SIDN Labs

https://sidnlabs.nl

February 13, 2017

Peer-reviewed Publication

Title: Domain names abuse and TLDs: from monetization towards mitigation

Authors: Giovane C. M. Moura, Moritz Müller, Marco Davids, Maarten Wullink, and Cristian Hesselman

Venue: 3rd IEEE/IFIP Workshop on Security for Emerging Distributed Network Technologies (DISSECT 2017), co-located with IFIP/IEEE International Symposium on Integrated Network Management (IM 2017), Lisbon, Portugal.

Conference dates: May 8th to 12th, 2017.

Citation:

- Moura, G.C. M., Müller, M., Davids, M., Wullink, M, Hesselman, C.: Domain names abuse and TLDs: from monetization towards mitigation. 3rd IEEE/IFIP Workshop on Security for Emerging Distributed Network Technologies (DISSECT 2017), co-located with IFIP/IEEE International Symposium on Integrated Network Management (IM 2017). Lisbon, Portugal, May 2017 (to appear)
- Bibtex:

```
@inproceedings{sidn-dissect-2017,
author = {{Giovane C. M. Moura,
Moritz Muller, Marco Davids, Maarten Wullink, and
Cristian Hesselman}},
booktitle={{ 3rd IEEE/IFIP Workshop on Security for
Emerging  Distributed Network Technologies (DISSECT 2017),
co-located with IFIP/IEEE International
Symposium on Integrated Network Management (IM 2017)}},
title={Domain names abuse and TLDs: from monetization
towards mitigation}},
year={2017},
month={May},
}
```

Domain names abuse and TLDs: from monetization towards mitigation

Giovane C. M. Moura, Moritz Müller, Marco Davids, Maarten Wullink, and Cristian Hesselman
SIDN Labs
Stichting Internet Domeinregistratie Nederland (SIDN)
Arnhem, The Netherlands
Email: {firstname.lastname}@sidn.nl

***Abstract*—Hidden behind domain names, there are lucrative (and ingenious) business models that misuse/abuse the DNS namespace and employ a diversified form of monetization. To curb some of those abuses, many research works have been proposed. However, while having a clear contribution and advancing the state-of-the-art, these works are constrained by their limited datasets and none of them present a survey on the forms of DNS abuse. In this paper, we address these limitations by presenting a case study in one top-level domain (TLD) operator (.nl) with diverse longitudinal datasets. We then cover eight business models that DNS abusers employ and their respective monetization form, and discuss how TLD operators can employ these datasets to detect these forms of abuse.**

I. Introduction

Domain names have long been misused for different types of abuse: phishing, malware distribution, spamming, and botnet command-and-control (C&C) are just some of them. Underlying each of these forms of abuse, we find profitable *business models*, which provide the *incentives* for these abusers to continue with such activities.

To curb such practices, the research community has been active in proposing various solutions, such as [1], [2], [3], [4], [5], [6]. While these works advance the state-of-the art and have a clear contribution, they are faced with two main shortcomings: (i) they are constrained by *type* and/or *duration* of their respectively available datasets (due to the difficulty in obtaining such datasets) and (ii) while these solutions cover different sorts of abuse, we lack a survey on domain-related abuses, which leaves the question of how much ground has *not* been covered yet unanswered.

This paper addresses both issues: by carrying out a case study on the top-level domain (TLD) of the Netherlands (.nl), we address the first issue by analyzing three longitudinal datasets readily available to TLD operators (§III): historical registration database (as opposed to hard-to-parse [7] and yet incomplete `whois` records), traffic to authoritative name servers (centralized view instead of DNS resolvers traffic), and the infrastructure used by the domains (obtained using DNS scans).

We address the second issue by presenting a survey on domain abuses (§IV) and discuss their underlying business models and respective monetization methods. We demonstrate how they create patterns in our datasets, and discuss how TLD operators [8] can leverage these to develop methods tailored to curb each form of abuse. The main contribution of this paper is, therefore, a survey of domain related abuses including their underlying business models, and a discussion on how TLD operators can use their datasets to mitigate them.

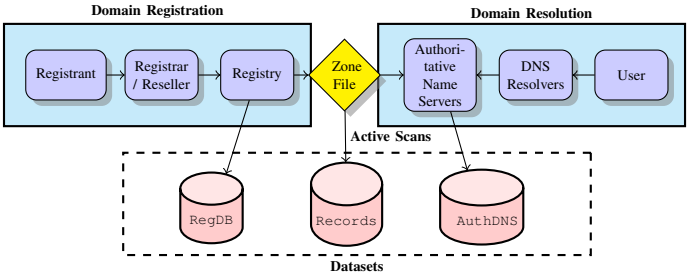

Fig. 1. TLD Operations: registration (left), domain name resolution (right), and derived datasets.

The remainder of this paper is divided as follows: we cover the two basic services provided by TLDs (domain registration and name resolution) in §II. Then, we introduce in §III the datasets we used . Following that, we present in §IV a survey on the types of DNS abuse and their underlying business models and their respective implications on the datasets. Finally, conclusions and future work are discussed in §V.

II. Background

A. Domain Registration

Domain registration consists of creating a unique domain name that ultimately is added to the zone file of a DNS zone. Typically, it involves the so-called triple-R: *registrant*, *registrar* (or reseller), and *registry* (or TLD operator). Figure 1 summarizes the process (left part). To register a domain in a specific TLD, first a registrant (a user) choses a registrar (e.g: GoDaddy) that is accredited by the TLD of his/her choice. Once the requirements of the registrar are fulfilled (personal data, payment), it contacts the registry and registers the requested domain on behalf of the user[1]. Different registrars have different registration interfaces, but the communications between the registrar and registry are typically performed using the Extensible Provisioning Protocol (EPP, RFC 5730).

[1]Some registries allow domain tasting, in which a user may try a domain for a few days for free (domain tasting), but is not the case for .nl.

Domains are registered for a certain period of time (depending on the registry), and after expiration, they can enter a Redemption Grade Period (RGP), in which the former registrant can still renew the domain. After this period has expired, the domain is deleted and other registrants can register it (this depends on the policy of the registry and .nl domains are made available after 40 days of the expiration date).

Each registry, in turn, maintains its own registration database, which then is used to generate a *Zone File* (Figure 1) that contains the list of all active and delegated domains under the respective TLD and their respective DNS records. Ultimately, the zone files are used as input files on the authoritative name servers for the particular TLD[2]. These zone files are also frequently updated, and each TLD operator chooses how often–.nl updates its zone files every hour.

B. Domain Resolution

Domain name resolution consists of resolving a domain name into, ultimately, its IP addresses or other specific types of DNS records [11]. We summarize this process on the right side of Figure 1. First a user attempts to access a web site (e.g,: example.nl). The stub DNS resolver on his/her computer sends a DNS request to its DNS resolver, typically provided by the ISP. The DNS resolver, in turn, contacts one of the Root DNS servers [12] (as provided by `root.hints` file) to obtain the authoritative name server for .nl. Then, it will send another request to one of the authoritative servers of .nl, which respond for example.nl. Caching on DNS resolvers [13] may eliminate some of these steps. Finally, the resolver responds to the user with the required DNS record.

III. Datasets and TLDs vantage point

`RegDB`: at domain registration side (Figure 1), we have access to the historical database of .nl, which contains historical information about registration and removal of domains from its respective zone files. We refer to this dataset as `RegDB`. This dataset contains complete information about registrant and registrar (and resellers, if applicable), as well as some of the DNS records of the respective domain (NS, DS, and DNS glue [11]) for a period of 20+ years. Due to privacy reasons, TLDs do not share this information. However, they make part of it available through a `whois` service, which is typically what researchers outside TLDs rely upon. However, this service has several shortcomings: incomplete data in comparison with `RegDB`, lack of historical data (only the current status is shown), lack of a standard data format[3](thus hard to parse [7]), and it is usually rate-limited (therefore hard to perform analysis on large number of domains).

`AuthDNS`: the other passive dataset is `AuthDNS`, which contains the incoming queries issued by resolvers to our .nl authoritative servers. This data provides a centralized but sampled view (due to caching on the resolvers [13]) of all queries issued to .nl. Similarly to the registration database, researchers usually do not have access to this type of data – when they have it typically covers a snapshot of it. We, on the other hand, have been continuously storing this data since May 2014. We use our open-source Hadoop-based ENTRADA [14] to store and process this dataset.

`Records`: last, .nl zone files contains information about all active domains, but not all the DNS records [11]. To obtain such information and types of records, we utilize the daily scans (Figure 1) to our zone [15].

Access to these datasets is regulated by our publicly available data privacy framework that conforms to both EU and Dutch legislation [16]. We refer the interested reader to [8] for a discussion on a security and stability role of TLDs, including privacy management.

[2]ICANN makes available both the Root DNS zone file [9] and the new generic top-level domains (gTLDs)[10]. Other TLD operators may share their zone files available upon request.

[3]RDAP (RFC7482) has been proposed and standardized to address `whois` limitations.

Business	Spam	RegDB	AuthDNS	Records	Lit
Phishing(0-day)	Yes	Weak	Strong	Weak	[3], [6]
Phishing(comp.)	Yes	None	Strong	Weak	[17]
Parking (Ads)	No	Strong	Weak	Strong	[18], [19]
Parking (Mal)	No	Strong	Weak	Strong	[18], [19]
Fake Goods	Yes	Weak	Weak	Medium	[6], [20]
Drop-Catch	No	Medium	Medium	Weak	[21]
Botnet C&C	No	Medium	Strong	?	[22]
Blackhat SEO	No	Medium	Medium	Strong	[23], [24]

TABLE I
Business Models and Datasets/signal "strength", and research works that cover those.

IV. Monetization methods and mitigation

How can one *monetize* using domain names? Answering this question allows us to understand the underlying business models employed by domain name abusers. These business models vary significantly, leaving an, often distinctive, "trace" on different types of data sets we discussed in §III, which can, in turn, be used to mitigate such abuses. Table I summarizes the relationship between commonly observed business models and the three datasets we covered in §III, and whether they use spam to advertise their domains.

There are, however, other business models that can be used to monetize on DNS – such as compromising a registrar or registrant, hijacking a domain. However, we primarily focus on abuses that we observe more often and discuss them in the next subsections.

A. Phishing (0-day)

Phishing is used to convince Internet users to "perform certain actions for the attacker's benefit" [25]. Attackers use phishing to steal banking/credit card/ID credentials, and may use them themselves or re-sell them in bulk.

There are two types of phishing, from a TLD point-of-view: *0-day* and *compromised*. 0-day, or newly registered domains [3], is a type of phishing attack in which an attacker first registers a domain (sometimes with a name that is a mimic or typo-squat of the impersonated one) for the sole purpose of the attack. Compromised phishing domains, on the other hand, are existing websites, typically running a content management system (CMS), whare are hacked and which host

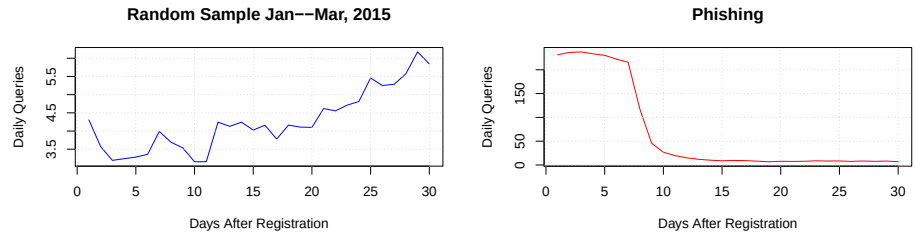

Fig. 2. .nl Random vs Phishing new domains average daily queries [6]

phishing websites. Also, phishers can leverage the previously built reputation of the compromised website.

The business model behind 0-day phishing consists in first (i) registering a domain name and setting up a malicious website, (ii) followed by a spam campaign to advertise the domains, (iii) performing ID theft (IDs, credit cards, bank credentials), and later (iv) selling this data or using it themselves. We have seen empirically that 0-day phishing domains are seen as "disposable" by attackers [6]: they remain online for a very short time (typically less than few days), since they raise flags in registrars/host providers that have an incentive to block them in order to avoid their own IP address ranges from being blacklisted, thus avoiding their other clients from also being penalized.

From the TLD datasets described in §III, we have seen that 0-day phishers tend to use the same subset of registrars, but they use fake credentials when registering the domains. To avoid detection, they use different networks/autonomous systems, so the `Records` dataset provides little evidence to detect them. The strongest signal comes from `AuthDNS`, since they employ spam to advertise their phishing sites, an abnormal number of queries is seen for the phishing web sites right after their registration [3]. In Figure 2 we show the average daily queries for newly registered domains to one authoritative server. The left side figure shows a sample of 20,000 randomly chosen domains on the .nl zone, while the right daily queries for 1,334 phishing sites as reported by Netcraft [26]. The random sample domains have less than 6 daily queries on the first days after registration, while phishing shows a large number of queries. To detect those, we have developed nDEWS [6], and classify new domains based on this pattern. Our work is based on the work of Hao *et al.* [3].

Different solutions have been proposed to address 0-day phishing [3], [6], but the key is to perform detection at *registration* time, which allows to preemptively protect users. With this in mind, Hao et. al [1] have proposed a registration time detection system for spamming domains – which are also used by phishing attackers and other attacks. Their work presents the most comprehensive evaluation of features to detect spamming domains at registration time, which can also be applied to phishing, since phishing most of the time employs spamming.

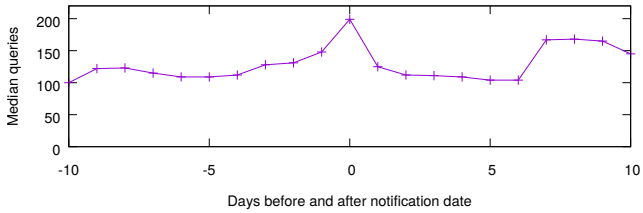

Fig. 3. Median daily queries for 1,374 compromised phishing sites on .nl

B. *Phishing (compromised websites)*

Compromised website phishing is the most common form of phishing attacks, and they rely upon the reputation of an already existing website [27]. They follow the same business model as 0-day phishing (§IV-A).

Since no domains are registered or changed, no changes are performed in `RegDB` for this attack. The same reasoning applies to `Records` dataset, since they also do not change. The only dataset that can be used in this case then is `AuthDNS`, since compromised websites exhibit an abnormal increase in traffic, as can be seen in Figure 3. We show in this figure the daily median queries for 1,334 phishing compromised domains (at least 7 days old) on .nl zone that were reported by Netcraft [26]. As can be seen, the notification day (0) coincides with a growth of the median number of queries for those domains, which can be due to the spam campaigns carried out by the attackers (and partially also due to other users of this blacklist). Also, some phishing result in DNS request for multi-level domains (e.g.: a.b.c.d.e.f.example.nl). However, as more resolvers implement QNAME minimization (RFC 7816), authoritative DNS servers will only see queries for second-level domains regardless the full-qualified domain name requested.

Detecting compromised phishing websites for an entire DNS zone poses a challenge as it requires (i) to monitor a large number of domains, (ii) choosing the right features, and (iii) to develop a solution that can differentiate true from false positives. For example, an increase on the median number of queries for a particular domain may be due to various reasons (for example, social networking campaigns, advertisement, etc.), but features related to the source of traffic (ASs, countries, timing) can be used in this process.

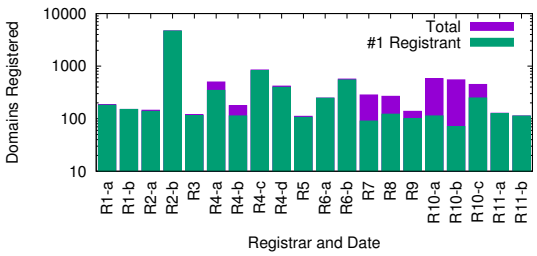

Fig. 4. Anomalous registrations for Registrars and Top 1 registrant – most of registrations are done in bulk by 1 registrant.

C. Parking (advertisement)

Domain parking consists of registering a domain and hosting solely advertisement content on it for profit purposes. Parking is not illegal in most TLDs, it has simply evolved as an unintended use of domains. Still, it is important to understand it in order to detect it from its malicious "cousin" (§IV-D): parking that redirect users to malicious websites.

The business model employed by "domainers" (users that perform parking) has been extensively covered by Vissers et. al [19], with an entire industry build around it. It consists of registering a large number of domains that may attract traffic, for example, from typo-squatting to drop-catch (re-registering expired domains). Then, they point these domains to advertisement networks and wait for incoming traffic, which is ultimately used for monetization.

Since often parking relies upon bulk registrations, we should be able to identify those at `RegDB`. As discussed in §II-A, a registrant needs a registrar to carry out registrations. To perform bulk registrations, they choose registrars that support it and that also give discounts for bulk registrations. Therefore, bulk registration should change the "normal" number of daily registration by a registrar.

To test this hypothesis, we employ a robust time-series anomaly detection proposed by Vallis et. al [28] on the daily registrations for all the registrars accredited to .nl using 152 days of registrations, obtained from `RegDB`. In this period, more than 150 thousand domains were registered, by roughly 1,000 registrars. Figure 4 shows the top 20 registrar/date tuples in which the number of registrations was at least 100 above the median for the registrar. In this figure, registrar/date are represented in Rn-δ, in which Rn is a registrar, and δ represent different dates.

As can be seen, only 11 registrars (R1–R11) are responsible for those bulk registrations over the period, having up to 4000 daily registrations above their median number. In the same figure, we also show the number of registrations done by only one registrant for each day. As can be seen, for most cases, a single registrant is responsible for the spike in the number of registrations. We believe this is due to the fact that parking is not illegal, therefore users have little incentive to conceal their identities. We use this information to distinguish ads parking from malicious redirect.

D. Parking (malicious redirect)

Parking domains with the intend of redirecting users to malicious websites employ the same business model as parking for advertisement (§IV-C) – the main difference lies in how money is made. Instead of monetizing from legitimate traffic and advertisement, users are redirected to websites that contain malware or other malicious content. After they are infected, these users may be exposed to extortion (ransomware) or have their computers being used as bots in botnets, which allow for various types of monetization, such as spamming campaigns, DDoS attacks on demand (such as booters [29]).

We have seen in Figure 4 that registrants of parked domains that employ ad networks use the same credentials when registering domains in bulk – since their business model is legitimate. Parking for the purposes of malicious redirection, on the other hand, is malicious, and the registrant would have incentives to change its credentials – thus avoiding their other websites to be detected. Therefore, we need to investigate whether other registration time features – such as time of registration, registrant – can be used in the detection.

E. Fake Goods

When we developed nDEWS [6] to detect 0-day phishing using `RegDB` and `AuthDNS` datasets, we came across 148 websites (in a period of 8 months) that showed the same pattern: sales of sneakers at high discount, and while they differ, they exhibited similar structures (layout, rate of discount, no TLS enabled). At a first glance, they look like any other online store, except for large discounts. One may even underestimate the risks posed by these websites in comparison with "traditional" banking/credit card phishing.

Before dismissing such websites, it is important to understand the counterfeit industry. According to the World Customs Organization, counterfeit goods account for nearly 10% of worldwide trade, an estimated $500 billion annually [30]. Sneakers are the *number one* products seized by the U.S. Customs and Border Protection [30].

In the literature, such websites are referred to as concocted stores, i.e., deceptive websites that appear to be legitimate commercial ones [31], and either fail to ship their ordered goods or ship different/counterfeit products. They differ from spoofed sites, which are intended to deceive authentic site's costumers [32].

From what we have seen, these stores follow a similar business model as 0-day phishing (§IV-A): registration of domain names that resemble the main brand, spam campaigns, and sales or credit card/ID theft. It is hard to determine with 100% certainty if a website is fraudulent or not – ultimately this involves a trial purchase. The problem with fake good sites is that they operate in a "grey area", and may remain online for months before being taken down, potentially causing more damage to more customers.

F. Drop-catch malware

As discussed in §II-A, domains can expire and be re-registered by other registrants, which enables reusing the

same name space for different registrants/companies over time. When a domain expires, the new registrant can profit from the *residual trust* that had been built up by the previous registrant – e.g. by a reputable company.

With this in mind, Lever *et al.* [21] carried out a comprehensive study with almost 180 million expired domains and showed a class of abuse that focuses on leveraging the residual trust of these domains. They uncovered a business model that revolves around re-registering expired domains and using it to redirect users to malware-hosting websites. This, in turn, leads to infected users, that are then prone to become part of botnets (that can be used in various forms of monetization) or even ransomware.

The business model is similar to parking domains with malicious redirects (§IV-D) – however instead of relying upon typo-squatting [33] , re-registration of domains builds on the reputation of the previous website.

To curb such abuse, the authors [21] have also proposed an algorithm that uses several features collected from passive data of resolvers in an ISP network to detect this sort of abuse, and validated it against `whois` data. Features included a passive recursive resolve DNS feed, as well as change on DNS records for such domains. As discussed in §III, `RegDB` has many advantages in comparison to `whois` which could improve their algorithm further.

G. Botnet C&C

A botnet is a network of compromised devices under control of a "botmaster", who uses it to carry out various types of attacks, including DDoS, malware hosting, spam campaigns, among others. Recently, an Internet-of-Things (IoT) botnet was involved in one of the largest DDoS attack recorded: 620 Gbps of direct attack traffic [34].

Botmaster use different techniques to control the bots and IRC, peer-to-peer, hard-coded domain names and domain generated algorithms (DGA) [35] are four common forms of botnet command and control (C&C) architecture.

Botnets with hard coded C&C addresses mostly only rely on a small number of domain names. In DGA botnets, in contrast, bots query thousands of possible domains, but attackers register only one or few daily. In case one of the registered domains is taken down, daily new lists of DGA domain names guarantee that preemptive registration is not feasible to take a botnet down.

Both mechanisms create different signals at the `RegDB` as well as at `AuthDNS`. While one or few domains are expected to be daily registered by a DGA botnet operator, `AuthDNS` should exhibit an abnormal traffic for non-existing domains (randomly generated but not registered) coming from the bots. We observe this on .nl: Figure 5 shows the number of queries for domain names that have been used by the Flashback botnet. Note, that a rapid increase of queries can be observed when the botnet becomes active. Additionally, domain names created by DGAs often have certain lexical attributes that can be observed in the `RegDB` as well.

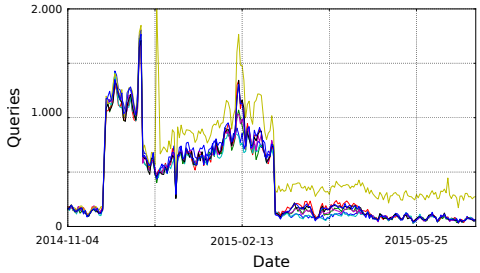

Fig. 5. Number of queries per day for C&C domain names of the Flashback botnet, as observed at two .nl authoritatve name servers.

Another technique used by botnets to increase its resilience and to hide the C&C infrastructure is the use of Fast-Flux Service Networks (FFSN) [36], which adds a proxy layer between the bots and the C&C servers. FFSNs leave a large footprint in the "Records" dataset because of continuously changing A and AAAA records of the C&C domain name.

TLD operators need to take into account that botmasters can spreads the C&C domains across multiple TLDs. Therefore, it may be necessary collaboration cross TLDs to detect C&C domains to curb botnet C&C domains.

H. Blackhat Search Engine Optimization (SEO)

Blackhat SEO refers to the practice of artificially improving the visibility of a website by adopting practices that target search engine algorithms but that are irrelevant to human users, such as employing invisible text or spamming links on other websites. This differs from SEO methods that are encouraged by search engines and that have the goal to improve the user experience. Blackhat SEO is also used by attackers, and the traffic drawn to these websites can be used for various forms of monetization: e.g. advertisement, fake shoes or malware [37]. While Blackhat SEO may not be illegal, as for parking domains for advertisement, its practice is debatable and we see it as a means to reach more victims for the aforementioned abusive activities.

We discuss two Blackhat SEO techniques that mainly employ domain names. The first one consists in re-registering an expired domain (as in §IV-F) and leveraging its residual trust, then to form a Private Blog Network (PBN) [23]. The dropped domains under a PBN are then filled with links and other information pointing to the website of which the search engine results should be improved.

The second technique makes use of DNS wildcard (RFC 4592) to "entrap" search engine crawlers in a circle of automatically-generated random domain names [38]. These automatically-generated pages are then set-up to include links to the monetization website, ultimately increasing its ranking in search engines.

To detect those practices, a TLD operator could analyze both `RegDB` and `Records` datasets. Queries for random subdomains appear in `AuthDNS` as well. It is still an open research question, as in §IV-F, to detect these from a TLD vantage point.

V. Conclusions

Hidden behind domain names, there are lucrative (and ingenious) business models that employ a diversified form of monetization to the benefit of attackers. In this paper, we have presented a survey of existing forms of domain abuse, and discussed how TLD operators can detect those, leveraging their readily available datasets that most researchers do not have access to. We hope this work can be used by other TLD operators to detect abuses in their respective DNS zone, relying on their already available data sets.

As future work, we will be developing specific solutions that address each business model presented in this paper, as we have already done for 0-day phishing.

References

[1] S. Hao, A. Kantchelian, B. Miller, V. Paxson, and N. Feamster, " PREDATOR: Proactive Recognition and Elimination of Domain Abuse at Time-Of-Registration," in *Proceedings of the 2016 ACM CCS*, October 2016.

[2] S. Hao, M. Thomas, V. Paxson, N. Feamster, C. Kreibich, C. Grier, and S. Hollenbeck, "Understanding the Domain Registration Behavior of Spammers," in *Proceedings of the 2013 Conference on Internet Measurement Conference*, ser. IMC '13. New York, NY, USA: ACM, 2013, pp. 63–76.

[3] Hao, Shuang and Feamster, Nick and Pandrangi, Ramakant, "Monitoring the Initial DNS Behavior of Malicious Domains," in *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, ser. IMC '11. New York, NY, USA: ACM, 2011.

[4] M. Antonakakis, R. Perdisci, W. Lee, N. Vasiloglou II, and D. Dagon, "Detecting Malware Domains at the Upper DNS Hierarchy." in *USENIX Security Symposium*, 2011, pp. 16–32.

[5] M. Antonakakis, R. Perdisci, D. Dagon, W. Lee, and N. Feamster, "Building a Dynamic Reputation System for DNS." in *USENIX security symposium*, 2010, pp. 273–290.

[6] Giovane C. M. Moura, Moritz Muller, Maarten Wullink, and Cristian Hesselman, "nDEWS: a New Domains Early Warning System for TLDs," in *IEEE/IFIP International Workshop on Analytics for Network and Service Management (AnNet 2016), co-located with IEEE/IFIP Network Operations and Management Symposium (NOMS 2016)*, April 2016.

[7] S. Liu, I. Foster, S. Savage, G. M. Voelker, and L. K. Saul, "Who is. com?: Learning to Parse WHOIS Records," in *Proceedings of the 2015 ACM Conference on Internet Measurement Conference*. ACM, 2015, pp. 369–380.

[8] C. Hesselman, G. C. M. Moura, R. d. O. Schmidt, and C. Toet, "Increasing DNS Security and Stability through a Control Plane for Top-Level Domain Operators," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 197–203, January 2017.

[9] IANA, "IANA — Root Files," https://www.iana.org/domains/root/files, 2016.

[10] ICANN, "Centralized Zone Data Service," https://czds.icann.org, 2016.

[11] IANA, "Domain Name System (DNS) Parameters," 2016. [Online]. Available: https://www.iana.org/assignments/dns-parameters/dns-parameters.xml

[12] G. C. M. Moura, R. de O. Schmidt, J. Heidemann, W. B. de Vries, M. Müller, L. Wei, and C. Hesselman, "Anycast vs. DDoS: Evaluating the November 2015 Root DNS Event," in *Proceedings of the ACM Internet Measurement Conference (IMC 2016)*, Nov. 2016.

[13] Y. Yu, D. Wessels, M. Larson, and L. Zhang, "Authority Server Selection in DNS Caching Resolvers," *SIGCOMM Computer Communication Review*, vol. 42, no. 2, pp. pp. 80–86, Mar. 2012.

[14] Maarten Wullink, Giovane C. M. Moura, Müller, M, and Cristian Hesselman, "ENTRADA: a High Performance Network Traffic Data Streaming Warehouse," in *Network Operations and Management Symposium (NOMS), 2016 IEEE* , April 2016.

[15] R. van Rijswijk-Deij, M. Jonker, A. Sperotto, and A. Pras, "A High-Performance, Scalable Infrastructure for Large-Scale Active DNS Measurements," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 6, pp. 1877–1888, June 2016.

[16] C. Hesselman, J. Jansen, M. Wullink, K. Vink, and M. Simon, "A privacy framework for DNS big data applications," Tech. Rep., 2014. [Online]. Available: https://www.sidnlabs.nl/downloads/whitepapers/SIDN_Labs_Privacyraamwerk_Position_Paper_V1.4_ENG.pdf

[17] A. Noroozian, M. Korczynski, S. Tajalizadehkhoob, and M. van Eeten, "Developing security reputation metrics for hosting providers," in *8th Workshop on Cyber Security Experimentation and Test (CSET 15)*, 2015.

[18] S. Alrwais, K. Yuan, E. Alowaisheq, Z. Li, and X. Wang, "Understanding the Dark Side of Domain Parking," in *23rd USENIX Security Symposium (USENIX Security 14)*. San Diego, CA: USENIX Association, Aug. 2014, pp. 207–222.

[19] T. Vissers, W. Joosen, and N. Nikiforakis, "Parking Sensors: Analyzing and Detecting Parked Domains." in *Proceedings of the 2015 Network and Distributed System Security Symposium (NDSS 2015), San Diego, California, USA.*, 2015.

[20] D. McCoy, A. Pitsillidis, G. Jordan, N. Weaver, C. Kreibich, B. Krebs, G. M. Voelker, S. Savage, and K. Levchenko, "PharmaLeaks: Understanding the Business of Online Pharmaceutical Affiliate Programs," in *Proceedings of the 21st USENIX Security Symposium*. Bellevue, Washington, USA: USENIX Association, August 2012.

[21] C. Lever, R. Walls, Y. Nadji, D. Dagon, P. McDaniel, and M. Antonakakis, "Domain-Z: 28 Registrations Later," In: *Proceedings of the 37th IEEE Symposium on Security and Privacy*. San Jose, California., 2016.

[22] B. Stone-Gross, M. Cova, L. Cavallaro, B. Gilbert, M. Szydlowski, R. Kemmerer, C. Kruegel, and G. Vigna, "Your botnet is my botnet: analysis of a botnet takeover," in *Proceedings of the 16th ACM conference on Computer and communications security*. ACM, 2009.

[23] niche site project, "Private Blog Networks http://nichesiteproject.com/private-blog-networks/," Sep. 2016.

[24] K. Du, H. Yang, Z. Li, H. Duan, and K. Zhang, "The Ever-Changing Labyrinth: A Large-Scale Analysis of Wildcard DNS Powered Blackhat SEO," in *25th USENIX Security Symposium (USENIX Security 16)*. USENIX Association.

[25] M. Khonji, Y. Iraqi, and A. Jones, "Phishing Detection: A Literature Survey," *IEEE Communications Surveys Tutorials*, vol. 15, no. 4, pp. 2091–2121, Fourth 2013.

[26] Netcraft, "Phishing Site Feed," http://www.netcraft.com/anti-phishing/phishing-site-feed/, 2015.

[27] M. Müller, "SIDekICk: SuspIcious DomaIn Classication in the .nl Zone," Master's thesis, University of Twente, the Netherlands, 2015. [Online]. Available: http://eprints.eemcs.utwente.nl/26196/

[28] O. Vallis, J. Hochenbaum, and A. Kejariwal, "A novel technique for long-term anomaly detection in the cloud," in *6th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 14)*, 2014.

[29] J. J. Santanna, R. van Rijswijk-Deij, R. Hofstede, A. Sperotto, M. Wierbosch, L. Zambenedetti Granville, and A. Pras, "Booters-An analysis of DDoS-as-a-service attacks," in *IFIP/IEEE Intl. Symposium on Integrated Network Management (IM)*. IEEE, May 2015.

[30] B. Burnsed, "The Most Counterfeited Products – Bloomberg Business," http://www.bloomberg.com/ss/08/10/1002_counterfeit/1.htm, 2015.

[31] A. Abbasi and H. Chen, "A comparison of tools for detecting fake websites," *Computer*, no. 10, pp. 78–86, 2009.

[32] N. Chou, R. Ledesma, Y. Teraguchi, J. C. Mitchell *et al.*, "Client-Side Defense Against Web-Based Identity Theft," in *NDSS*, 2004.

[33] T. Moore and B. Edelman, "Measuring the perpetrators and funders of typosquatting," in *International Conference on Financial Cryptography and Data Security*. Springer, 2010.

[34] B. Krebs, "KrebsOnSecurity Hit With Record DDoS https://krebsonsecurity.com/2016/09/krebsonsecurity-hit-with-record-ddos/," Sep. 2016.

[35] M. Antonakakis, R. Perdisci, Y. Nadji, N. Vasiloglou II, S. Abu-Nimeh, W. Lee, and D. Dagon, "From Throw-Away Traffic to Bots: Detecting the Rise of DGA-Based Malware." in *USENIX Security Symposium*, 2012, pp. 491–506.

[36] J. Nazario and T. Holz, "As the net churns: Fast-flux botnet observations," in *Malicious and Unwanted Software, 2008. MALWARE 2008. 3rd International Conference on*. IEEE, 2008.

[37] D. Wang, S. Savage, and G. M. Voelker, "Juice: A longitudinal study of an seo campaign," in *Proceedings of the NDSS Symposium*, 2013.

[38] K. Du, H. Yang, Z. Li, H. Duan, and K. Zhang, "The ever-changing labyrinth: A large-scale analysis of wildcard dns powered blackhat seo," in *25th USENIX Security Symposium (USENIX Security 16)*. Austin, TX: USENIX Association, Aug. 2016, pp. 245–262.