# Statistical Analysis of DNS Abuse in gTLDs: Intermediate Draft Report

Maciej Korczyński*, Maarten Wullink†, Giovane C.M. Moura†, Cristian Hesselman†

*Delft University of Technology, The Netherlands
†SIDN Labs, The Netherlands
Email: maciej.korczynski@tudeft.nl, {maaarten.wullink, giovane.moura, cristian.hesselman}@sidn.nl

## I. Introduction

Commissioned by the Competition, Consumer Trust, and Consumer Choice Review Team (CCTRT) with the support of ICANN, this report is focused on determining rates of common forms of abusive activities in the domain name system. The study aims to compare rates of these activities between new and legacy gTLDs. This intermediate draft report is meant as a preview of the study's data, methodology, and findings for review by the CCTRT. The final report is expected in July 2017.

## II. Background

### A. Domain Name Ecosystem

The Internet Corporation for Assigned Names and Numbers (ICANN) [1] approves all top-level domains (TLDs) and delegates the responsibility to a particular organization ("registry operators", "sponsors", or "delegees") to maintain an authoritative source for registered domain names within a TLD [2]. Domain *registries* manage the registration of domain names within their TLDs and generate zone files that list all available domain names with their authoritative name servers.

TLDs can be categorized into three main groups [3]: *i*) country code TLDs (ccTLDs) created for more than 250 countries and country codes such as .pl or .fr. The country registry sets the registration and delegation policies for a ccTLD, *ii*) generic TLDs (gTLDs) such as .com or .amsterdam, and *iii*) .arpa – a special TLD that is used for technical infrastructure.

Several other entities play a role for a domain name to be registered, secured and maintained on the Web. Domain registrars manage the registration of Internet domain names. They are generally accredited by TLD registries and may be accredited by ICANN. Web hosting providers host server infrastructure, which is used to host content and/or services for the domain. Domain Name System (DNS) providers operate DNS servers that map domain and host names to the corresponding Internet Protocol (IP) addresses.

### B. Generic TLDs

The first group of generic top-level domains (gTLDs) was defined by RFC 920 [4] in October 1984 – in the early development of the domain name system of the Internet – and introduced a few months later. The initial group of gTLDs (.gov, .edu, .com, .mil, .org, and .net) were distinct from country-code TLDs. Until 2012, several gTLDs were approved and further introduced by ICANN, including a set of sponsored gTLDs such as .asia, .jobs, .travel, or .mobi. In this paper, we refer to all gTLDs introduced before the New gTLD Program [5] initiated by ICANN in late 2013 as *legacy* gTLDs. In this study, we analyze a set of 18 legacy gTLDs (.aero, .asia, .biz, .cat, .com, .coop, .info, .jobs, .mobi, .museum, .name, .net, .org, .post, .pro, .tel, .travel, and .xxx) for which we were able to obtain zone files and we perform a more fine-grained analysis using a set of 9 legacy gTLDs (.asia, .biz, .com, .coop, .info, .mobi, .net, .org, and pro) for which we obtained the WHOIS data. We contrast them with the *new gTLDs*.

### C. New gTLDs

ICANN's New gTLD Program [5] started in 2012 and expanded the root zone by delegating more than 1,200 new gTLDs since October 2013.

To obtain new gTLDs, applicants are required to undergo an intensive application and evaluation process [6] that includes screening applicants for the technical and financial capabilities necessary for operating a top-level domain.

Ultimately, after a new gTLD is assigned to an applicant, it will then be delegated to the root zone. Following initial delegation, each new gTLD registry is required to have a "sunrise" period of at least 30 days, during which trademark holders have an advance opportunity to register domain names corresponding to their marks before names are generally available to the public.

New gTLDs can be classified into four broad categories[1] [8]:

- Standard or generic gTLD [9]: is a TLD that is generally open for public registration, e.g. .movie, .xyz, or .family. While most of these TLDs are open to public registration, some registries may impose restrictions on who or which entities can register in their domains.
- Community gTLD [10]: this category covers TLDs that are restricted to a specific community, such as .thai, .audi or .pharmacy.
- Geographic gTLD [11]: this type of TLD covers cities, states, or regions, e.g. .amsterdam or .berlin.

[1] Note that some gTLDs cross categories. For example, some community gTLDs such as .madrid are also geographic gTLDs [7].

- Brand gTLD [12]: for companies seeking to have their specific brand as a TLD, such as .google or .hitachi.

In our study, we analyze new gTLDs that are intended for public use. Therefore, we excluded the great majority of brand gTLDs for which domains cannot be registered by regular users[2], in particular for malicious purposes. This report covers new gTLDs for which registries have submitted their sunrise date information requested by ICANN. In the first quarter of 2014, there were 77 new gTLDs for which the sunrise period ended and domain names were available for public registration. For comparison, by the end of 2016 the group consisted of 522 new gTLDs.

## III. Data Collection

### A. Blacklists

To asses the prevalence of maliciously registered and compromised domains per gTLD and registrar, we use 10 heterogeneous blacklists representing malware, phishing and spam generously provided to us by Spamhaus [13], the Anti-Phishing Working Group (APWG) [14], StopBadware [15], SURBL [16], and CleanMX [17]. All five organizations provide reputable domain or URL blacklists used in operational environments. The domain blacklist provided to us by Spamhaus consist of domains with low reputation collected from spam payload URLs, spam senders and sources, known spammers, phishing, virus and malware-related websites. The list is built mainly using spamtraps and by monitoring emails. Spamhaus does a number of checks to prevent legitimate domains being listed. As it is a near zero false positive list it is safe to use by production mail systems [18]. The APWG feed consists of online phishing URL block/white lists with accompanying confidence level indicators submitted by accredited users through the eCrime Exchange (eCX) platform. Note that starting from September 2015 Facebook data, which represented a significant part of URLs, was excluded from the feed and it got a module of its own. The StopBadware Data Sharing Program (DSP) feed consists of URL blacklists shared by ESET, Fortinet, and Sophos security companies [19]–[21], Internet Identity, Google's Safe Browsing appeals results, the StopBadware community, and other contributors [22]. In our study we also use three domain blacklists generously provided by SURBL. *SURBL ph* is a phishing domain blacklist comprised of data supplied by MailSecurity, PhishTank, OITC phishing, PhishLabs, US DHS, NATO as well as data from various corporations and numerous other sources including proprietary data as well as information from traps [23]. *SURBL jp* blacklist contains domains analyzed and categorized as spam (e.g. uncategorized unsolicited) by jwSpamSpy software, traps, and participating mail servers. *SURBL ws* is similar and contains mainly spam domains from SpamAssassin, ASSP as well as information from other data sources including internal and external trap networks. *SURBL mw* list contains data from multiple sources that cover malicious domains used to host malware websites, payloads or associated redirectors. This feed includes the DNS blackhole malicious site data from malwaredomains.com, OITC, Malware Domain List, US DHS, internal and external DGAs, Impact, trap data using static and dynamic filtering and more [23]. Note that unlike the other data feeds, SURBL data covers the 2,5-year study period between July 2014 and December 2016. Finally, CleanMX provided us three URL blacklists containing phishing, malware websites, as well as the "portals" feed that contain defaced, spamvertized, hacked, and other types of abused websites.

Table I
Overview of blacklists: unique blacklisted gTLD domain names, FQDNs, and URLs, for the APWG, StopBadware SDP, Spamhaus, CleanMX, and SURBL datasets for 2014, 2015, 2016.

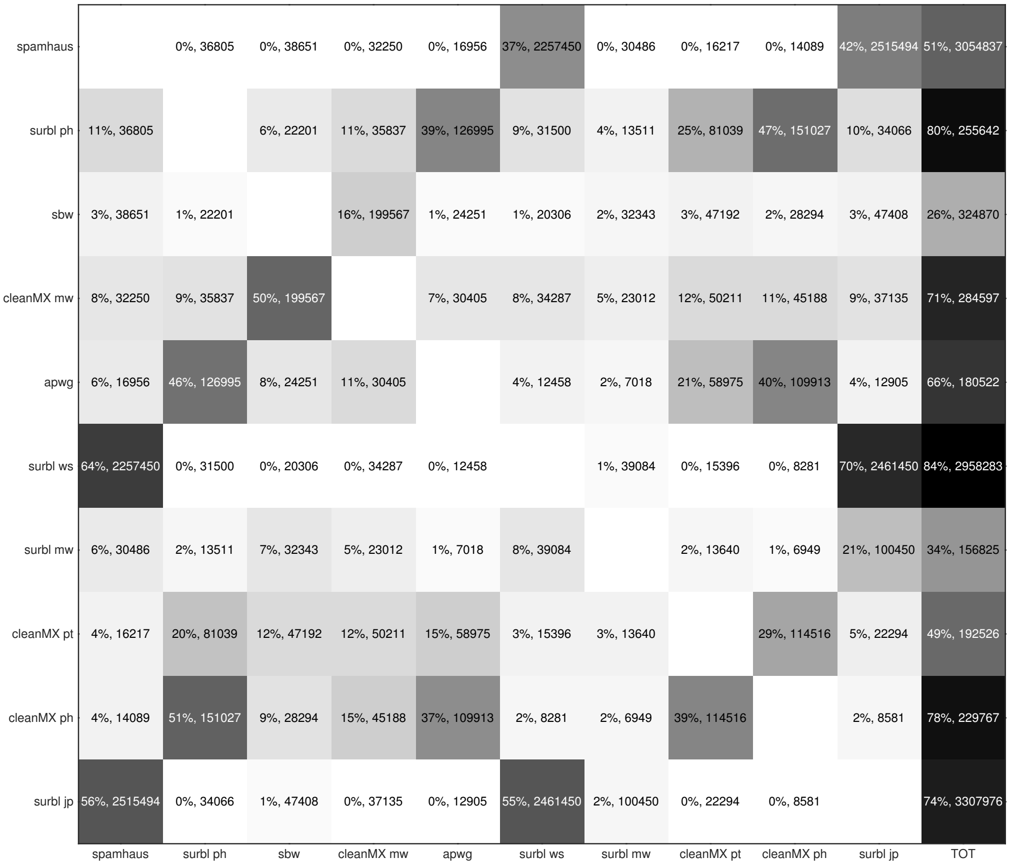
| Year | Dataset | # domains | # FQDNs | # URLs |
|---|---|---|---|---|
| 2014 | StopBadware | 403,347 | 728,007 | 1,522,548 |
| | APWG | 60,681 | 891,996 | 4,993,966 |
| | Spamhaus | 1,901,970 | – | – |
| | CleanMX ph | 68,523 | 86,838 | 269,770 |
| | CleanMX mw | 169,237 | 533,142 | 2,628,295 |
| | CleanMX pt | 205,051 | 251,181 | 526,599 |
| | SURBL ph | 68,208 | – | – |
| | SURBL mw | 289,664 | – | – |
| | SURBL ws | 1,229,698 | – | – |
| | SURBL jp | 1,484,807 | – | – |
| 2015 | StopBadware | 501,982 | 652,549 | 5,744,669 |
| | APWG | 139,538 | 1,665,839 | 20,221,682 |
| | Spamhaus | 2,505,407 | – | – |
| | CleanMX ph | 98,112 | 150,396 | 478,259 |
| | CleanMX mw | 117,140 | 263,218 | 1,002,658 |
| | CleanMX pt | 124,608 | 197,703 | 469,410 |
| | SURBL ph | 134,591 | – | – |
| | SURBL mw | 220,073 | – | – |
| | SURBL ws | 1,813,858 | – | – |
| | SURBL jp | 2,475,745 | – | – |
| 2016 | StopBadware | 502,579 | 586,181 | 2,998,978 |
| | APWG | 83,215 | 103,190 | 230,636 |
| | Spamhaus | 3,944,684 | – | – |
| | CleanMX ph | 138,869 | 207,984 | 738,385 |
| | CleanMX mw | 149,632 | 203,419 | 1,076,547 |
| | CleanMX pt | 68,413 | 108,145 | 829,533 |
| | SURBL ph | 173,326 | – | – |
| | SURBL mw | 106,819 | – | – |
| | SURBL ws | 2,023,178 | – | – |
| | SURBL jp | 2,442,592 | – | – |

Table I shows the number of unique gTLD domain names, fully-qualified domain names (FQDNs)[3] and URLs in these data feeds for 2014, 2015 and 2016. Notice that we define domain names as 2nd-level or 3rd-level, or even $n$th-level domain names, if a given TLD registry provides such registrations, e.g. *.gov.uk, *.co.uk, *.ac.uk, etc. To extract domain names from our feeds, we use a modified version of the public suffix list maintained by Mozilla [24]. Note that new gTLD registries offer uniquely 2nd-level domain registrations.

The distinction between different types of blacklists is very important for the registry operators and other intermediaries such as hosting providers or registrars. As already explained

[2] with a few exceptions such as .allfinanz or .forex brand gTLDs for which the sunrise period has been announced and ended.

[3] FQDN is the name for a specific host that includes both a hostname and a domain name. For example, a FQDN for a hypothetical dns server might be ns1.domain.gov.uk, where ns1 is the hostname and domain.gov.uk is the domain name.



| | spamhaus | surbl ph | sbw | cleanMX mw | apwg | surbl ws | surbl mw | cleanMX pt | cleanMX ph | surbl jp | TOT |
|---|---|---|---|---|---|---|---|---|---|---|---|
| spamhaus | | 0%, 36805 | 0%, 38651 | 0%, 32250 | 0%, 16956 | 37%, 2257450 | 0%, 30486 | 0%, 16217 | 0%, 14089 | 42%, 2515494 | 51%, 3054837 |
| surbl ph | 11%, 36805 | | 6%, 22201 | 11%, 35837 | 39%, 126995 | 9%, 31500 | 4%, 13511 | 25%, 81039 | 47%, 151027 | 10%, 34066 | 80%, 255642 |
| sbw | 3%, 38651 | 1%, 22201 | | 16%, 199567 | 1%, 24251 | 1%, 20306 | 2%, 32343 | 3%, 47192 | 2%, 28294 | 3%, 47408 | 26%, 324870 |
| cleanMX mw | 8%, 32250 | 9%, 35837 | 50%, 199567 | | 7%, 30405 | 8%, 34287 | 5%, 23012 | 12%, 50211 | 11%, 45188 | 9%, 37135 | 71%, 284597 |
| apwg | 6%, 16956 | 46%, 126995 | 8%, 24251 | 11%, 30405 | | 4%, 12458 | 2%, 7018 | 21%, 58975 | 40%, 109913 | 4%, 12905 | 66%, 180522 |
| surbl ws | 64%, 2257450 | 0%, 31500 | 0%, 20306 | 0%, 34287 | 0%, 12458 | | 1%, 39084 | 0%, 15396 | 0%, 8281 | 70%, 2461450 | 84%, 2958283 |
| surbl mw | 6%, 30486 | 2%, 13511 | 7%, 32343 | 5%, 23012 | 1%, 7018 | 8%, 39084 | | 2%, 13640 | 1%, 6949 | 21%, 100450 | 34%, 156825 |
| cleanMX pt | 4%, 16217 | 20%, 81039 | 12%, 47192 | 12%, 50211 | 15%, 58975 | 3%, 15396 | 3%, 13640 | | 29%, 114516 | 5%, 22294 | 49%, 192526 |
| cleanMX ph | 4%, 14089 | 51%, 151027 | 9%, 28294 | 15%, 45188 | 37%, 109913 | 2%, 8281 | 2%, 6949 | 39%, 114516 | | 2%, 8581 | 78%, 229767 |
| surbl jp | 56%, 2515494 | 0%, 34066 | 1%, 47408 | 0%, 37135 | 0%, 12905 | 55%, 2461450 | 2%, 100450 | 0%, 22294 | 0%, 8581 | | 74%, 3307976 |

Figure 1. Pairwise overlap of feeds with unique domains as unit of abuse (2014-2016)

StopBadware or APWG provide blacklists that focus on URLs. Some domain names in the URLs are registered by miscreants for malicious purposes only. The majority of domain names in the URLs are however compromised domains that were registered by legitimate users (see e.g. global phishing survey reports [25], [26]). From the operational point of view blocking domain name element of a blacklisted URL might harm legitimate operations. On the other hand, Spamhaus and other data providers maintain blacklists of domain names and perform extensive sanity checks to prevent legitimate domain names being listed. As a result, the domain blacklists can be used by production systems to, for example, block emails that contain malicious domain names. In this paper, we refer to both domains that appear on the domain blacklists and domain name elements of blacklisted URLs as "abused domains". Table II provides an overview of the blacklists used in our study and their corresponding types.

Figure 1 illustrates pairwise feed intersections as a matrix, with unique domain name as the unit for abuse. Note that darker shades of grey represent higher overlaps. For example, the overlap between Spamhaus and SURBL ws indicates 2,257,450 domain names in common within the observation period. This overlap constitutes 37% of the Spamhaus feed. In comparison, 2,257,450 domain names represent 64% of the SURBL ws feed. This is to be expected as both blacklists contain the same type of abuse, i.e. spam (see Table II). The rightmost column indicates the absolute number and the percentage of samples that the blacklist has in common with all other feeds combined. For instance the overlap between Spamhaus and all other blacklists is equal to 3,054,837 and indicates that as many as 51% of all domains blacklisted by Spamhaus occured at least on one other blacklist.

Table II
OVERVIEW OF BLACKLIST TYPES

| | |
|---|---|
| StopBadware | Malware URLs |
| APWG | Phishing URLs |
| Spamhaus | Spam domains |
| CleanMX phishing | Phishing URLs |
| CleanMX malware | Malware URLs |
| CleanMX portals | Other URLs |
| SURBL ph | Phishing domains |
| SURBL mw | Malware domains |
| SURBL ws | Spam domains |
| SURBL jp | Spam domains |

### B. WHOIS data

Not all blacklists used in this study contain additional domain name attributes such as registrar information or date of registration. These attributes are provided by an additional data source, WHOIS data. ICANN has provided a WHOIS database covering the 3-year study period (2014-2016). The database was compiled by a commercial vendor, Whois XML API [27]. The database contains domain names for 9 legacy gTLDs: .asia, .biz, .com, .coop, .info, .mobi, .net, .org, and pro. It also contains the domain names for 1182 new gTLDs that have been delegated in the study period [28].

The database uses temporal versioning, in which every domain is scanned once in a 3 month period. Each scan period corresponds to a database version. For this study, which spans 36 months, we have used 12 sequential versions of the WHOIS database. Table III lists each database version (Version) and the number of TLDs (#TLDs) and domains (#Domains) found in the version. The versioning timestamps are used to map the correct version of WHOIS data to a domain name extracted from blacklisted URL. We extract the <domain, registrar name> tuples from the WHOIS data and use these tuples to map the domain name element from a blacklisted URL to a sponsoring registrar. This registrar name is used to determine the amount of abuse related to the registrar. We also extract the <domain, creation date> tuples to determine if the domain has been maliciously registered or compromised (see subsection IV-C for more details).

Table III
WHOIS DATA OVERVIEW: THE NUMBER OF TLDS (# TLD) AND DOMAIN NAMES (# DOMAINS) COVERED FROM 2014, 2015, AND 2016.

| Version | #TLDs | #Domains |
|---|---|---|
| 7 | 9 | 149,391,635 |
| 8 | 9 | 149,994,294 |
| 9 | 9 | 148,048,806 |
| 10 | 369 | 157,677,494 |
| 11 | 369 | 159,494,214 |
| 12 | 565 | 159,254,213 |
| 13 | 598 | 163,348,556 |
| 14 | 713 | 166,608,406 |
| 15 | 777 | 179,238,074 |
| 16 | 947 | 183,951,585 |
| 17 | 1,014 | 190,223,971 |
| 18 | 1,191 | 193,521,942 |

## C. DNS zone files

In order to calculate sizes and the DNSSEC deployment rate for each gTLD, we processed DNS zone files provided by ICANN and extracted the unique domains. These files contain data for every delegated new gTLD and for the following legacy gTLDs: .aero, .asia, .biz, .cat, .com, .coop, .info, .jobs, .mobi, .museum, .name, .net, .org, .post, .pro, .tel, .travel, and .xxx. A zone file describes a DNS zone and contains an authoritative list of registered domains for the particular zone (gTLD). Since the list of domains contained in a zone is usually dynamic (domains are registered and expired, or their records change), the respective zone file is also dynamic. Different registries apply different zone publication policies. For example, .com updates its zone every 5 minutes, while .nl updates its zone every 30 minutes.

ICANN has provided us with daily zone files for the 3-year study period. Figure 2 shows a time series of daily unique zone files we have used for this study. Note that some drops indicate days that not all zone files were available due to operational problems.

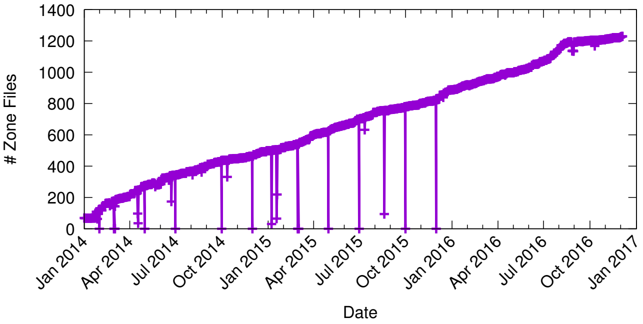


Figure 2. Number of daily zone files obtained for this study.

In this study we analyze new gTLDs whose domain names became available for the public registration within the study period. As the time between the delegation of a new gTLD and the end of the sunrise period takes even several months[4], in our analysis we include new gTLDs after their sunrise periods. This data has also been provided by ICANN via their public portal [28]. It contains 522 new gTLDs with sunrise periods that ended by the end of the study period.

# IV. METHODOLOGY

## A. Security metrics

To determine the distribution of abusive activities across the gTLDs and registrars we build on our previously proposed three occurrence security metrics [30]. First, we proposed to analyze the occurrence of *unique abused domains*.

Although, it is the most intuitive metric, it also has its limitations. It may not give an indication of the amount of abuse coming from a given domain name. For example, modern botnets extensively employ domain generation algorithms (DGAs) to generate a daily list of domain names and register a subset of those generated names as rendezvous points between compromised end users' machines and command-and-control servers (e.g. 123.malicious.com, 234.malicious.com, 432.malicious.com) [31]. Or, a single domain name registered for malicious purposes only (e.g. somedomain.com) may be used in several phishing campaigns against, for example, different banks (e.g. bankofamerica.somedomain.com, us.hsbc.com.somedomain.com, connect.secure.wellsfargo.somedomain.com) [26].

In terms of the number of unique domains (somedomain.com), the dynamic reputation system would assign the reputation score equal to 1. To overcome this limitation, we further proposed a second, complementary metric: the number of *unique fully qualified domain names* (*FQDNs*). In both examples, the reputation system based on the number of FQDNs would assign a score equal to 3 as we would observe three FQDNs generated by the attacker.

[4] E.g. delegation of .zuerich: December 25, 2014 [29], zone file seen for the first time: January 1, 2015, sunrise period termination: June 5, 2017 [28]

We encounter, however, some limitations using the second approach as well. A single FQDN of a compromised website could be used, for example, to distribute malware configuration and binary files or serve as dropzones, etc. using distinctive paths (e.g. malicious.com/wp-content/file.php, malicious.com/wp-content/gate.php, etc.) [32].

This is why we proposed a third, complementary abuse occurrence metric: *unique blacklisted URLs* aggregated by TLDs. It reveals information that is not captured by other two metrics, namely the amount of abuse associated with unique FQDNs. It stems from our previous work with the Dutch national police [33]. Our analysis of URLs used to distribute child abuse material revealed that some FQDNs are used more extensively by miscreants. In fact, one FQDN can be used to share one abusive image, whereas another can distribute tens or hundreds of images. Our manual analysis of other types of abuse such as malware or phishing confirms this trend.

Reliable reputation metrics have to account for a commonly observed trend that larger market players such as broadband or hosting providers tend to experience a larger amount of abuse [33]. For that reason, each of the previously proposed metrics are normalized by the size of the corresponding gTLDs or registrars which we discuss in the following section.

### B. Size

*1) Top-level domains:* To obtain a meaningful, quantitative security metric representing the distribution of domains listed in blacklists per gTLD, we first need to estimate their sizes. The obtained sizes can be used as a normalization factor for the amount of abuse in each gTLD or as an explanatory factor for the concentrations of abused domains. Once normalized, gTLDs can be compared in terms of the prevalence of abusive domains, FQDNs, and URLs.

We calculate the size of each gTLD by counting the number of 2$^{nd}$-level domain names present in a zone file of each gTLD at the end of the observation period. We utilized zone files obtained from ICANN as they are the most accurate for gTLD sizes. For example, to calculate abuse rates for the first quarter of 2014, we used the number of domains present in the zone files on March 31, 2014. An alternative would be to use the ICANN monthly reports that summarize domain activity for all registered domains. Some registrants, however, purchase domains and do not associate them with the name servers. As a result, they are not found in the zone files but are included in the monthly ICANN summaries. As the number of domains in a TLD registry can be seen as an approximation of the attack "surface size" for cybercriminals, the number of domains found in a zone file is more accurate.

One limitation of our approach is that it is unclear what portion of the domains are in use and contain content. More specifically, Halvorson et al., for example, show that there are as many as 16% of domains in new gTLDs with NS records that do not even resolve yet [6]. It remains unknown, however, if the trend is the same over our study period and if it is the same for all gTLDs. For example, we randomly selected 20 .bank and 20 .xyz domains and we visited them manually. 13 .bank websites resolved to the same "click through landing page", with the goal of encouraging the visitor to click through to another non-.bank page of a company. The next 4 pages did not resolve, 2 redirected the visitor to the corresponding website of a bank and only one served an actual online banking website. Furthermore, 11 .xyz domains did not resolve, 2 were parking websites, 2 redirected the visitor to an external website, 2 served no content, and 3 pointed to a domain reseller or a hosting provider website informing that the domain is for sale.

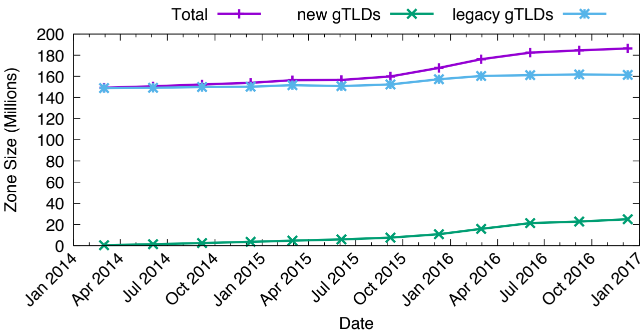


Figure 3. Absolute growth of **legacy** gTLD, **new** gTLDs and **all** gTLDs.

As we mentioned before, the size of a TLD can be interpreted as the "attack surface" for cybercriminals. In other words, the more domains managed by a hosting provider or registry the bigger the chance of getting compromised [34]–[36]. For domains registered by miscreants for malicious purposes, the TLD size may serve as a proxy for the popularity of a TLD. Some TLDs might be more popular among cybercriminals be due to, for example, lower registration prices.

Figure 3 shows the absolute growth of legacy and new gTLDs during the 3-year study period between January 2014 and December 2016. Starting from the first quarter of 2016 the number of domains in new gTLDs grows considerably in comparison to the legacy gTLDs, for which the size stays relatively constant. However, as the gTLD market share remains highly disproportionate (there are many more legacy gTLD domains, in particular .com domains), we expect the absolute number of abused .com domains to be significantly higher in comparison to the rest of the market. For completeness, Figure 4 shows the absolute growth of the top 5 largest new gTLDs respectively at the end of 2016. We do not present the absolute growth of the top 5 largest legacy gTLDs (.com, .net, .org, .info, .biz) as they remain stable during the entire study period (approximately 127M, 15.5M, 10.5M, 5.4M, and 1M, at the end of 2016, respectively).

*2) Registrars:* We calculate the registrars' size from the WHOIS data by counting the number of distinct domain names linked to each registrar name. A problem with this method is that the WHOIS data may contain multiple name variants for a registrar, each of these names may slightly differ. For example, GoDaddy is found as a registrar using 52 distinct name variations, e.g. "GODADDY.COM, LLC", "GoDaddy.com, LLC (R91-LROR)" and "GoDaddy.com, Inc.". This means

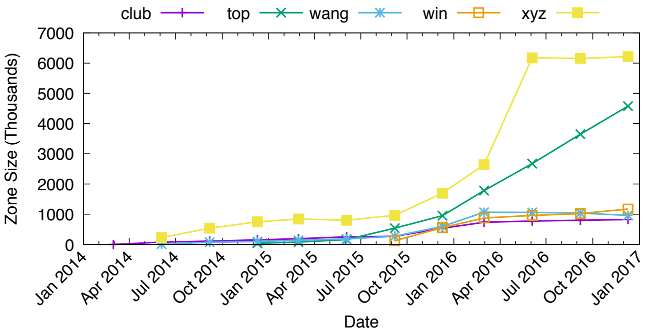


Figure 4. Absolute growth of top 5 largest **new** gTLDs as of end of 2016.

we need to perform an additional entity resolution step to be able to group together all the different registrar name variants as a single registrar. We also used the IANA Registrar ID which is assigned to ICANN accredited registrars. The IANA website [37] lists every accredited registrar together with the corresponding ID.

Using a script, this list of registrar names was automatically matched against every registrar name found in the WHOIS data. After this step we still needed to manually map the registrar variants that could not be mapped automatically.

The main limitation of our approach is that the database contains 9 legacy gTLDs and all new gTLDs. This means that we are missing registrar information for all ccTLDs needed to estimate the size of each registrar. According to our previous research, there are at least 139M domains operated by registries of ccTLDs [30]. This is, however, just an approximation as the great majority of ccTLD registries do not make their zone files available to third parties. Another limitation is that the "registrarname" attribute in the available WHOIS data contains an empty string for 0.5% of all records [5].

To determine the amount of abuse related to a registrar, we map each domain found in a blacklist to its respective WHOIS record which contains the registrar information. The WHOIS data uses temporal versioning, which means it may contain multiple versions of each domain, with each version authoritative for a distinct time period. To determine which version of a domain we should use, we use the date a domain was added to the blacklist and try to find the WHOIS version with the closest enclosing time-window[6].

### C. Compromised versus maliciously registered domains

Distinguishing between compromised and maliciously registered domains is critical because they require different mitigation actions by different intermediaries. For example, hosting providers have a larger role to play in cleaning up content of compromised websites whereas domain registrars are more responsible for suspending domains registered by miscreants for malicious purposes. Note that in practice, many large market players play multiple roles. For example, GoDaddy offers registration, web hosting, and DNS services.

Our analysis is based on the assumption that maliciously registered domains are involved in a criminal activity within a short time after the registration. This hypothesis has been previously considered by Hao et al. [38]. They examined the delay between the time when spam domains, appearing in spam messages, were initially registered and when they were ultimately used in attacks. They concluded that more than 99% of the domains used in spam campaigns were maliciously registered within 25 days of the attack.

To estimate the time between original registration and blacklisting, we analyze domain WHOIS information and extract the domain *creation date*. According to the Registrar Accreditation Agreement [39], the creation date of the domain registration cannot be changed as long as the domain does not expire.

In our study, we use a threshold of 25 days between a domain registration and blacklisting dates. If this time period is less than or equal to 25 days, then we label a domain as maliciously registered. Otherwise, it is considered to be compromised. If the registration information is not available then the domain is unlabeled.

[In the final version of this study, we plan to add a number of additional heuristics to distinguish maliciously registered from compromised domains, e.g. if a given domain name contains a string of a brand name or its misspelled version indicating malicious registration, etc.]

We maintain a list of domains of legitimate services (11,075 domains) that tend to be misused by miscreants. For example, bit.ly – a domain used by a legitimate URL shortener service – could be used by an attacker to create a malicious URL (e.g. bit.ly/dcsahy) that may further be used to redirect a legitimate user to a phishing website. In fact, previous research shows that miscreants extensively abuse a variety of services with good reputations, affecting not only the reputation of those services, but of entire TLDs [34].

The list is composed of the 10,000 most popular domains according to Alexa rankings [40] and our own, manually maintained lists of domains of legitimate services (332 domains of URL shorteners and 840 domains of free hosting providers). These represent a separate, third group of domains that are neither maliciously registered nor hacked (i.e. third party domains). This group includes:

- **Free hosting and dynamic DNS (DDNS)** services offering shared higher-level domains, such as Hostinger, a free hosting provider offering subdomains such as *.22.vc, *.pe.hu, or No-IP free DDNS providing e.g. *.no-ip.net subdomains.
- **Content delivery network (CDN)** services providing downloadable content, such as CloudFront offered by Amazon Web Services *.cloudfront.net.
- **Cloud-based file sharing** services such as Google Drive cloud storage and file backup (googledrive.com/*) or Dropbox (dl.dropbox.com/*) and their shortened versions such as db.tt/*, or the simple file sharing service providing

[5] The lack of registrar name is due to two reasons: the WHOIS database contains domains that are reserved and domains with missing WHOIS records due to the domains having expired.

[6] We do not differentiate these domains from domains that have been re-registered for malicious purposes ("recidivist").

URL shortening, ge.tt/*.

- **Other legitimate applications** such as URL shortener services like Google's goo.gl/* or bit.ly/* operated by Bitly, or blog post services, etc.

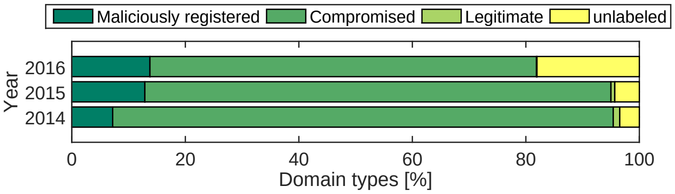


Figure 5. Categorization results: the fraction of maliciously registered, compromised, legitimate, and unlabelled domains for APWG feed in 2014, 2015, and 2016.

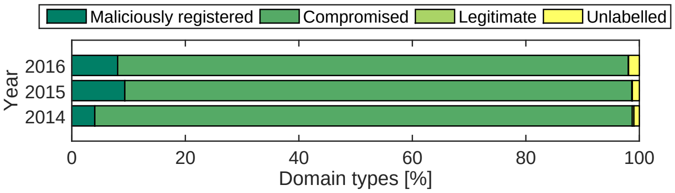


Figure 6. Categorization results: the fraction of maliciously registered, compromised, legitimate, and unlabelled domains for StopBadware DSP feed in 2014, 2015, and 2016.

Figure 5 and Figure 6 show the categorization of domains blacklisted by APWG and StopBadware respectively during the study period (2014, 2015, and 2016). Note that up to 1% of all domains submitted to the APWG have been pre-filtered based on the maintained list of domains corresponding to legitimate services. For comparison, we have excluded less than 0.3% of the StopBadware domains. A previous study showed that a large portion of legitimate domains are misused by miscreants to distribute malware or used in phishing campaigns [30]. However, some may also represent legitimate domains that were incorrectly blacklisted.

We note a limitation to this method: up to 18% and 1.9% of the APWG and StopBadware domains, respectively, are not categorized. This is mainly because the corresponding WHOIS data was not available or the registration date was after blacklisting, suggesting a domain suspension or sinkholing. However, a fraction of categorized domain instances allow us to draw general conclusions about the prevalence of maliciously registered and compromised domains, respectively.

When we excluded unlabelled domains (see Figure 5 and Figure 6), in 2014, we found that 91.4% of abused phishing and 95.6% of malware domains (listed on a URL blacklist) were compromised by criminals. In 2015, those percentages were 85.8% and 90% for phishing and malware, respectively. Finally, in 2016, 85.8% of phishing and 91.7% of malware domains were compromised. Our findings confirm others' analysis of phishing and Zeus C&C server feeds [25], [32], namely that domains listed in URL blacklists are predominantly *compromised* rather than *maliciously registered* domains. Note that the number of malicious phishing registrations might be undercounted. If an attacker does not use a maliciously registered domain within 25 days or the malicious activity is detected more than 25 days after the domain creation then the domain will be miscategorized as "compromised". Moreover, the APWG feed consists of an increased number of URL shortening links which potentially hide maliciously registered domains. We manually inspected a sample of the APWG feed and did not observe "double reports" of shortened URLs and their landing pages (websites actually hosting malicious content). Finally, we manually analyzed unclassified domain names for which the WHOIS data was not available and found that the majority of them were registered in the fourth quarter of 2016 and contained misspelled versions of brand names indicating malicious registration.

For completeness, the majority of Spamhaus and SURBL domain blacklists contain maliciously registered rather than compromised domains. This is because they perform a number of sanity checks to prevent legitimate domain names being listed.

### D. DNSSEC deployment

[Forthcoming pending final results in July 2017.]

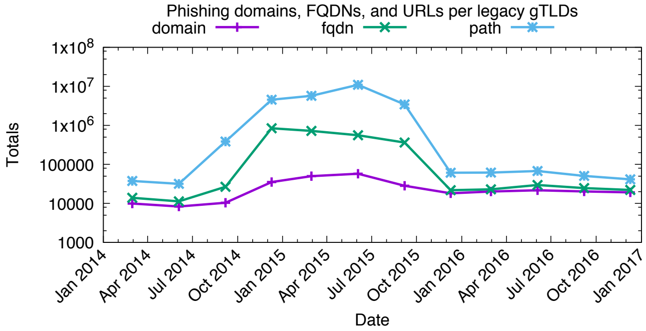


Figure 7. Time series of counts of phishing domains, FQDNs, and URLs (paths) in **legacy** gTLD based on the Anti-Phishing Working Group feed (2014-2016).

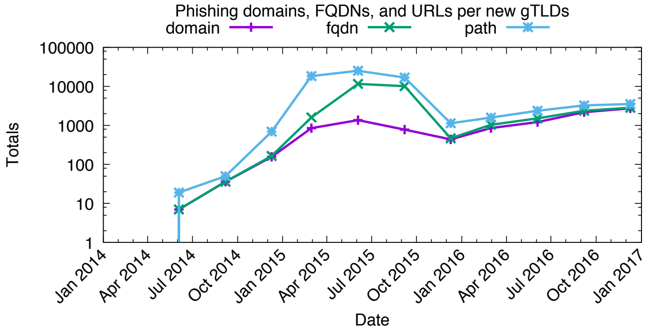


Figure 8. Time series of counts of phishing domains, FQDNs, and URLs (paths) in **new** gTLD based on the Anti-Phishing Working Group feed (2014-2016).

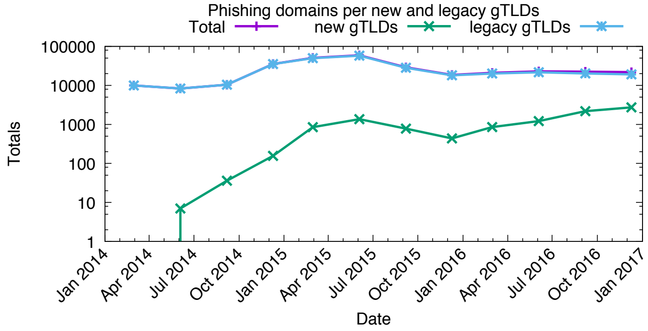


Figure 9. Time series of counts of phishing domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the Anti-Phishing Working Group feed (2014-2016). Please notice $y$ axis in log scale and overlapping lines.

## V. Results

### A. TLD reputation

*1) Phishing reputation:* We first present the three occurrence security metrics that provide insight into the distribution of abuse across legacy gTLDs (Figure 7) and new gTLDs (Figure 8) over time. We aggregate the phishing incidents on a quarterly basis (x-axis) and present the results using a logarithmic scale (y-axis). Note that the observed "decrease" in the amount of abused domains, FQDNs, and URLs (paths) in the fourth quarter of 2015 is caused by the changes in the organization of APWG URL blacklists and not by the decrease in criminal activity. As explained in section III, starting from September 2015, Facebook data, which represented a significant part of URLs, was excluded from the feed.

We observe a significant difference between three metrics based on concentration of abused domains, FQDNs, and URLs which were blacklisted by APWG. This is because the second and third one are mainly affected by legitimate services such as file storage web services or popular URL shortening services [30]. For example, in our previous work [30], we found 44,856 unique *.s3.amazonaws.com FQDNs that correspond to an online file storage web service offered by Amazon Web Services (AWS), or 377,726 unique t.co/* URLs, where t.co is a popular URL shortener operated by Twitter. The results confirm that the two complementary occurrence metrics (number of FQDN and URLs) are useful and reveal information that is not captured by the number of unique abused domains. Please compare Figure 7 and Figure 8 with the corresponding Figure 43 and Figure 44 in the Appendix section representing the cleanMX phishing dataset.

In the remainder of this subsection, we will only consider the number of unique abused domains. Figure 9 presents a time series of counts of phishing domains observed in both legacy gTLDs, new gTLDs, and the total number of all phishing domains. Similarly, we aggregate the phishing incidents on a quarterly basis and present the phishing counts using a logarithmic scale. Note, that the *total* number of phishing domains (purple line) has been driven by phishing domains in legacy gTLDs (mainly .com domains). While the number of abused domains remains approximately constant in legacy gTLDs, we observe a clear upward trend in the absolute number of phishing domains in new gTLDs. The trend is confirmed by other phishing datasets (see Figure 35 for SURBL phishing and Figure 45 for CleanMX phishing datasets).

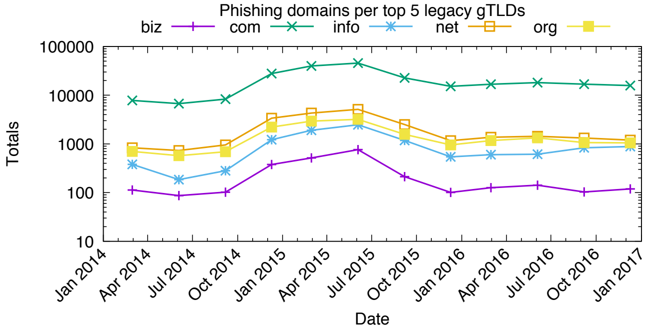


Figure 10. Time series of counts of phishing domains in the top 5 most abused **legacy** gTLDs in the last quarter of 2016 based on the Anti-Phishing Working Group feed (2014-2016).

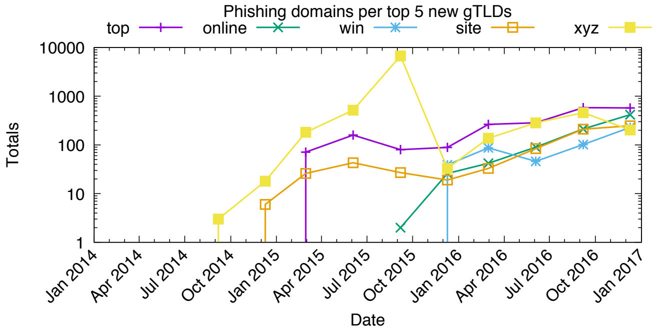


Figure 11. Time series of counts of phishing domains in the top 5 most abused **new** gTLDs in the last quarter of 2016 based on the Anti-Phishing Working Group feed (2014-2016).

Figure 10 and Figure 11 show the top 5 most abused legacy and new gTLDs with the highest absolute number of unique phishing domains at the end of 2016, respectively[7]. The number of abused phishing domains in legacy gTLDs is mainly driven by the .com gTLD and at the end of 2016 represents 82.5% (15,795 of 19,157) of all abused legacy gTLD domains considered in this study.

In comparison, in the .top TLD - the second largest new gTLD (see Figure 4) - we find the highest concentration of all phishing domains (21%, which represents 574 out of 2,738 new gTLD domains blacklisted by APWG). The upward trend in the number of phishing domains in new gTLDs (see Figure 9) is consistent with the rising trend of the top 5 new gTLDs in terms of the absolute number of abused domains listed by APWG. In fact, the five new gTLDs suffering from

[7] In Figure 11, we see that .top and .xyz, for example, starts at $y = 0$, while .online starts with $y > 0$ on its first data point. This is because differently from the others, .online had a small number of blacklisted URLs after its sunrise period, i.e., right after it became available for public registration. A similar behavior can be observed, for example, in Figure 8 and Figure 9.

the highest concentrations of domain names used in phishing attacks listed on the APWG domain blacklist in the last quarter of 2016 collectively owned 58.7% of all blacklisted domains in all new gTLDs.

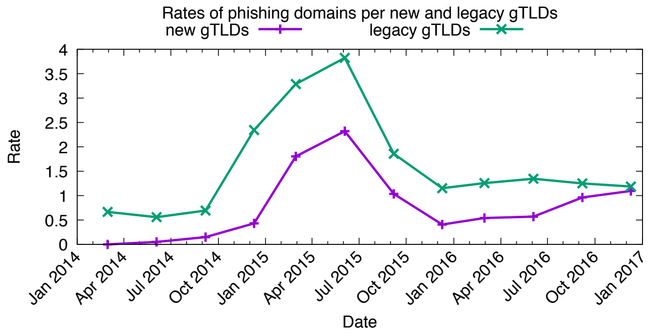


Figure 12. Time series of abuse rates of phishing domains in **legacy** gTLDs and **new** gTLDs based on the Anti-Phishing Working Group feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

As discussed before, reliable reputation metrics have to account for a commonly observed trend that larger market players experience a larger amount of domain abuse [30], [33], [36]. Figure 12 shows a time series of abuse rates of phishing domains of legacy gTLDs and new gTLDs based on the APWG feed (for comparison, see Figure 46 for abused CleanMX phishing domains and Figure 36 for SURBL phishing domains). The abuse rates are presented in a linear scale. For example, in the second quarter of 2015 the domain abuse rate for legacy gTLDs is equal to 3.82503. This means that, on average, legacy gTLDs had 3.8 blacklisted phishing domains per 10,000. Interestingly, the phishing abuse rates in legacy and new gTLDs are converging with time and were almost the same at the end of 2016. In the early stage of the New gTLD Program, phishing abuse rates were equal to 0.56 and 0.05 for legacy and new gTLDs, respectively (see the second quarter of 2014 in Figure 12). We observed 7 abused domains out of approximately 1,355,000 domains registered by the general public. For comparison, in the fourth quarter of 2016, abuse rates were equal to 1.19 and 1.1 for legacy and new gTLDs, respectively.

Up to this point, our descriptive statistical analysis of phishing abuse rates in new and legacy gTLDs has conflated compromised and maliciously registered domains. Now we compare rates of compromised domains.

Figure 13 shows a time series of abuse rates of compromised phishing domains of legacy gTLDs and new gTLDs based on the APWG feed. As expected, curves in Figure 12 (all blacklisted phishing domains) and Figure 13 (compromised phishing domains) have similar shapes due to a disproportionate concentration of compromised domains (for more details see Figure 5).

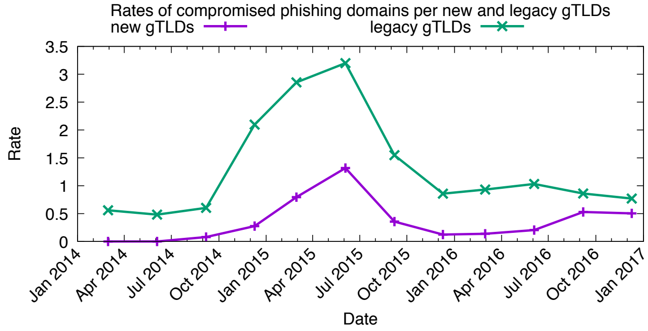


Figure 13. Time series of abuse rates of **compromised** phishing domains in **legacy** gTLDs and **new** gTLDs based on the Anti-Phishing Working Group feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#compromised\ domains/\#all\ domains$.

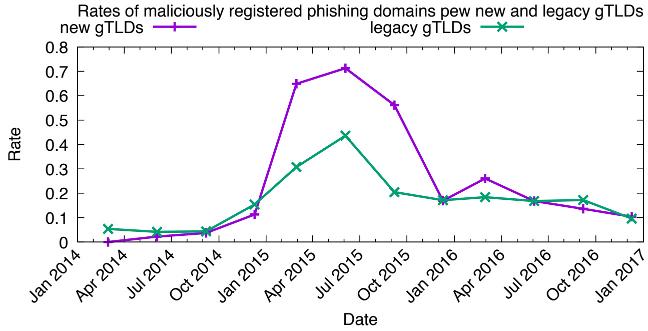


Figure 14. Time series of abuse rates of **maliciously** registered phishing domains in **legacy** gTLDs and **new** gTLDs based on the Anti-Phishing Working Group feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#maliciously\ registered\ domains/\#all\ domains$.

Figure 14 shows a time series of abuse rates of maliciously registered phishing domains in legacy and new gTLDs based on the APWG feed. The results indicate that, in relative terms, cybercriminals do not really have a preference between new and old gTLDs. We observe, however, relatively higher rates of maliciously registered new gTLD domains in the first three quarters of 2015. By manual analysis of malicious domains blacklisted in the third quarter of 2015, we find 3,542 domains registered in 53 gTLDs. The majority are .com domains (63%). We find 423 abused new gTLD domains. Interestingly, we observe as many as 175 and 88 abused .work and .xyz domains, respectively. The results indicates that the majority of .work domains were registered by the same person. 139 domains were registered on the same day and the names were composed of similar strings. Note that only 139 abused domains, blacklisted in the third quarter of 2015, influenced significantly the security reputation of all new gTLDs (see Figure 14).

Moreover, miscreants often seem able to maliciously register strings containing trademarked terms. For example, by manual analysis of maliciously registered domains in the fourth quarter of 2015 we find as many as 56 abused .top domains. 48 out of 56 contain the following strings: apple, icloud, iphone, their combinations, or misspelled versions of these strings suggesting that they were all used in the same phishing campaign against users of products of Apple Inc.

*2) Malware reputation:* We now analyze the malware activity reported by the StopBadware DSP. We refer the reader

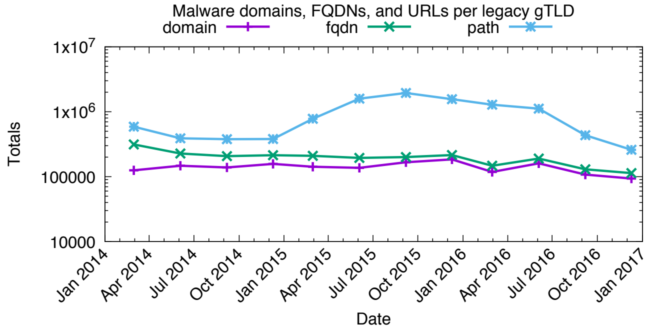


Figure 15. Time series of counts of malware domains, FQDNs, and URLs (paths) in **legacy** gTLD based on the StopBadware DSP feed (2014-2016).

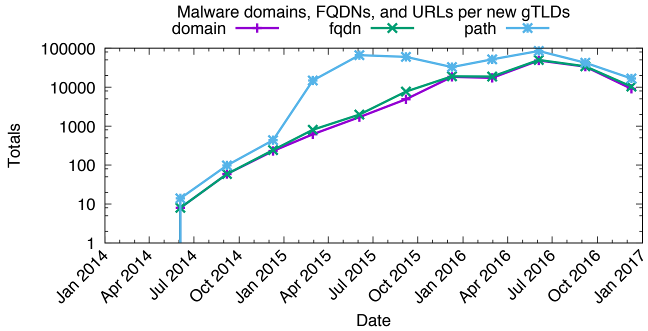


Figure 16. Time series of counts of malware domains, FQDNs, and URLs (paths) in **new** gTLD based on the StopBadware DSP feed (2014-2016).

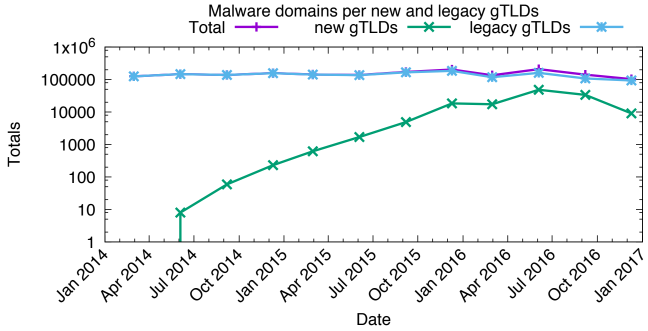


Figure 17. Time series of counts of malware domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the StopBadware DSP feed (2014-2016).

to Figure 15 and Figure 16 for overall absolute occurrence security metrics (see also Figure 47 and Figure 48 for the corresponding CleanMX malware datasets). More specifically, we present time series of counts of domains, FQDNs, and URLs (paths) of legacy gTLDs and new gTLD, respectively, aggregated on a quarterly basis. Y-axis are expressed in a logarithmic scale. Similarly to phishing, we observe a significant difference between the three occurrence metrics, especially between concentrations of URLs and the other two security metrics (domains and FQDNs).

From this point forward, we only consider the number of unique domains. Figure 17 presents a time series of counts of malware domains in legacy gTLD, new gTLDs, and all gTLDs (Total) based on the StopBadware feed between 2014 and 2016. Similar to phishing, the total number of malware incidents in all gTLDs is mainly driven by incidents in legacy gTLDs (88.6%). Again, in legacy gTLDs the number of abused domains remains approximately constant, whereas there is an upward trend in the absolute number of malware domains in new gTLDs. Figure 33 and Figure 49 presenting malware domains in legacy and new gTLDs for SURBL mw and CleanMX malware datasets confirm this trend.

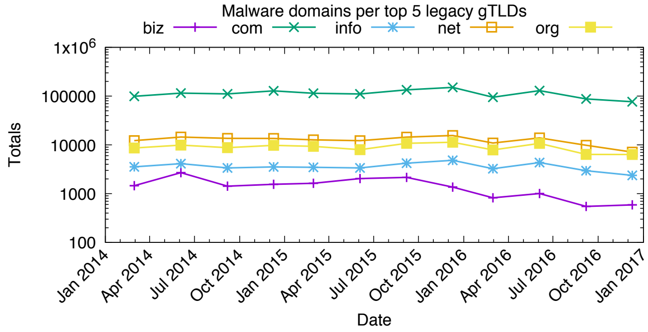


Figure 18. Time series of counts of malware domains in the top most abused 5 **legacy** gTLDs in the last quarter of 2016 based on the StopBadware DSP feed (2014-2016).

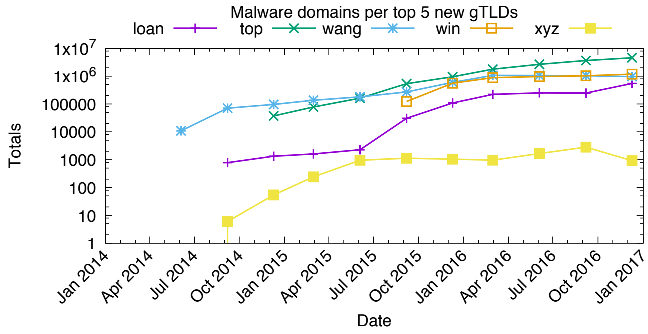


Figure 19. Time series of counts of malware domains in the top 5 most abused **new** gTLDs in the last quarter of 2016 based on the StopBadware DSP feed (2014-2016).

Figure 18 and Figure 19 show the top 5 most abused legacy gTLDs and new gTLDs with the highest absolute number of unique malware domains at the end of 2016, respectively. As the majority of domains are compromised rather than maliciously registered (see Figure 6), the distribution of malware by legacy gTLDs has very similar gTLD market share. The top 5 legacy gTLDs in terms of phishing and malware domains are the same. While the .xyz TLD is the largest new gTLD (see Figure 4), the absolute and therefore relative number of domains listed in blacklists is much lower in comparison to other new gTLDs depicted in Figure 19. Specifically, in the fourth quarter of 2016, the relative score of the .xyz TLD is equal to 1.5 malware domain per 10,000 domains. For comparison, the relative score of the .top gTLD (which in absolute terms consistently suffers from the highest concentration of blacklisted malware domains since the fourth quarter of 2015) is equal to 8.4.

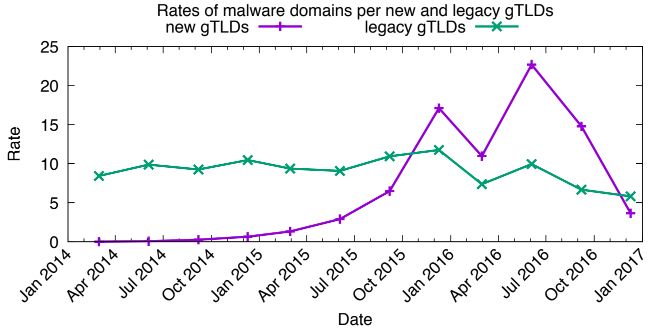


Figure 20. Time series of abuse rates of **malware domains** in legacy gTLDs and new gTLDs based on the StopBadware DSP feed (2014-2016). Rates are calculated as follows: $S = 10,000 * (\#blacklisted\ domains/ \#all\ domains)$.

We now account for gTLD sizes and plot a time series of abuse rates of malware domains in legacy and new gTLDs based on the StopBadware feed (see Figure 20). As before, the abuse rates are presented in a linear scale. Interestingly, between the second quarter of 2014 and the first quarter of 2016, we observe an exponential growth of abuse rates in the new gTLDs. In the second quarter of 2016 the difference between malware abuse rates in legacy and new gTLDs is the most significant. While legacy gTLDs collectively had a malware-domains-per-10,000 rate of 9.9, the new gTLDs experienced a rate of 22.7. In absolute terms, malware domains in new gTLDs constitute 23% of all gTLD domains blacklisted by StopBadware in that period. SURBL and CleanMX malware datasets confirm the upward trend in terms of the malware-domains-per-10,000 rates in new gTLDs in comparison to legacy gTLDs. We refer the reader to Figure 34 and Figure 50.

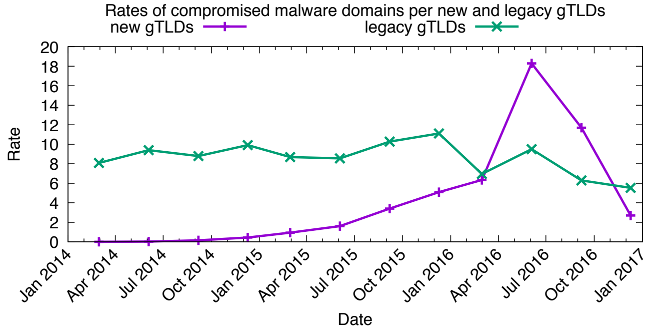


Figure 21. Time series of abuse rates of **compromised malware domains** in legacy gTLDs and new gTLDs based on the StopBadware DSP feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#compromised\ domains/\#all\ domains$.

In our descriptive analysis, we will now differentiate between maliciously registered and compromised domains to further make an attempt to distill factors that drive higher abuse rates in new gTLDs. Figure 21 and Figure 22 show time series of abuse rates of compromised and maliciously registered malware domains, respectively, in legacy gTLDs and new gTLDs. As expected, malware abuse rates in legacy gTLDs are mainly driven by compromised domains. More interestingly, the results suggest that the attackers apply more diverse methods to abuse domains in new gTLDs. For example, in the second quarter of 2016, the malware rates are driven by compromised domains (compare Figure 21 and Figure 20), whereas in the last quarter of 2015 by maliciously registered domains (compare Figure 22 and Figure 20). Nevertheless, manual analysis reveals something different. The spike in malware rates in new gTLDs in the last quarter of 2015 can indeed be explained by an increased number of malicious registrations. Specifically, we found that 7,868 out of 12,805 domains (61.4%) were registered in the .win gTLD and blacklisted within a very short time. However, our manual analysis of new gTLD domains in the second quarter of 2016 provides evidence that those domains were, in fact, maliciously registered rather than compromised. First, we found that the overwhelming majority of malware domains, which were categorized as compromised, belong to one of four new gTLDs: .win, .loan, .top, and .link (74%, which represents 28,801 out of 38,940 domains). Note that their distribution does not correspond to the new gTLD market share. Second, we find common patterns in domain names. Third, the registries of those four new gTLDs compete on price, and in the second quarter of 2016 their registration prices were below $1, which was lower than the registration fee for a .com domain. Therefore, we conclude that those domains were either registered by the attacker(s) earlier for later use or blacklisted after several weeks of being used for malicious purposes.

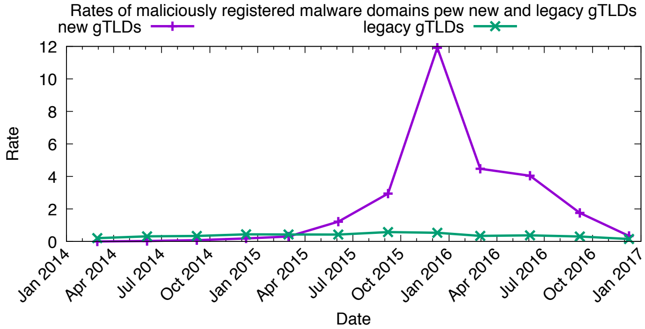


Figure 22. Time series of abuse rates of **maliciously registered malware domains** in legacy gTLDs and new gTLDs based on the StopBadware DSP feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#maliciously\ registered\ domains/\#all\ domains$.

### B. Privacy or Proxy services

In this section we present the results of an analysis to determine if there is a difference in the usage of `WHOIS` privacy or proxy services for abused domains in legacy gTLDs and new gTLDs. WHOIS privacy or proxy services are designed to hide the actual owner of a domain name, in practice this works by replacing the registrant information in WHOIS with the information of the WHOIS privacy or proxy service.

There are many legitimate reasons why someone may want to hide being the owner of a domain name. The usage of a WHOIS privacy or proxy services by itself is, therefore not a reliable single indicator of malicious activity. A previous study by National Physical Laboratories [41], however did find that a significant portion of abusive domains use privacy or proxy services.

There are numerous WHOIS privacy or proxy services available, which can be used by domain owners. To identify the most commonly used services we used the following methodology.

1) Using the WHOIS data, we aggregated all distinct domains by "registrant name" and "registrant organization" attributes and created a list with the top 5,000 registrants.
2) A keyword search on the top 5,000 "registrant name" and "registrant organization" attributes, trying to match any registrant with keywords such as: "privacy", "proxy", "protect", "private", "whois" etc.
3) A manual inspection of the suspect "registrant name" and "registrant organization" attributes to decide if the registrant is a privacy or proxy service, when this is not immediately clear from the name itself we use an internet search to find additional information.

Using the above described method we identified 570 "registrant name" and "registrant organizations" attribute combinations used by WHOIS privacy or proxy services.

Each blacklist abuse incident contains metadata such as the date when the domain was added to the blacklist, we used this date to identify the correct historical WHOIS record for an abused domain. By comparing the "registrant name" and "registrant organization" attributes from the domain WHOIS record to the list of known WHOIS privacy or proxy services, we are able to correctly identify abusive domains that were using a WHOIS privacy or proxy service at the time the domain was added to a blacklist.

To get an indication of how common WHOIS privacy or proxy service usage is, we aggregated all domains from the WHOIS data by their create date. This shows us the number of newly added domains per month for legacy and new gTLDs. After checking how many of these domains were using a privacy or proxy service when the domain was registered, we calculated what percentage of the total number of newly registered domains is using a privacy or proxy service (see Figure 23). We find that the for legacy gTLDs the usage is stable with a mean of 24%, and a standard deviation of 1.6. For new gTLDs the usage is generally below that of legacy gTLDs with a mean of 19% and a standard deviation of 9.6, which is visualized by the larger spikes and the increase to above the level of legacy gTLDs near the end of the study period.

For each blacklist used in this study we analysed the proportion of domains that were using a privacy or proxy service at the time the domain was found to be abusive and included in the blacklist. Here again we make a distinction between legacy and new gTLD domains.

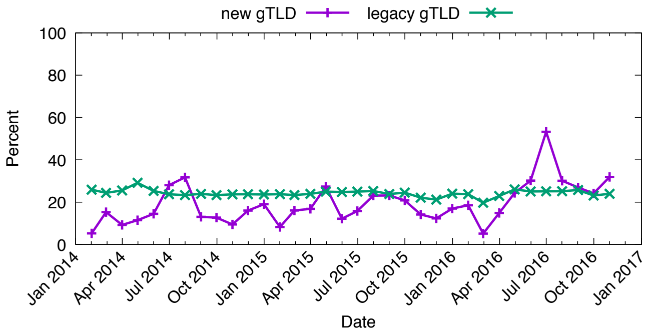


Figure 23. Usage percentage of privacy or proxy services for newly registered domains

When look at two blacklist mainly driven by maliciously registered domains, all SURBL feeds combined (see Figure 24) and Spamhaus (see Figure 25), we find that the usage of privacy or proxy services has been increasing from the start of the New gTLD Program and reached the same level of usage in late 2015. In 2016 the usage for new gTLDs has mainly followed the same pattern, but at a lower level, as is seen for legacy gTLDs.

In 2016 the mean usage per month of privacy and proxy services by abusive domains in new gTLD observed is 4,649 with a standard deviation of 1,872 (see Figure 24), while for legacy gTLDs the mean usage per month is 19,544 with a standard deviation of 9,788. For Spamhaus (see Figure 25) the 2016 new gTLDs mean usage per month is 6,758 with a standard deviation of 1,597, while for legacy gTLDs the mean usage per month is 15,453 with a standard deviation of 4,227.

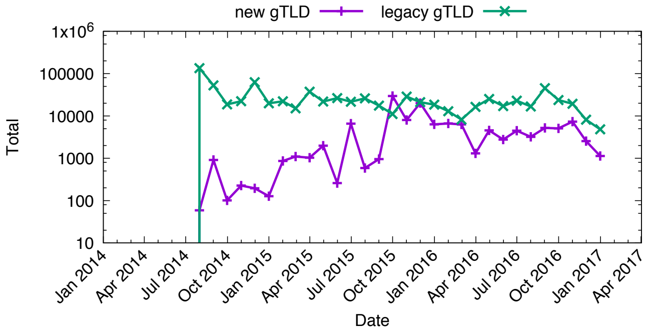


Figure 24. Usage of privacy or proxy services for abusive domains, reported by SURBL

The APWG feed is mainly composed of compromised domains, see Figure 5. The same is true for the StopBadware DSP feed, see Figure 6. The first abused new gTLD domains are reported by APWG starting in October 2014, which is later than for the other blacklist, although it is not clear why.

When we investigate the use of privacy or proxy services for abused domains in 2016 reported by APWG (Figure 5) we find that, for new gTLDs, APWG has a mean usage of 69 with a standard deviation of 41, for legacy gTLDs this is a mean usage of 831 with a standard deviation of 177. This results in a factor 12 difference between legacy and new gTLD domains

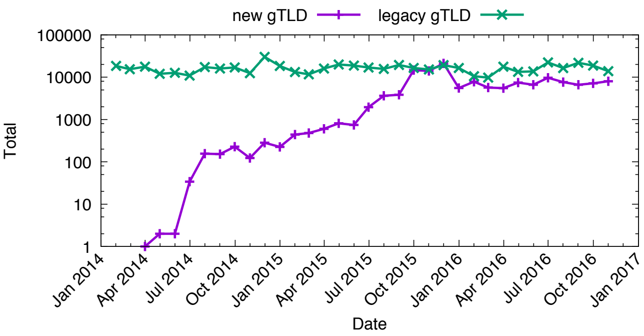


Figure 25. Usage of privacy or proxy services for abusive domains, reported by Spamhaus

for APWG, Overall APWG contains 14 times more legacy domains than new gTLD domains, which is an indication that the usage of privacy or proxy services for abusive domains reported by APWG is not excessive.

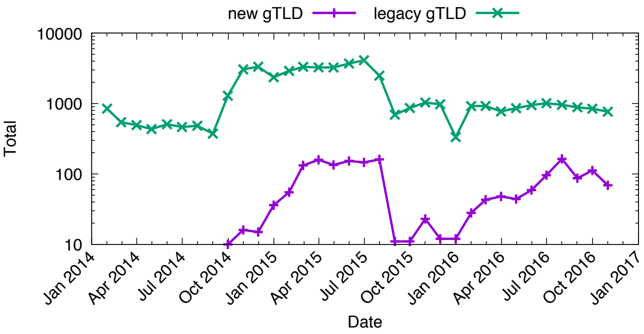


Figure 26. Usage of privacy or proxy services for abusive domains, reported by APWG

When we investigate the use of privacy or proxy services for abused domains in 2016 reported by StopBadware (Figure 6) we find that, for new gTLDs, StopBadware shows a mean usage of 314 with a standard deviation of 142, for legacy gTLDs this is a mean usage of 4622 with a standard deviation of 1846. This results in a factor 15 difference between legacy and new gTLD domains for StopBadware, overall StopBadware contains 3 times more legacy domains as new gTLD domains, which is an indication that the usage of privacy or proxy services for abusive domains reported by StopBadware is lower than expected.

### C. Geographic region

For each blacklist we present a comparison of the geographical locations of abused domains to determine if there is a difference in the location of abuse between legacy gTLD and new gTLD domain locations. To determine the geographical location of an abused domain we are using the address of the domain's sponsoring registrar. Table IV lists the 10 countries hosting most registrars, almost 54% of the identified registrars are located in the United States, which is almost 1 order of magnitude more than the number of registrars located in the next country, China. With such a high proportion of registrars located in a single country, the general hypothesis is that most of the abused domains will probably also be located in this country. We find that while this is true for legacy gTLDs, for new gTLDs however there are a number of cases where this is not the case. For example, when we take look at the new gTLD countries for StopBadware and SURBL in Table VI and Table VII, we find that the United States occupies the 3rd and 4th place.

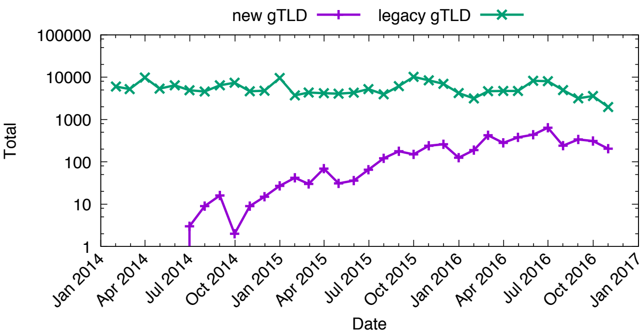


Figure 27. Usage of privacy or proxy services for abusive domains, reported by StopBadware

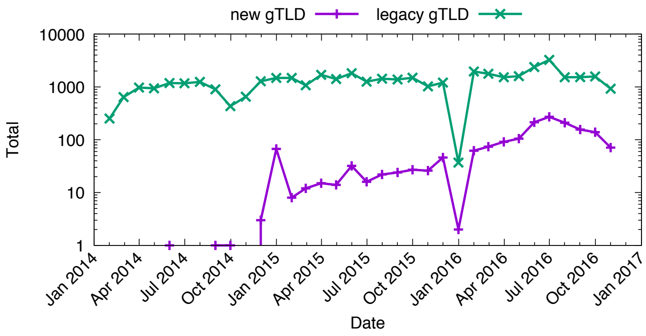


Figure 28. Usage of privacy or proxy services for abusive domains, reported by CleanMX Phishing

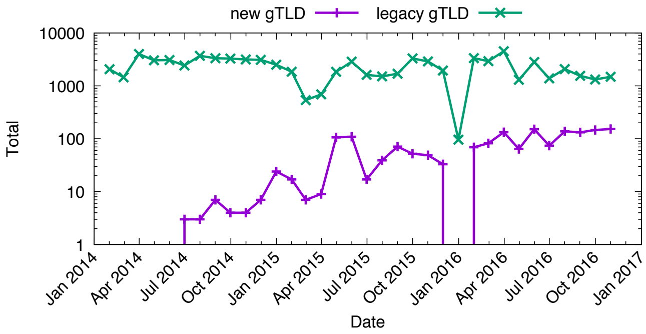


Figure 29. Usage of privacy or proxy services for abusive domains, reported by CleanMX Viruses

Although the majority of the registrars is located in the United States, the story might be different when we look at the number of registered domains. There are a small number of very large registrars and many smaller registrars and. These registrars are not uniformly distributed across countries,

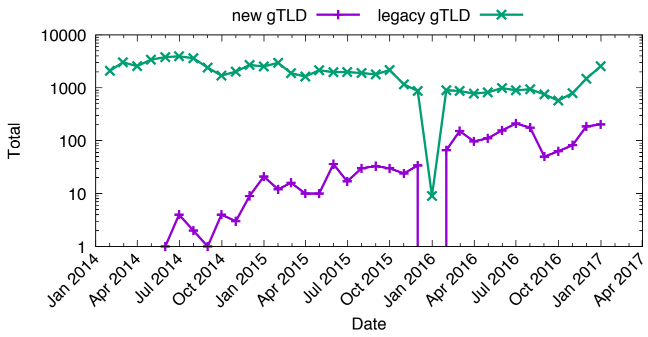


Figure 30. Usage of privacy or proxy services for abusive domains, reported by CleanMX Portals

Table IV
Top10 registrar countries

| Country | #Registrars | share |
|---|---|---|
| United States | 2,682 | 53.88 |
| China | 281 | 5.64 |
| Germany | 201 | 4.04 |
| Canada | 177 | 3.56 |
| United Kingdom | 160 | 3.21 |
| India | 144 | 2.89 |
| France | 116 | 2.33 |
| Australia | 111 | 2.23 |
| Spain | 105 | 2.11 |
| Japan | 95 | 1.91 |

Table V
Top10 locations of new and legacy gTLD domains

| New | #Domains | Share | Legacy | #Domains | Share |
|---|---|---|---|---|---|
| China | 7,832,264 | 28.57 | USA | 145,652,390 | 58.81 |
| USA | 6,114,944 | 22.31 | China | 22,409,117 | 9.05 |
| Gibraltar | 2,603,236 | 9.5 | Germany | 16,574,944 | 6.69 |
| Cayman Islands | 1,959,580 | 7.15 | Canada | 14,198,455 | 5.73 |
| Singapore | 1,700,985 | 6.2 | India | 9,509,405 | 3.84 |
| Japan | 1,667,079 | 6.08 | Japan | 6,400,530 | 2.58 |
| India | 1,274,622 | 4.65 | Australia | 5,950,392 | 2.4 |
| Germany | 1,056,541 | 3.85 | France | 4,573,133 | 1.85 |
| Hong Kong | 815,039 | 2.97 | UK | 3,670,192 | 1.48 |
| Canada | 422,834 | 1.54 | Turkey | 2,216,396 | 0.89 |

meaning that a relatively small number of larger registrars located outside of the US, may skew results to show many domains registered outside the US. Table V lists the countries where most of the legacy and new gTLD domains are located. The number or registered legacy gTLD domains per country is heavily influenced by the distribution of registrars across countries. The top countries are an exact match. For legacy gTLDs the major player is the Unites States, with 6,5 times more domains compared to number 2, China. For new gTLDs however, we find that the country distribution has changed most new gTLD domains are now located in China followed by the US and Gibraltar. The difference between the top countries is less extreme for new gTLDs than it is for legacy gTLDs.

The WHOIS data used for this study contains a "registrar name" attribute for each domain record, however there is no geographical information for the registrar available in the WHOIS data. To map each registrar to a geographical location we used the following method:

1) Extract every unique "registrar name" attribute from the WHOIS data.
2) Using an automated process combine the extracted "registrar name" attribute with the country information for ICANN-Accredited Registrars, available from the ICANN website [42].
3) Manually match remaining name variants (the automated process is not able to match every registrar name variant to a country) to their corresponding countries.
4) Manually lookup the country information for registrars that could not be found automatically (not every registrar is accredited by ICANN) using publicly available information from the corporate website of the registrar or domain industry websites [43].

This method resulted in a list containing 5,985 registrars (and name variants) with their geographical location. Together these registrars manage over 99.99% of all the domains found in the WHOIS data.

For each blacklist we calculated two abuse metrics, the "percentage" and "rate". The "percentage" is used to indicate the proportion of the total number of abused domains from a blacklist that can be attributed to a country. The "rate" is the ratio between the number of legacy or new gTLDs in the blacklist attributed to a country multiplied by 10,000, and divided by the total number of domains managed by registrars located in that country. For example, Table VI shows that for 43.74% of the abused new gTLD domains reported by StopBadware, the sponsoring registrar is located in Gibraltar. Almost 218 abused new gTLD domains per 10,000 located in Gibraltar are abusive.

The results in Table VI, Table VII and Table VIII all show a high amount of abuse for Gibraltar. When we investigate why Gibraltar has such a high number of abused new gTLD domains, we find that the abuse is driven by a single registrar: Alpnames Limited. For example, during the study period this registrar has acted as the sponsoring registrar for 43.74% (57,141) of the new gTLD domains that have been blacklisted by StopBadware. Moreover, note that for new gTLDs, the spam-domains-per-10,000 rate reported by Spamhaus for Gibraltar is equal to 3,122 (Table VIII), whereas for example for APWG only 0.47 (Table IX). This is mainly because Spamhaus and APWG capture different attackers' dynamics and therefore give a very complementary view of domain abuse. While the majority of URLs blacklisted by APWG represent hacked domains registered by legitimate users, the Spamhaus domain blacklist is composed of domains used purely for malicious purposes.

### D. Registrar reputation

Here we present the distribution of abused domains across ICANN accredited registrars. In subsection IV-C we show that domains listed in blacklists are predominantly compromised rather than maliciously registered. We assume that

Table VI
StopBadware top10 Legacy gTLD and new gTLD rate between all domains listed in blacklist and both the blacklist (percentage) and registrar country (rate) total number of domains.

| # | new gTLD Country | #Incidents | percentage | rate | Legacy gTLD country | #Incidents | percentage | rate |
|---|---|---|---|---|---|---|---|---|
| 1 | Gibraltar | 57,141 | 43.74 | 217.81 | United States | 575,839 | 51.44 | 37.79 |
| 2 | China | 54,684 | 41.86 | 208.44 | China | 198,315 | 17.72 | 13.01 |
| 3 | United States | 7,354 | 5.63 | 28.03 | India | 85,205 | 7.61 | 5.59 |
| 4 | India | 5,401 | 4.13 | 20.59 | Canada | 49,192 | 4.39 | 3.23 |
| 5 | Singapore | 1,492 | 1.14 | 5.69 | Germany | 45,959 | 4.11 | 3.02 |
| 6 | United Kingdom | 858 | 0.66 | 3.27 | France | 20,699 | 1.85 | 1.36 |
| 7 | Hong Kong | 829 | 0.63 | 3.16 | United Kingdom | 16,093 | 1.44 | 1.06 |
| 8 | Barbados | 468 | 0.36 | 1.78 | Spain | 14,309 | 1.28 | 0.94 |
| 9 | France | 413 | 0.32 | 1.57 | Turkey | 14,253 | 1.27 | 0.94 |
| 10 | Germany | 375 | 0.29 | 1.43 | Hong Kong | 12,666 | 1.13 | 0.83 |

Table VII
SURBL top10 Legacy gTLD and new gTLD ratio between all domains listed in blacklist and both the blacklist (percentage) and registrar country (rate) total number of domains.

| # | new gTLD Country | #Incidents | percentage | rate | Legacy gTLD country | #Incidents | percentage | rate |
|---|---|---|---|---|---|---|---|---|
| 1 | Gibraltar | 585,839 | 47.4 | 2233.07 | United States | 1,893,528 | 47.87 | 124.27 |
| 2 | Japan | 249,426 | 20.18 | 950.75 | Japan | 1,074,165 | 27.15 | 70.49 |
| 3 | China | 201,869 | 16.33 | 769.47 | China | 312,560 | 7.9 | 20.51 |
| 4 | United States | 87,139 | 7.05 | 332.15 | India | 243,127 | 6.15 | 15.96 |
| 5 | India | 45,059 | 3.65 | 171.75 | Germany | 66,075 | 1.67 | 4.34 |
| 6 | United Kingdom | 19,775 | 1.6 | 75.38 | Ireland | 58,226 | 1.47 | 3.82 |
| 7 | United Arab Emirates | 11,746 | 0.95 | 44.77 | Canada | 37,861 | 0.96 | 2.48 |
| 8 | Canada | 6,110 | 0.49 | 23.29 | Turkey | 32,222 | 0.81 | 2.11 |
| 9 | France | 6,073 | 0.49 | 23.15 | Australia | 30,870 | 0.78 | 2.03 |
| 10 | Australia | 5,852 | 0.47 | 22.31 | Bahamas | 28,762 | 0.73 | 1.89 |

Table VIII
Spamhaus top10 Legacy gTLD and new gTLD rate between all domains listed in blacklist and both the blacklist (percentage) and registrar country (rate) total number of domains.

| # | new gTLD Country | #Incidents | percentage | rate | Legacy gTLD country | #Incidents | percentage | rate |
|---|---|---|---|---|---|---|---|---|
| 1 | Gibraltar | 819,097 | 53.35 | 3122.19 | United States | 2,004,414 | 47.65 | 131.54 |
| 2 | Japan | 22,0144 | 14.34 | 839.13 | Japan | 1,059,177 | 25.18 | 69.51 |
| 3 | United States | 170,781 | 11.12 | 650.97 | China | 349,610 | 8.31 | 22.94 |
| 4 | China | 169,239 | 11.02 | 645.1 | India | 257,244 | 6.12 | 16.88 |
| 5 | India | 48,518 | 3.16 | 184.94 | Turkey | 91,019 | 2.16 | 5.97 |
| 6 | Singapore | 30,743 | 2.0 | 117.18 | Bahamas | 72,904 | 1.73 | 4.78 |
| 7 | United Kingdom | 19,725 | 1.28 | 75.19 | Germany | 68,506 | 1.63 | 4.5 |
| 8 | Cayman Islands | 13,671 | 0.89 | 52.11 | Canada | 68,463 | 1.63 | 4.49 |
| 9 | France | 11,384 | 0.74 | 43.39 | Australia | 37,996 | 0.9 | 2.49 |
| 10 | Australia | 5,410 | 0.35 | 20.62 | United Kingdom | 28,317 | 0.67 | 1.86 |

Table IX
APWG top10 Legacy gTLD and new gTLD rate between all domains listed in blacklist and both the blacklist (percentage) and registrar country (rate) total number of domains.

| # | new gTLD Country | #Incidents | percentage | rate | Legacy gTLD country | #Incidents | percentage | rate |
|---|---|---|---|---|---|---|---|---|
| 1 | United States | 2,610 | 36.51 | 4.19 | United States | 148,545 | 61.54 | 9.75 |
| 2 | China | 2,127 | 29.76 | 3.42 | India | 25,863 | 10.72 | 1.7 |
| 3 | India | 320 | 4.48 | 0.51 | Canada | 12,036 | 4.99 | 0.79 |
| 4 | Gibraltar | 292 | 4.09 | 0.47 | Germany | 8,366 | 3.47 | 0.55 |
| 5 | Germany | 289 | 4.04 | 0.46 | China | 7,052 | 2.92 | 0.46 |
| 6 | Singapore | 207 | 2.9 | 0.33 | Australia | 4,991 | 2.07 | 0.33 |
| 7 | Japan | 200 | 2.8 | 0.32 | United Kingdom | 4,689 | 1.94 | 0.31 |
| 8 | United Kingdom | 163 | 2.28 | 0.26 | France | 3,966 | 1.64 | 0.26 |
| 9 | Turkey | 162 | 2.27 | 0.26 | Turkey | 3,852 | 1.6 | 0.25 |
| 10 | Canada | 155 | 2.17 | 0.25 | Bahamas | 2,498 | 1.03 | 0.16 |

the miscreants responsible for compromising domains have automated scanners to analyze web based software for known vulnerabilities at scale. When a vulnerable domain is detected, it is compromised regardless of the TLD or registrar.

Table X
CleanMX Phishing top10 Legacy gTLD and new gTLD rate between all domains listed in blacklist and both the blacklist (percentage) and registrar country (rate) total number of domains.

| # | new gTLD Country | #Incidents | percentage | rate | Legacy gTLD country | #Incidents | percentage | rate |
|---|---|---|---|---|---|---|---|---|
| 1 | United States | 3,638 | 55.2 | 5.84 | United States | 163,173 | 60.78 | 10.71 |
| 2 | Gibraltar | 691 | 10.48 | 1.11 | India | 32,760 | 12.2 | 2.15 |
| 3 | China | 594 | 9.01 | 0.95 | Canada | 16,542 | 6.16 | 1.09 |
| 4 | India | 533 | 8.09 | 0.86 | Germany | 8,636 | 3.22 | 0.57 |
| 5 | United Kingdom | 140 | 2.12 | 0.22 | China | 6,797 | 2.53 | 0.45 |
| 6 | Canada | 135 | 2.05 | 0.22 | Australia | 6,282 | 2.34 | 0.41 |
| 7 | Germany | 132 | 2.0 | 0.21 | United Kingdom | 4,858 | 1.81 | 0.32 |
| 8 | Singapore | 104 | 1.58 | 0.17 | France | 4,193 | 1.56 | 0.28 |
| 9 | France | 78 | 1.18 | 0.13 | Turkey | 3,871 | 1.44 | 0.25 |
| 10 | Cayman Islands | 71 | 1.08 | 0.11 | Bahamas | 2,157 | 0.8 | 0.14 |

Table XI
CleanMX Portals top10 Legacy gTLD and new gTLD rate between all domains listed in blacklist and both the blacklist (percentage) and registrar country (rate) total number of domains.

| # | new gTLD Country | #Incidents | percentage | rate | Legacy gTLD country | #Incidents | percentage | rate |
|---|---|---|---|---|---|---|---|---|
| 1 | United States | 2,658 | 45.34 | 4.27 | United States | 213,877 | 57.57 | 14.04 |
| 2 | India | 840 | 14.33 | 1.35 | India | 47,377 | 12.75 | 3.11 |
| 3 | Gibraltar | 786 | 13.41 | 1.26 | Canada | 21,551 | 5.8 | 1.41 |
| 4 | China | 615 | 10.49 | 0.99 | China | 19,738 | 5.31 | 1.3 |
| 5 | France | 225 | 3.84 | 0.36 | Germany | 11,930 | 3.21 | 0.78 |
| 6 | United Kingdom | 101 | 1.72 | 0.16 | France | 6,915 | 1.86 | 0.45 |
| 7 | Cayman Islands | 86 | 1.47 | 0.14 | Turkey | 6,708 | 1.81 | 0.44 |
| 8 | Germany | 86 | 1.47 | 0.14 | United Kingdom | 6,097 | 1.64 | 0.4 |
| 9 | Russian Federation | 78 | 1.33 | 0.13 | Australia | 5,457 | 1.47 | 0.36 |
| 10 | Singapore | 69 | 1.18 | 0.11 | Spain | 3,829 | 1.03 | 0.25 |

Table XII
CleanMX Viruses top10 Legacy gTLD and new gTLD rate between all domains listed in blacklist and both the blacklist (percentage) and registrar country (rate) total number of domains.

| # | new gTLD Country | #Incidents | percentage | rate | Legacy gTLD country | #Incidents | percentage | rate |
|---|---|---|---|---|---|---|---|---|
| 1 | China | 3,117 | 34.56 | 3.89 | United States | 209,255 | 55.44 | 13.73 |
| 2 | United States | 2,477 | 27.47 | 3.09 | China | 41,315 | 10.95 | 2.71 |
| 3 | Gibraltar | 1,807 | 20.04 | 2.26 | India | 30,435 | 8.06 | 2.0 |
| 4 | India | 340 | 3.77 | 0.42 | Canada | 17,083 | 4.53 | 1.12 |
| 5 | United Kingdom | 307 | 3.4 | 0.38 | Germany | 14,031 | 3.72 | 0.92 |
| 6 | Cayman Islands | 195 | 2.16 | 0.24 | France | 7,555 | 2.0 | 0.5 |
| 7 | Singapore | 136 | 1.51 | 0.17 | Spain | 6,337 | 1.68 | 0.42 |
| 8 | France | 109 | 1.21 | 0.14 | Turkey | 5,905 | 1.56 | 0.39 |
| 9 | Germany | 108 | 1.2 | 0.13 | United Kingdom | 5,717 | 1.51 | 0.38 |
| 10 | Japan | 108 | 1.2 | 0.13 | Japan | 5,089 | 1.35 | 0.33 |

For each registrar we find how many (#Incidents) can be attributed to the registrar and the total number of domains sponsored by that registrar (#Domains). We than calculate what proportion (Percentage) of all domains managed by the registrar is reported as abusive by a blacklist. An outlier with a relatively high rate compared to its peers may be caused by registrar-specific policies or operational practices.

Note, sinkholing of confiscated abusive domains or preventive registration of botnet C&C infrastructure domains is a common practice and special registrars have been created for this purpose e.g. "Afilias Special Projects" or "Verisign Security and Stability". These registrars have high numbers of abuse and have been filtered out during the analysis because they are not regular registrars.

This section contains a table for each blacklist and the sponsoring registrars with most abusive new gTLD and legacy gTLD domains (#Domains). For reach registrar the total number of abused domains (#Incidents) reported by the blacklist and the proportion (Percent) of the registrar portfolio reported by the blacklist. For Example, Table XIII lists the number reported incidents for "Nanjing Imperiosus Technology" as 25,991, with a total number of 26,096 under its management, this 99.6% of all new gTLDs of this registrar are reported by the SURBL blacklist.

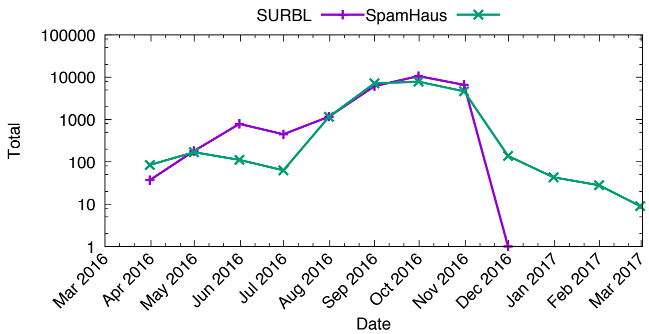


Figure 31. Abusive domains managed by Nanjing Imperiosus Technology Co. Ltd

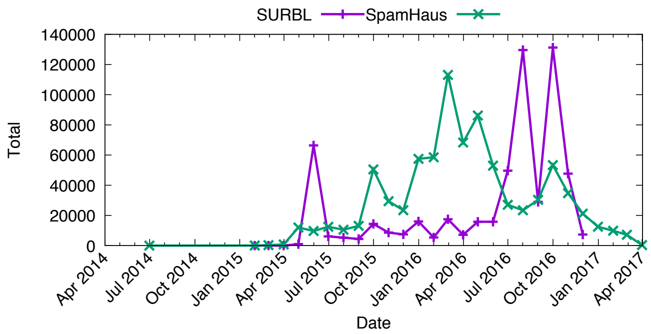


Figure 32. Abusive domains managed by Alpnames Limited

Table XIII and Table XIV list the registar "Nanjing Imperiosus Technology Co. Ltd." as an outlier, almost 100% of its domains are reported as abusive by SURBL and 82% by SpamHaus. Figure 31 shows that both blacklists have marked domains managed by this registrar as abusive starting from early 2016. Starting from November 2016 we see a sharp decline in domains reported by SpamHaus and SURBL has not reported any new abused domains after November 2016 at all. This can be explained by the fact that ICANN has terminated the registrar accreditation [44] for this registrar, as it was determined that the registrar was in breach of the Registrar Accreditation Agreement (RAA). Termination of the RAA had an immediate and dramatic effect on the amount of abuse linked to this registrar.

Figure 32 shows one registrar, Alpnames Limited, having a high volume of abusive new gTLD domains reported by both Spamhaus and SURBL. The SURBL feed shows 2 distinctive peaks with a high number of abuse reports in 2016. After more detailed analysis, we find that these peaks correspond with 103,758 reports of abusive domains in the .top gTLD in August 2016. In October 2016 we find a second peak, which is caused by 120,669 reports of abusive domains in the .science gTLD. This registrar is known for its very low pricing or giving domains away for free. In 2016 it did have promotions for domains using the .science gTLD for $1 or less. We did not find corresponding increases in the size of the .top and .science zone files, indicating the abusive domains have been registered over a longer period of time.

## References

[1] "Internet Corporation for Assigned Names and Numbers (ICANN)," https://www.icann.org.

[2] B. Rechterman and T. Ruiz, "Method for Gathering Domain Name Registration Information From a Registrant via a Registrar's Web Site," 2004, US Patent App. 10/408,050. [Online]. Available: http://www.google.com/patents/US20040199608

[3] "Top-Level Domains (gTLDs)," http://archive.icann.org/en/tlds.

[4] J. Postel and J. Reynolds, "Domain requirements," Internet Requests for Comments, RFC Editor, RFC 920, October 1984.

[5] ICANN, "New gTLD Program," https://icannwiki.com/New_gTLD_Program, 2017.

[6] T. Halvorson, M. F. Der, I. Foster, S. Savage, L. K. Saul, and G. M. Voelker, "From .Academy to .Zone: An Analysis of the New TLD Land Rush," in *Proceedings of the 2015 ACM Conference on Internet Measurement Conference*, ser. IMC'15. ACM, 2015, pp. 381–394.

[7] ICANN, ".madrid," https://icannwiki.org/.madrid, March 2015.

[8] ——, "New gTLD Program," https://icannwiki.org/New_gTLD_Program, February 2017.

[9] ——, "New gTLD Program," https://icannwiki.org/New_gTLD_Generic_Applications, February 2017.

[10] ——, "New gTLD Program," https://icannwiki.org/Community_TLD, February 2017.

[11] ——, "New gTLD Program," https://icannwiki.org/New_gTLD_Geographic_Applications, February 2017.

[12] ——, "New gTLD Program," https://icannwiki.org/New_gTLD_Brand_Applications, February 2017.

[13] "The Spamhaus Project," www.spamhaus.org.

[14] "Anti-Phishing Working Group (APWG): Cross-industry Global Group Supporting Tackling the Phishing Menace," http://www.antiphishing.org.

[15] "StopBadware: A Nonprofit Anti-malware Organization." https://www.stopbadware.org.

[16] "SURBL - URI reputation data," http://www.surbl.org.

[17] "Spam-Filter Anti-Spam Virenschutz," http://clean-mx.de.

[18] "The Domain Block List," https://www.spamhaus.org/dbl.

[19] "ESET: Security Software," http://www.eset.com.

[20] "Fortinet: Network & Computer Security," http://www.fortinet.com.

[21] "Sophos: Computer Security, Antivirus," http://www.sophos.com.

[22] "StopBadware: Data Sharing Program," https://www.stopbadware.org/data-sharing.

[23] "SURBL Lists," http://www.surbl.org/lists.

[24] "Public Suffix List," https://publicsuffix.org.

[25] G. Aaron and R. Rasmussen, "Anti-Phishing Working Group (APWG) Global Phishing Survey: Trends and Domain Name Use in 1H2014," http://docs.apwg.org/reports/APWG_Global_Phishing_Report_1H_2014.pdf.

[26] ——, "Anti-Phishing Working Group (APWG) Global Phishing Survey: Trends and Domain Name Use in 2H2014," http://docs.apwg.org/reports/APWG_Global_Phishing_Report_2H_2014.pdf, 2015.

[27] Whois, "Whois XML API – Whois Lookup – Domain Name Search," https://www.whoisxmlapi.com, February 2017.

[28] ICANN, "TLD Startup Information," https://newgtlds.icann.org/en/program-status/sunrise-claims-periods, Retrieved on February 2017.

[29] "ICANN: .zuerich TLD," https://icannwiki.org/.zuerich.

[30] M. Korczyński, S. Tajalizadehkhoob, A. Noroozian, M. Wullink, C. Hesselman, and M. van Eeten, "Reputation metrics design to improve intermediary incentives for security of tlds," in *2017 IEEE European Symposium on Security and Privacy (Euro SP)*, April 2017.

[31] D. Plohmann, K. Yakdan, M. Klatt, J. Bader, and E. Gerhards-Padilla, "A Comprehensive Measurement Study of Domain Generating Malware," in *25th USENIX Security Symposium (USENIX Security 16)*. USENIX Association, Aug. 2016, pp. 263–278.

[32] "ZeusTracker: A Nonprofit Organization Tracking ZeuS C&C Servers." https://zeustracker.abuse.ch.

[33] A. Noroozian, M. Korczyński, S. Tajalizadehkhoob, and M. van Eeten, "Developing security reputation metrics for hosting providers," in *Proceedings of the 8th USENIX CSET*, 2015, pp. 1–8.

[34] M. Korczyński, S. Tajalizadehkhoob, A. Noroozian, M. Wullink, C. Hesselman, and M. van Eeten, "Reputation metrics design to improve intermediary incentives for security of tlds," in *2017 IEEE European Symposium on Security and Privacy (Euro SP)*, April 2017.

Table XIII
SURBL top10 percentage between blacklisted new and legacy gTLD domains (#Incidents) and total number of registrar gTLD domains (#Domains).

| # | new gTLD registrar | #Domains | #Incidents | Percent | Legacy gTLD registrar | #Domains | #Incidents | Percent |
|---|---|---|---|---|---|---|---|---|
| 1 | Nanjing Imperiosus Technology | 26,096 | 25,991 | 99.6 | HOAPDI INC. | 141 | 126 | 89.36 |
| 2 | Intracom Middle East FZE | 20,639 | 11,254 | 54.53 | asia registry r2-asia (700000) | 1,379 | 598 | 43.36 |
| 3 | Dot Holding Inc. | 153 | 76 | 49.67 | Nanjing Imperiosus Technology | 35,309 | 10,892 | 30.85 |
| 4 | Alpnames Limited | 2,623,443 | 585,839 | 22.33 | Paknic (Private) Limited | 10,512 | 3,081 | 29.31 |
| 5 | Todaynic.com, Inc. | 317,534 | 69,330 | 21.83 | Intracom Middle East FZE | 67 | 16 | 23.88 |
| 6 | Web Werks India d/b/a ZenRegistry.com | 784 | 146 | 18.62 | AFRIREGISTER S.A. | 1,540 | 266 | 17.27 |
| 7 | Xiamen Nawang Technology Co., Ltd | 281,148 | 42,067 | 14.96 | Minds and Machines LLC | 1,115 | 171 | 15.34 |
| 8 | GMO Internet d/b/a Onamae.com | 1,673,447 | 249,420 | 14.9 | OwnRegistrar, Inc. | 19,745 | 2,933 | 14.85 |
| 9 | TLD Registrar Solutions Ltd. | 148,915 | 19,542 | 13.12 | GMO Internet d/b/a Onamae.com | 7,171,201 | 1,061,902 | 14.81 |
| 10 | Instra Corporation Pty Ltd. | 76,079 | 5,814 | 7.64 | GoName.com, Inc | 2,662 | 384 | 14.43 |

Table XIV
Spamhaus top10 rate between blacklisted new and legacy gTLD domains (#Incidents) and total number of registrar gTLD domains (#Domains).

| # | new gTLD registrar | #Domains | #Incidents | Percent | Legacy gTLD registrar | #Domains | #Incidents | Percent |
|---|---|---|---|---|---|---|---|---|
| 1 | Nanjing Imperiosus Technology | 26,096 | 21435 | 82.14 | ABSYSTEMS dba yourname... | 688 | 632 | 91.86 |
| 2 | NameCentral, Inc. | 9 | 3 | 33.33 | Ednit Software Private Limited | 522 | 283 | 54.21 |
| 3 | Dot Holding Inc. | 153 | 50 | 32.68 | Dynamic Dolphin, Inc. | 12,515 | 5,870 | 46.9 |
| 4 | Shanghai Best Oray Information S&T | 3,357 | 1,081 | 32.2 | Webair Internet Development, Inc. | 19,599 | 7,483 | 38.18 |
| 5 | Alpnames Limited | 2,623,443 | 819,097 | 31.22 | asia registry r2-asia (700000) | 1,379 | 460 | 33.36 |
| 6 | NameSilo, LLC | 30,777 | 6,456 | 20.98 | Nanjing Imperiosus Technology | 35,309 | 11,487 | 32.53 |
| 7 | Zhengzhou Century Connect Electronic... | 15,558 | 2,737 | 17.59 | Eranet International Limited | 2,287 | 737 | 32.23 |
| 8 | Netowl , Inc. | 1,128 | 165 | 14.63 | GoName-TN.com, Inc. | 7,088 | 1,815 | 25.61 |
| 9 | GMO Internet d/b/a Onamae.com | 1,673,447 | 219967 | 13.14 | Paknic (Private) Limited | 10,512 | 2,545 | 24.21 |
| 10 | TLD Registrar Solutions Ltd. | 148,915 | 19,456 | 13.07 | Alpnames Limited | 25,597 | 5,807 | 22.69 |

Table XV
APWG top10 rate between blacklisted new and legacy gTLD domains (#Incidents) and total number of registrar gTLD domains (#Domains).

| # | new gTLD registrar | #Domains | #Incidents | Percent | Legacy gTLD registrar | #Domains | #Incidents | Percent |
|---|---|---|---|---|---|---|---|---|
| 1 | Key-Systems GmbH | 8,077 | 148 | 1.83 | Minds and Machines LLC | 1,115 | 117 | 10.49 |
| 2 | AB Name ISP | 1,069 | 9 | 0.84 | Tecnologia & Desarrollo Y Mercado... | 2,027 | 128 | 6.31 |
| 3 | Shenzhen HuLianXianFeng Technology | 6,115 | 19 | 0.31 | Abu-Ghazaleh Intellectual... | 1,365 | 27 | 1.98 |
| 4 | FBS Inc. | 56,340 | 162 | 0.29 | Rethem Hosting LLC | 3,840 | 62 | 1.61 |
| 5 | Shanghai Meicheng Technology... | 50,122 | 114 | 0.23 | Shinjiru Technology Sdn Bhd | 15,986 | 242 | 1.51 |
| 6 | CV. Rumahweb Indonesia | 10,751 | 23 | 0.21 | Naugus Limited LLC | 7,803 | 102 | 1.31 |
| 7 | Jiangsu Bangning Science & technology | 186,390 | 323 | 0.17 | Upperlink Limited | 4,519 | 55 | 1.22 |
| 8 | Paragon Internet Group Ltd | 3,640 | 6 | 0.16 | Danesco Trading Ltd. | 184,206 | 1,205 | 0.65 |
| 9 | DOTSERVE INC. | 10,738 | 17 | 0.16 | Bottle Domains ,Inc. | 1,3121 | 81 | 0.62 |
| 10 | 101domain GRS Limited | 1,38,812 | 214 | 0.15 | Pheenix 100 ,LLC | 355 | 2 | 0.56 |

Table XVI
StopBadware top10 rate between blacklisted new and legacy gTLD domains (#Incidents) and total number of registrar gTLD domains (#Domains).

| # | new gTLD registrar | #Domains | #Incidents | Percent | Legacy gTLD registrar | #Domains | #Incidents | Percent |
|---|---|---|---|---|---|---|---|---|
| 1 | Xiamen Nawang Technology Co. ,Ltd | 281,148 | 12,396 | 4.41 | BoteroSolutions.com S.A. | 4 | 2 | 50.0 |
| 2 | Foshan YiDong Network Co. , LTD | 45,460 | 1,694 | 3.73 | Rethem Hosting LLC | 3,840 | 773 | 20.13 |
| 3 | Super Registry Ltd | 21,244 | 468 | 2.2 | RESERVED-IANA | 26 | 4 | 15.38 |
| 4 | Alpnames Limited | 2,623,443 | 57,141 | 2.18 | 0101 Internet ,Inc. | 8,315 | 576 | 6.93 |
| 5 | Netowl ,Inc. | 1,128 | 20 | 1.77 | Zhengzhou Zitian Network Technology | 12,235 | 555 | 4.54 |
| 6 | Todaynic.com ,Inc. | 31,7534 | 4,932 | 1.55 | Xiamen Nawang Technology Co., Ltd | 206,661 | 5,762 | 2.79 |
| 7 | Jiangsu Bangning Science & technology | 186,390 | 2,776 | 1.49 | Minds and Machines LLC | 1,115 | 26 | 2.33 |
| 8 | Web Werks India d/b/a ZenRegistry.com | 784 | 7 | 0.89 | Danesco Trading Ltd. | 184,206 | 4,263 | 2.31 |
| 9 | CV. Rumahweb Indonesia | 10,751 | 91 | 0.85 | In2net Network Inc. | 106,987 | 2431 | 2.27 |
| 10 | Chengdu West Dimension Digital... | 4,874,061 | 29,219 | 0.6 | Shanghai Oweb Network Co. , Ltd | 149 | 3 | 2.01 |

Table XVII
CleanMX Phishing top10 rate between blacklisted new and legacy gTLD domains (#Incidents) and total number of registrar gTLD domains (#Domains).

| # | new gTLD registrar | #Domains | #Incidents | Percent | Legacy gTLD registrar | #Domains | #Incidents | Percent |
|---|---|---|---|---|---|---|---|---|
| 1 | AB Name ISP | 1,069 | 3 | 0.28 | Minds and Machines LLC | 1,115 | 108 | 9.69 |
| 2 | CV. Rumahweb Indonesia | 10,751 | 28 | 0.26 | Upperlink Limited | 4,519 | 74 | 1.64 |
| 3 | Web4Africa Inc. | 2,428 | 5 | 0.21 | Shinjiru Technology Sdn Bhd | 15,986 | 257 | 1.61 |
| 4 | Shenzhen HuLianXianFeng Technology Co., LTD | 6,115 | 10 | 0.16 | Rethem Hosting LLC | 3,840 | 61 | 1.59 |
| 5 | 10dencehispahard, S.L. | 6,455 | 10 | 0.15 | Launchpad.com Inc. | 1,110,124 | 9,959 | 0.9 |
| 6 | Marcaria.com International, Inc. | 14,885 | 23 | 0.15 | Web4Africa Inc. | 22,339 | 169 | 0.76 |
| 7 | ZNet Technologies Pvt Ltd. | 1,365 | 2 | 0.15 | Dattatec.com SRL | 196,917 | 1,299 | 0.66 |
| 8 | BigRock Solutions Ltd. | 3,453 | 5 | 0.14 | Name121, Inc. | 17,626 | 113 | 0.64 |
| 9 | Network Information Center Mexico, S.C. | 1,491 | 2 | 0.13 | CCI REG S.A. | 29,004 | 177 | 0.61 |
| 10 | One.com A/S | 21,837 | 29 | 0.13 | Enetica Pty Ltd | 36,708 | 200 | 0.54 |

Table XVIII
CleanMX Viruses top10 rate between blacklisted new and legacy gTLD domains (#Incidents) and total number of registrar gTLD domains (#Domains).

| # | new gTLD registrar | #Domains | #Incidents | Percent | Legacy gTLD registrar | #Domains | #Incidents | Percent |
|---|---|---|---|---|---|---|---|---|
| 1 | Danesco Trading Ltd. | 137 | 2 | 1.46 | BoteroSolutions.com S.A. | 4 | 2 | 50.0 |
| 2 | Foshan YiDong Network Co., LTD | 45,460 | 209 | 0.46 | 0101 Internet, Inc. | 8,315 | 262 | 3.15 |
| 3 | Xiamen Nawang Technology Co., Ltd | 281,148 | 899 | 0.32 | Minds and Machines LLC | 1,115 | 26 | 2.33 |
| 4 | Authentic Web Inc. | 1,179 | 3 | 0.25 | Rethem Hosting LLC | 3,840 | 73 | 1.9 |
| 5 | TLD Registrar Solutions Ltd. | 148,915 | 286 | 0.19 | Soluciones Corporativas IP, SL | 197,859 | 3,036 | 1.53 |
| 6 | Dynadot, LLC | 93,116 | 124 | 0.13 | Pheenix 7, LLC | 314 | 4 | 1.27 |
| 7 | CV. Rumahweb Indonesia | 10,751 | 11 | 0.1 | Danesco Trading Ltd. | 184,206 | 1,671 | 0.91 |
| 8 | Eranet International Limited | 39,600 | 29 | 0.07 | CloudFlare, Inc. | 221 | 2 | 0.9 |
| 9 | FBS Inc. | 56,340 | 39 | 0.07 | Paknic (Private) Limited | 10,512 | 93 | 0.88 |
| 10 | Alpnames Limited | 2,623,443 | 1,807 | 0.07 | IPNIC, Inc. | 687 | 6 | 0.87 |

Table XIX
CleanMX Portals top10 rate between blacklisted new and legacy gTLD domains (#Incidents) and total number of registrar gTLD domains (#Domains).

| # | new gTLD registrar | #Domains | #Incidents | Percent | Legacy gTLD registrar | #Domains | #Incidents | Percent |
|---|---|---|---|---|---|---|---|---|
| 1 | Marcaria.com International, Inc. | 14,885 | 19 | 0.13 | Minds and Machines LLC | 1,115 | 65 | 5.83 |
| 2 | Gandi SAS | 177,962 | 159 | 0.09 | 0101 Internet, Inc. | 8,315 | 122 | 1.47 |
| 3 | Register NV dba Register.eu | 26,272 | 23 | 0.09 | Rethem Hosting LLC | 3,840 | 49 | 1.28 |
| 4 | NameCheap, Inc. | 1,912,822 | 1664 | 0.09 | Shinjiru Technology Sdn Bhd | 15,986 | 157 | 0.98 |
| 5 | BigRock Solutions Ltd. | 3,453 | 3 | 0.09 | ZNet Technologies Pvt Ltd. | 50,381 | 466 | 0.92 |
| 6 | FBS Inc. | 56,340 | 48 | 0.09 | Name121, Inc. | 17,626 | 151 | 0.86 |
| 7 | MAT BAO CORPORATION | 2,511 | 2 | 0.08 | OwnRegistrar, Inc. | 19,745 | 156 | 0.79 |
| 8 | CV. Rumahweb Indonesia | 10,751 | 8 | 0.07 | Upperlink Limited | 4,519 | 35 | 0.77 |
| 9 | Todaynic.com, Inc. | 317,534 | 221 | 0.07 | Web Site Source, Inc. | 5,526 | 41 | 0.74 |
| 10 | Online SAS | 2,984 | 2 | 0.07 | Catalog.com | 28,737 | 213 | 0.74 |

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

[36] S. Tajalizadehkhoob, R. Böhme, C. Gañán, M. Korczyński, and M. van Eeten, “Rotten Apples or Bad Harvest? What We Are Measuring When We Are Measuring Abuse,” 2017. [Online]. Available: https://arxiv.org/abs/1702.01624

[37] “IANA: Registrar IDs,” https://www.iana.org/assignments/registrar-ids/registrar-ids.xhtml.

[38] S. Hao, N. Feamster, and R. Pandrangi, “Monitoring the Initial DNS Behavior of Malicious Domains,” in *Proceedings of the 2011 Conference on Internet Measurement Conference (IMC’11)*. ACM, 2011, pp. 269–278.

[39] “Registrar Accreditation Agreement,” 2013. [Online]. Available: https://www.icann.org/resources/pages/approved-with-specs-2013-09-17-en#whois-accuracy

[40] “Alexa: Actionable Analytics for the Web,” http://www.alexa.com.

[41] “National Physical Laboratory: A Study of Whois Privacy and Proxy Service Abuse,” https://gnso.icann.org/en/issues/whois/pp-abuse-study-20sep13-en.pdf.

[42] “ICANN: ICANN-Accredited Registrars,” https://www.icann.org/registrar-reports/accredited-list.html.

[43] “Registrar OWL: Domain registrar statistics, pricing, reviews and comparisons,” http://www.registrarowl.com/.

[44] “ICANN: NOTICE OF TERMINATION OF ACCREDITATION AGREEMENT,” https://www.icann.org/uploads/compliance_notice/attachment/895/serad-to-hansmann-4jan17.pdf.

# Appendices

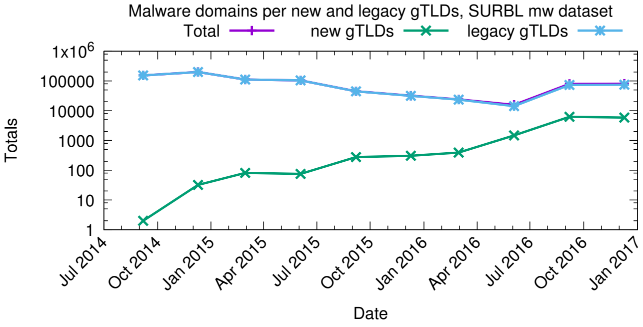


Figure 33. Time series of counts of malware domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the **SURBL mw** feed (2014-2016). Please notice $y$ axis in log scale and overlapping lines.

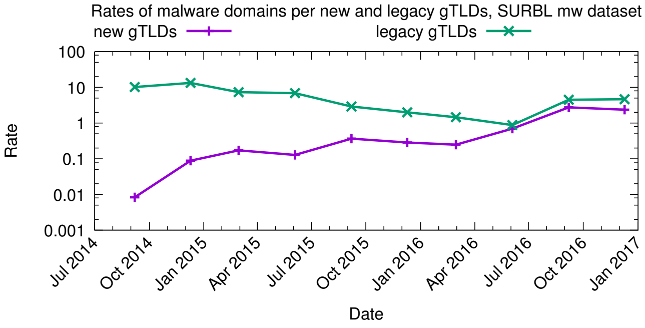


Figure 34. Time series of abuse rates of malware domains in **legacy** gTLDs and **new** gTLDs based on the **SURBL mw** feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

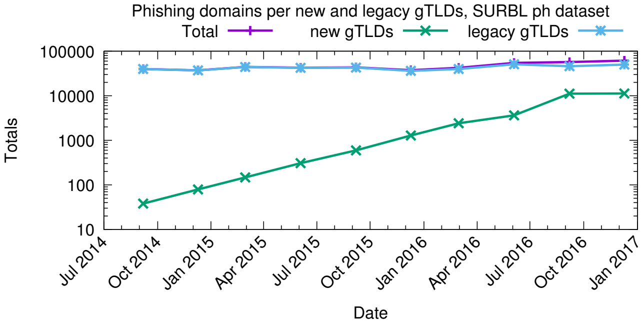


Figure 35. Time series of counts of phishing domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the **SURBL ph** feed (2014-2016). Please notice $y$ axis in log scale and overlapping lines.

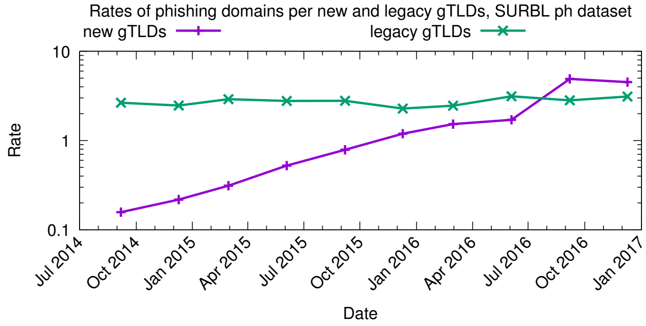


Figure 36. Time series of abuse rates of phishing domains in **legacy** gTLDs and **new** gTLDs based on the **SURBL ph** feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

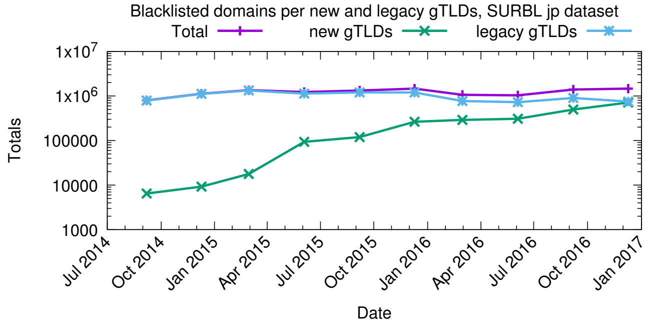


Figure 37. Time series of counts of blacklisted domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the **SURBL jp** feed (2014-2016). Please notice $y$ axis in log scale and overlapping lines.

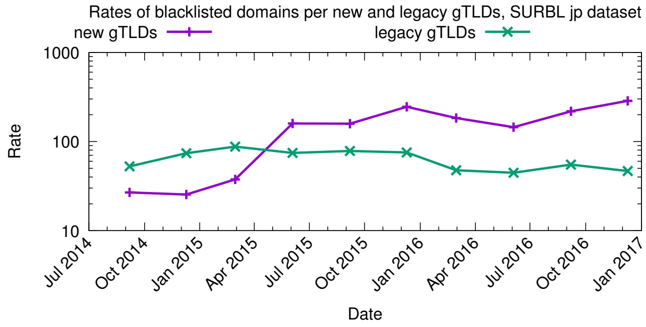


Figure 38. Time series of abuse rates of blacklisted domains in **legacy** gTLDs and **new** gTLDs based on the **SURBL jp** feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

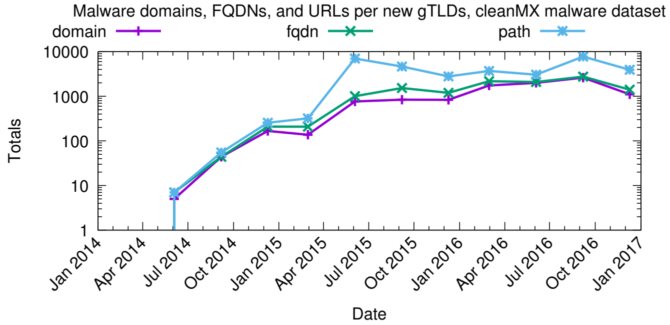


Figure 48. Time series of counts of blacklisted malware domains, FQDNs, and URLs (paths) in **new** gTLD based on the **cleanMX phishing** feed (2014-2016). Please notice $y$ axis in log scale and overlapping lines.

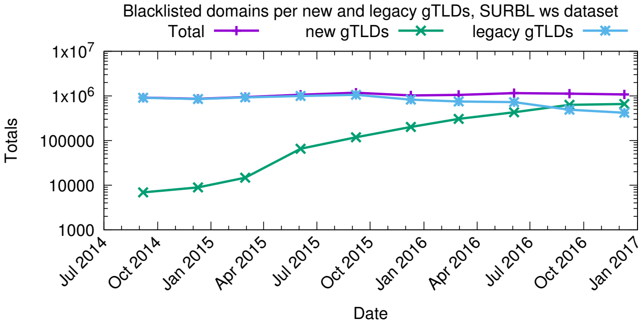


Figure 39. Time series of counts of blacklisted domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the **SURBL ws** feed (2014-2016). Please notice $y$ axis in log scale and overlapping lines.

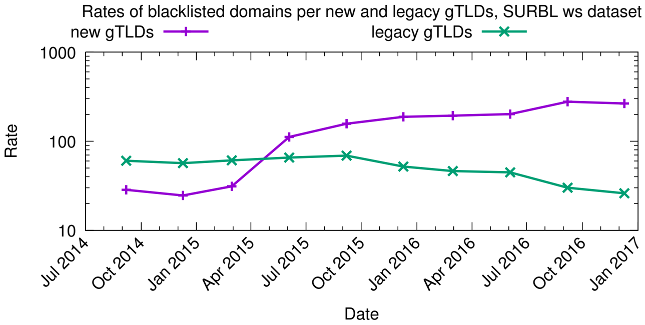


Figure 40. Time series of abuse rates of blacklisted domains in **legacy** gTLDs and **new** gTLDs based on the **SURBL ws** feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

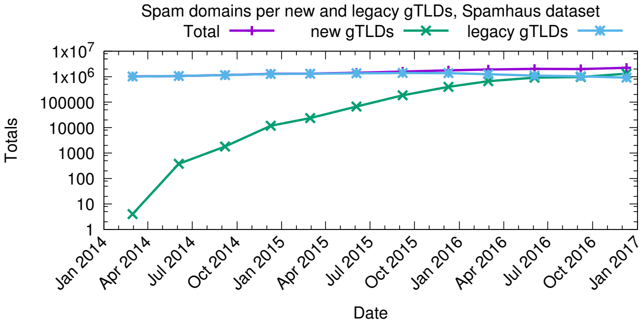


Figure 41. Time series of counts of blacklisted domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the **spamhaus** feed (2014-2016). Please notice $y$ axis in log scale and overlapping lines.

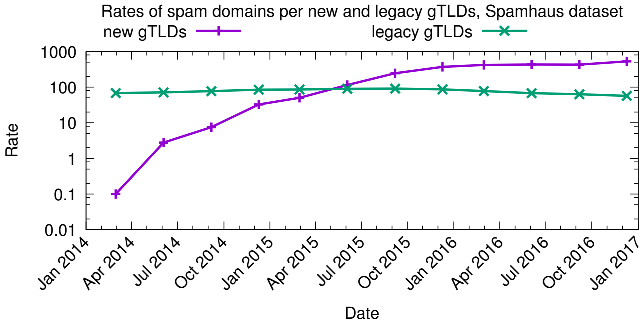


Figure 42. Time series of abuse rates of blacklisted domains in **legacy** gTLDs and **new** gTLDs based on the **spamhaus** feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

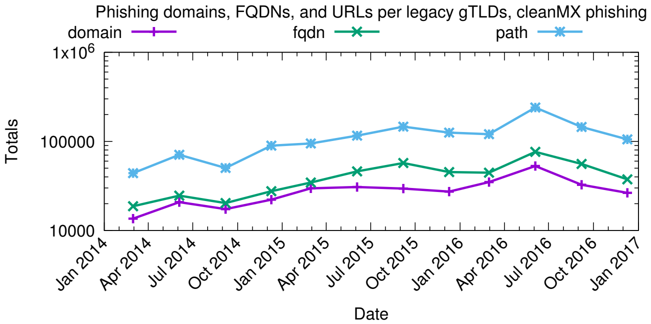


Figure 43. Time series of counts of blacklisted phishing domains, FQDNs, and URLs (paths) in **legacy** gTLD based on the **cleanMX phishing** feed (2014-2016). Please notice $y$ axis in log scale and overlapping lines.

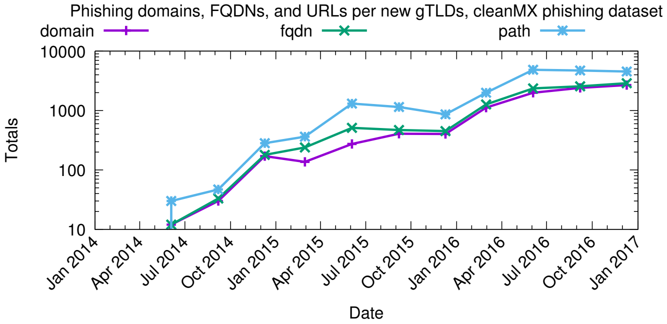


Figure 44. Time series of counts of phishing domains, FQDNs, and URLs (paths) in **new** gTLD based on the **cleanMX phishing** feed (2014-2016). Please notice $y$ axis in log scale and overlapping lines.

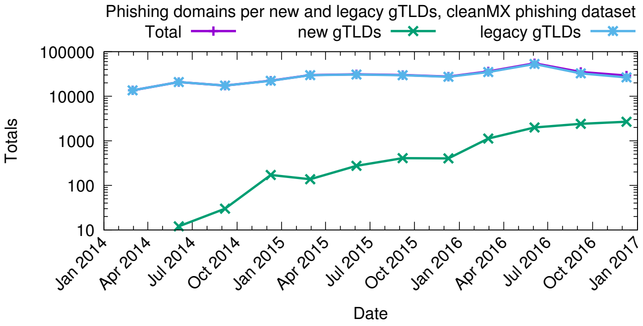


Figure 45. Time series of counts of blacklisted phishing domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the **cleanMX phishing** feed (2014-2016). Please notice $y$ axis in log scale and overlapping lines.

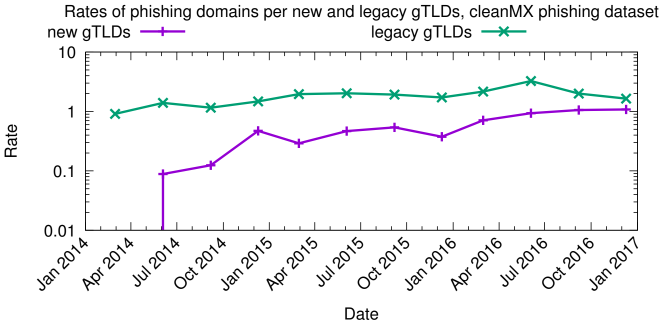


Figure 46. Time series of abuse rates of blacklisted phishing domains in **legacy** gTLDs and **new** gTLDs based on the **cleanMX phishing** feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

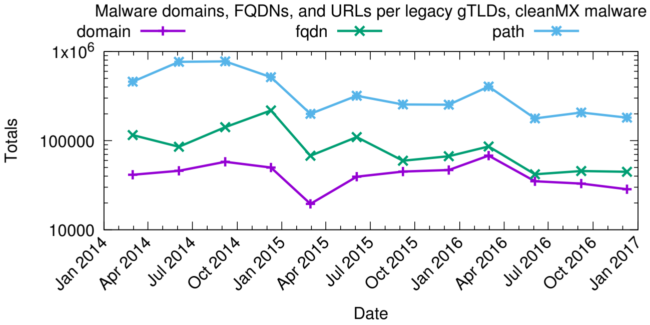


Figure 47. Time series of counts of blacklisted malware domains, FQDNs, and URLs (paths) in **legacy** gTLD based on the **cleanMX malware** feed (2014-2016). Please notice $y$ axis in log scale and overlapping lines.

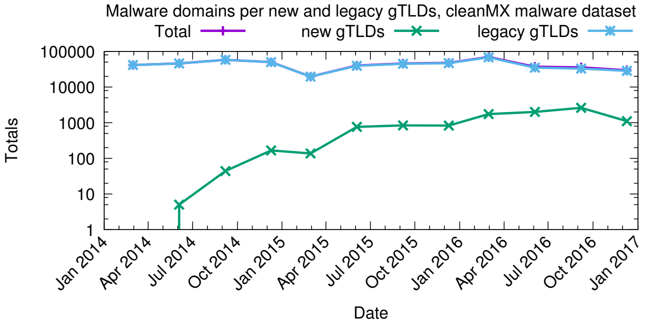


Figure 49. Time series of counts of blacklisted malware domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the **cleanMX malware** feed (2014-2016). Please notice $y$ axis in log scale and overlapping lines.

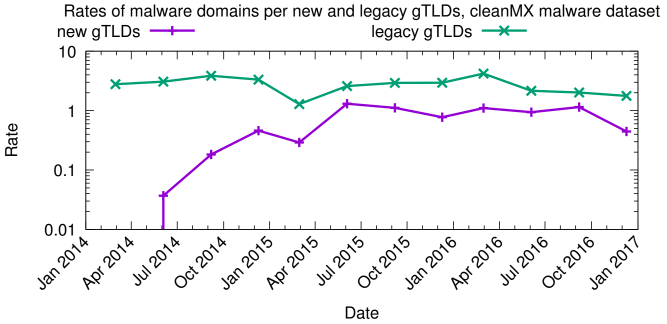


Figure 50. Time series of abuse rates of blacklisted malware domains in **legacy** gTLDs and **new** gTLDs based on the **cleanMX malware** feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.