55% of scam domains occur in attacks at least one day after registration, suggesting the potential for early discovery of malicious domains, solely based on properties of the DNS infrastructure that resolves those domains. We also find that there are a few regions of IP address space that host name servers and other types of servers for only malicious domains. Malicious domains have resource records that are distributed more widely across IP address space, and they are more quickly looked up by a variety of different networks. We also identify a set of "tainted" ASes that are used heavily by bad domains to host resource records. The features we observe are often evident before any attack even takes place; ultimately, they might serve as the basis for a DNS-based early warning system for attacks.

Categories and Subject Descriptors

C.2.3 [**Computer-Communication Networks**]: Network Operations—*Network monitoring*; K.6.5 [**Management of Computing and Information Systems**]: Security and Protection

General Terms

Measurement, Security

Keywords

DNS, Domain Registration, Spam, Malicious Domain

IMC'11, November 2–4, 2011, Berlin, Germany.

1. INTRODUCTION

The Domain Name System (DNS), the Internet's lookup service for mapping names to IP addresses, provides a critical service for Internet applications; unfortunately, it also allows attackers to direct victims to Web sites that host scams, malware, and other malicious content. To mitigate these threats, network operators try to derive a reputation for each domain that reflects the likelihood that the domain is associated with a particular type of attack (e.g., scam, phishing, malware hosting). The rate at which new domains appear makes quickly developing a reputation for these domains particularly challenging: in our analysis, we find that over tens of thousands of new domains are registered every day. Existing DNS reputation systems use the characteristics of DNS lookups from resolvers that look up a domain to distinguish legitimate from malicious domains [1, 2]. Unfortunately, these systems must observe a significant volume of DNS lookups before determining the reputation for a domain, which only occurs *after* compromise has taken place.

Towards facilitating pre-attack detection of malicious domains, we study the initial DNS activity for each domain and characterize how the observable behavior for a malicious domain differs from that of legitimate domains. We study two aspects of initial DNS behavior associated with domains: (1) the *DNS infrastructure* used to resolve the domains to IP addresses; and (2) the *DNS lookup patterns* from the networks that perform initial lookups to the domain. Certain characteristics of the DNS infrastructure may be unique to malicious domains, such as the IP address ranges and ASes that host either the authoritative name servers for the sites, or the sites themselves. Identifying infrastructure that is common across malicious domains may provide hints for identifying malicious domains before the attacks themselves are mounted. Characteristics of early DNS lookups can help network operators discover valuable information about the nature of the domains that are being looked up. Notably, we find that domains that are registered for malicious purposes are initially queried from a much more diverse set of subnets than legitimate domains.

Our study of DNS behavior early in a domain's life cycle is motivated by our ultimate desire to perform early detection of malicious domains. We use domains collected at several large spam traps as a source of domains associated with spam campaigns. To characterize the resource record behavior of each domain, we perform periodic iterative queries of newly registered domains in March 2011. To characterize DNS lookup patterns across networks, we use information about DNS lookups collected from the Verisign top-level domain servers, coupled with registration information about these domains.

We focus exclusively on the early DNS behavior of a domain, which is enabled by two important pieces of information. First,

registration records alert us when a domain is registered, and allow us to begin querying it immediately, before attacks. Second, we study a global view of early DNS lookup patterns across the entire Internet for `.com` and `.net` domains. Our study reveals the following findings:

- *Domain registration and resource record establishment happens before attacks take place.* As many as 55% of spam campaigns may occur at least one day after the domain referenced in the spam messages were registered, offering the potential for early discovery of malicious domains based on initial DNS behavior.
- *DNS infrastructure for malicious domains is located in different address space regions and autonomous systems than the infrastructure for legitimate domains.* A few autonomous systems and IP address regions host infrastructure only for domains that are associated with malicious activity. Identifying these at domain registration time can potentially enable early detection.
- *Early lookup patterns for a newly registered malicious domains differ significantly from the patterns for a legitimate domain.* Domains associated with spam campaigns are initially looked up by a more diverse set of network address regions than legitimate domains. Especially, the newly registered spam domains become "popular" more quickly.

These features may ultimately be used to develop unique fingerprints for distinguishing legitimate domains from those that are associated with Internet attacks.

The rest of this paper is organized as follows. Section 2 surveys the problem context and related work. Section 3 describes the data sets that we use for our analysis. Section 4 studies the characteristics of resource records for newly registered legitimate and malicious domains. Section 5 studies the lookup characteristics for different types of domains, and Section 6 concludes.

2. CONTEXT AND RELATED WORK

We provide a brief overview of DNS records and lookups, as well as an overview of recent DNS studies.

2.1 DNS Resource Records and Lookups

When an entity registers a DNS domain, domain name registries insert several basic entries into the zone files to refer to the services for the domain. NS records point to the authoritative name servers for the zone, MX records point to the domain's mail servers, and A records point to the hosts. The NS and MX records can be further resolved to IP addresses. A single domain is typically assigned multiple server records for redundancy, but the number of IP addresses associated with the records is typically much less than the number of the domains being registered.

In March 2011, three million second-level domains under `.com` and `.net` were newly registered with NS records, but the number of distinct IPs mapped from NS records was just 150 thousand; A records and MX records have similar statistics. This observation indicates that the same server has been repeatedly used by many different domains to host DNS infrastructure (e.g., the most "heavily" used IP addresses carry NS records for around 300 scam domains).

Recursive DNS servers relay the users' queries to the zone's authoritative servers to acquire resource records, which reduces DNS traffic in the wide area. Recursive servers commonly respond to the hosts' requests within their respective networks, so the set of recursive servers querying for a domain can be a reasonable approximation for the networks that have attempted to reach the domain. The top-level domain (TLD) name servers thus provide a natural vantage point for monitoring the lookups directed to the second-level domains (the direct sub-domains below a TLD). Although DNS caching prevents us from determining the volume of lookups to a domain, the distribution of the recursive servers contains rich information about which networks have issued lookups for a domain; this statistic is particularly useful during the early part of a domain's life cycle, when it is initially registered and no caching has yet taken place.

2.2 Related Work

Monitoring and analysis on zones' resource records Previous studies have used the mechanism of querying the DNS servers to check the zones' resource records. Holz *et al.* investigated the diversity of the A records returned in the lookups to identify fast-flux service networks [10]. Konte *et al.* studied the changing rates of the IP address in the DNS records of scam domains [12]. Our work, on the other hand, tracks DNS records of newly registered domains to infer spatial and temporal characteristics. Anax scanned the recursive servers to find out anomaly in the cached records and detect poisoning attacks [1]. In contrast, we monitor the records in the zones' authoritative servers to discover the characteristics in the malicious domains' registration.

DNS lookup patterns The first studies of DNS lookup behavior at a local resolver were performed by Danzig *et al.* [8] and Jung *et al.* [11]; both of these studies examined lookup behavior from the vantage point of lookups to a single local resolver, and did not attempt to characterize how these lookup patterns differed for malicious domains. Notos [2] and EXPOSURE [4] studied DNS lookup behavior within a local domain below the DNS resolvers to build the domains' reputation. Such a view of DNS lookup behavior is valuable, but this vantage point cannot reveal coordinated behavior across multiple networks, and it relies first on an attack to take place or hosts being compromised before it can detect any malicious domains. Antonakakis *et al.* also monitored the DNS traffic from authoritative servers or top-level domain servers to detect malware domains [3], but the zones' dynamics were still reconstructed from DNS request and response messages; they did not focus on the behavioral analysis of newly registered domains. Other work has examined DNS lookup behavior at a DNS root server [5–7]. The focus of these studies was different from this paper. Castro *et al.* [7] and Brownlee *et al.* [6] attempted to characterize how much DNS traffic at the DNS root server was illegitimate. Broido *et al.* identified misconfigured hosts using spectrography to identify machines that were mistakenly issuing automatically configured DNS queries [5]. In contrast, we study DNS lookup patterns from the perspective of a top-level domain, and examines the behavior of lookups as seen from recursive resolvers, as opposed to lookups from individual hosts.

Domain registration inference Recently, a number of research efforts have studied domain registration patterns. Kreibich *et al.* [13] investigated the time from a domain's registration to its use in spam. Spring *et al.* examined the delay between registration of a malware domain and the first successfully resolved response in DNS traffic [17]. We make a similar observation, but on a much larger set of domains under `.com` and `.net`; further, we explore the DNS characteristics in domains' early life cycle. Felegyhazi *et al.* [9] proposed to automatically identify malicious domains based on `WHOIS` and name server information. In contrast, we actively collect different types of resource records to track the changes, and monitor the networks querying the domains. Some of the charac-

type	*example*
DNZA entry	add-new example.com NS ns1.example.com
Query record	example.com 111.111.111.0 , 22.22.22.0

Table 1: Data format examples.

teristics that we observe about malicious domains could be used to build efficient dynamic reputation systems.

3. DATA COLLECTION

We describe our data and the process of probing for resource records and correlating with spam messages.

DNS data The top-level domain servers are responsible for maintaining the zone information (more specific, second-level domains) and answer the queries for the registered domains. Verisign, Inc. operates the generic top-level domains (gTLDs) for `.com` and `.net`, which account for over 45% of registered domain names on the Internet [16]. The servers maintain two kinds of dynamics about the second-level domains. The first type of information is the *Domain Name Zone Alert (DNZA)*. This information includes changes about the zone, such as whether a domain name was newly registered or a name server's IP address was modified. The DNZA files keep track of these changes.

The second type of information concerns the *DNS queries* issued by the recursive servers. After the recursive servers sent queries to the TLD name servers for resolving the second-level domains names, Verisign's systems aggregated the source IP addresses into /24 subnets for logging and the TLD name servers recorded the querying subnets each day. The query records show the relationship between the domain names and the queriers. Verisign deploys multiple TLD name servers to resolve second-level domain names, and we collected the logs of querying /24s from all servers for analysis. Table 1 shows the example format of each type of data. The DNZA entry indicates that an "add-new" command created a new domain `example.com` and the NS record was `ns1.example.com`; The query record means that there were queries from /24s of "111.111.111.0" and "22.22.22.0" for the domain. The DNZA files and the query data were collected at Verisign's `.com` and `.net` TLD name servers during the period of March 2011. On average, about 80 million domains were queried each day.

Resource records The DNZA entries with "add-new" commands show what domains are newly registered. To get the new zone's resource records, we must perform active queries to their authority servers, since the second-level domains' records are not available within the TLD name servers. After a second-level domain under `.com` or `.net` is registered, we probe the domain once a day to discover the resource records and the resolved IPs. As mentioned in Section 2, we collect NS, MX and A records. We performed the probing procedure during March 2011. For example, 190 thousand domains were created on March 1, 2011; we continually queried those domains over the next 30 days. At the end of March 2011, we accumulated 4 million domains for monitoring. We use the PlanetLab platform [14] to make it feasible to query a large set of domains. Each PlanetLab node is responsible to query a subset of domains, and deliver the collected information back to the central monitor. Eventually, we deployed around 150 PlanetLab nodes to perform the probing procedure throughout the month. Though the daily querying does not capture all the changes in the resource records, it continually tracks the "snapshots" of DNS infrastructure and implies the change trends.

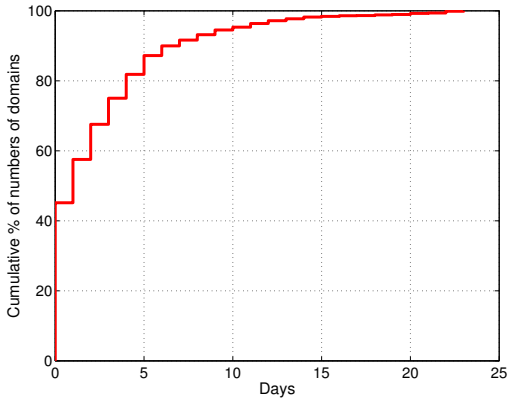


Figure 1: Days between a malicious domain's registration (in March 2011) and the time when the domain showed up in spam.

Spamming Scam domains appearing in spam messages are the major targets in our study, since timestamps in each email help explicitly identify when the spamming activities occurred, and spam is related to many different attacks, such as phishing. We used spam trap to capture emails sent from spammers during March 2011. Because the domains for the spam trap have no legitimate email addresses, emails received at the mail server were all spam. The second-level domains appearing in the messages' URLs were extracted as being involved in spamming activities (overall, 40% of unique second-level domains were found under `.com` and `.net`). In the context of this paper, we use "scam domains" and "malicious domains" interchangeably to refer to the second-level domains identified being associated with spam. From spam traps, we identified 2,045 scam domains as newly registered during March 2011. We also checked the domains with Spamhaus [15], and identified 4,587 blacklisted second-level domains. The union of these two sets yielded a total of 5,988 `.com` and `.net` second-level domains that we considered spamming-related.

To obtain a representative set of legitimate domains for comparison, we sampled 6,000 domains registered during March 2011 that have not yet appeared in any blacklist.

4. REGISTRATION & RESOURCE RECORDS

We first check the time between the registration of a domain and the subsequent attack to investigate the potential for early detection. Then, we explore how DNS behavior associated with infrastructure—where a domain's resolvers initially reside—can be an early signal for malicious domains.

4.1 Time Between Registration and Attack

We hypothesize that there may be some time between when spammers register new domains and when they send spam. We examine the extent of the delay between the time when a domain is initially registered and when it is ultimately used in an attack. If such a delay exists, it might allow blacklist operators to list the malicious domains, possibly before the spam campaign occurs.

How much time occurs between the domain registration and attack? Figure 1 shows the distribution between the time when we start to observe records about the malicious domains registered

in March 2011, and the earliest time when the domains appeared in the spam messages. We take the timestamps in our spam traps, as well as emails received at the Yahoo! mail servers. Yahoo! Inc. provides the received time of all email messages and the URLs contained in the messages. The Yahoo! email gives a broader coverage of monitoring around the world. We take the earliest time points when seeing a "bad" domain in email messages (either in Yahoo! data or in spam trap) as the estimated start of the spamming attack about that domain. The x-axis represents the delay between when a domain was registered and when we first witnessed the domain associated with a spam campaign, and the y-axis is the percentage of the malicious domains registered in March 2011.

Finding 4.1 (Delay until attack) *More than* 55% *of the malicious domains appeared in spam campaigns more than one day after they were registered.*

We define the first five days after domain registration as *pre-attack period*. About 20% of domains might not be used in attacks during this period, and the time windows for other domains being explored in spamming are also limited. In the rest of the paper, we will analyze the characteristics of DNS infrastructure for malicious domains both throughout their lifetime (i.e., after the domains' registration) and within the pre-attack period. In Section 5, we further investigate the lookup behavior during the early stage.

4.2 Location of DNS Infrastructure

When determining the IP address that maps to each DNS record, we find that the assigned records for spam domains across IP space are far from uniform.

How is the DNS infrastructure that hosts a domain initially distributed across IP address space? The initial distribution of domain records across IP address space may provide clues as to a domain's reputation. Figure 2 shows how the IPs associated with NS, MX, and A records from malicious and legitimate domains are distributed across IP address space. The x-axis represents IPv4 space. If an IP maps to multiple records from different domains, we count it only once in the figure. The y-axis indicates the percentage of addresses less than or equal to the IP value on the x-axis. The solid blue curves plot the distribution of legitimate sample domains, the red dashed curves show the outcome of malicious domains, and the green dash-dot curves represent observed records for the malicious domains during pre-attack period. Interestingly, we observe that the DNS records associated with malicious domains are distributed differently than the records associated with legitimate ones.

Finding 4.2 (Distribution across IP address space) *The IP addresses used by malicious domains in the NS, MX and A records are distributed densely in a small fraction of IP address space.*

The IP addresses associated with DNS resource records are not distributed evenly across the IP address space. Some network range has more IPs pointed from NS, MX or A records; while the record IPs in other fraction of address space are distributed sparsely. Particularly two network blocks carried records from malicious domains, 96.45.0.0/16 and 216.162.0.0/16. The prefix 173.213.0.0/16 has many IPs in spamming domains, but the same range hosts legitimate domains, too. This observation indicates that if IPs corresponding to different domains' records reside close to each other in a network block, those domains may appear in spam in the future.

Is the DNS infrastructure for malicious domains located in particular ASes? Of course, the IP addresses of the records are not sufficient to confirm that a domain is scam-related. We examined the distribution of the resource records across ASes and compared the distribution of legitimate and malicious domains. Table 2 shows the top three ASes ranked by the percentage of domains ever having records being resolved into ASes.

(a) Legitimate domains

Type	*AS*	*domain ratio*	*AS Name*	*Country*
	8560	15.9%	1&1 Internet AG	Germany
	26496	10.9%	GoDaddy.com, Inc.	U.S.
NS	4134	10.1%	Chinanet Backbone	China
	26496	30.5%	GoDaddy.com, Inc.	U.S.
	15169	7.3%	Google Inc.	U.S.
MX	21844	7.0%	ThePlanet.com	U.S.
	26496	31.8%	GoDaddy.com, Inc.	U.S.
	8560	4.3%	1&1 Internet AG	Germany
A	21844	4.1%	ThePlanet.com	U.S.

(b) Malicious domains

Type	*AS*	*domain ratio*	*AS Name*	*Country*
	4134	33.6%	Chinanet Backbone	China
	28753	17.0%	Leaseweb De	Germany
NS	31365	16.3%	SGSTelekom	Turkey
	197088	23.9%	Colohost LLC	Latvija
	3292	19.3%	TDC Data Networks	U.S.
MX	5632	12.3%	3dgwebhosting.com Inc	U.S.
	4134	19.3%	Chinanet Backbone	China
	197088	14.3%	Colohost LLC	Latvia
A	30890	13.8%	Evolva Telecom	Romania

(c) Malicious domains in pre-attack period

Type	*AS*	*domain ratio*	*AS Name*	*Country*
	4134	37.8%	Chinanet Backbone	China
	28753	20.4%	Leaseweb De	Germany
NS	27699	11.3%	Tel. De Sao Paulo S.A.	Brazil
	197088	14.3%	Colohost LLC	Latvija
	3292	21.8%	TDC Data Networks	U.S.
MX	5632	12.3%	3dgwebhosting.com Inc	U.S.
	4134	19.7%	Chinanet Backbone	China
	197088	15.0%	Colohost LLC	Latvia
A	28753	11.6%	Leaseweb De	Germany

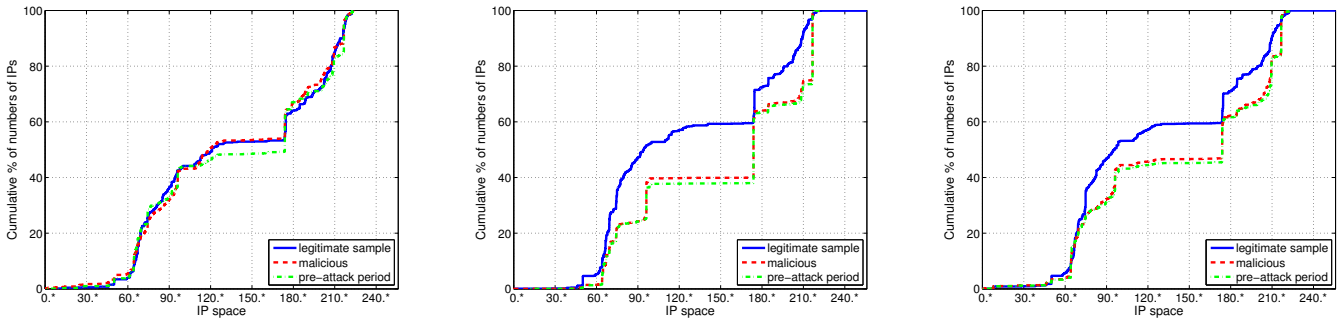
Table 2: Top three ASes containing domains' records.

Finding 4.3 (Distribution across ASes) *More than* 30% *of the malicious domains have at least one record resolving to one or two particular ASes, which are different from those ASes mostly used by legitimate domains.*

We observe that many of the new legitimate domains have larger registrars like `GoDaddy` operate their DNS, and host their service infrastructure with well-known provider, like `Google`. On the other hand, spamming domains' records are scattered across multiple ASes and countries. Spammers appear to prefer certain specific ASes to host their DNS infrastructure.

Are there "bad" ASes that host DNS infrastructure exclusively for malicious domains? We define an AS as "tainted" once the number of malicious domains whose DNS records are resolved within the AS exceeds a threshold. The set of tainted ASes represent the networks that attackers most heavily use, as indicated by the malicious domains' registration. After a domain's registration, attackers create DNS entries for the domain, and the records resolve to different IP addresses. We then check whether the resulting IPs belong to the tainted ASes. If a domain accumulates many records that resolve to tainted ASes, we suspect that the domain is related to the observed attacks.

Finding 4.4 (Domains hosted by "bad" ASes) *Most legitimate domains have A, MX, and NS records that are hosted*



(a) NS records (b) MX records (c) A records

Figure 2: Fraction of IP addresses associated with malicious domains and comparison with legitimate domains.

almost entirely in untainted ASes. On the other hand, the majority of spam domains have records hosted in tainted ASes, even during the pre-attack period.

We derive the tainted AS set by including an AS that has hosted records for more than 100 spam domains. Figure 3 shows the ratio of the tainted record number to the number of all records for the domain. More than 90% of legitimate new domains have zero records belonging to the tainted AS set.

5. EARLY LOOKUP BEHAVIOR

The recursive DNS resolvers initially query the TLD name servers to get referrals to second-level domains. In this section, we explore the characteristics in the lookup networks to different types of domains. Queries to a malicious domain may signal the onset of attack, and the abnormal pattern in the global DNS traffic could help to detect the attack campaign in its infancy.

5.1 Network-Wide Patterns

We first investigate the querying patterns across different domains, to see whether similar sets of networks were looking up different domains. Our intuition is that domains that are used for malicious purposes may be looked up by similar groups of networks as well. For example, a user clicking on a URL in spam might click on other spam URLs. If two domains are queried by the same set of recursive DNS servers, they may be the same type of domain.

Are the networks querying different domains distributed similarly? We measure the similarity using an average pairwise similarity of querying /24 network blocks over n days. Suppose two domains A and B who have sequences of querying /24 set $\{a_1, a_2, \ldots, a_n\}$ and $\{b_1, b_2, \ldots, b_n\}$ over n days. The similarity between domains A and B is

$$S(A,B) = \frac{\sum_{i=1}^{n} J(a_i, b_i)}{n}, \text{ where } J(a_i, b_i) = \frac{|a_i \cap b_i|}{|a_i \cup b_i|}$$

where $J(a_i, b_i)$ is the Jaccard index of set a_i and b_i: the size of the set intersection divided by the size of union. Based on this pairwise similarity, we aggregate the domains into different groups using single-linkage clustering, a simple and efficient clustering method [18]. We considered a 5-day time period from March 1–5, 2011, during which there were 804 malicious domains and 1,104 sampled legitimate domains registered. We terminate the clustering after 50,000 comparisons, which places 1,631 domains into 17 clusters that have more than a single domain. We expect domains to fall into distinct clusters. Tables 3 shows the statistics for the five largest clusters. The first three columns show the domain counts in each cluster. The last column means the percentage of the malicious domains in the cluster.

total	*malicious*	*legitimate*	*% malicious*
1404	463	941	33.0%
157	156	1	99.4%
16	16	0	100.0%
10	10	0	100.0%
10	10	0	100.0%

Table 3: Five largest clusters based on lookup networks.

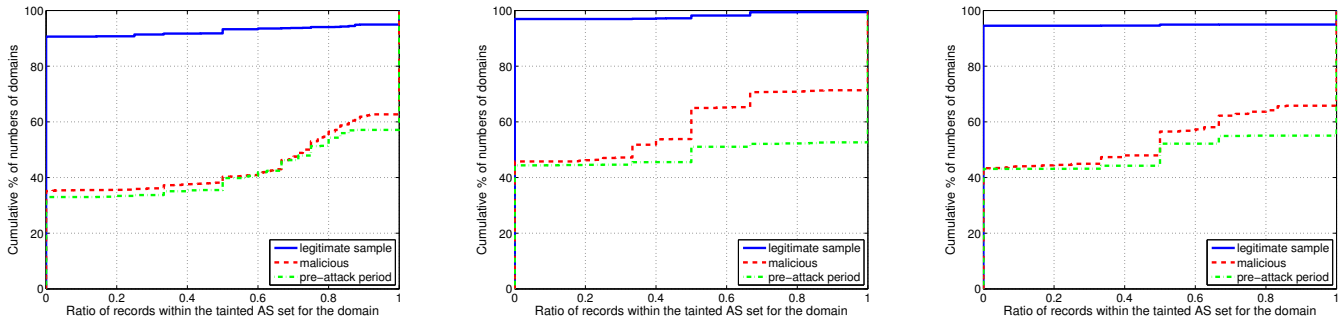
Finding 5.1 (Similarity in lookups) *Different malicious domains are looked up by similar group of network blocks, which may indicate that they are part of the same spamming campaign.*

The results show that clustering often works well: many of the clusters contain either only all good or all bad domains. The legitimate domains contained in the large cluster are only queried by a small number of networks, which a detection system could easily filter. These results suggest that domains of certain types do share similar network-wide spatial lookup patterns that may ultimately be used as input to a blacklist.

5.2 Evolution of Lookup Traffic

The numbers of distinct networks querying the TLD servers for the second-level domains approximate how widely around the world the users try to connect to these domains. Although there might be multiple connection attempts behind one recursive server, counting all querying recursive servers is a good indicator for the domain's initial "popularity". Our intuition is that once deployed, malicious domains may receive a lot of traffic in a short time, but visits to legitimate domains will increase relatively more slowly.

How quickly do the newly registered domains become popular? Figure 4 shows the average lookup volume from /24s for domains in different categories over time. The x-axis shows the number of days after a domains' registration. The y-axis shows the average number of querying /24s over the domains with error bars (i.e., standard error). The solid blue curve shows lookups for legitimate domains; the y-axis values are multiplied by 10 to make the figure more readable. The dashed red curve shows the patterns of malicious domains.



(a) NS records (b) MX records (c) A records

Figure 3: The distribution for the ratio of domains' records falling in the "tainted" AS set.

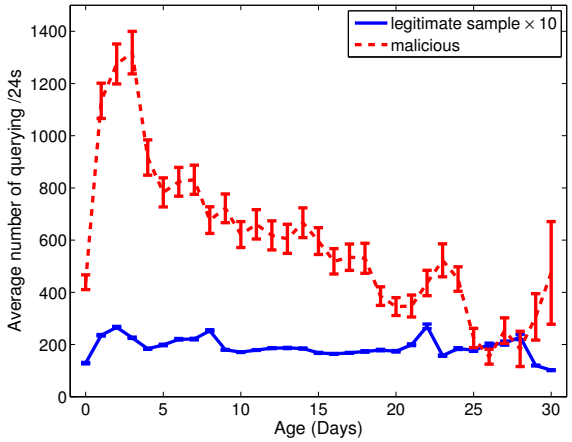


Figure 4: Number of querying /24s after domains' registration.

Finding 5.2 (Initial lookup trends) *Queries to the malicious domains increased quickly after the domains were registered, and usually reached the peak in the first 3–4 days.*

On the other hand, /24s querying for domains not reported as malicious increased slowly and stayed relatively low over the 30-day period. The markedly different lookup patterns of likely legitimate domains and those involved in spamming activities might ultimately help blacklist operators quickly detect bad domains, by watching for newly registered domains that suddenly become popular. The changes to malicious domains also indicate that the initial five-day period may contain valuable information, since these attack domains are heavily queried at the beginning, but lookups quickly trail off after that.

6. CONCLUSION AND FUTURE WORK

We have monitored DNS resource records for second-level domains newly registered in March 2011 and examined the lookup traffic to large authoritative top-level domain servers. We show the DNS characteristics observed at TLD name servers and extracted from zones' resource records for malicious domains are different than those for legitimate domains. Resource records of malicious domains tend to resolve to specific IP address range and ASes. Once we identify a set of "tainted" autonomous systems that host many scam domains, the legitimate domains rarely have resource records within the tainted AS set. We also discover that miscreant domains exhibit distinct clusters, in terms of the networks that look up these domains. Finally, we find that these domains become widely popular considerably more quickly after their initial registration time.

The distinct DNS characteristics and their tendency on different types of domains suggest that it may ultimately be possible to fingerprint domains based on their resource records and lookup traffic close to TLD name servers before an attack ever takes place. Although a single pattern in DNS might have limited power to identify malicious domains, the combination of our findings may ultimately guide the design of future "early warning" systems for DNS.

Acknowledgments

We thank Yahoo! Inc. for providing the received time of the email messages and the contained URLs. We also thank our shepherd, Christian Kreibich, as well as David Dagon and Anirudh Ramachandran for helpful comments on the paper. This work was funded by NSF CAREER Award CNS-0643974 and NSF Awards CNS-0716278 and CNS-0721581.

REFERENCES

[1] M. Antonakakis, D. Dagon, X. Luo, R. Perdisci, W. Lee, and J. Bellmor. A Centralized Monitoring Infrastructure for Improving DNS Security. In *Proc. 13th International Symposium on Recent Advances in Intrusion Detection (RAID)*, Ottawa, Ontario, Canada, Sept. 2010.

[2] M. Antonakakis, R. Perdisci, D. Dagon, W. Lee, and N. Feamster. Building a Dynamic Reputation System for DNS. In *Proc. 19th USENIX Security Symposium*, Washington, DC, Aug. 2010.

[3] M. Antonakakis, R. Perdisci, W. Lee, N. V. II, and D. Dagon. Detecting Malware Domains at the Upper DNS Hierarchy. In *Proc. 20th USENIX Security Symposium*, San Francisco, CA, Aug. 2011.

[4] L. Bilge, E. Kirda, C. Kruegel, and M. Balduzzi. EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis. In *Proc. 18th Annual Network and Distributed System Security Symposium (NDSS)*, San Diego, CA, Feb. 2011.

[5] A. Broido, E. Nemeth, and K. Claffy. Spectroscopy of DNS Update Traffic. *ACM SIGMETRICS Performance Evaluation Review*, 31(1):321, June 2003.

[6] N. Brownlee, K. Claffy, and E. Nemeth. DNS Measurements at a Root Server. In *Proc. IEEE Conference on Global Communications (GlobeCom)*, San Antonio, TX, Nov. 2001.

[7] S. Castro, D. Wessels, M. Fomenkov, and K. Claffy. A Day at the Root of the Internet. *ACM SIGCOMM Computer Communication Review*, 38(5):41–46, Oct. 2008.

[8] P. Danzig, K. Obraczka, and A. Kumar. An Analysis of Wide-Area Name Server Traffic: A Study of the Internet Domain Name System. *ACM SIGCOMM Computer Communication Review*, 22(4):292, Oct. 1992.

[9] M. Felegyhazi, C. Kreibich, and V. Paxson. On the Potential of Proactive Domain Blacklisting. In *Proc. 3rd USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*, San Jose, CA, Apr. 2010.

[10] T. Holz, C. Gorecki, K. Rieck, and F. C. Freiling. Measuring and Detecting Fast-Flux Service Networks. In *Proc. 16th Annual Network and Distributed System Security Symposium (NDSS)*, San Diego, CA, Feb. 2008.

[11] J. Jung, E. Sit, H. Balakrishnan, and R. Morris. DNS Performance and the Effectiveness of Caching. In *Proc. ACM SIGCOMM Internet Measurement Workshop*, San Fransisco, CA, Nov. 2001.

[12] M. Konte, N. Feamster, and J. Jung. Dynamics of Online Scam Hosting Infrastructure. In *Proc. Passive and Active Measurement (PAM)*, Seoul, South Korea, Apr. 2009.

[13] C. Kreibich, C. Kanich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage. Spamcraft: An Inside Look At Spam Campaign Orchestration. In *Proc. 2nd USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*, Boston, MA, Apr. 2009.

[14] PlanetLab. `http://www.planet-lab.org/`.

[15] Spamhaus. `http://www.spamhaus.org/`.

[16] Domain Name Industry Brief. `http://www.verisigninc.com/DNIB`, 2011.

[17] J. M. Spring, L. B. Metcalf, and E. Stoner. Correlating Domain Registrations and DNS First Activity in General and for Malware. In *Proc. Securing and Trusting Internet Names (SATIN)*, Teddington, United Kingdom, Apr. 2011.

[18] J. Zupan. *Clustering of Large Data Sets*. John Wiley and Sons, Ltd., 1982.

Summary Review Documentation for

"Monitoring the Initial DNS Behavior of Malicious Domains"

Authors: S. Hao, N. Feamster, R. Pandrangi

Reviewer #1

Strengths: Unique dataset coupled with thorough analysis. The conclusions about tainted address blocks is interesting.

Weaknesses: Paper does not develop/evaluate a (new) technique for predicting whether a new domain is spam or not, but to be fair, that is probably out of scope for a short paper.

Comments to Authors: This is well-written paper based on a relatively unique data set. There are no real "surprises", but the data is analyzed and presented well. It is not clear how useful the observations about spam domains are: for instance, knowing that most spam domains are not immediately used (Figure 1), and that the distribution for MX/A/NS records of spam versus legitimate domains is different (Figure 2) is interesting but does not lead to a "test" for whether a single new domain is legitimate or not. Perhaps there is a method or perhaps a legitimate-use probability can be associated with a domain based on its initial behavior, but such a technique was not developed or evaluated in this paper.

In this short paper, we investigated the potential for using DNS lookup patterns for detection, since the paper is mainly to measure different features and leaves the detection system as future work. To address this comment, we added a paragraph in the conclusion section to discuss future work on detecting malicious versus legitimate domains.

In the absence of such a method, the paper does a competent job of presenting the data cleanly and drawing initial conclusions.

Reviewer #2

Strengths: The topic of the paper is still a cool topic, and I like the combination of datasets and features the authors study.

Weaknesses: The piece is quite rushed and in parts unpleasant to read. The authors miss relevant related work, and each of the three features they use has been touched in previous work.

Comments to Authors: The first finding is not novel. Check the LEET 2009 Spamcraft paper for a plot documenting delays from registration time to time of use.

We have added this paper to the references, although our findings and conclusions are different from those from the Spamcraft paper. The previous work sees similar delays between registration and first lookup for a single campaign, but we focus on the time delay between registration and attack for .com and .net domains to investigate the possibility of early detection before the attack occurs.

Your second finding isn't particularly novel either. The proactive domain blacklisting paper from LEET 2010 is clearly related, as its authors tried to predict malicious domain use from the patterns of registration and the infrastructure touched by a domain's hosting. I'm surprised you don't cite it. You should also cite Spamhaus' DBL, as it too is a predictive spam blacklist.

We have added the paper "On the potential of proactive domain blacklisting" from LEET 2010 to the references. We have already referred to Spamhaus in Section 3, so there is no need to refer to it again.

The third finding is interesting but it too has been covered in previous work, see Spring et al.'s paper at the SATIN workshop.

We have added the paper "Correlating Domain Registrations and DNS First Activity in General and for Malware" (Securing and Trusting Internet Names 2011) to the references. This paper reports some findings that are similar to ours. They examined the delay between registration of a malware domain and the first successfully resolved response in DNS traffic.

Given that all three features are promising I would suggest you actually try to build a predictor for malicious domains so you can report on its accuracy - it would be interesting (particularly its real-time aspects), and I am pretty sure it would work quite well.

Specific comments:

- Sec. 2.1: "When a DNS domain is registered, several basic entries are inserted to refer to the services for the domain." Please make the writing more active -- registered by whom, inserted where?

We changed this sentence to "When an entity registers a DNS domain, domain name registries insert several basic entries into the zone files to refer to the services for the domain."

- Sec 3: again passive voice: "the source IP addresses were aggregated into /24 subnets and the TLD name servers recorded the number of queries for each domain". Is this something you did, or do Verisign's systems provide that?

Verisign provided the information. We changed this sentence to "Verisign's systems aggregated the source IP addresses into /24 subnets for logging and the TLD name servers recorded querying subnets each day."

- How do the query record entries scale? It seems that the number of /24s must be huge for popular domains...

Since we recorded the querying /24s (not the individual queries), the data for each domain has an upper limit (at most the number of all /24s). Our work focuses on the newly registered domains, so fewer domains would get huge queries in a short time.

- Did you build your own PlanetLab experiment or did you use CoDNS?

We designed and built our own experimental platform. We must collect different types of records, including A, NS, and MX records, so we have built and deployed our own system to perform the tasks of resolving the domain names.

- Why would spammers care about their domains' MX records?

We did not make any changes to the paper, but we will briefly clarify this question here. MX records could exhibit certain patterns:

- *When considering email spam, attackers don't hope if someone replies the email, the message gets bounced back. The MX-type could be setup anyway.*
- *If spammers don't care about the MX records, that could be a pattern itself, since the legitimate domains do care to configure correct MX records.*

- I can't make out in Sec 4.2 whether the NS record alone would suffice in detecting malice-related patterns. Per the above papers and my reading of Table 2(b), I am virtually certain they do.) If so, you do npt need to conduct A-record lookups as the NSs are in the zone file, which obviously simplifies the process.

We did not make any changes to the paper, but there are two reasons why we probed for NS-type and A-type records:

- *The NS records in domains' authority name servers might be different from glue NS records in TLD name servers.*
- *The overhead to fetch other types of records (like A-type or MX-type) is marginal. Since we didn't actually evaluate the detection performance (this is a measurement paper), it is still not clear which feature could work well. We hope to show all analysis results.*

Reviewer #3

Strengths: Interesting measurements and conclusions. The proposal to use these conclusions to detect malicious domains early seems promising.

Weaknesses: I do not think that the conclusions are new. The proposal (of how to use these conclusions) is not clear.

Comments to Authors: My main concern is that, despite being useful and interesting, the conclusions are not new. For instance, it is known that machines that participate in malicious activities tend to be clustered ("Observed structure of addresses in IP traffic", IEEE/ACM Trans. on Networking, Dec. 2006).

We did not make any changes to the paper, since the paper "Observed structure of addresses in IP traffic" is not actually that related to our work:
1) The paper did not focus on DNS analysis
2)The traffic monitoring did not occur anywhere near the root DNS resolvers.

Section 5 is confusing to the non-expert, in the following way. The paper is supposed to be looking at the characteristics of malicious domains *before* they are involved in malicious activity. But then, what does "early lookup behavior" refer to? Are malicious domains looked up before they are involved in malicious activity, i.e., before they appear in spam messages? This is what I thought at first, but then the following comment confused me: In Section 5.1, it says that the reason why malicious domains are looked up by the same resolvers is that a user clicking on one spam URL is likely to click on another and/or malicious domains may be participating in the same spamming campaigns. This comment implies that we are talking about lookups resulting from spam, i.e., a point in time after the malicious domains have started demonstrating malicious behavior. Hence, it is not clear how this lookup-clustering characteristic could be used to detect malicious domains *early*. I recommend that the authors clarify this issue.

We have added more explanation in Section 5: "Queries to a malicious domain may signal the onset of attack, and the abnormal pattern in the global DNS traffic could help to detect the attack campaign in its infancy."

Section 4 is clearer in this respect, i.e., it says that malicious domains tend to be served by infrastructure that uses specific address blocks and few "tainted" ASes, which are rarely used by legitimate domains -- implying that an early detection system could investigate domains that are served by these few tainted ASes. This makes sense. However, the numbers in Table 3 (on which this conclusion is based) are confusing: Apparently, the AS mostly used by malicious domains is the Chinanet Backbone AS. My understanding is that this AS hosts a large fraction of the Chinese Internet infrastructure. Hence, if we consider this AS "tainted", a large fraction of legitimate Chinese domains will have to be investigated as suspect of malicious behavior. This does not sound right.

We didn't try to build a detection system and evaluate its performance in this paper. Regarding the "Chinanet backbone" case, if an AS has a large mixture of good and bad domains, that AS should be handled as a special case, since obviously blocking the entire AS is not feasible. In such cases, using other auxiliary patterns to distinguish malicious from legitimate domains (e.g., lookup patterns) may be necessary. Even if a detection method simply suspected all domains involved in "Chinanet backbone", based on our findings in Table 2(a), it saves effort to investigate 10% of legitimate domains compared to all domains.

Reviewer #4

Strengths: - Problem is important and the paper presents some interesting new insights about the DNS dynamics of malware sites.
- Unique datasets from the TLD servers.

Weaknesses: - Not sure how unbiased the data set is.

Comments to Authors: I think the problem is important and the authors do a good job in the paper showing evidence that often there is not much lag between a domain getting registered and beginning to send spam. I like the paper overall.

I have only some minor comments.

The authors mention the use of Yahoo mail servers for collecting timestamps. It might be good to elaborate what exact data they had for this purpose.

We added the following clarification: "Yahoo! Inc. provides the received time of the email messages and the URLs contained in the messages."

Finding 4.1 is not intuitive. I think you may want to say the contrapositive, that 45% of malicious domains appeared within a day, if you want this finding to be reflective of the early detectability. The sentences following this finding also need to be rephrased properly; right now they are confusing.

We made no changes to the paper. We stated that "55% of the malicious domains appeared in spam campaigns more than one day after they were registered", because the intention is to check what is the buffer time people have to detect a bad domain before it is used in malicious activity.

I would use benign as an opposite of malicious. Legitimate is not accurate since all domains are legitimate since they have been registered with the registrar (and not hijacked, which is yet another study).

We do not agree with this suggestion and have not made the change in the paper. It is much clearer to use "legitimate", since this corresponds with standard parlance in the literature (cf. "legitimate" email vs. spam).

Isn't the non-uniform distribution of IP addresses across the NS, MX and A records an artifact of the data you have. Presumably if

you have only a partial dataset, you would potentially see only some attack IPs.

We arguably have the most representative dataset possible; it is one of the most expansive datasets that it is possible to obtain: .com (and .net) is a good representative of TLD, and the data have big coverage already. It is true the particular IP blocks will be different for domains under different TLDs, but they will have similar patterns as we see in .com (and .net).
We added some statistics in Section 3, e.g. ".com and .net accounts for over 45% registered domain names on the Internet".

I think 2(a) is not in line with the finding 4.3 and it is not clearly explained. I see very little difference between the benign and malicious domains. MX/A records, I can see some difference, but not so much in 2(a).

The MX and A records of malicious domains are still relatively concentrated in one or two particular ASes (about 30% of the records reside in these ASes). Although the reviewer is correct that some ASes appear in both sets, we are not suggesting that this feature alone would be sufficient to detect bad domains. In some cases, ASes predominantly appear in one list or the other, and this information will be useful for establishing a prior.

I am not sure if the query data collected using recursive queries at the TLD server is unbiased between malicious and benign domains.

This comment does not make sense. We have already emphasized in Section 3: "Verisign deploys multiple TLD name servers to resolve second-level domain names, and we collected the logs of querying /24s from all servers for analysis."

Reviewer #5

Strengths: Interesting insights from a great data set (top level domain DNS request volume).

Weaknesses: - Interesting observation, but can this be turned into some practical policies to blacklist malicious domains?
- Data set only available to the owner of top level domain managers-
- Observation only based on .com and .net. Is this representative?

Comments to Authors: I enjoyed reading the paper. Let me focus on some of the weaknesses:

- The main weakness is that it is not clear how to turn your findings into some practical policies to blacklist domains? You clearly show differences between legitimate and malicious domains, but can you use this information in practice? Would it be easy for domains to avoid the new blacklisting rules identified? I understand this is a short paper and I am hoping that you will cover this in future work.

- Your work is based on .com and .net data. How representative is this? Could you comment on some statistics to say X% of spams or blacklisted domains are .com or .net? Is it 90% or 2%?

Based on analysis of our spam trap data, we see that around 40% of unique second-level domains reside under .com and .net. These figures are derived from all second-level domains in the spamtrap email.

- Your analysis of /24 reminded me of a work that you did not reference: S. Venkataram, A. Blum, D. Song. S. Sen, and O. Spatscheck,"Tracking dynamic sources of malicious activity at Internet-scale", in Proc. NIPS, 2009.

This paper might be a useful clustering method if we were designing a detection scheme, but since our work is not mainly to develop a new clustering algorithm, we do not think this paper is particularly related to our work.

- Regarding Figure 1, do you have finer grain data to show what happens during the first couple of hours for the other 45% of the domains?

Because the active resource-record probing and the lookup data are aggregated on a daily basis, analyzing the data on a more fine-grained basis is not possible.

- Could you add a candle chart on Figure 4 to show the distribution of legitimate traffic to see how practical a policy based on this metric would work? e.g. do a lot of legitimate domain exhibit that same pattern as the malicious domains? You just show the differences between the means.

We included standard error in the figure for both malicious and legitimate domains. Because the value for legitimate domains is comparatively small, these values might be difficult to see in the figure.

Response from the Authors

The review comments fall into three categories.

1) *Some related work is not included, and the differences from our paper are not described carefully*. To correct this concern, we added most of the related papers pointed out by the reviewers, including "On the potential of proactive domain blacklisting" from LEET 2010 and "Correlating Domain Registrations and DNS First Activity in General and for Malware" from the SATIN 2011 workshop. There are three major differences of our work from earlier studies: we focus on newly registered second-level domains, the DNS records are actively probed instead of from passively monitoring, and we observe the lookup pattern at the TLD name servers (with a global view from *all* /24 networks across the Internet). We did not include several of the papers that reviewers mentioned that do not appear to be related to our paper.

2) *It is unclear whether the DNS patterns could be eventually developed into a detection system*. Since the intention of our work is to reveal DNS characteristics that could shed light on identifying different patterns for malicious and legitimate domains, the paper does not actually design a detection system based on the features (which is our next step). We changed text to make the claim clear, e.g. adding a paragraph in the conclusion section to state detection algorithms as future work.

3) *Some details of the data and the findings need more clarification*. To present the Verisign data more clearly, we checked the representativeness of the .com and .net domains and added more details to the paper about the collection process. We also added extra explanation explaining how we correlate second-level domains appearing in spamtrap URLs with those that appear in a Yahoo! email trace.