Review on Passive DNS Analysis for Finding Malicious Domain

Ku. Vijayalaxmi Janardan Tayade
Student
Sanmati Engineering College Washim

ABSTRACT
The domain name service (DNS) plays an important role in the operation of the Internet, providing a two-way mapping between domain names and their numerical identifiers. Given its fundamental role, it is not surprising that a wide variety of malicious activities involve the domain name service in one way or another. For example, bots resolve DNS names to locate their command and control servers, and spam mails contain URLs that link to domains that resolve to scam servers. Thus, it seems beneficial to monitor the use of the DNS system for signs that indicate that a certain name is used as part of a malicious operation. In this paper, we introduce EXPOSURE, a system that employs large-scale, passive DNS analysis techniques to detect domains that are involved in malicious activity. We use 15 features that we extract from the DNS traffic that allow us to characterize different properties of DNS names and the ways that they are queried. Our experiments with a large, real-world data set consisting of 100 billion DNS requests, and a real-life deployment for two weeks in an ISP show that our approach is scalable and that we are able to automatically identify unknown malicious domains that are misused in a variety of malicious activity (such as for botnet command and control, spamming, and phishing).

Keywords
DNS, Domain Registration, Spam, Malicious Domain.

1. INTRODUCTION
The Domain Name System (DNS), the Internet's lookup administration for mapping names to IP locations, gives a discriminating support of Internet applications; tragically, it additionally permits assailants to direct casualties to Web locales that host tricks, malware, and different pernicious substance. To alleviate these dangers, system administrators attempt to determine a notoriety for every space that mirrors the probability that the area is connected with a specific sort of assault (e.g., trick, phishing, malware facilitating). The rate at which new areas show up makes rapidly building up a notoriety for these spaces especially difficult: in our investigation, we observe that over countless new areas are enrolled each day. Existing DNS notoriety frameworks utilize the attributes of DNS lookups from resolvers that gaze upward an area to recognize genuine from noxious spaces [1, 2]. Shockingly, these frameworks must watch a critical volume of DNS lookups before deciding the notoriety for a space, which just happens after trade off has occurred. Towards encouraging pre-assault identification of malevolent spaces, we think about the beginning DNS movement for every area and portray how the detectable conduct for a vindictive area contrasts from that of real areas. We think about two parts of beginning DNS conduct connected with areas: (1) the DNS base used to determine the spaces to IP locations; and (2) the DNS lookup designs from the systems that perform starting lookups to the area. Certain qualities of the DNS foundation may be one of a kind to pernicious spaces, for example, the IP location extents and ASes that host either the legitimate name servers for the locales, or the destinations themselves. Distinguishing foundation that is regular crosswise over malevolent spaces may give clues to recognizing pernicious areas before the assaults themselves are mounted. Attributes of right on time DNS lookups can help system administrators find important data about the areas' way that are being turned upward. Strikingly, we find that spaces that are enrolled for noxious reasons for existing are at first questioned from an a great deal more differing arrangement of subnets than real areas. Our investigation of DNS conduct right on time in an area's life cycle is spurred by our definitive craving to perform early discovery of malevolent spaces. We utilize spaces gathered at a few substantial spam traps as a wellspring of areas connected with spam battles. To portray the asset record conduct of every space, we perform intermittent iterative inquiries of recently enrolled areas in March 2011. To describe DNS lookup designs crosswise over systems, we utilize data about DNS lookups gathered from the Verisign top-level space servers, combined with enlistment data about these areas. We concentrate only on the early DNS conduct of a space, which is empowered by two essential bits of data. To start with, enlistment records alarm us when an area is enrolled, and permit us to start questioning it promptly, before assaults. Second, here examine a worldwide perspective of ahead of schedule DNS lookup designs over the whole Internet for .com and .net spaces. Our study uncovers the accompanying discoveries:

- Domain enrollment and asset record foundation happens before assaults occur. Upwards of 55% of spam crusades may happen no less than one day after the space referenced in the spam messages were enrolled, offering the potential for right on time disclosure of malignant areas taking into account beginning DNS conduct.
- DNS base for malevolent areas is situated in diverse location space districts and independent frameworks than the foundation for true blue spaces. A couple of self-ruling frameworks and IP address locales host foundation just for areas that are connected with pernicious movement. Recognizing these at space enlistment time can conceivably empower early discovery.
- Early lookup designs for a recently enrolled pernicious spaces vary essentially from the examples for a true blue area. Spaces connected with spam crusades are at first gazed upward by a more differing arrangement of system location locales than honest to goodness areas. Particularly, the recently enlisted spam spaces get to be "mainstream" all the more rapidly. These

components might eventually be utilized to create interesting fingerprints for recognizing real spaces from those that are connected with Internet assaults. Whatever remains of this paper is sorted out as takes after. Area 2 studies the issue connection and related work. Segment 3 portrays the information sets that we use for our examination. Area 4 contemplates the attributes of asset records for recently enlisted authentic and vindictive areas. Area 5 thinks about the lookup attributes for changed sorts of spaces, and Section 6 finish

2. RELATED WORK

Observing and examination on zones' asset records Previous studies have utilized the component of questioning the DNS servers to check the zones' asset records. Holz et al. researched the A's differences records returned in the lookups to distinguish fast flux administration systems [10]. Konteetal. concentrated on the changing rates of the IP address in the DNS records of trick areas [12]. Our work, then again, tracks DNS records of recently enrolled areas to surmise spatial and worldly attributes. Anax checked the recursive servers to figure out peculiarity in the reserved records and identify harming assaults [1]. Conversely, we screen the records in the zones' definitive servers to find the attributes in the malevolent spaces' enlistment. DNS lookup designs The first investigations of DNS lookup conduct at a neighborhood resolver were performed by Danzig et al. [8] and Jung et al. [11]; both of these studies analyzed lookup conduct from the vantage purpose of lookups to a solitary neighborhood resolver, and did not endeavor to portray what these lookup like examples contrasted for noxious spaces. Notos [2] and EXPOSURE [4] contemplated DNS lookup conduct inside of a neighborhood space beneath the DNS resolvers to manufacture the areas' notoriety. Such a perspective of DNS lookup conduct is significant, however this vantage point can't uncover facilitated conduct over different systems, and it depends first on an assault to occur or has being traded off before it can recognize any vindictive areas. Antonakakis et al. likewise observed the DNS movement from legitimate servers or top-level space servers to recognize malware areas [3], however the zones' elements were still recreated from DNS solicitation and reaction messages; they didn't concentrate on the behavioral examination of recently enlisted areas. Other work has analyzed DNS lookup conduct at a DNS root server [5–7]. The center of these studies was not quite the same as this paper. Castro et al. [7] and Brownlee et al. [6] endeavored to portray the amount DNS activity at the DNS root server was illegitimate. Broad et al. recognized misconfigured hosts utilizing spectro graphy to distinguish machines that were erroneously issuing consequently arranged DNS inquiries [5]. Conversely, we concentrate on DNS lookup designs from the point of view of a top-level space, and inspects the conduct of lookups as seen from recursive resolvers, instead of lookups from individual hosts. Area enlistment deduction Recently, various examination endeavors have concentrated on space enrollment designs. Kreibich et al. [13] examined the time from an area's enrollment to its utilization in spam. Spring et al. analyzed the deferral between enrollment of a malware area and the first effectively determined reaction in DNS activity [17]. We mention a comparative objective fact, yet on a much bigger arrangement of areas under .com and .net; further, we investigate the DNS qualities in spaces' initial life cycle. Felegyhazi et al. [9] proposed to consequently recognize pernicious spaces in view of WHOIS and name server data. Interestingly, we effectively gather distinctive sorts of asset records to track the progressions, and screen the systems questioning the spaces. An attributes' percentage that we see about malignant spaces could be utilized to construct productive element notoriety framework

Table 1: Data format examples.

Type	Example
DNZA entry	Add-new example.com NS ns1.example.com
Query record	Example.com 111.111.111.0,22.22.22.0

3. REGISTRATION & RESOURCE RECORDS

We first check the time between the enrollment of a space and the consequent assault to examine the potential for ahead of schedule recognition. At that point, we investigate how DNS conduct connected with framework—where a space's resolvers at first live—can be an early flag for malignant ar

3.1 Timebetween Registration And Attack

We theorize that there may be some time between when spammers enlist new spaces and when they send spam. We inspect the delay's degree between the time when an area is at first enlisted and when it is eventually utilized as a part of an assault. On the off chance that such a deferral exists, it may permit boycott administrators to list the vindictive areas, potentially before the spam crusade happens. What amount of time happens between the area enlistment and assault? Figure 1 demonstrates the conveyance between the time when we begin to watch records about the pernicious spaces enlisted in March 2011, and the soonest time when the areas showed up in the spam messages. We take the timestamps in our spam traps, and additionally messages got at the Yahoo! mail servers. Yippee! Inc. gives the got time of all email messages and the URLs contained in the messages. The Yahoo! email gives a more extensive scope of checking far and wide. Here take the most punctual time focuses when seeing an "awful" space in email messages (either in Yahoo! information or in spam trap) as the assessed begin of the spamming assault about that space. The x-hub speaks to the deferral between when a space was enlisted and when we initially seen the area connected with a spam battle, and the y-pivot is the rate of the noxious areas enrolled in March 2011. Discovering 4.1 (Delay until assault) More than 55% of the malevolent spaces showed up in spam crusades over one day after they were enrolled. Hence characterize the initial five days after area enrollment as pre-assault period. Around 20% of spaces won't not be utilized as a part of assaults amid this period, and the time windows for different areas being investigated in spamming are likewise constrained. In whatever is left of the paper, we will dissect the attributes of DNS base for malevolent areas both all through their lifetime (i.e., after the spaces' enlistment) and inside of the pre-assault period. In Section 5, here further examine the lookup conduct amid the early stage.

3.2 Location Of Dns Infrastructure

At the point when deciding the IP address that maps to each DNS record, here find that the doled out records for spam areas crosswise over IP space are a long way from uniform. How is the DNS framework that has an area at first circulated crosswise over IP location space? The starting conveyance of area records crosswise over IP location space may give pieces of information as to a space's notoriety. Figure 2 indicates how the IPs connected with NS, MX, and A records from malevolent and honest to goodness areas are conveyed crosswise over IP location space. The x-pivot speaks to IPv4 space. In the event that an IP maps to numerous records from

diverse areas, we include it just once the figure. The y-hub shows the rate of addresses not exactly or equivalent to the IP esteem on the x-pivot. The strong blue bends plot the circulation of genuine example areas, the red dashed bends demonstrate the result of malignant spaces, and the green dash-speck bends speak to watched records for the pernicious areas amid pre-assault period. Interestingly, in this watch that the DNS records connected with vindictive areas are appropriated uniquely in contrast to the records connected with honest to goodness ones.Discovering 4.2 (Distribution crosswise over IP location space) The IP locations utilized by vindictive areas as a part of the NS, MX and A records are dispersed thickly in a little portion of IP location space. The IP locations connected with DNS asset records are not conveyed equally over the IP location space. Some system extent has more IPs pointed from NS, MX or A records; while the record IPs in other part of location space are dispersed scantily. Especially two system squares conveyed records from malignant spaces, 96.45.0.0/16 and 216.162.0.0/16. The prefix 173.213.0.0/16 has numerous IPs in spamming spaces, yet the same reach hosts authentic areas, as well. This perception demonstrates that if IPs comparing to distinctive areas' records dwell near one another in a system square, those spaces may show up in spam later on. Is the DNS foundation for pernicious areas situated specifically ASes? Obviously, the IP locations of the records are definitely not.

Table 2: Main three ASes containing areas' records. adequate to affirm that an area is trick related.

(a)**Legitimate domains**

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

(a) Malicious domains

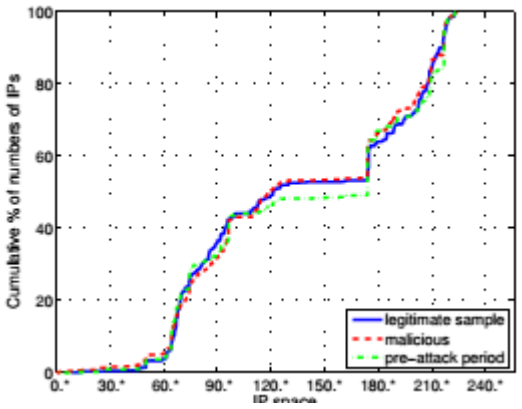
Type	AS	Domain ratio	AS Name	Country
NS	4134 28753 31365	33.6% 17.0% 16.3%	Chinanet Backbone Leaseweb De SGSTelekom	China Germany Turkey
MX	197088 3292 5632	23.9% 19.3% 12.3%	Colorhost LLC TDC Data Networks 3dgwebhosting.com Inc	Latvija U.S. U.S.
A	4134 197088 30890	19.3% 14.3% 13.3%	Chinanet Backbone Colorhost LLC Evolva Telecom	China Latvia Romania

(b) Malicious domains in pre-attack period

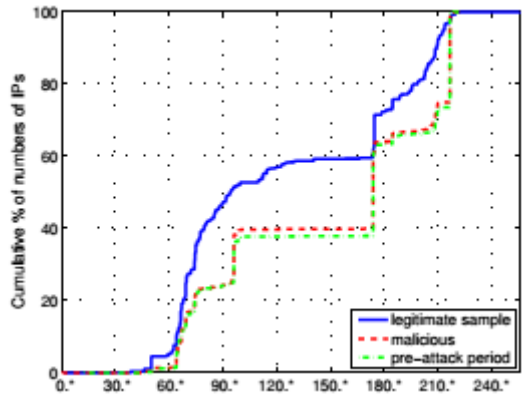
Type	AS	Domain ratio	AS Name	Country
NS	4134 28753 27699	37.8% 20.4% 11.3%	Chinanet Backbone Leaseweb De Tel. De Sao Paulo S.A.	China Germany Brazil
MX	197088 3292 5632	14.3% 21.8% 12.3%	Colorhost LLC TDC Data Networks 3dgwebhosting.com Inc	Latvija U.S. U.S.
A	4134 197088 28753	19.7% 15.0% 11.6%	Chinanet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

Now inspected the asset's circulation records crosswise over ASes and analyzed the conveyance of genuine and malignant areas. Table 2 demonstrates the main three ASes positioned by the rate of areas regularly having records being determined into ASes. Discovering 4.3 (Distribution crosswise over ASes) More than 30% of the malevolent areas have no less than one record setting out to maybe a couple specific ASes, which are unique in relation to those ASes for the most part utilized by honest to goodness spaces. We watch that a significant number of the new real areas have bigger enlistment centers like GoDaddy work their DNS, and host their administration foundation with understood supplier, as Google. Then again, spamming areas' records are scattered over numerous ASes and nations. Spammers seem to incline toward certain particular ASes to have their DNS base. Are there "terrible" ASes that host DNS base solely for pernicious areas? We characterize an as "spoiled" once the quantity of noxious areas whose DNS records are determined inside of the AS surpasses a limit. The arrangement of polluted ASes speak to the systems that assailants most vigorously use, as showed by the noxious spaces' enrollment. After an area's enrollment, assailants make DNS sections for the space, and the records resolve to distinctive IP addresses. We then check whether the subsequent IPs have a place with the corrupted ASes. In the event that a space collects numerous records that set out to spoiled ASes, we think that the area is identified with the watched assaults. Discovering 4.4 (Domains facilitated by "awful" ASes) Most real spaces have A, MX, and NS records that are facilitated totally in untainted ASes. Then again, the lion's share of spam areas have records facilitated in spoiled ASes, notwithstanding amid the pre-assault period.

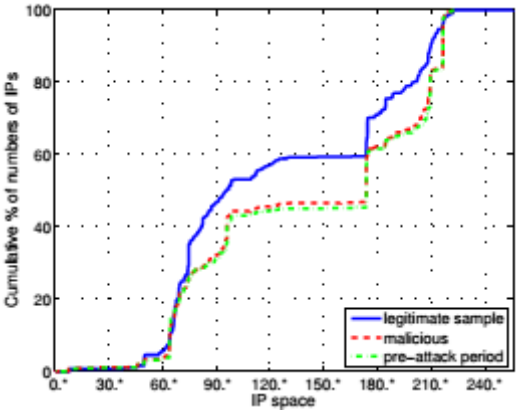
We determine the polluted AS set by including an AS that has facilitated records for more than 100 spam spaces. Figure 3 demonstrates the proportion of the polluted record number to the quantity of all records for the area. More than 90% of authentic new areas have zero records having a place with the polluted AS set.



(a) NS records



(b) MX records



(c) A records

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

Table 3: Five largest clusters based on lookup networks.

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%

4. LIMITATIONS

A decided assailant who knows howEXPOSURE works and who is educated about the elements we are searching for in DNS movement may attempt to dodge discovery. To dodge EXPOSURE, theattackers could attempt to maintain a strategic distance from the particular elements and conduct that we are searching for in DNS activity. For instance, an assailant could choose to relegate uniform TTL values over all traded off machines. Notwithstanding, this would imply that the aggressors would not have the capacity to recognize more solid, and less dependable has any longer and would take an unwavering quality hit on their vindictive bases. As another illustration, the assailants could attempt to lessen the quantity of DNS lookups for a malignant space so that just a solitary lookup is performed each hour (i.e., so that the vindictive area is boycotted). Be that as it may, this is not trifling to execute, decreases the assault's effect, and obliges a high level of coordination on the aggressor's side. Despite the fact that it is feasible for an aggressor to stay underneath our location radar by maintaining a strategic distance from the utilization of these elements, we trust that this accompanies an expense for the assailant. Thus, our frameworks helps expand the trouble bar for the aggressors, compels them to surrender the utilization of elements that are valuable for them by and by, and makes it more mind boggling for them to deal with their foundations. Obviously, our identification rate likewise relies on upon the preparation set. Here don't prepare for the group of malignant spaces that constitute assaults that are reasonably obscure and have not been experienced before in the wild by malware analyzers, apparatuses, or specialists. In any case, the more malignant spaces are encouraged to the framework, the more exhaustive our methodology gets to be after some time. Note that if the systems that we are observing and preparing our framework on are not contaminated, clearly, we won't see any malevolent spaces. We trust that we can enhance our capacity to see more noxious assaults by having admittance to bigger systems and having more establishments of EXPOSURE.

5. CONCLUSION AND FUTURE WORK

We have observed DNS asset records for second-level spaces recently enrolled in March 2011 and analyzed the lookup movement to substantial definitive top-level area servers. We demonstrate the DNS attributes saw at TLD name servers and removed from zones' asset records for vindictive spaces are not the same as those for genuine areas. Asset records of malevolent spaces tend to set out to particular IP location reach and ASes. When we recognize an arrangement of "corrupted" independent frameworks that host numerous trick areas, the authentic spaces once in a while have asset records inside of the polluted AS set. We likewise find that scalawag areas display unmistakable bunches, as far as the systems that turn upward these spaces. At last, we find that these areas turn out to be generally prominent extensively all the more rapidly after their beginning enlistment time. The unmistakable DNS qualities and their inclination on distinctive sorts of spaces propose that it might eventually be conceivable to unique finger impression areas taking into account their asset records and lookup movement near TLD name servers before an assault ever happens. Despite the fact that a solitary example in DNS may have constrained energy to distinguish pernicious areas, the mix of our discoveries might eventually control the configuration of future "early cautioning" frameworks for DNS.

6. 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] M. Antonakakis, R. Perdisci, D. Dagon, W. Lee, and N. Feamster. Building a Dynamic Reputation System for DNS. In19th Usenix Security Symposium, 2010.

[7] Michle Basseville and Igor V. Nikiforov. Detection of Abrupt Changes - Theory and Application. PrenticeHall, 1993.

[8] Ulrich Bayer, Christopher Kruegel, and Engin Kirda. TTAnalyze: A Tool for Analyzing Malware. In15th EICAR Conference, Hamburg, Germany, 2006.

[9] Pavel Berkhin. Survey of clustering data mining techniques. Technical report, 2002.

[10] A. P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. In Pattern Recognition, volume 30, pages 1145–1159, 1997.