Statistical Analysis of DNS Abuse in gTLDs Final Report

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

*Delft University of Technology, The Netherlands
†SIDN Labs, The Netherlands
{Maciej.Korczynski, S.T.Tajalizadehkhoob}@tudelft.nl
{Maarten.Wullink, Giovane.Moura, Cristian.Hesselman}@sidn.nl

***Abstract*—Commissioned by the Competition, Consumer Trust, and Consumer Choice Review Team with the support of ICANN, this study is focused on measuring rates of common forms of abusive activities in the domain name system. We conduct a comprehensive study examining malicious behavior in the global DNS and compare abuse rates in new and legacy gTLDs. We combine data sets from many sources, including zone files, domain `WHOIS` information, data obtained through our active measurements, and 11 reputable blacklists representing malware, phishing, and spam. We find that the new gTLDs have impacted spam counts of the legacy gTLDs: abused domains in the new gTLDs do not increase the number of total malicious registrations but instead, we observe a decrease in the number of malicious registrations in legacy gTLDs. While legacy gTLDs collectively had a spam-domains-per-10,000 rate of 56.9, in the last quarter of 2016, the new gTLDs experienced a rate of 526.6–which is almost one order of magnitude higher. In this study, we also analyze the relationship between the collected security indicators and the structural properties of new gTLDs, and abuse, at the level of gTLDs. The results indicate that abuse counts primarily correlate with stricter registration policies. Our findings suggest that some new gTLDs have become a growing target for malicious actors. While the analysis of spam blacklists reveals that approximately one third of the new gTLDs available for registration did not experience a single incident in the last quarter of 2016, Spamhaus blacklisted at least 10% of all registered domains in 15 new gTLDs.**

I. Introduction

As a result of a many-year multi-stakeholder policy development process, the Internet Corporation for Assigned Names and Numbers (ICANN) introduced the New Generic Top-Level Domain (gTLD) Program (the Program), which has enabled hundreds of new gTLDs to enter the domain name system (DNS) since the first delegations occurred in October 2013[1]. The Program was developed to increase competition and choice in the domain name space. More than 1,900 applications for new gTLDs were filed after the process opened in 2012. To date, more than 1,200 new gTLDs have been delegated to the DNS' root zone. However, while the New gTLD Program may increase the range of available gTLDs available to consumers, it may also create new opportunities for cybercriminals.

A number of safeguards were built into the Program that were intended to mitigate the rates of abusive, malicious, and criminal activity in these new gTLDs, such as phishing, spam, and malware distribution. ICANN is currently engaged in a review of these safeguards and their effects on rates of DNS abuse as an aspect of the Competition, Consumer Trust, and Consumer Choice Review[2]. In this paper, we conduct a comprehensive study examining rates of malicious and abusive behavior in the global DNS and compare abuse rates in new gTLDs and legacy gTLDs. As the DNS represents a large ecosystem of registries, registrars, privacy/proxy service providers, etc. the study aims to capture inputs in a representative manner from across the multitude of players relevant to abusive practices.

Previous research studied the impact of the New gTLD Program on the domain name ecosystem [1]. Halvorson *et al.* concluded that speculative and defensive registrations dominate the growth of registrations in new gTLDs. They also found that the new gTLDs have yet to have significant impact on the legacy gTLDs. Their work, however, provides a very little empirical information about the security of new gTLDs. In this paper, we analyze the impact of the New gTLD Program on the DNS abuse landscape and assess if legacy and new gTLDs are seen by miscreants as interchangeable. Overall, our main contributions can be summarized as follows:

- We make a comprehensive descriptive statistical comparison of rates of DNS abuse in new and legacy gTLDs as they pertain to spam, phishing, and malware distribution.
- Using regression modelling we perform inferential statistical analysis testing the correlation between passively and actively measured properties of new gTLDs as predictors of rates of abuse.
- We analyze proportions of abusive domains across other relevant to abusive practices players, i.e. registrars and privacy/proxy service providers.

Our findings demonstrate a number of notable trends in relation to the new gTLD landscape and cybercriminal activity. While we find higher concentrations of compromised (hacked) domains in legacy gTLDs, miscreants frequently choose to

[1] https://gnso.icann.org/en/group-activities/inactive/2007/new-gtld-intro

[2] https://newgtlds.icann.org/en/reviews/cct/dns-abuse

register domain names using one of the new gTLDs. We also find that the new gTLDs have significant impact on abuse counts of the legacy gTLDs. Interestingly, maliciously registered spam domains in the new gTLDs do not increase the number of total malicious registrations. Instead, we witness a decrease in the number of malicious registrations in legacy gTLDs. An analysis of the Spamhaus blacklist reveals that in the last quarter of 2016, new gTLDs collectively had approximately one order of magnitude higher spam-domains-per-10,000 rate in comparison to legacy gTLDs.

We also systematically analyze how different structural and security-related properties of new gTLD operators influence abuse counts. Our inferential analysis reveals that abuse counts primarily correlate with stricter registration policies. Finally, the analysis of proportions of abusive domains across new gTLDs, registrars, and privacy/proxy service providers reveals entities suffering from very high concentrations of abused domains. We find new gTLDs and registrars with concentrations of blacklisted domains above 50%, in particular a registrar with over 90% of its domains reported as abusive by SURBL.

II. Background

ICANN [2] is responsible for maintaining the root namespace and its expansion with new top-level domains (TLDs), in particular new gTLDs, and delegates the responsibility to maintain an authoritative source for registered domain names within a TLD to *registry operators*. ICANN holds registries responsible for complying with the terms of a registry agreement. Domain registries manage the registration and delegation of domain names within their TLDs.

The DNS represents a large ecosystem and 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* maintain server infrastructure that is used to host content for the domain. *DNS providers* operate DNS servers that map domain and host names to the corresponding IP addresses. The *WHOIS Privacy and Proxy* service providers conceal certain personal data of domain name registrants. *Registrants* are individuals or organisations that hold or manage domain names. The aim of this study is to capture inputs in a representative manner from across this multitude of players relevant to abusive practices.

A. Generic TLDs

The first group of generic top-level domains (gTLDs) was defined by RFC 920 [3] in October 1984 and introduced a few months later. The initial group of gTLDs (.gov, .edu, .com, .mil, .org, and .net) were distinct from country-code TLDs (ccTLDs). 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 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 WHOIS data. We contrast them with the *new gTLDs*.

B. New gTLDs

ICANN's New gTLD Program started in 2012 and expanded the root zone by delegating more than 1,200 new gTLDs since October 2013 [4]. To obtain a new gTLD, applicants are required to undergo an intensive application and evaluation process [1] that includes screening the applicants technical and financial capabilities for operating a new gTLD.

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 [5][3]:

- Standard or generic gTLD [7]: is a gTLD that is generally open for public registration, e.g. .movie, .xyz, or .family. While most of these gTLDs are open to public registration, some registries may impose restrictions on who or which entities can register in their domains.
- Community gTLD [8]: this category covers gTLDs that are restricted to a specific community, such as .thai, .radio or .pharmacy.
- Geographic gTLD [9]: this type of gTLD covers cities, states, or regions, e.g. .amsterdam or .berlin.
- Brand gTLD [10]: for companies seeking to have their specific brand as a gTLD, 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[4], in particular for malicious purposes. This study 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.

C. Safeguards Against DNS Abuse

In preparation for the New gTLD Program, ICANN sought advice from different DNS abuse and security experts to examine the potential for increases in criminal activity in an expanded DNS in order to further determine a number of possible measures to preemptively mitigate abusive and malicious activities [11], [12]. As a result of broad discussion with multiple stakeholders such as Anti Phishing Working

[3]Note that some gTLDs cross categories. For example, some community gTLDs such as .madrid are also geographic gTLDs [6].

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

Group (APWG), Registry Internet Safety Group (RISG), the Security and Stability Advisory Committee (SSAC), Computer Emergency Response Teams (CERTs) and members of the banking/financial, and Internet security communities, ICANN proposed 9 safeguards that can be summarized under the following four key subject categories [11]:

- How do we ensure that bad actors do not run Registries?
- How do we ensure integrity and utility of registry information?
- How do we ensure more effective effort to combat identified abuse?
- How do we provide an enhanced control framework for TLDs with intrinsic potential for abuse?

In this paper, we conduct a first comprehensive study examining rates of malicious and abusive behavior in the global DNS and compare abuse rates in new and legacy gTLDs. We aim to provide research to assess the general effectiveness of the proposed safeguards to mitigate DNS abuse that were described in ICANN's "New gTLD Program Explanatory Memorandum: Mitigating Malicious Conduct" released in 2009 [11] and "New gTLD Program Safeguards Against DNS Abuse" [12].

III. Data Collection

A. Blacklists

To assess the prevalence of maliciously registered[5] and compromised domains[6] per gTLD and registrar, we use 11 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], the Secure Domain Foundation (SDF) [17] and CleanMX [18]. All six 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 for production mail systems [19]. 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 [20]–[22], Internet Identity, Google's Safe Browsing appeals results, the StopBadware community, and other contributors [23]. In our study we also use four 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 [24]. *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 [24]. The SDF feed contains domains and URLs classified as phishing or malware. The domain names were queried against the Secure Domain Foundation's Luminous API which aggregates data from open source blacklist feeds and registrar suspension lists. At the time of the queries, the SDF data included suspended domain names provided by registrars and SDF-vetted data from Alien Vault, APWG, Binary Defense Systems, Charles Haley, Chaos Reigns, Dragon Research Group, Malwarebytes [25], Malcode Block List, MultiProxy, Malware Domain List, OpenBL.org, and pfBlockNG. Note that unlike the other data feeds, SURBL and SDF feeds cover 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 shows the number of unique gTLD domain names, fully-qualified domain names (FQDNs)[7] 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 [26]. 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 previously explained, StopBadware and 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, i.e. they were registered by legitimate users and hacked (see e.g. global phishing survey reports [27], [28]). From the operational point of view blocking the domain name element of a blacklisted URL might harm legitimate operations. On the other hand, Spamhaus and other data providers maintain

[5]Domains registered by miscreants for the purpose of malicious activity

[6]Domains hacked by miscreants, exploited through vulnerable web hosting

[7]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.

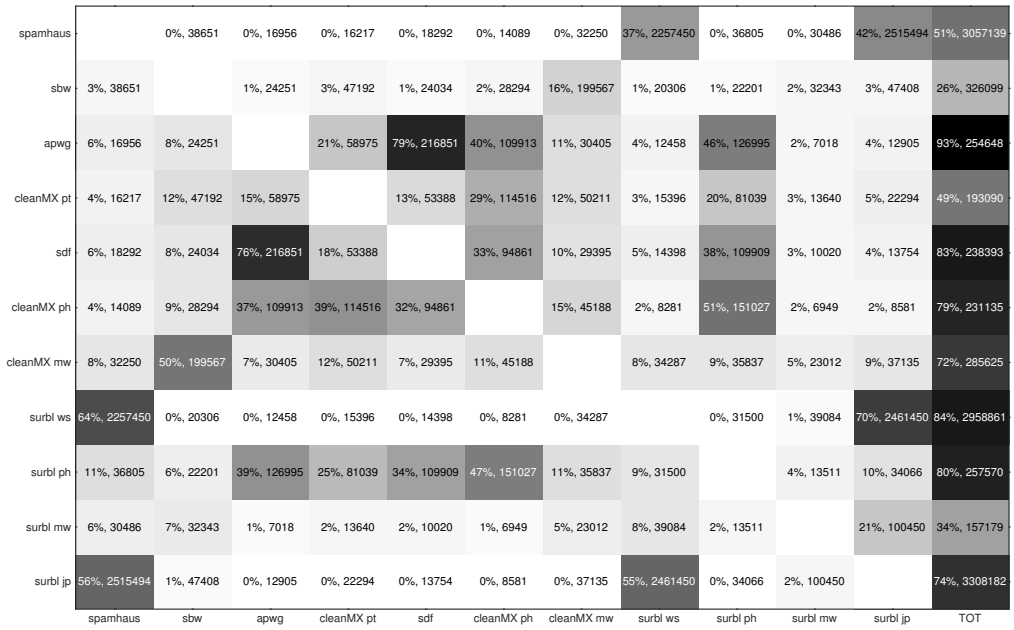

Figure 1. Pairwise overlap of feeds with unique domains as unit of abuse (2014-2016)

Table I
OVERVIEW OF BLACKLISTS: UNIQUE BLACKLISTED GTLD DOMAIN NAMES, FQDNS, AND URLS, FOR THE APWG, STOPBADWARE SDP, SPAMHAUS, CLEANMX, SDF 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	–	–
	SDF	41,094	93,324	723,523
	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	–	–
	SDF	142,285	535,406	4,391,796
	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	–	–
	SDF	110,687	122,326	264,465
	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	–	–

blacklists of domain names and perform extensive 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 in the domain blacklists and domain name elements of blacklisted URLs as "abused domains" or "blacklisted domains". Table II provides an overview of the blacklists used in our study and their corresponding types.

Table II
OVERVIEW OF BLACKLIST TYPES

StopBadware	Malware URLs
APWG	Phishing URLs
Spamhaus	Spam domains
SDF	Other URLs
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

Figure 1 illustrates pairwise feed intersections as a matrix, with unique domain names as the unit for abuse. Note that darker shades of grey represent higher overlaps. For example, the overlap between Spamhaus and SURBL ws indicates that they have 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 are blacklisted by at least one other blacklist.

B. WHOIS Data

Most of the blacklists used for this study contain no additional domain name attributes such as registrar name or date of registration. For the purposes of this study, these attributes were obtained via a WHOIS database provided by a third-party vendor [29] covering the 3-year study period (2014-2016). The database contains WHOIS information for the domains of 18 legacy gTLDs: .aero, .asia, .biz, .cat, .com, .coop, .info, .jobs, .mobi, .museum, .name, .net, .org, .post, .pro, .tel, .travel, and .xxx. It also contains WHOIS information for the domains of 1196 new gTLDs that have been delegated during the study period [30].

The database uses temporal versioning, meaning that 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. For example, 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. The 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-D for more details).

During the domain name to WHOIS record mapping process we found that a significant number of abusive domains could not be found in the available WHOIS data. We asked DomainTools [31] to provide the WHOIS data for the domains for which we did not have WHOIS details. Using the DomainTools data we could map an additional 6,081,870 abusive domains.

C. DNS Zone Files

In order to calculate sizes for each gTLD, we processed DNS zone files provided by ICANN and extracted the unique domains. The zone 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 and delegated domains for the particular zone (gTLD). Since the list of domains contained in a zone is usually dynamic (domains are registered, expire, or have their records changed), the respective zone file is also dynamic. Different registries also apply different zone publication policies. For example, .com updates its zone every 5 minutes, while .nl updates its zone every 30 minutes.

Table III
WHOIS DATA OVERVIEW: THE NUMBER OF TLDS (# TLD) AND DOMAIN NAMES (# DOMAINS) 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

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

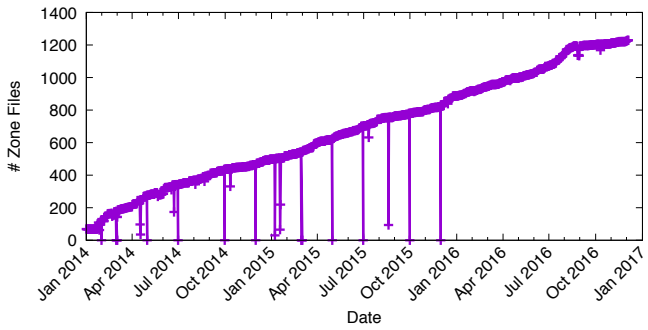

Figure 2. Number of daily zone files obtained for this study.

We also relied upon zone files to determine the number of DNS Security Extensions (DNSSEC)-signed domains for each gTLD. One of the New gTLD Program safeguards require that all new gTLD applicants have a specific plan for DNSSEC deployment [12]. We used this data in the inferential analysis as a proxy for security efforts of registries of new gTLDs. Using regular expressions we matched the DS records in the zone files and counted the distinct number of domains with DS records. The DS record is kept in the parent (TLD) zone and is used to prove the validity of cryptographic DNSSEC chain. If there is a DS record then this indicates that the domain has support for DNSSEC.

D. Active Web Scan

Using an existing Web crawling platform developed at SIDN Labs we have crawled each new gTLD domain found in the zone files generated on May 2, 2017 (24,2 million domains). The number of legacy gTLD domains is too large to scan in the time available for this study, therefore we decided to create a representative sample of 16,7 million domains (from the same date) to scan.

The domains have been extracted from zone files that have been published shortly before we started crawling the domains.

We could have used the historical zone files to extract the domains but using older zone files would result in more error domains because domains could have been deleted in the meantime.

First, the crawler attempts to fetch the main website for each domain by prepending the 'www' label to the domain and crawling the resulting FQDN, e.g. www.example.com. If this results in an NXDOMAIN or another error the crawler will then try to crawl the apex (naked domain) e.g. example.com. If both variants return an error then the domain is considered non-responsive. If the crawler detects a redirect in either the retrieved HTML code or the HTTP headers then these redirects are followed, a hard maximum limit of 5 redirects has been configured. Any domain resulting in a crawl chain of more than 5 redirects is marked as non-responsive.

The crawler is designed to have a minimal impact on the servers that are crawled. For this reason only the main page is retrieved instead of the entire website. The data captured for each domain includes the HTML code, HTTP headers and status codes.

To determine if a domain is parked the HTML code is analysed using pattern matching to search for strings, which might indicate the domain is for sale. The analyzer also looks for URLs that are linked to known parking services provider. Any redirects to domains belonging to parking providers are also recognized.

E. Active DNS Scan

During the domain scan process we also query the DNS to retrieve the A, AAAA and SOA records. The DNS crawler sends queries to a dedicated instance of the Unbound DNS resolver and analyzes the results. We use the SOA record to determine if the primary authoritative name server for the domain is linked to a known parking services provider.

F. Passive Data for Registries

In this study we analyze new gTLDs whose domain names became available for public registration within the study period. As the time between the delegation of a new gTLD and the end of the sunrise period might take several months[8], in our analysis we include new gTLDs after their sunrise periods. This data has been provided by ICANN via their public portal [30]. It contains 522 new gTLDs with sunrise periods that ended before the end of the study period.

We also used a list of registry operators, their affiliates, and associated new gTLDs provided to us by ICANN. We mapped gTLDs to related registry operators regardless of what name they are operating under. We used the mapping of parent companies of registry operators and the corresponding new gTLDs in the inferential analysis as a proxy for registration practices.

We relied upon ICANN's listings of new generic [7], community [8], geographic [9], and brand [10] gTLD registry applications. We used this data in the inferential analysis as a proxy for restricted registration. We assigned registration "levels" to new gTLDs, from least to most restricted group: 1 generic, 2 geographic, 3 community, and 4 brand. Intuitively, while generic gTLDs are normally unrestricted and open for public registration, registration policies of community or brand gTLDs are strict and may prevent miscreants from malicious registrations.

[8]E.g. delegation of .zuerich: December 25, 2014 [32], zone file seen for the first time: January 1, 2015, sunrise period termination: June 5, 2017 [30]

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 [33]. First, we 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) [34]. 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) [28].

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 analyze 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.

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.) [35].

This is why we analyze 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 [36]. 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

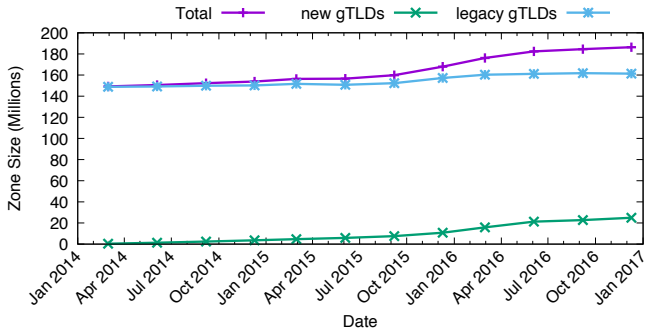

Figure 3. Absolute growth of **legacy** gTLD, **new** gTLDs and **all** gTLDs.

[36]. 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 Estimate of TLDs

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 2nd-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 [37]. Some registrants, however, purchase domains and do not associate them with the name servers. Therefore, they are not present 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 [33], [36], [38], the number of domains found in a zone file is more accurate.

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), one might 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).

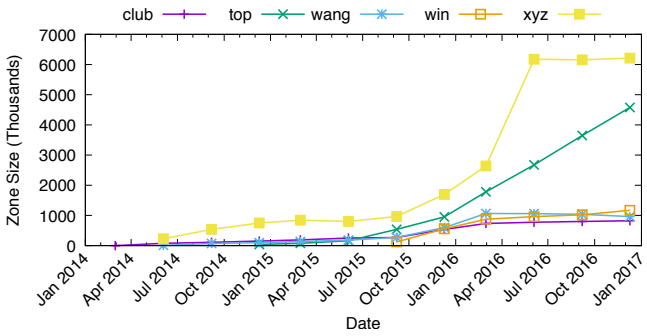

Figure 4. Absolute growth of top 5 largest **new** gTLDs as of end of 2016.

The TLD size can also be used as an explanatory factor for the concentrations of abused domains [33], [36], [38]. However, it is unclear what portion of the domains are in use and serve content. Halvorson *et al.* have shown, for example, that in 2015 there were as many as 16% of domains in new gTLDs with `NS` records that did not resolve [1]. Using our Web and DNS crawling platform, we performed a new scan and classified each domain to one of the five groups: *i) No DNS* domains that do not resolve when queried by our DNS crawler, *ii) Parked* domains that are owned by parking services, advertisement syndicators, and advertisers. We follow the classification methodology outlined by Vissers *et al.* [39], *iii) HTTP Error* domains for which authoritative NS servers return valid responses but the corresponding websites do not return an HTTP 200, *vi) Redirect* domains are redirected to a different domain, and *v) Content* domains serve a valid Web content to their users.

Figure 5 shows the categorization results for all domains in the new gTLDs and a random sample of the legacy gTLDs. Interestingly, we find a significant increase in erroneous domains in new gTLDs ("No DNS" and "HTTP Error" categories) in comparison to legacy gTLDs. "No DNS" domains account for about a quarter of all domains (24.2%), whereas domains for which the corresponding websites serve an HTTP error account for another 12.2%.

Note that we use the measurement data in the inferential analysis to correct for TLD sizes. Intuitively, only the domains serving content are exposed to certain types of vulnerabilities and can be hacked. On the other hand, parked domains may be used to scam users or to distribute malware. One might therefore expect a positive correlation between the number of parked domains and maliciously registered domains.

C. Size Estimate of 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,

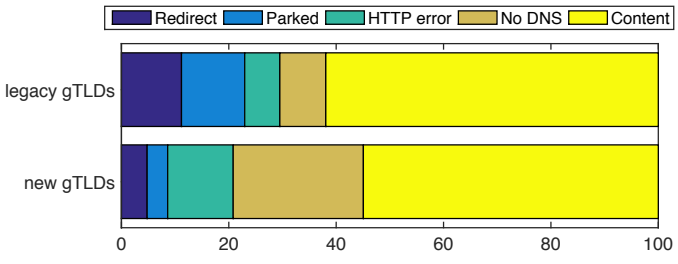

Figure 5. Categorization for all domains in the new TLDs and a random sample of the legacy TLDs.

e.g. "GODADDY.COM, LLC", "GoDaddy.com, LLC (R91-LROR)" and "GoDaddy.com, Inc.". This means 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 [40] 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.

A limitation of our approach is that the WHOIS database only contains domains for legacy gTLDs and 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 [33]. 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 [9].

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[10].

D. 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.

To distinguish between compromised and maliciously registered domains we build on three heuristics previously used in domain abuse surveys (e.g. global phishing survey by Aaron and Rasmussen [41]). More specifically, we label a domain as maliciously registered if it was involved in criminal activity within a relatively short time after its registration or if it contains a brand name or a misspelled variant of brand name. We flag a domain as malicious if it is blacklisted within 3 months after its registration. Aaron and Rasmussen have recently examined the delay between the time when phishing domains were initially registered and when they were ultimately used in attacks [41]. Their analysis indicates that miscreants tend to age the malicious domains they register to ensure a higher reputation score from security organizations. They concluded that the great majority of the domains used for phishing were maliciously registered within three months before they were used in an 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 [42], the creation date of the domain registration cannot be changed as long as the domain does not expire.

Furthermore, Aaron and Rasmussen identified 783 unique phishing target organizations in 2015 and 679 in 2016, among which the most popular ones were PayPal, Yahoo!, Apple, and Taobao [41]. We use this information to create a list of keywords that the attackers may incorporate in maliciously registered domain names. As the great majority of phishing attacks target the most popular organizations, we extracted 300 keywords of the most popular domains according to the Alexa ranking and we labeled each blacklisted domain as maliciously registered if it contains an extracted string or its misspelled version. For example, 0paypalpayment.com would be labeled as malicious as it contains a string *paypal*. To test if the domain contains a misspelled keyword, we first remove all digits from a domain name and split the resulting string into words with the "–" character. We compute the Levenshtein edit distances between the predefined keywords and a set of words derived from a domain name. If any Levenshtein edit distance is smaller than 2, we label the domain as maliciously registered. We provide a discussion on limitations of these heuristics later.

Note that from the categorization process we exclude a list of 11,075 domains of legitimate services that tend to be misused by miscreants. These represent a separate, third group of domains that are neither maliciously registered nor hacked (i.e. third party domains). 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,

[9]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.

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

but of entire TLDs [33]. The list is composed of the 10,000 most popular domains according to Alexa [43] and our own, manually maintained lists of domains of legitimate services (332 domains of URL shorteners and 840 domains of free hosting providers). This group includes:

- **Free hosting and dynamic DNS (DDNS)** services offering shared higher-level domains, such as Hostinger, a free hosting provider offering subdomains, 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 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.

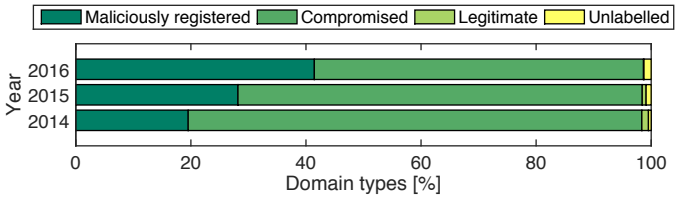

Figure 6. Categorization results: the fraction of maliciously registered, compromised, legitimate, and unlabelled domains for **APWG** feed in 2014, 2015, and 2016.

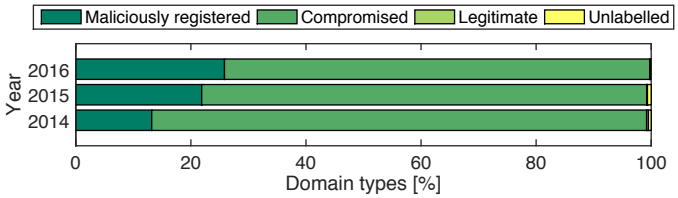

Figure 7. Categorization results: the fraction of maliciously registered, compromised, legitimate, and unlabelled domains for **StopBadware DSP** feed in 2014, 2015, and 2016.

Figure 6 and Figure 7 show the categorization of domains blacklisted by APWG and StopBadware respectively during the study period (2014, 2015, and 2016). Note that up to 1.1% of all domains submitted to the APWG have been pre-filtered based on the maintained list of domains corresponding to legitimate services and labeled as "legitimate". For comparison, we have excluded less than 0.3% of the StopBadware domains. A previous study showed that domains of legitimate services are often misused by miscreants to distribute malware or used in phishing campaigns [33]. However, some may also represent legitimate domains that were incorrectly blacklisted.

We note a limitation to this method: up to 1.2% and 0.6% of the APWG and StopBadware domains, respectively, are unlabelled. This is mainly because the corresponding WHOIS data was not available. However, a large fraction of labelled domain instances allow us to draw general conclusions about the prevalence of maliciously registered and compromised domains, respectively.

The results indicate that 78.8% of abused phishing and 86% of malware domains (listed on URL blacklists in 2014) were compromised by criminals (see Figure 6 and Figure 7). In 2016, those percentages were smaller: 57.2% and 73.9% of phishing and malware domains were labeled as compromised. Although domains listed in URL blacklists are predominantly compromised, their number has been gradually decreasing. The miscreants tend to choose to register domains more often. We find that 19.5%, 28.2%, 41.5% and 13.2%, 21.9%, 25.8% (in 2014, 2015, and 2016) of all phishing and malware domains respectively were presumbly maliciously registered by criminals. These reveal the changing over time profit-maximizing behavior of a large proportion of criminals that prefer to register rather than compromise domains.

Note that the number of malicious registrations might be undercounted. If an attacker does not use a maliciously registered domain within three months or the malicious activity is detected more than three months after the domain creation then the domain might 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).

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.

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 8) and new gTLDs (Figure 9), over time. We aggregate the count of phishing incidents on a quarterly basis (x-axis) and present the results in a logarithmic scale (y-axis). Note that the observed "drop" 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 a decline in criminal activities. 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 in APWG. This is because the metrics based on FQDNs and URLs are heavily affected by legitimate services such as file storage web services or popular URL shortening services [33]. For example, in our previous work [33], we found

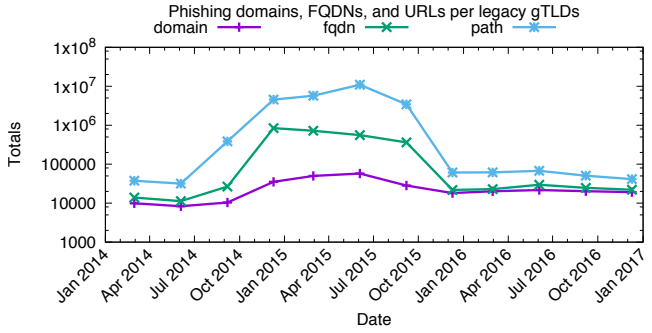

Figure 8. Time series of counts of phishing domains, FQDNs, and URLs (paths) in **legacy** gTLD based on the Anti-Phishing Working Group feed (2014-2016).

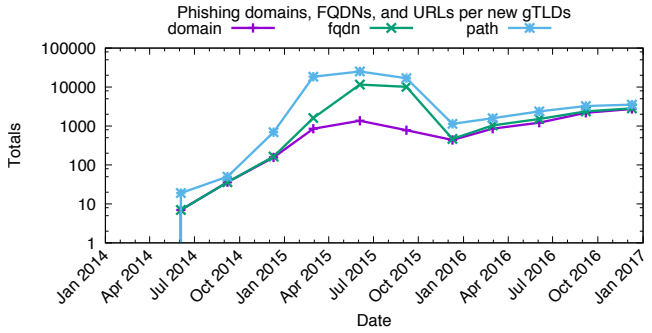

Figure 9. Time series of counts of phishing domains, FQDNs, and URLs (paths) in **new** gTLD based on the Anti-Phishing Working Group feed (2014-2016).

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 8 and Figure 9 with the corresponding Figure 40 and Figure 41 in the Appendix section representing the CleanMX phishing dataset.

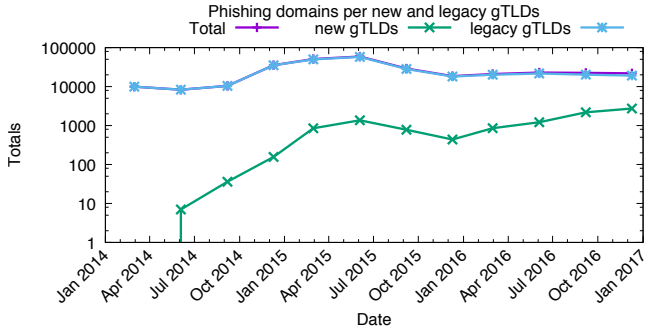

Figure 10. Time series of counts of phishing domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the Anti-Phishing Working Group feed (2014-2016). Please notice y axis in log scale and overlapping lines.

In the remainder of this subsection, we will only focus on the number of unique abused domains, as our metric. Figure 10 presents a time series of total counts of phishing domains, and those observed in legacy gTLDs and new gTLDs. Similar to before, we aggregate the phishing incidents on a quarterly basis and present the phishing counts using a logarithmic scale. As it is clear from the figure, the pink line overlaps largely with the blue line. That is mainly because the *total* number of phishing domains (purple line) has been driven by phishing domains in legacy gTLDs due to its disproportionate market share. 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 34 for SURBL phishing and Figure 42 for CleanMX phishing datasets).

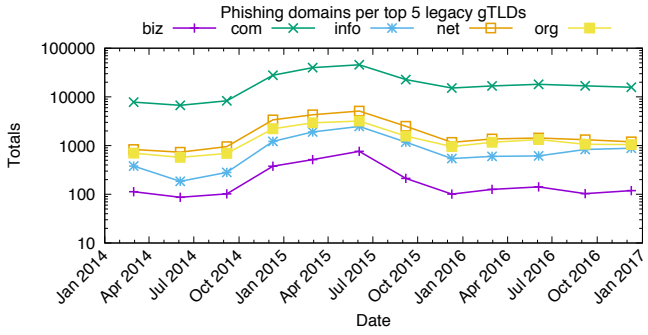

Figure 11. Time series of counts of phishing domains in the top 5 most abused **legacy** gTLDs in the last quarter of 2016 based on the Anti-Phishing Working Group feed (2014-2016).

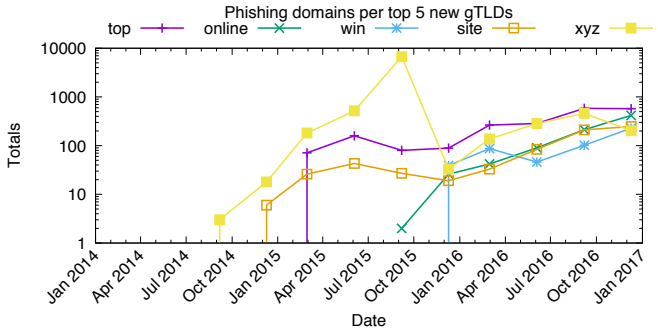

Figure 12. Time series of counts of phishing domains in the top 5 most abused **new** gTLDs in the last quarter of 2016 based on the Anti-Phishing Working Group feed (2014-2016).

Figure 11 and Figure 12 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[11]. The number of abused phishing domains in legacy gTLDs is mainly

[11] In Figure 12, 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 9 and Figure 10.

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 10) 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 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.

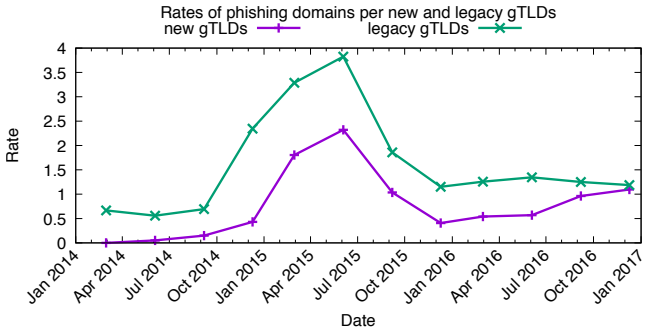

Figure 13. Time series of abuse rates of phishing domains in **legacy** gTLDs and **new** gTLDs based on the Anti-Phishing Working Group feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

As discussed before, reliable reputation metrics have to account for a commonly observed trend of the size that larger market players experience a larger amount of domain abuse [33], [36], [38]. Figure 13 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 43 for abused CleanMX phishing domains and Figure 35 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 13). 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 abuse rates for these two types separately. Figure 14

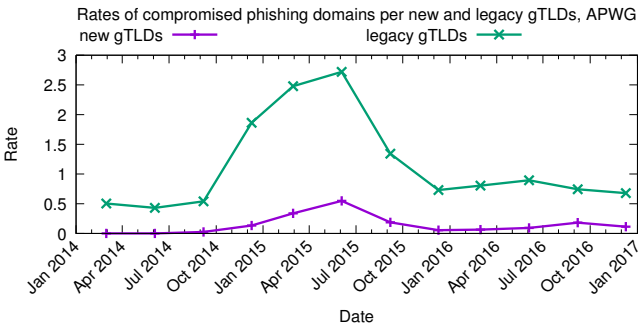

Figure 14. Time series of abuse rates of **compromised** phishing domains in **legacy** gTLDs and **new** gTLDs based on the Anti-Phishing Working Group feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#compromised\ domains/\#all\ domains$.

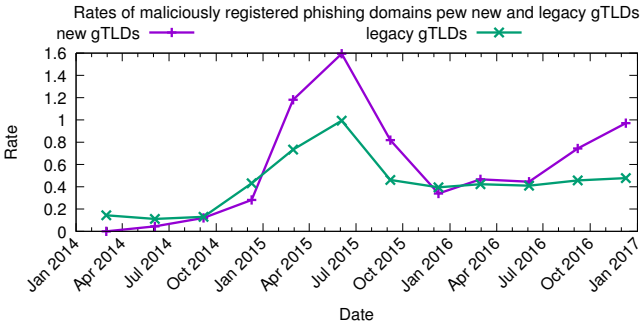

Figure 15. Time series of abuse rates of **maliciously** registered phishing domains in **legacy** gTLDs and **new** gTLDs based on the Anti-Phishing Working Group feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#maliciously\ registered\ domains/\#all\ domains$.

shows a time series of abuse rates for compromised phishing domains of legacy gTLDs and new gTLDs, based on the APWG feed. The curves corresponding to all blacklisted phishing domains and compromised phishing domains of legacy gTLDs (compare Figure 13 and Figure 14) follow a similar pattern due to a disproportionate concentration of compromised domains in legacy gTLDs.

Figure 15 shows a time series of abuse rates for maliciously registered phishing domains in legacy and new gTLDs in APWG feed. When we compare the rates of all blacklisted domains of new gTLDs with rates of maliciously registered domains (compare Figure 13 and Figure 15), we conclude that while there are higher relative concentrations of compromised domains in legacy gTLDs, the miscreants frequently choose to maliciously register domain names using one of the new gTLDs.

Moreover, we observe relatively higher rates of maliciously registered new gTLD domains in the first three quarters of 2015. By manual analysis of, for example, malicious domains blacklisted in the third quarter of 2015, we find 7,630 domains registered in 75 gTLDs (65 new gTLDs and 10 legacy gTLDs). The majority are .com domains (68.3%). We find 616 abused new gTLD domains. Interestingly, we observe as many as 182 and 111 abused .work and .xyz domains, respectively.

The results indicate that the majority of .work domains were registered by the same person. 150 domains were registered on the same day using the same registrant information, the same registrar, and the domain names were composed of similar strings. Note that only 150 abused domains, blacklisted in the third quarter of 2015, influenced the security reputation of all new gTLDs (see Figure 15).

Moreover, attackers 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 88 abused .top domains. 75 out of 88 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.

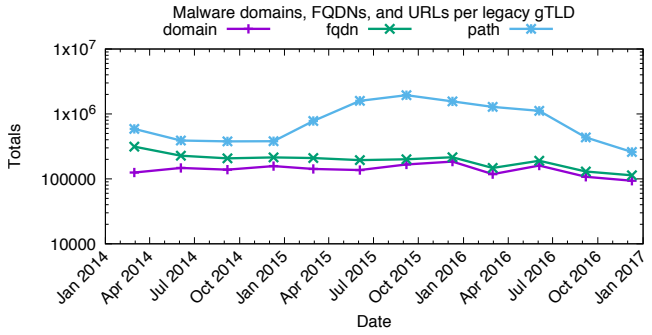

Figure 16. Time series of counts of malware domains, FQDNs, and URLs (paths) in **legacy** gTLD based on the StopBadware DSP feed (2014-2016).

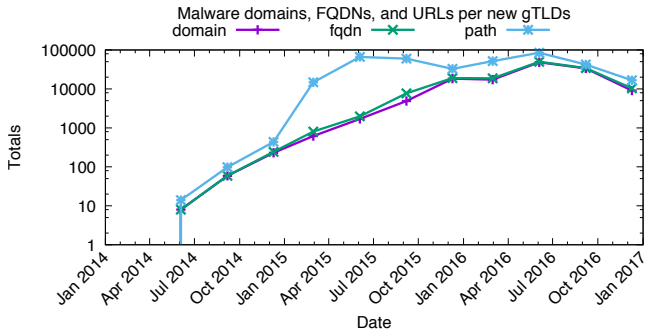

Figure 17. Time series of counts of malware domains, FQDNs, and URLs (paths) in **new** gTLD based on the StopBadware DSP feed (2014-2016).

2) Malware Reputation: We now analyze the malware activity reported by the StopBadware DSP. We refer the reader to Figure 16 and Figure 17 for overall absolute occurrence security metrics (see also Figure 44 and Figure 45 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).

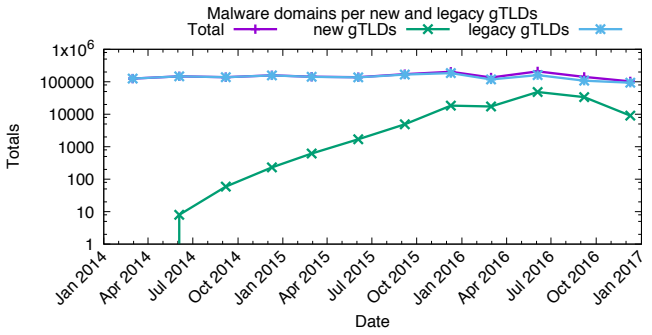

Figure 18. Time series of counts of malware domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the StopBadware DSP feed (2014-2016).

From this point forward, we only consider the number of unique domains. Figure 18 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 32 and Figure 46 presenting malware domains in legacy and new gTLDs for SURBL mw and CleanMX malware datasets confirm this trend.

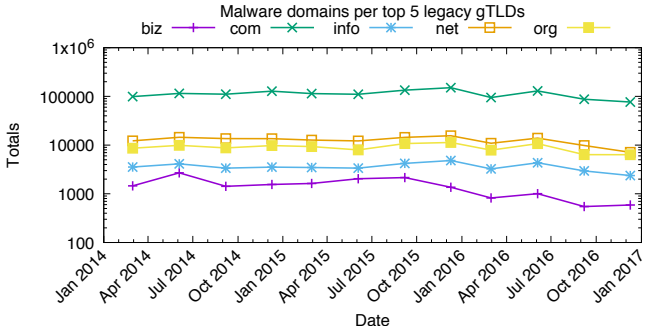

Figure 19. Time series of counts of malware domains in the top most abused 5 **legacy** gTLDs in the last quarter of 2016 based on the StopBadware DSP feed (2014-2016).

Figure 19 and Figure 20 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 7), 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 20. Specifically, in the fourth quarter of 2016, the relative score of the .xyz TLD is equal to 1.5 malware domain per 10,000

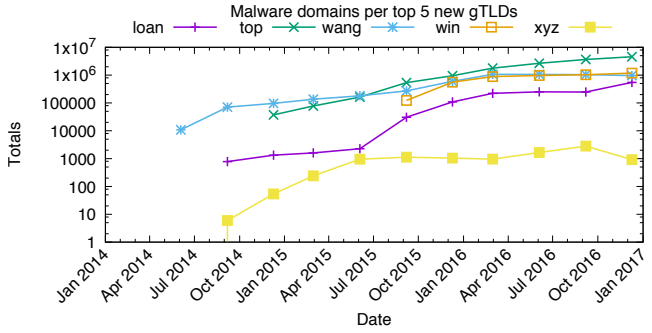

Figure 20. Time series of counts of malware domains in the top 5 most abused **new** gTLDs in the last quarter of 2016 based on the StopBadware DSP feed (2014-2016).

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.

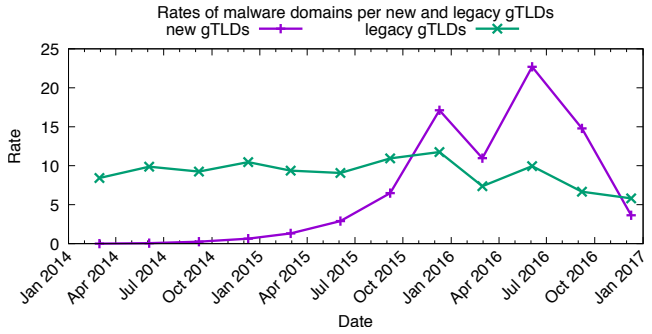

Figure 21. Time series of abuse rates of **malware domains** in legacy gTLDs and new gTLDs based on the StopBadware DSP feed (2014-2016). Rates are calculated as follows: $S = 10,000 * (\#blacklisted\ domains/\#all\ domains)$.

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 21). 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 33 and Figure 47.

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 22 and Figure 23 show

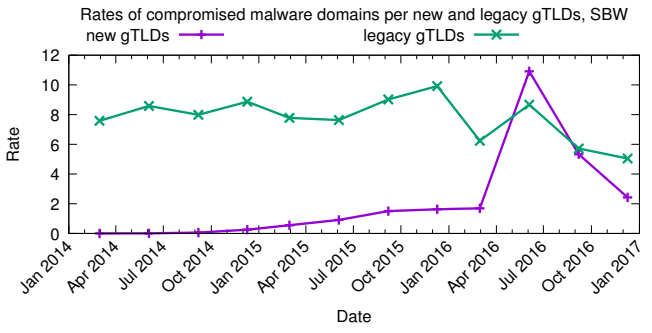

Figure 22. Time series of abuse rates of **compromised malware domains** in legacy gTLDs and new gTLDs based on the StopBadware DSP feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#compromised\ domains/\#all\ domains$.

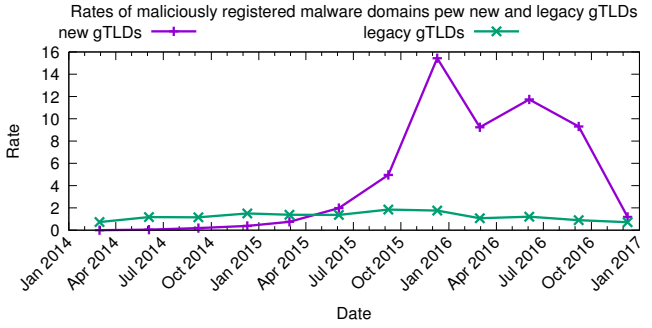

Figure 23. Time series of abuse rates of **maliciously registered malware domains** in legacy gTLDs and new gTLDs based on the StopBadware DSP feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#maliciously\ registered\ domains/\#all\ domains$.

time series of abuse rates of compromised and maliciously registered malware domains, respectively, in legacy gTLDs and new gTLDs. The results suggests that similarly to phishing, malware abuse rates in legacy gTLDs are mainly driven by compromised domains (compare Figure 21 with Figure 22). As expected, the malware rates for new gTLDs are driven by maliciously registered domains (compare Figure 21 with Figure 23).

More on details, 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 in new gTLDs (see Figure 23). Specifically, we found that 10,014 out of 16,591 domains (60.4%) labeled as malicious 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 (see Figure 22). 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 (77.1%, which represents 19,261 out of 24,987 domains). We find distinctive common patterns in domain name registration further suggesting malicious registrations. For example, we find 9,376 .link domains of which 9,256 were created in the

first quarter of 2016 and 9,253 were registered with Alpnames Limited registrar. 8,381 of all .link domains were registered using two registrar names only. Moreover, 8,205 and 1,027 were composed of 5 and 6 randomly generated characters, respectively. We created a user account with Alpnames Limited and tested bulk domain registration options. In fact, it is possible to randomly generate up to 2,000 domains at once from the selection of 27 new gTLDs using different patterns like letters, time, cities, zip codes, etc. Finally, note that the registries of .win, .loan, .top, and .link new gTLDs compete on price, and in 2016 their registration prices were occasionally 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.

3) Spam Reputation: We briefly discuss the spam activity in the new and legacy gTLDs reported by Spamhaus. Note that Spamhaus provides *domain* rather than *URL* blacklist, which means that the great majority of listed domains are maliciously registered. Figure 24 presents a time series of counts of spam domains observed in legacy gTLDs, new gTLDs, and the total number of all spam domains. We aggregate the unique spam domains on a quarterly basis and present the spam counts using a logarithmic scale. While previously we observed a clear upward trend in the absolute number of *phishing* and *malware* domains in new gTLDs, alarmingly, we now witness that the absolute number of malicious *spam* domains in new gTLDs is actually higher than in legacy gTLDs at the end of 2016. Note that the total number of spam incidents in all gTLDs is approximately constant and in the last quarter of 2016 is driven by incidents in new gTLDs (58.8%). Figure 36 and Figure 38 presenting spam domains in legacy and new gTLDs for SURBL jp and SURBL ws spam datasets, respectively, confirm the observed trend. In fact, the results reveal a new tendency: the attackers seems to switch from abusing legacy to new gTLDs.

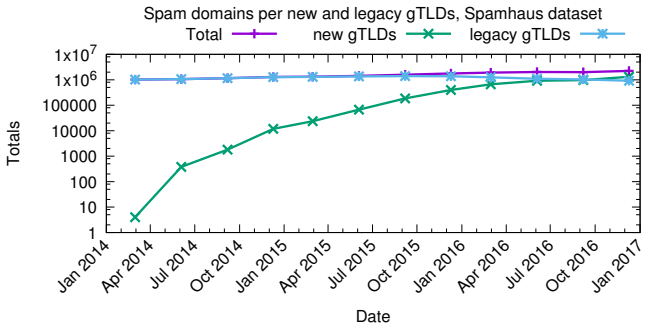

Figure 24. Time series of counts of blacklisted domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the **spamhaus** feed (2014-2016). Please notice y axis in log scale and overlapping lines.

Figure 25 shows a time series of abuse rates of spam domains of legacy gTLDs and new gTLDs based on the Spamhaus feed. For comparison, see Figure 37 and Figure 39 for spam domains blacklisted by SURBL jp and SURBL ws, respectively. As expected, the difference between spam abuse rates in legacy and new gTLDs is very significant. While legacy gTLDs collectively had a spam-domains-per-10,000 rate of 56.9, in the last quarter of 2016, the new gTLDs experienced a rate of 526.6–which is almost one order of magnitude higher. When comparing abuse rates of SURBL jp and SURBL ws spam feeds in the same period we observe a spam-domains-per-10,000 rates of 46.6 and 26 for legacy gTLDs, whereas for new gTLDs a spam-domains-per-10,000 rates are 286.3, and 265.2, respectively.

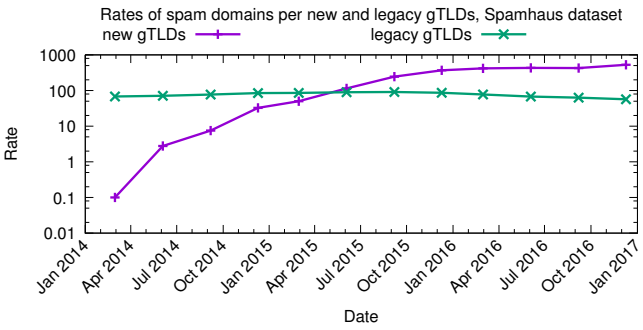

Figure 25. Time series of abuse rates of blacklisted domains in **legacy** gTLDs and **new** gTLDs based on the **spamhaus** feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

Table XXIV and Table XXV show top 10 new and legacy gTLDs, respectively, with the highest relative concentrations of blacklisted domains for all feeds in the fourth quarter of 2016. For more details we refer the reader to the appendix section. For example, spam-domains-per-10,000 rates calculated using Spamhaus feed for .science, .stream, and .study are equal to 5154, 4756 and 3343, respectively. In other words, as many as 51.5%, 47.6% and 33.4% of all domains in the corresponding zones were abused by cybercriminals and blacklisted by Spamhaus. Note that our results clearly indicate that the problem is not caused by a single or a few abused new gTLDs. For example, as many as 15 most abused new gTLDs had spam-domains-per-10,000 rates calculated using Spamhaus feed higher than 1,000 at the end of 2016.

B. Regression Analysis of Abuse in New gTLDs

One of the main goals of this study is to analyze the relationship between the collected security indicators and the structural properties of new gTLDs, and abuse, at the level of gTLDs. We use regression analysis to examine the amount of variance that those properties can collectively explain, from the total observed variance in the abuse counts.

Regression analysis has been used before to study the variation in phishing abuse across the population of various intermediaries such as hosting providers [38] and TLDs [33]. The intermediaries can be broken down into different potential drivers of variation in abuse counts, such as size, pricing, domain popularity index, or the amount of WordPress sites on phishing abuse [33], [38]. In this paper, we apply

Table IV
NEGATIVE BINOMIAL GLM FOR COUNT OF ABUSED DOMAINS PER NEW GTLD

	Dependent variable:						
	apwg	sbw	cmx ph	cmx pt	cmx mw	surbl ph	surbl mw
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
New gTLD size	0.00002***	0.00001***	0.00002***	0.00003***	0.00001***	0.00002***	0.00002***
	(0.00001)	(0.00000)	(0.00001)	(0.00001)	(0.00000)	(0.00001)	(0.00001)
Parked	0.0003***	0.0001***	0.0002***	0.00003	0.0001***	0.0002***	0.00001
	(0.00004)	(0.00003)	(0.00003)	(0.00004)	(0.00003)	(0.00004)	(0.00004)
DNSSEC	0.00001***	0.00002***	0.00002***	0.00002***	0.00001***	0.00002***	0.00002***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
No DNS	−0.00004***	−0.00003***	−0.00005***	−0.00005***	−0.00002***	−0.00004***	−0.00004***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00000)	(0.00001)	(0.00001)
HTTP Error	−0.00002	−0.00004***	−0.0001***	−0.00003*	−0.00004***	−0.0001***	−0.0001***
	(0.00002)	(0.00001)	(0.00001)	(0.00002)	(0.00001)	(0.00002)	(0.00002)
Type	−0.540**	−0.150	−0.400**	−0.120	−0.190	−0.760***	−0.170
	(0.220)	(0.120)	(0.180)	(0.170)	(0.160)	(0.190)	(0.220)
Constant	−0.630**	−0.390**	−0.960***	−1.200***	−1.600***	0.330	−2.200***
	(0.280)	(0.170)	(0.230)	(0.230)	(0.220)	(0.230)	(0.290)
Observations	521	521	521	521	521	521	521
Log Likelihood	−566.000	−792.000	−508.000	−546.000	−392.000	−786.000	−284.000
θ	0.140***	0.330***	0.240***	0.200***	0.470***	0.190***	0.240***
	(0.017)	(0.035)	(0.034)	(0.024)	(0.087)	(0.019)	(0.051)
AIC	1,149.000	1,600.000	1,031.000	1,109.000	800.000	1,588.000	583.000

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors in brackets

a similar statistical approach in order to analyze how the different properties of new gTLD operators relate to distinct types of abuse. We model the number of abused domains as a dependent variable using negative binomial generalized linear model (GLM) with a Log link function. We applied the negative binomial distribution because the abuse counts aggregated per new gTLDs proved to be overdispersed with respect to a Poisson distribution, for which the variance is equal to the mean. Note that negative binomial distribution is especially suitable for discrete data over a positive range whose sample variance exceeds significantly the sample mean.

Our regression models are built using the datasets explained in subsection III-A. We define our dependent variable as the number of abused domains (i.e. blacklisted domains or domain name elements of blacklisted URLs). Depending on the model, we use the total number of abused domains or treat maliciously registered and compromised domains separately (details follow later). The independent variables in the models are the following properties of new gTLDs: *"new gTLD size"*: number of domains in TLD, *"Parked"*: number of parked domains, *"No DNS"*: number of domains that do not resolve, *"HTTP Error"*: number of domains for which their websites return an HTTP error, *"DNSSEC"*: number of DNSSEC-signed domains, *"Type"*: an integer corresponding to the type of new gTLD, from least to most restricted group: 1 generic, 2 geographic, 3 community, and 4 brand, *"Registry"*: name of the registry operator that the TLD is operating under.

Table IV and Table V contain the summary of the regression models, i.e., the estimated coefficients, and their significance levels together with the goodness-of-fit measures such as the maximum Log likelihood, θ values and minimum Akaike information criterion (AIC) value (for more details, we refer the reader to the relevant literature). Table IV illustrates negative binomial GLMs for number of all abused domains per new gTLD. Table V illustrates GLM for number of maliciously registered and compromised phishing domains separately per new gTLD. Note that the presented models are chosen from a stepwise addition of the variables into a baseline model with a single explanatory variable.

Table IV indicates a positive and statistically significant correlation between new gTLD size and abuse counts. The results are very consistent for all the analyzed abuse feeds. The coefficients are, however, very weak. We suspect that this is because the majority of abused domains in the new gTLDs are maliciously registered rather than compromised.

As expected, two variables indicating the number of domains that do not serve valid Web content to their users, i.e. "No DNS" and "HTTP Error" show a weak significant relationship with abuse counts. That means, the more domains labelled as "No DNS" and/or "HTTP Error", the less abused domains. Those two variables also correspond to the count of compromised domains rather than maliciously

Table V
NEGATIVE BINOMIAL GLM FOR COUNT OF COMPROMISED AND MALICIOUSLY REGISTERED PHISHING DOMAINS PER NEW GTLD

	Response Variable: Count of domains in APWG data			
	Comp.	Mal.	Comp.	Mal.
	(1)	(2)	(3)	(4)
gTLD size			−0.00001***	0.00000
			(0.00000)	(0.00001)
Parked			0.0001***	0.0003***
			(0.00003)	(0.00005)
HTTP Err.			0.00003***	0.00001
			(0.00001)	(0.00002)
No DNS			0.00000	−0.00002***
			(0.00001)	(0.00001)
DNSSEC			0.00001***	0.00001***
			(0.00000)	(0.00000)
Type			−0.230	−0.780***
			(0.210)	(0.280)
Constant	−0.610***	1.500***	−1.800***	−0.190
	(0.180)	(0.220)	(0.270)	(0.350)
Obs.	521	521	521	521
Log Like.	−352	−545	−295	−495
θ	0.069***	0.039***	0.240***	0.080***
	(0.010)	(0.005)	(0.046)	(0.011)
AIC	705	1,091	604	1,004
Note:				*p<0.1; **p<0.05; ***p<0.01
				Standard errors in brackets

registered counts.

Moreover, the number of parked domains in new gTLDs plays a weak statistically significant role in explaining the variance in phishing and malware domains. The more parked domains in a new gTLD, the more abused domains. This is to be expected as landing pages of parked domains may serve malware on a large scale. Note that the coefficients are very small. For example, if we hold the other independent variables constant and increase the number of parked domains by one unit (which is the equivalent to multiplying the number of parked domains of a gTLD by 10 since it is in the log_{10} scale), the number of phishing domains in APWG is multiplied by $e^{0.0003} = 1.0003$.

Previous research indicates a negative significant relation between the DNSSEC deployment and the count of phishing domains [33]. The authors used DNSSEC deployment as a proxy for the security efforts of both ccTLD and gTLD registries. In our analysis we test the relationship between the number of DNSSEC domains and abuse counts from various types of blacklists for new gTLDs. Note that ICANN requires each new gTLD to demonstrate a plan for DNSSEC deployment to ensure integrity and utility of registry information. Therefore, in our analysis, the number of DNSSEC-signed domains cannot serve as a proxy for registry efforts and obviously it does not prevent phishing attacks. One may suspect that attackers could be interested in deploying DNSSEC and signing their maliciously registered domains. Although it is not clear if that is the case, we indeed observe a weak but positive and statistically significant correlation between the number of DNNSEC-signed domains and the number of abused domains.

The regression results consistently show a negative correlation between the "Type" variable reflecting strict registrations and count of phishing domains. In fact, in comparison to other variables, the obtained coefficients indicate the strongest statistically significant negative correlation for APWG, CleanMX phishing, and SURBL phishing datasets: -0.54, -0.4, and -0.76, respectively (see Table IV). Note that for all other considered datasets, in particular malware, we also observe negative but not statistically significant correlations. When we consider separately maliciously registered and compromised domains the "Type" of new gTLD plays a significant role in explaining phishing abuse counts only for malicious registrations (see Table V). Again, the results are intuitive. For example, it is much easier to register domains in the .top *standard* gTLD than it is for the .pharmacy *community* gTLD, for which the registration policy restricts the sale of domains to legitimate pharmacies only.

We have also considered other models that contain "Registry" as a fixed effect to capture systematic differences in the policies of registries across new gTLDs such as pricing, bulk registration options, etc. Interestingly, our results indicate that none of the registry operators have statistically significant effect on the abuse counts.

C. Privacy and 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 and Proxy services for abused domains in legacy gTLDs and new gTLDs. WHOIS Privacy and Proxy services are designed to conceal certain personal data of domain name registrants who use them. In practice this works by replacing the registrant information in `WHOIS` with the information of the WHOIS Privacy and Proxy service.

There are many legitimate reasons why someone may want to conceal possession of a domain name. The usage of a WHOIS Privacy and Proxy services by itself is, therefore not a reliable single indicator of malicious activity. A previous study by National Physical Laboratories [44], however did find that a significant portion of abusive domains use Privacy and Proxy services.

There are numerous WHOIS Privacy and 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 and 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 and 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 and Proxy services, we are able to correctly identify abusive domains that were using a WHOIS Privacy and Proxy service at the time the domain was added to a blacklist.

To get an indication of how common WHOIS Privacy and 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 and Proxy service when the domain was registered, we calculated what percentage of the total number of newly registered domains is using a Privacy and Proxy service (see Figure 26). We find that 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 18% and a standard deviation of 9.3, which is visualized by the larger spikes and the increase to above the level of legacy gTLDs near the end of the study period.

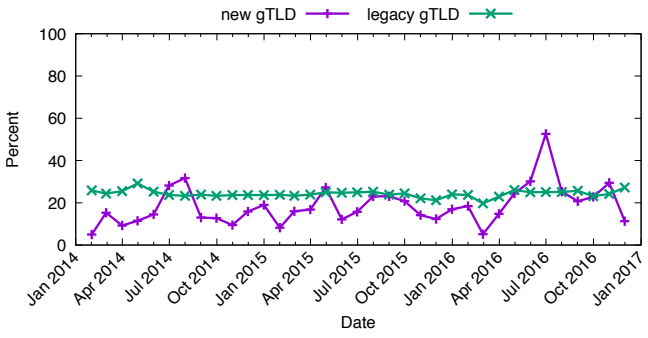

Figure 26. Usage percentage of Privacy and Proxy services for newly registered domains

Figure 27 shows the percentage of all newly created domains using Privacy and Proxy service, that have been reported to the Spamhaus or SURBL blacklist on or after the registration date. We have chosen to use Spamhaus and SURBL for this figure because these blacklists mainly contain maliciously registered domains. Here again, just as seen in Figure 26, we find that the variability for the new gTLDs is higher than compared to the legacy gTLDs. At the end of 2016 we find that both the new gTLD and legacy gTLD line show a similar increase.

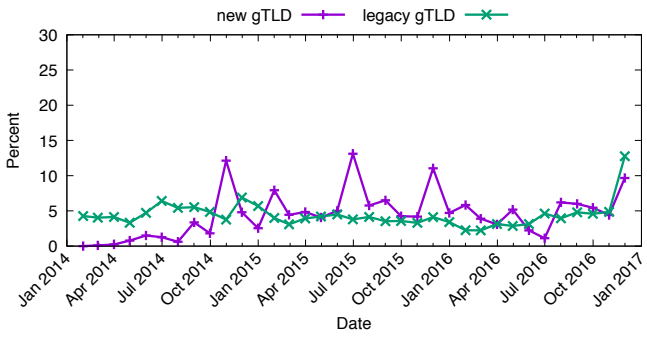

Figure 27. Percentage of abusive newly registered domains using Privacy and Proxy services

For each blacklist used in this study we analysed the proportion of domains that were using a Privacy and 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.

When look at two blacklist mainly driven by maliciously registered domains, all SURBL feeds combined (see Figure 28) and Spamhaus (see Figure 29), we find that the usage of Privacy and 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.

For SURBL in 2016 the mean usage per month of privacy and proxy services by abusive domains in new gTLD observed is 5,874 with a standard deviation of 1,984 (see Figure 28), while for legacy gTLDs the mean usage per month is 21,744 with a standard deviation of 9,475. For Spamhaus (see Figure 29) the 2016 new gTLDs mean usage per month is 8,951 with a standard deviation of 2,892, while for legacy gTLDs the mean usage per month is 16,569 with a standard deviation of 3,843.

In the SURBL data we find 2 large peaks (see Figure 28) of abusive new gTLD domains using Privacy and Proxy services. Both of these peaks are driven by the .xyz, .click and .link new gTLDs. We attempted to find peaks in new registration that correspond to the two peaks seen in Figure 28. In the 7-15 day period leading up to a peak we do see an increase in the number of new registrations for the .xyz, .click and .link new gTLDs with the same registrar. However, we do not find strong evidence that the malicious registrations belong to a single or multiple campaigns using WHOIS Privacy and Proxy services.

The analysis of the use of WHOIS Privacy and Proxy services leads us to conclude that the usage of a WHOIS Privacy and Proxy services by itself is not a reliable indicator of malicious activity. Apart from the peaks, the usage of Privacy and Proxy services for abusive domains is not that high (see Figure 28, Figure 29). The usage of Privacy proxy seems to be higher in legacy gTLDs.

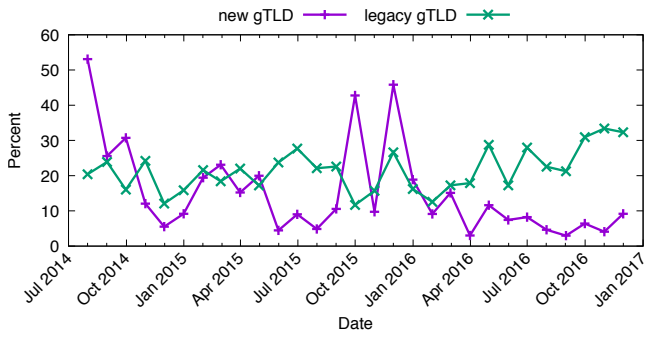

Figure 28. Usage of Privacy and Proxy services for abusive domains, reported by SURBL

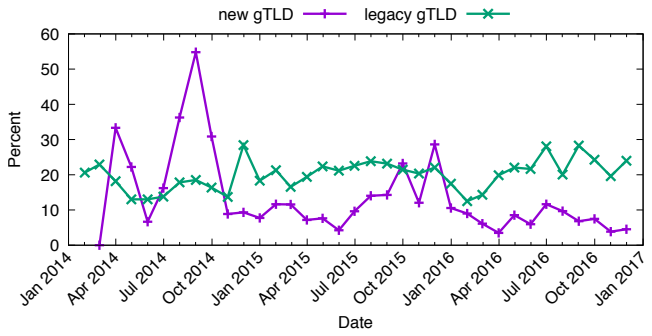

Figure 29. Usage of Privacy and Proxy services for abusive domains, reported by Spamhaus

D. 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 use the address of the domain's sponsoring registrar. Table VI lists the 10 countries hosting the most registrars, almost 54% of the identified registrars are located in the United States, which is almost 1 order of magnitude more compared to the number of registrars located in the second largest 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 VIII and Table IX, we find that the United States occupies the 3rd and 4th place.

Although the majority of the registrars are 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. These registrars are not uniformly distributed across countries, 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 VII 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 no geographical information for the registrar is 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 [45].
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.

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.

Table VI
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 VII
TOP10 LOCATIONS OF NEW AND LEGACY GTLD DOMAINS

New	#Domains	Share	Legacy	#Domains	Share
China	8,076,776	27.92	US	152,527,872	56.72
US	6,283,269	21.72	China	24,098,150	8.96
Gibraltar	3,028,035	10.47	Germany	18,044,735	6.71
Cayman Is.	2,069,919	7.16	Canada	16,704,693	6.21
Singapore	1,870,886	6.47	India	11,135,408	4.14
Japan	1,741,228	6.02	Japan	7,935,585	2.95
India	1,323,117	4.57	Australia	6,425,896	2.39
Germany	1,105,708	3.82	France	4,988,581	1.86
Hong Kong	836,069	2.89	UK	4,511,714	1.68
France	450,371	1.56	Turkey	2,418,232	0.9

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 VIII shows that for 37.09% of the abused new gTLD domains reported by StopBadware, the sponsoring registrar is located in Gibraltar. Almost 195 abused new gTLD domains per 10,000 located in Gibraltar are abusive.

The results in Table VIII, Table IX and Table X 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 53.97% (59,044) of the new gTLD domains that have been blacklisted by Spamhaus. Moreover, note that for new gTLDs, the spam-domains-per-10,000 rate reported by Spamhaus for Gibraltar is equal to 3,991 (Table X), whereas for example for APWG only 0.56 (Table XI). 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.

E. Registrar Reputation

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

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 XVI lists the number reported incidents for "Nanjing Imperiosus Technology" as 35,502, with a total number of 38,025 under its management, this 93.36% of all new gTLDs of this registrar in the WHOIS data are reported by the SURBL blacklist.

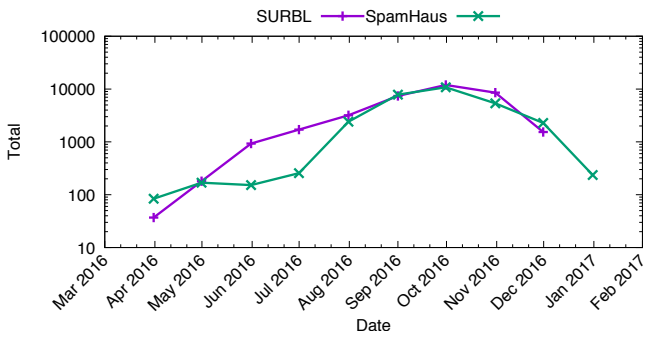

Figure 30. Abusive domains managed by Nanjing Imperiosus Technology Co. Ltd

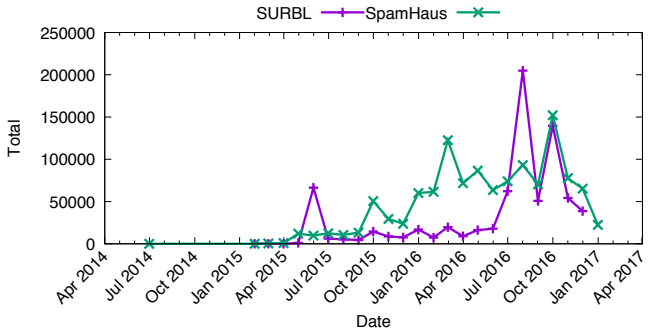

Figure 31. Abusive domains managed by Alpnames Limited

Table XVI and Table XVII list the registar "Nanjing Imperiosus Technology Co. Ltd." as an outlier, over 90% of its domains are reported as abusive by SURBL and 78% by Spamhaus. Figure 30 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 [46] 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 31 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

Table VIII
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	59,044	37.09	194.99	United States	578,538	51.42	37.93
2	China	55,957	35.15	184.8	China	198,995	17.69	13.05
3	Singapore	24,475	15.38	80.83	India	86,293	7.67	5.66
4	United States	7,786	4.89	25.71	Canada	49,274	4.38	3.23
5	India	6,969	4.38	23.01	Germany	46,097	4.1	3.02
6	United Kingdom	915	0.57	3.02	France	20,712	1.84	1.36
7	Hong Kong	859	0.54	2.84	United Kingdom	16,099	1.43	1.06
8	Barbados	515	0.32	1.7	Spain	14,317	1.27	0.94
9	France	420	0.26	1.39	Turkey	14,261	1.27	0.93
10	Japan	399	0.25	1.32	Hong Kong	13,174	1.17	0.86

Table IX
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	751,748	49.44	2482.63	United States	1,985,574	47.06	130.18
2	Japan	295,647	19.44	976.37	Japan	1,190,409	28.21	78.05
3	China	214,332	14.1	707.83	China	319,546	7.57	20.95
4	United States	109,989	7.23	363.24	India	268,812	6.37	17.62
5	India	54,782	3.6	180.92	Germany	73,185	1.73	4.8
6	United Kingdom	24,955	1.64	82.41	Ireland	58,292	1.38	3.82
7	France	20,121	1.32	66.45	Canada	40,355	0.96	2.65
8	United Arab Emirates	11,793	0.78	38.95	Australia	33,080	0.78	2.17
9	Cayman Islands	8,912	0.59	29.43	Turkey	32,266	0.76	2.12
10	Canada	6,494	0.43	21.45	Bahamas	28,918	0.69	1.9

Table X
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	1,208,509	53.97	3991.07	United States	2,118,994	47.09	138.93
2	Japan	263,899	11.78	871.52	Japan	1,131,316	25.14	74.17
3	China	228,728	10.21	755.37	China	390,860	8.69	25.63
4	United States	221,577	9.89	731.75	India	291,398	6.48	19.1
5	Singapore	123,290	5.51	407.16	Turkey	92,475	2.06	6.06
6	India	75,040	3.35	247.82	Germany	74,919	1.66	4.91
7	United Kingdom	32,352	1.44	106.84	Bahamas	73,290	1.63	4.81
8	France	25,807	1.15	85.23	Canada	70,700	1.57	4.64
9	Cayman Islands	17,891	0.8	59.08	Australia	40,620	0.9	2.66
10	Hong Kong	7,592	0.34	25.07	United Kingdom	28,631	0.64	1.88

Table XI
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	3,551	36.51	5.65	United States	155,928	60.95	10.22
2	China	2,673	27.48	4.25	India	27,027	10.56	1.77
3	India	695	7.15	1.11	Canada	14,096	5.51	0.92
4	Germany	363	3.73	0.58	Germany	9,174	3.59	0.6
5	Gibraltar	350	3.6	0.56	China	8,085	3.16	0.53
6	United Kingdom	297	3.05	0.47	Australia	5,406	2.11	0.35
7	Canada	291	2.99	0.46	United Kingdom	4,946	1.93	0.32
8	Singapore	281	2.89	0.45	France	4,056	1.59	0.27
9	Japan	201	2.07	0.32	Turkey	3,876	1.52	0.25
10	Hong Kong	177	1.82	0.28	Bahamas	2,528	0.99	0.17

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 US $1 or less. We did not find corresponding peaks in the size of the

Table XII
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	4,970	55.16	7.91	United States	169,251	60.57	11.1
2	Gibraltar	1,040	11.54	1.66	India	34,101	12.2	2.24
3	India	793	8.8	1.26	Canada	17,609	6.3	1.15
4	China	719	7.98	1.14	Germany	9,184	3.29	0.6
5	United Kingdom	198	2.2	0.32	China	7,329	2.62	0.48
6	Canada	197	2.19	0.31	Australia	6,502	2.33	0.43
7	Germany	168	1.86	0.27	United Kingdom	4,996	1.79	0.33
8	Singapore	133	1.48	0.21	France	4,281	1.53	0.28
9	Russian Federation	105	1.17	0.17	Turkey	3,878	1.39	0.25
10	France	84	0.93	0.13	Bahamas	2,164	0.77	0.14

Table XIII
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	3,124	41.89	4.97	United States	216,414	57.4	14.19
2	India	1,359	18.22	2.16	India	48,177	12.78	3.16
3	Gibraltar	1,018	13.65	1.62	Canada	21,706	5.76	1.42
4	China	691	9.27	1.1	China	20,226	5.36	1.33
5	France	244	3.27	0.39	Germany	12,229	3.24	0.8
6	Singapore	149	2.0	0.24	Turkey	6,983	1.85	0.46
7	United Kingdom	119	1.6	0.19	France	6,966	1.85	0.46
8	Germany	109	1.46	0.17	United Kingdom	6,165	1.64	0.4
9	Netherlands	109	1.46	0.17	Australia	5,545	1.47	0.36
10	Russian Federation	99	1.33	0.16	Spain	3,849	1.02	0.25

Table XIV
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
2	United States	2,652	28.19	3.28	China	41,962	10.9	2.75
3	Gibraltar	1,839	19.55	2.28	India	30,765	7.99	2.02
4	India	397	4.22	0.49	Canada	17,149	4.46	1.12
5	United Kingdom	312	3.32	0.39	Germany	14,168	3.68	0.93
6	Cayman Islands	195	2.07	0.24	France	7,576	1.97	0.5
7	Singapore	148	1.57	0.18	Spain	6,333	1.65	0.42
8	Japan	110	1.17	0.14	Turkey	5,907	1.54	0.39
9	France	110	1.17	0.14	United Kingdom	5,762	1.5	0.38
10	Germany	109	1.16	0.13	Japan	5,134	1.33	0.34

Table XV
SDF 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	18,675	68.93	29.72	United States	151,517	61.51	9.93
2	China	2,233	8.24	3.55	India	24,777	10.06	1.62
3	Cayman Islands	1,779	6.57	2.83	Canada	12,764	5.18	0.84
4	India	875	3.23	1.39	Germany	9,030	3.67	0.59
5	United Kingdom	584	2.16	0.93	China	8,620	3.5	0.57
6	Gibraltar	470	1.73	0.75	Australia	4,746	1.93	0.31
7	Germany	449	1.66	0.71	United Kingdom	4,696	1.91	0.31
8	Japan	323	1.19	0.51	France	3,783	1.54	0.25
9	Canada	309	1.14	0.49	Turkey	3,417	1.39	0.22
10	Singapore	230	0.85	0.37	Bahamas	2,504	1.02	0.16

.top and .science zone files, indicating the abusive domains have been registered over a longer period of time.

VI. RELATED WORK

To mitigate domain name abuse more effectively different classes of intermediaries such as registries, registrars or host-

Table XVI
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	38,025	35,502	93.36	HOAPDI INC.	141	126	89.36
2	Intracom Middle East FZE	20,640	11,255	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	10,834	30.68
4	Alpnames Limited	3,028,011	751,748	24.83	Paknic (Private) Limited	10,525	3,083	29.29
5	Todaynic.com, Inc.	329,399	69,404	21.07	OwnRegistrar, Inc.	22,188	5,238	23.61
6	Web Werks India Pvt. Ltd	785	146	18.6	Eranet International Limited	6,109	1,339	21.92
7	GMO Internet, Inc. d/b/a Onamae.com	1,734,775	295,641	17.04	BR domain Inc. dba namegear.co	847	158	18.65
8	TLD Registrar Solutions Ltd.	163,988	24,700	15.06	Netlynx Inc.	17,612	3,030	17.2
9	Xiamen Nawang Technology Co., Ltd	282,925	42,089	14.88	AFRIREGISTER S.A.	1,551	266	17.15
10	Instra Corporation Pty Ltd.	77,642	6,200	7.99	GMO Internet, Inc. d/b/a Onamae.com	7,306,312	1,177,886	16.12

Table XVII
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	38,025	29,682	78.06	ABSYSTEMS INC	688	632	91.86
2	Alpnames Limited	3,028,011	1,208,509	39.91	Eranet International Limited	6,109	4,074	66.69
3	Shanghai Best Oray Information S&T	3,600	1,324	36.78	Ednit Software Private Limited	524	285	54.39
4	Dot Holding Inc.	153	50	32.68	Dynamic Dolphin, Inc.	12,515	5,870	46.9
5	MAT BAO CORPORATION	3,116	746	23.94	Webair Internet Development, Inc.	19,607	7,484	38.17
6	NameSilo, LLC	31,084	6,718	21.61	asia registry r2-asia (700000)	1,379	460	33.36
7	Zhengzhou Century Connect Elec. Tech. Dev.	16,057	3,235	20.15	Nanjing Imperiosus Technology	35,309	11,475	32.5
8	TLD Registrar Solutions Ltd.	163,988	32,043	19.54	Alpnames Limited	27,558	7,604	27.59
9	Netowl, Inc.	1,190	206	17.31	GoName-TN.com, Inc.	7,088	1,815	25.61
10	GMO Internet, Inc. d/b/a Onamae.com	1,734,775	263,681	15.2	Paknic (Private) Limited	10,525	2,553	24.26

Table XVIII
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	8,078	148	1.83	Minds and Machines LLC	1,115	117	10.49
2	AB Name ISP	1,074	14	1.3	Tecnologia, Desarrollo Y Mercado S.	2,027	128	6.31
3	OpenTLD B.V.	793	5	0.63	BR domain Inc. dba namegear.co	847	18	2.13
4	BigRock Solutions Ltd.	3,465	11	0.32	Abu-Ghazaleh Intellectual Property	1,365	27	1.98
5	DOTSERVE INC.	10,782	34	0.32	Shinjiru Technology Sdn Bhd	16,134	256	1.59
6	Shenzhen HuLianXianFeng Technology	6,125	19	0.31	Naugus Limited LLC	7,803	102	1.31
7	Shanghai Meicheng Technology Inf. Dev.	50,501	151	0.3	Upperlink Limited	4,527	56	1.24
8	FBS Inc.	56,438	164	0.29	Rethem Hosting LLC	3,841	38	0.99
9	Paragon Internet Group Ltd	3,645	10	0.27	Eranet International Limited	6,109	45	0.74
10	CV. Rumahweb Indonesia	10,822	26	0.24	Danesco Trading Ltd.	186,035	1,214	0.65

Table XIX
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	282,925	12,427	4.39	Rethem Hosting LLC	3,841	772	20.1
2	Foshan YiDong Network Co., LTD	45,556	1,698	3.73	0101 Internet, Inc.	8,317	576	6.93
3	Netowl, Inc.	1,190	43	3.61	Zhengzhou Zitian Network Technology	12,235	555	4.54
4	Super Registry Ltd	21,322	515	2.42	Xiamen Nawang Technology Co., Ltd	206,700	5,762	2.79
5	Alpnames Limited	3,028,011	59,044	1.95	Minds and Machines LLC	1,115	26	2.33
6	Jiangsu Bangning Science & technology	200,323	3,310	1.65	Danesco Trading Ltd.	186,035	4,297	2.31
7	Todaynic.com, Inc.	329,399	4,999	1.52	In2net Network Inc.	106,992	2,432	2.27
8	Alibaba Cloud Computing	1,859,602	24,474	1.32	Shanghai Oweb Network Co., Ltd	149	3	2.01
9	CV. Rumahweb Indonesia	10,822	98	0.91	CyanDomains, Inc.	16,965	337	1.99
10	Web Werks India Pvt. Ltd	785	7	0.89	Key-Systems, LLC	161	3	1.86

Table XX
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	OpenTLD B.V.	793	3	0.38	Minds and Machines LLC	1,115	108	9.69
2	AB Name ISP	1,074	4	0.37	Shinjiru Technology Sdn Bhd	16,134	280	1.74
3	Web4Africa Inc.	2,445	9	0.37	Upperlink Limited	4,527	74	1.63
4	CV. Rumahweb Indonesia	10,822	33	0.3	BR domain Inc. dba namegear.co	847	8	0.94
5	BigRock Solutions Ltd.	3,465	8	0.23	Launchpad.com Inc.	1,110,454	10,069	0.91
6	DOTSERVE INC.	10,782	24	0.22	Eranet International Limited	6,109	51	0.83
7	Shenzhen HuLianXianFeng Technology	6,125	10	0.16	Web4Africa Inc.	22,418	187	0.83
8	10dencehispahard, S.L.	6,455	10	0.15	Rethem Hosting LLC	3,841	32	0.83
9	Marcaria.com International, Inc.	14,886	23	0.15	Dattatec.com SRL	196,950	1,307	0.66
10	ZNet Technologies Pvt Ltd.	1,365	2	0.15	Name121, Inc.	17,626	113	0.64

Table XXI
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	0101 Internet, Inc.	8,317	262	3.15
2	Foshan YiDong Network Co., LTD	45,556	209	0.46	Minds and Machines LLC	1,115	26	2.33
3	Xiamen Nawang Technology Co., Ltd	282,925	899	0.32	Soluciones Corporativas IP, SL	197,864	3,029	1.53
4	Authentic Web Inc.	1,179	3	0.25	Rethem Hosting LLC	3,841	51	1.33
5	Netowl, Inc.	1,190	3	0.25	Pheenix 7, LLC	314	4	1.27
6	TLD Registrar Solutions Ltd.	163,988	289	0.18	Danesco Trading Ltd.	186,035	1,702	0.91
7	Eranet International Limited	42,154	58	0.14	CloudFlare, Inc.	221	2	0.9
8	Dynadot, LLC	94,891	125	0.13	Paknic (Private) Limited	10,525	93	0.88
9	CV. Rumahweb Indonesia	10,822	13	0.12	IPNIC, Inc.	687	6	0.87
10	Jiangsu Bangning Science & technology	200,323	145	0.07	UKRNAMES	69,480	558	0.8

Table XXII
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	Hosting Concepts B.V. d/b/a Openprovider	32,341	105	0.32	Minds and Machines LLC	1,115	65	5.83
2	Marcaria.com International, Inc.	14,886	19	0.13	0101 Internet, Inc.	8,317	122	1.47
3	PDR Ltd. d/b/a PublicDomainRegistry.com	1,299,611	1,354	0.1	Shinjiru Technology Sdn Bhd	16,134	164	1.02
4	NameCheap, Inc.	1,961,146	1,979	0.1	ZNet Technologies Pvt Ltd.	50,396	466	0.92
5	Gandi SAS	178,282	160	0.09	Name121, Inc.	17,626	151	0.86
6	BigRock Solutions Ltd.	3,465	3	0.09	Upperlink Limited	4,527	36	0.8
7	FBS Inc.	56,438	48	0.09	Web Site Source, Inc.	5,528	41	0.74
8	Register NV dba Register.eu	26,287	22	0.08	Catalog.com	28,738	213	0.74
9	CV. Rumahweb Indonesia	10,822	9	0.08	OwnRegistrar, Inc.	22,188	161	0.73
10	Regtime Ltd.	12,199	10	0.08	Mps Infotecnics Limited	8,352	59	0.71

Table XXIII
SDF 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	8,078	155	1.92	Minds and Machines LLC	1,115	115	10.31
2	AB Name ISP	1,074	12	1.12	Tecnologia, Desarrollo Y Mercado S.	2,027	123	6.07
3	OpenTLD B.V.	793	5	0.63	Abu-Ghazaleh Intellectual Property	1,365	30	2.2
4	NameCheap, Inc.	1,961,146	12,161	0.62	BR domain Inc. dba namegear.co	847	17	2.01
5	Bizcn.com, Inc.	124,345	464	0.37	Shinjiru Technology Sdn Bhd	16,134	271	1.68
6	Nanjing Imperiosus Technology	38,025	133	0.35	Eranet International Limited	6,109	91	1.49
7	BigRock Solutions Ltd.	3,465	11	0.32	Naugus Limited LLC	7,803	105	1.35
8	TLD Registrar Solutions Ltd.	163,988	476	0.29	Upperlink Limited	4,527	55	1.21
9	FBS Inc.	56,438	158	0.28	Rethem Hosting LLC	3,841	29	0.76
10	Shenzhen HuLianXianFeng Technology	6,125	17	0.28	Paknic (Private) Limited	10,525	74	0.7

ing providers should be able to benchmark themselves against their market. Currently, there exists very little empirical information about the security of TLDs, in particular new gTLDs. However, a number of studies included security metrics as a part of their analysis.

Levchenko *et al.* found some registrars, ASes and banks which are disproportionately popular among criminals, possibly due to their security practices [47]. Moore and Edelman found a concentration of typosquatted domains on a small number of name servers [48]. Korczyński *et al.* illuminated the problem of non-secure DNS dynamic updates, which allow a cybercriminal to manipulate DNS entries in zone files of authoritative name servers. They found that 66.2% of vulnerable domains are hosted on the infrastructure of a single broadband Internet Service Provider (ISP). Reconfiguring zone files of just 10 providers would reduce the prevalence of the problem with 88.6% [49]. Ma *et al.* used name server and registrar information to distinguish malicious URLs from benign ones [50]. Hao *et al.* observed that 46% of the spam domains come from just two registrars [51]. However, they considered only the `.com` TLD and did not consider the size estimate for smaller registrars which might register a disproportionate amount of malicious domains. Antonakakis *et al.* developed a dynamic reputation system using passive DNS data to classify legitimate and malicious domains and assign a reputation score to the new domains [52]. Our work does not rank individual domains but rather designs reputation metrics for the legacy and new gTLDs.

Numerous studies attributed security incidents to hosting providers by equating them with ASes. The number of incidents is often normalized by the AS size [36], [53]. Mahjoub investigated the concentration of abuse in ASes by analyzing hosted content, AS topology and IP space reservation [54]. Noorozian *et al.* presented a systematic approach for metrics development and identify the main challenges that plague metric design [36]. In the process, they answer an urgent question posed to them by the Dutch police: "Which are the worst hosting providers under our jurisdiction?". In their follow-up work, Noorozian *et al.* presented a causal model of security incidents using seven abuse datasets and then proposed a new modelling approach to enable better measurement of security performance from abuse data [55]. Other studies identified malicious ASes using AS topology, BGP-related features and by exploring ASes providing transit for malicious ASes [56]–[58].

Only a few studies on DNS abuse in TLDs have been conducted. Rasmussen and Aaron regularly release Anti-Phishing Working Group Global Phishing Reports, in which they examine phishing datasets collected by APWG and several other supplementary phishing feeds. Recently, they concluded that phishing in the new gTLDs is rising, but is not yet as pervasive as it is in the domain space as a whole [41]. Halvorson *et al.* found that new gTLD domains are more than twice as likely as legacy TLDs to appear on a domain blacklist within their first month of registration [1]. In the most similar study to our paper, Korczyński *et al.* designed security metrics to measure and benchmark entire TLDs against their market [33]. They explicitly distinguished the metrics from the idea of measuring security performance because the measured values of the proposed metrics are driven by multiple factors, not just by the performance of the particular market player. They found that next to TLD size, abuse primarily correlates with domain pricing (free versus paid registrations), efforts of intermediaries (measured through the proxy of their DNSSEC deployment rate), and strict registration policies [33]. In this paper, we extend their methodology and perform the first comprehensive statistical comparison of rates of DNS abuse in new and legacy gTLDs as they pertain to spam, phishing, and malware distribution.

VII. Discussion and conclusions

The ICANN New gTLD Program enabled hundreds of new gTLDs to enter into the DNS since the first delegations occurred in late 2013. A number of security measures were built into the Program to preemptively mitigate the rates of abusive, malicious, and criminal activity in these new gTLDs. In this paper, we performed the first comprehensive study examining rates of malicious and abusive behavior in new and legacy gTLDs to evaluate the effectiveness of the proposed safeguards. We used data sets from many sources, including zone files, domain `WHOIS` information, data obtained through our active measurements, and 11 heterogeneous domain and URL blacklists representing malware, phishing, and spam generously provided to us by five reputable organisations.

We found that the *absolute* number of phishing domains 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 and malware domains in new gTLDs. The phishing and malware abuse *rates* in legacy and new gTLDs, however, are converging with time and are very similar at the end of 2016.

The analysis of spam feeds revealed that the *absolute* number of spam domains in new gTLDs is higher than in legacy gTLDs at the end of 2016. Interestingly, we find that the new gTLDs have impacted spam counts of the legacy gTLDs: abused domains in the new gTLDs do not increase the number of total malicious registrations. Instead, we observed a decrease in the number of malicious registrations in legacy gTLDs. Moreover, while legacy gTLDs collectively had a spam-domains-per-10,000 rate of 56.9, in the last quarter of 2016, the new gTLDs experienced a rate of 526.6–which is almost one order of magnitude higher. The analysis of the three most abused new gTLDs show that 51.5%, 47.6% and 33.4% of all registered domains were abused by cybercriminals and blacklisted by Spamhaus in the last quarter of 2016.

Does the problem affect all new gTLDs? No. Our analysis of Spamhaus and SURBL blacklists reveals that approximately 32% and 36% of all new gTLDs available for registration did not experience a single incident in the last quarter of 2016. On the other hand, Spamhaus blacklisted at least 10% of all

registered domains in as many as 15 new gTLDs at the end of 2016.

While we found higher concentrations of *compromised* domains in legacy gTLDs, miscreants frequently choose to *maliciously register* domain names using one of the new gTLDs. The registry operators of the most abused new gTLDs compete on price. We found that their retail registration prices were occasionally below US $1 or even US $0.50, which was lower than the registration fee for .com domains. It is not clear, however, if pricing is the only factor driving high concentrations of maliciously registered domains. We created a user account with one registrar suffering from disproportionately high concentration of abused domains and tested bulk domain registration options. It is possible, for example, to randomly generate up to 2,000 domains at once for a selection of 27 new gTLDs using different patterns including letters, time, cities, zip codes, etc.

We also systematically analyzed how different structural and security-related properties of new gTLD operators influence abuse counts. As expected, we found that the number of domains in new gTLDs, the number of parked domains, or the number of DNSSEC-signed domains play a statistically significant but very weak role in explaining the differences in abuse counts between different new gTLDs. Our inferential analysis revealed that abuse counts primarily correlate with strict registrations. Miscreants prefer to register, for example, *standard* new gTLD domain names, which are generally open for public registration, rather than *community* new gTLDs for which registries may impose restrictions on who or which entities can register their domains. In our models, we also considered the name of the registry operators to capture systematic differences in the policies of registries across new gTLDs such as pricing, bulk registration options, etc. In other words, we tested the correlation between registry operators and domain abuse counts. However, we did not find any statistically significant effects on the abuse counts. In future work, we plan to collect detailed data on registration policies across all new gTLDs and perform a more fine-grained analysis on factors that may also influence abuse counts.

Our findings suggest that some new gTLDs have become a growing target for malicious actors. Competitive domain registration prices, unrestrictive registration practices, a variety of other registration options such as available payment methods, free services such as DNS or WHOIS privacy, and finally the increased availability of domain names decrease barriers to abuse and may make some new gTLDs targets for cybercriminals.

ACKNOWLEDGEMENTS

This study was commissioned by the Competition, Consumer Trust, and Consumer Choice Review Team with the support of ICANN. We would like to thank ICANN, DomainTools, Whois XML API, Spamhaus, SURBL, StopBadware, CleanMX, Secure Domain Foundation, Anti-Phishing Working Group for providing access to their data. Authors also thank Roland van Rijswijk for his help in obtaining additional domain data and anonymous reviewers for their constructive and valuable comments.

REFERENCES

[1] 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.
[2] "Internet Corporation for Assigned Names and Numbers (ICANN)," https://www.icann.org.
[3] J. Postel and J. Reynolds, "Domain requirements," Internet Requests for Comments, RFC Editor, RFC 920, October 1984.
[4] ICANN, "New gTLD Program," https://icannwiki.com/New_gTLD_Program, 2017.
[5] ——, "New gTLD Program," https://icannwiki.org/New_gTLD_Program, February 2017.
[6] ——, ".madrid," https://icannwiki.org/.madrid, March 2015.
[7] ——, "New gTLD Program," https://icannwiki.org/New_gTLD_Generic_Applications, February 2017.
[8] ——, "New gTLD Program," https://icannwiki.org/Community_TLD, February 2017.
[9] ——, "New gTLD Program," https://icannwiki.org/New_gTLD_Geographic_Applications, February 2017.
[10] ——, "New gTLD Program," https://icannwiki.org/New_gTLD_Brand_Applications, February 2017.
[11] ——, "New gTLD Program Explanatory Memorandum: Mitigating Malicious Conduct," https://archive.icann.org/en/topics/new-gtlds/mitigating-malicious-conduct-04oct09-en.pdf, October 2009.
[12] ——, "New gTLD Program Safeguards Against DNS Abuse," https://newgtlds.icann.org/en/reviews/dns-abuse/safeguards-against-dns-abuse-18jul16-en.pdf, July 2016.
[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] "The Secure Domain Foundation," https://securedomain.org/.
[18] "Spam-Filter Anti-Spam Virenschutz," http://clean-mx.de.
[19] "The Domain Block List," https://www.spamhaus.org/dbl.
[20] "ESET: Security Software," http://www.eset.com.
[21] "Fortinet: Network & Computer Security," http://www.fortinet.com.
[22] "Sophos: Computer Security, Antivirus," http://www.sophos.com.
[23] "StopBadware: Data Sharing Program," https://www.stopbadware.org/data-sharing.
[24] "SURBL Lists," http://www.surbl.org/lists.
[25] "Malwarebytes," https://www.malwarebytes.com/.
[26] "Public Suffix List," https://publicsuffix.org.
[27] 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.
[28] ——, "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.
[29] "Whois XML API," https://www.whoisxmlapi.com/.
[30] ICANN, "TLD Startup Information," https://newgtlds.icann.org/en/program-status/sunrise-claims-periods, Retrieved on February 2017.
[31] "DomainTools: Domain Whois Lookup, Whois API & DNS Data Research," http://www.domaintools.com.
[32] "ICANN: .zuerich TLD," https://icannwiki.org/.zuerich.
[33] 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.
[34] 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.
[35] "ZeusTracker: A Nonprofit Organization Tracking ZeuS C&C Servers." https://zeustracker.abuse.ch.

[36] 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.
[37] ICANN, "Monthly Registry Reports," https://www.icann.org/resources/pages/registry-reports.
[38] 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
[39] T. Vissers, W. Joosen, and N. Nikiforakis, "Parking sensors: Analyzing and detecting parked domains." in *Network and Distributed System Security Symposium (NDSS)*, 2015.
[40] "IANA: Registrar IDs," https://www.iana.org/assignments/registrar-ids/registrar-ids.xhtml.
[41] G. Aaron and R. Rasmussen, "Global Phishing Survey: Trends and Domain Name Use in 2016," http://docs.apwg.org/reports/APWG_Global_Phishing_Report_2015-2016.pdf.
[42] ICANN, "Registrar Accreditation Agreement," 2013. [Online]. Available: https://www.icann.org/resources/pages/approved-with-specs-2013-09-17-en#whois-accuracy
[43] "Alexa: Actionable Analytics for the Web," http://www.alexa.com.
[44] "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.
[45] "ICANN: ICANN-Accredited Registrars," https://www.icann.org/registrar-reports/accredited-list.html.
[46] "ICANN: NOTICE OF TERMINATION OF ACCREDITATION AGREEMENT," https://www.icann.org/uploads/compliance_notice/attachment/895/serad-to-hansmann-4jan17.pdf.
[47] K. Levchenko, A. Pitsillidis, N. Chachra, B. Enright, M. Félegyházi, C. Grier, T. Halvorson, C. Kanich, C. Kreibich, H. Liu, D. McCoy, N. Weaver, V. Paxson, G. M. Voelker, and S. Savage, "Click Trajectories: End-to-End Analysis of the Spam Value Chain," in *Proceedings of the 2011 IEEE Symposium on Security and Privacy*. IEEE Computer Society, 2011, pp. 431–446.
[48] T. Moore and B. Edelman, "Measuring the Perpetrators and Funders of Typosquatting," in *Proceedings of the 14th International Conference on Financial Cryptography and Data Security*, ser. FC'10. Springer-Verlag, 2010, pp. 175–191.
[49] M. Korczyński, M. Król, and M. van Eeten, "Zone Poisoning: The How and Where of Non-Secure DNS Dynamic Updates," in *Proceedings of the 2016 ACM on Internet Measurement Conference*, ser. IMC '16. ACM, 2016, pp. 271–278.
[50] J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Beyond Blacklists: Learning to Detect Malicious Web Sites from Suspicious URLs," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '09. ACM, 2009, pp. 1245–1254.
[51] S. Hao, M. Thomas, V. Paxson, N. Feamster, C. Kreibich, C. Grier, and S. Hollenbeck, "Understanding the Domain Registration Behavior of Spammers," in *Proceedings of the 2013 Conference on Internet Measurement Conference (IMC'13)*. ACM, 2013, pp. 63–76.
[52] M. Antonakakis, R. Perdisci, D. Dagon, W. Lee, and N. Feamster, "Building a Dynamic Reputation System for DNS," in *USENIX Security Symposium*. USENIX Association, 2010, pp. 273–290.
[53] C. A. Shue, A. J. Kalafut, and M. Gupta, "Abnormally Malicious Autonomous Systems and Their Internet Connectivity," *IEEE/ACM Transactions on Networking*, vol. 20, no. 1, pp. 220–230, 2012.
[54] D. Mahjoub, "Sweeping the IP Space: the Hunt for Evil on the Internet." Virus Bulletin Conference, 2014. [Online]. Available: https://www.virusbtn.com/pdf/conference/vb2014/VB2014-Mahjoub.pdf
[55] A. Noroozian, M. Ciere, M. Korczyński, S. Tajalizadehkhoob, and M. Eeten, "Inferring the Security Performance of Providers from Noisy and Heterogenous Abuse Datasets," in *WEIS 2017*, June 2017.
[56] B. Stone-Gross, C. Kruegel, K. Almeroth, A. Moser, and E. Kirda, "FIRE: FInding Rogue nEtworks," in *Proceedings of the 2009 Annual Computer Security Applications Conference*, ser. ACSAC '09. IEEE Computer Society, 2009, pp. 231–240.
[57] M. Konte, R. Perdisci, and N. Feamster, "ASwatch: An AS Reputation System to Expose Bulletproof Hosting ASes," in *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*. ACM, 2015, pp. 625–638.
[58] C. Wagner, J. François, R. State, A. Dulaunoy, T. Engel, and G. Massen, "ASMATRA: Ranking ASs Providing Transit Service to Malware Hosters," in *IFIP/IEEE International Symposium on Integrated Network Management*. IEEE, 2013, pp. 260–268.

APPENDIX

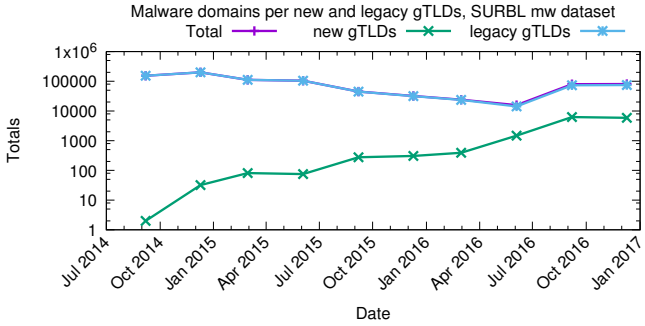

Figure 32. Time series of counts of malware domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the **SURBL mw** feed (2014-2016). Please notice y axis in log scale and overlapping lines.

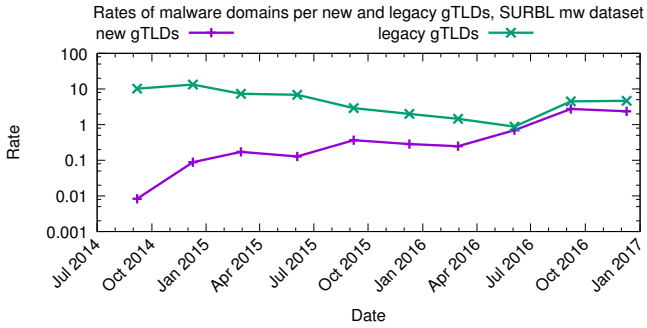

Figure 33. Time series of abuse rates of malware domains in **legacy** gTLDs and **new** gTLDs based on the **SURBL mw** feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

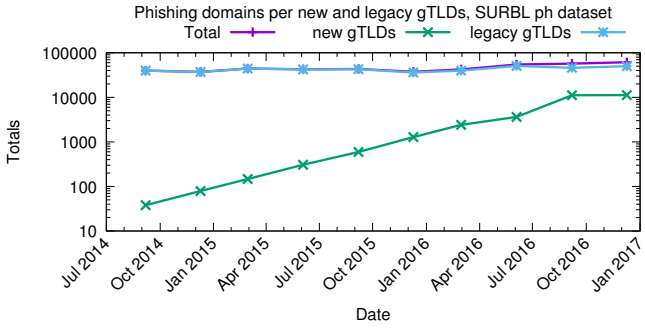

Figure 34. Time series of counts of phishing domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the **SURBL ph** feed (2014-2016). Please notice y axis in log scale and overlapping lines.

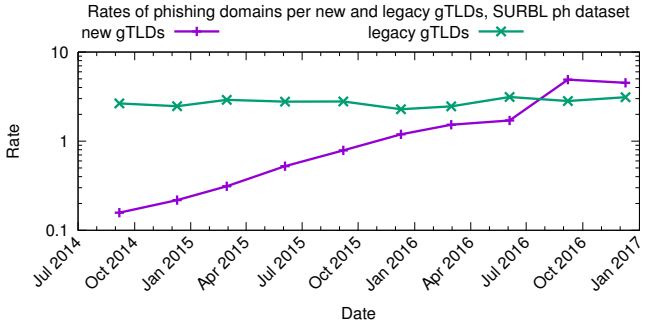

Figure 35. Time series of abuse rates of phishing domains in **legacy** gTLDs and **new** gTLDs based on the **SURBL ph** feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

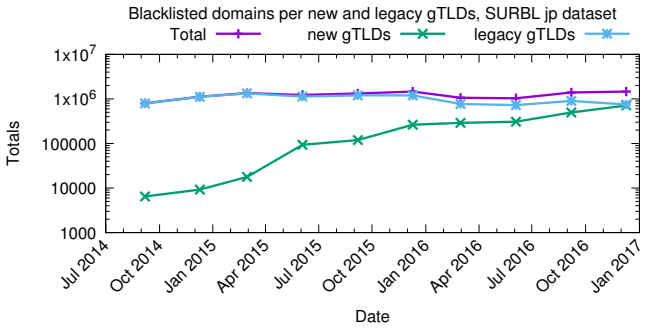

Figure 36. Time series of counts of blacklisted domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the **SURBL jp** feed (2014-2016). Please notice y axis in log scale and overlapping lines.

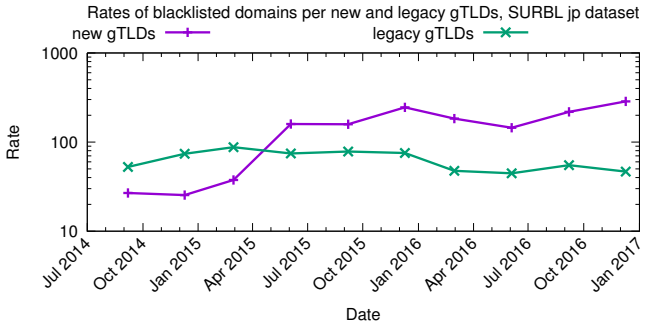

Figure 37. Time series of abuse rates of blacklisted domains in **legacy** gTLDs and **new** gTLDs based on the **SURBL jp** feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

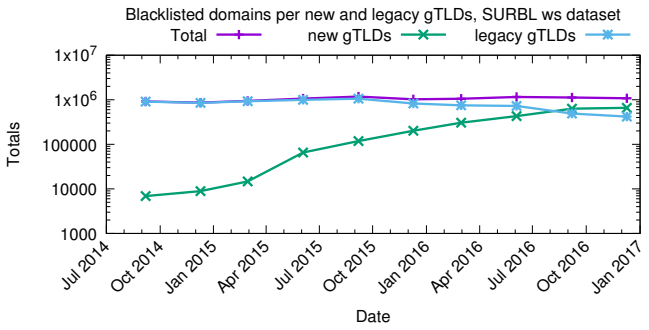

Figure 38. Time series of counts of blacklisted domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the **SURBL ws** feed (2014-2016). Please notice y axis in log scale and overlapping lines.

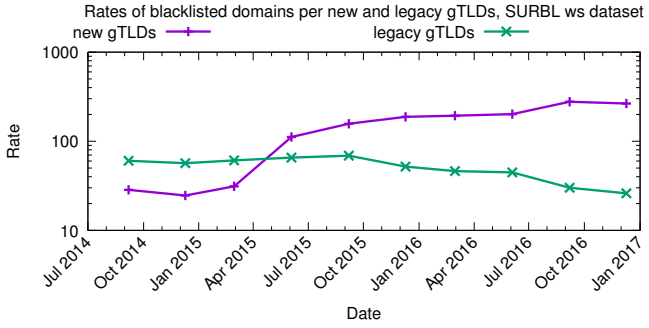

Figure 39. Time series of abuse rates of blacklisted domains in **legacy** gTLDs and **new** gTLDs based on the **SURBL ws** feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

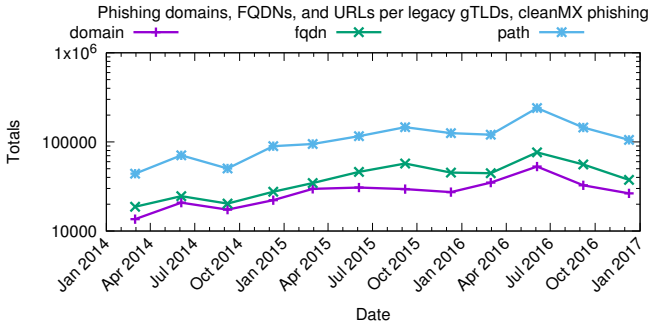

Figure 40. Time series of counts of blacklisted phishing domains, FQDNs, and URLs (paths) in **legacy** gTLD based on the **CleanMX phishing** feed (2014-2016). Please notice y axis in log scale and overlapping lines.

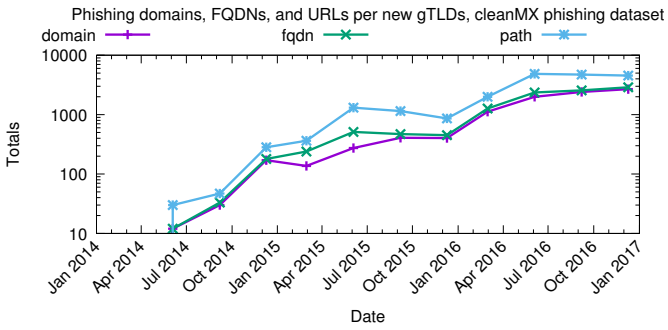

Figure 41. Time series of counts of phishing domains, FQDNs, and URLs (paths) in **new** gTLD based on the **CleanMX phishing** feed (2014-2016). Please notice y axis in log scale and overlapping lines.

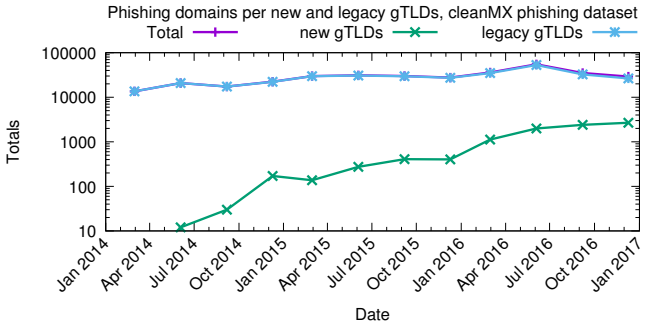

Figure 42. Time series of counts of blacklisted phishing domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the **CleanMX phishing** feed (2014-2016). Please notice y axis in log scale and overlapping lines.

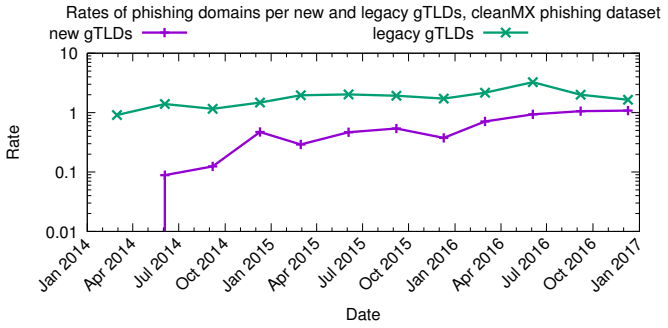

Figure 43. Time series of abuse rates of blacklisted phishing domains in **legacy** gTLDs and **new** gTLDs based on the **CleanMX phishing** feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

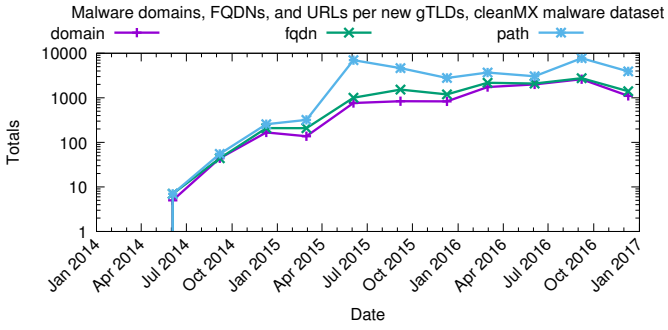

Figure 45. Time series of counts of blacklisted malware domains, FQDNs, and URLs (paths) in **new** gTLD based on the **CleanMX phishing** feed (2014-2016). Please notice y axis in log scale and overlapping lines.

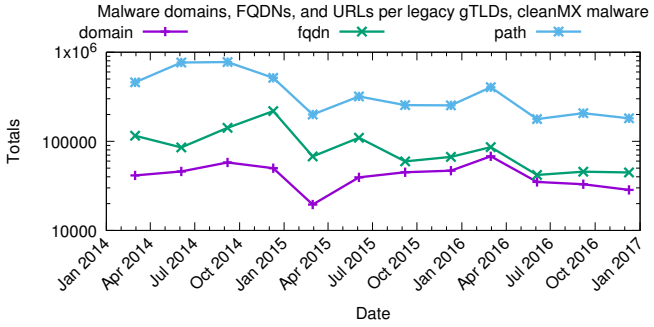

Figure 44. Time series of counts of blacklisted malware domains, FQDNs, and URLs (paths) in **legacy** gTLD based on the **CleanMX malware** feed (2014-2016). Please notice y axis in log scale and overlapping lines.

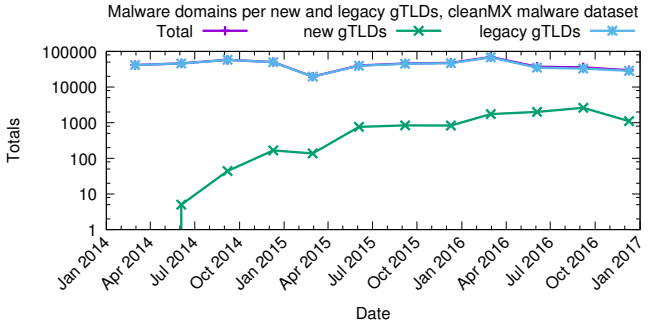

Figure 46. Time series of counts of blacklisted malware domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the **CleanMX malware** feed (2014-2016). Please notice y axis in log scale and overlapping lines.

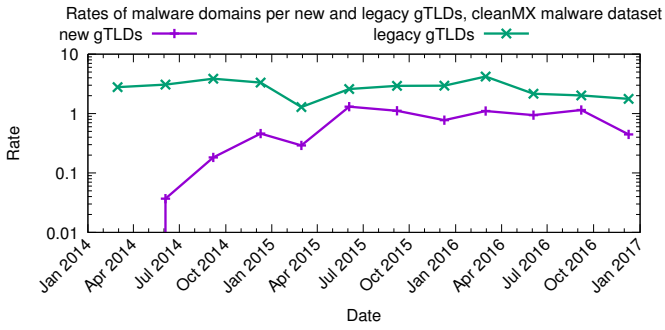

Figure 47. Time series of abuse rates of blacklisted malware domains in **legacy** gTLDs and **new** gTLDs based on the **CleanMX malware** feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

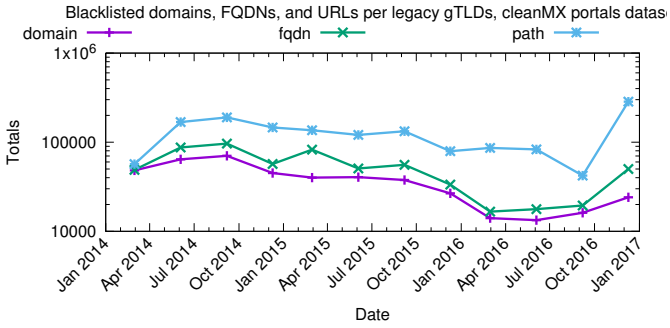

Figure 48. Time series of counts of blacklisted domains, FQDNs, and URLs (paths) in **legacy** gTLD based on the **CleanMX portals** feed (2014-2016). Please notice y axis in log scale and overlapping lines.

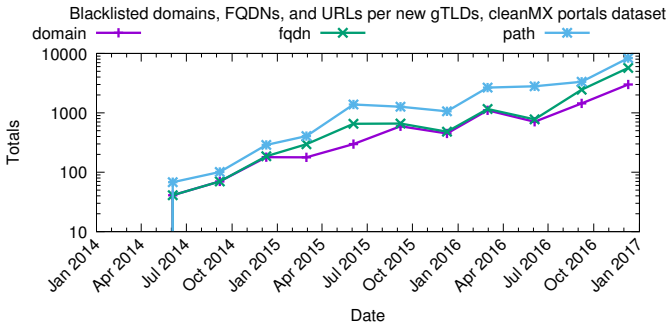

Figure 49. Time series of counts of blacklisted domains, FQDNs, and URLs (paths) in **new** gTLD based on the **CleanMX portals** feed (2014-2016). Please notice y axis in log scale and overlapping lines.

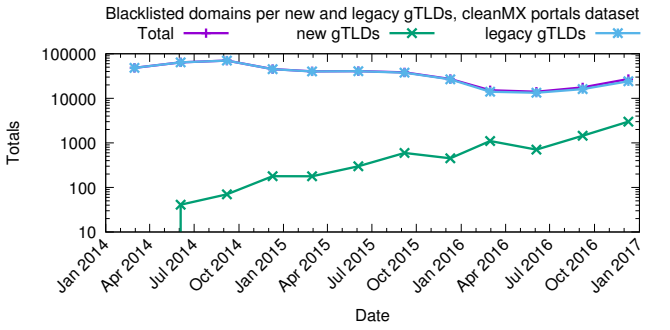

Figure 50. Time series of counts of blacklisted domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the **CleanMX portals** feed (2014-2016). Please notice y axis in log scale and overlapping lines.

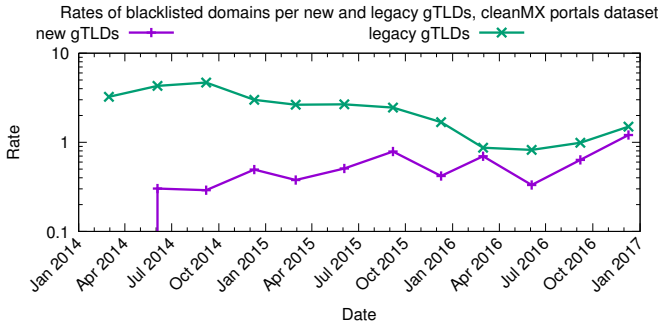

Figure 51. Time series of abuse rates of blacklisted domains in **legacy** gTLDs and **new** gTLDs based on the **CleanMX portals** feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

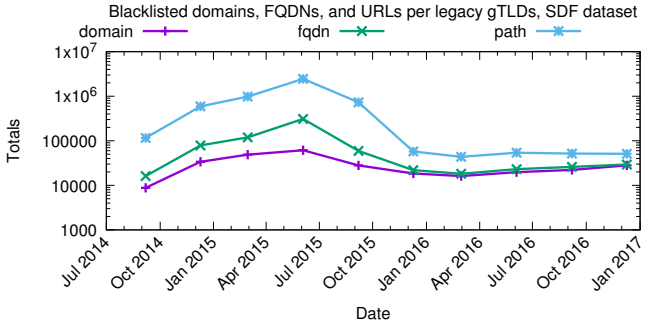

Figure 52. Time series of counts of blacklisted domains, FQDNs, and URLs (paths) in **legacy** gTLD based on the **Secure Domain Foundation** feed (2014-2016). Please notice y axis in log scale and overlapping lines.

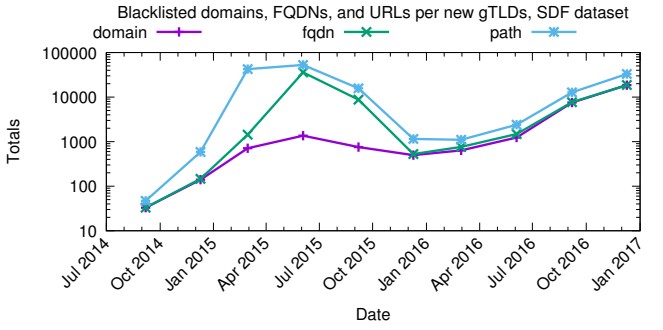

Figure 53. Time series of counts of blacklisted domains, FQDNs, and URLs (paths) in **new** gTLD based on the **Secure Domain Foundation** feed (2014-2016). Please notice y axis in log scale and overlapping lines.

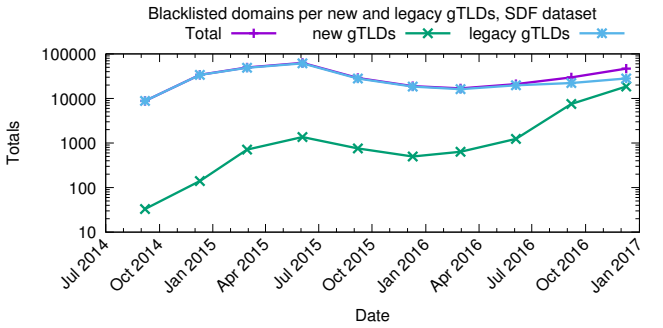

Figure 54. Time series of counts of blacklisted domains in **legacy** gTLD, **new** gTLDs, and **all** gTLDs (Total) based on the **Secure Domain Foundation** feed (2014-2016). Please notice y axis in log scale and overlapping lines.

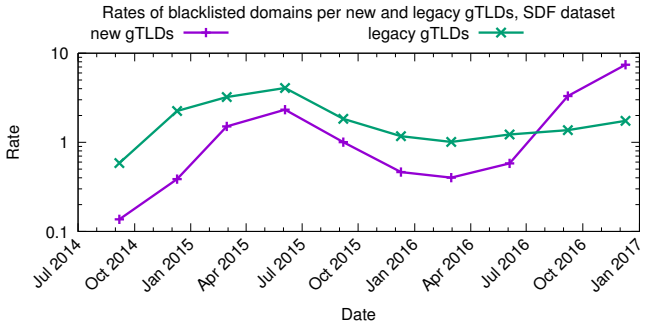

Figure 55. Time series of abuse rates of blacklisted domains in **legacy** gTLDs and **new** gTLDs based on the **Secure Domain Foundation** feed (2014-2016). Rates are calculated as follows: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

Table XXIV
TOP 10 NEW GTLDS WITH THE HIGHEST RELATIVE CONCENTRATION OF BLACKLISTED DOMAINS FOR STOPBADWARE SDP, APWG, SPAMHAUS, SECURE DOMAIN FOUNDATION, SURBL, AND CLEANMX DATASETS (FOURTH QUARTER OF 2016). SCORES ARE CALCULATED AS FOLLOWS: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

StopBadware			APWG			Spamhaus			SDF		
TLD	**# Domains**	**Score**	**TLD**	**# Domains**	**Score**	**TLD**	**# Domains**	**Score**	**TLD**	**# Domains**	**Score**
TOYS	32	78	LIMITED	31	66	SCIENCE	117,782	5,154	SUPPORT	510	294
TRADE	221	15	SUPPORT	43	24	STREAM	18,543	4,756	TECH	4,409	158
TATAR	1	11	CENTER	72	22	STUDY	1,118	3,343	ONLINE	4,179	83
WANG	1,086	11	CREDITCARD	1	13	DOWNLOAD	16,399	2,016	LIMITED	15	32
JUEGOS	1	9	SERVICES	24	10	CLICK	20,713	1,814	REVIEW	161	24
TOP	3,830	8	ONLINE	417	8	TOP	736,339	1,705	CLAIMS	3	19
MOE	5	8	MOE	5	8	GDN	45,547	1,602	PRESS	91	19
CAB	3	7	HOST	32	7	TRADE	23,581	1,521	FURNITURE	4	18
PICS	10	7	LEASE	1	6	REVIEW	9415	1,318	WEBSITE	298	15
TATTOO	2	7	REPORT	3	6	ACCOUNTANT	6,722	1,279	CREDITCARD	1	13

SURBL ph			SURBL mw			SURBL ws			SURBL jp		
TLD	**# Domains**	**Score**	**TLD**	**# Domains**	**Score**	**TLD**	**# Domains**	**Score**	**TLD**	**# Domains**	**Score**
LIMITED	51	109	FOOTBALL	7	16	RACING	51,443	3,812	SCIENCE	152,719	6,683
SUPPORT	82	46	TOP	5,066	11	DOWNLOAD	21,515	2,645	CLICK	27,871	2,441
CENTER	93	29	RIP	1	5	ACCOUNTANT	10,543	2,007	GDN	50,940	1,792
SERVICES	61	25	BID	200	3	REVIEW	12,615	1,766	STREAM	6,033	1,547
CRICKET	57	22	DENTIST	1	3	GDN	49,427	1,739	LINK	39,764	1,238
ONLINE	903	16	LGBT	1	3	FAITH	5,540	1,301	REVIEW	8,705	1,219
WEBSITE	318	14	ACCOUNTANT	11	2	TRADE	19,330	1,247	CRICKET	2,468	993
REPORT	7	14	CAB	1	2	CLICK	13,270	1,162	TRADE	14,535	937
HOST	65	13	SUPPORT	5	2	STREAM	4,406	1,130	FAITH	3,130	735
CREDITCARD	1	13	POKER	1	2	DATE	1,3851	999	TOP	285,488	661

CleanMX ph			CleanMX mw			CleanMX pt		
TLD	**# Domains**	**Score**	**TLD**	**# Domains**	**Score**	**TLD**	**# Domains**	**Score**
SARL	4	46	RODEO	1	12	QPON	1	20
LIMITED	9	19	TATAR	1	11	GRATIS	5	13
SUPPORT	19	10	MOE	6	10	CRICKET	32	12
ONLINE	493	9	HOW	1	4	TATAR	1	11
REPORT	4	8	ONLINE	183	3	DURBAN	2	8
MOE	4	6	CASINO	1	3	CLICK	72	6
CENTER	21	6	CHEAP	1	3	WEBCAM	18	4
REST	1	5	TAX	1	2	TAXI	2	4
SERVICES	13	5	CAB	1	2	WEBSITE	105	4
LAT	1	4	COMPUTER	1	2	LIMITED	2	4

Table XXV
TOP 10 LEGACY GTLDS WITH THE HIGHEST RELATIVE CONCENTRATION OF BLACKLISTED DOMAINS FOR STOPBADWARE SDP, APWG, SPAMHAUS, SECURE DOMAIN FOUNDATION, SURBL, AND CLEANMX DATASETS (FOURTH QUARTER OF 2016). RATES ARE CALCULATED AS FOLLOWS: $S = 10,000 * \#blacklisted\ domains/\#all\ domains$.

StopBadware			APWG			Spamhaus			SDF		
TLD	**# Domains**	**Score**	**TLD**	**# Domains**	**Score**	**TLD**	**# Domains**	**Score**	**TLD**	**# Domains**	**Score**
CAT	90	8	INFO	884	1	BIZ	57,234	251	INFO	2,317	4
MOBI	459	7	ASIA	24	1	NET	172,323	113	COOP	3	3
TRAVEL	11	6	CAT	17	1	ASIA	2,192	98	CAT	35	3
COM	76,544	6	BIZ	119	1	INFO	26,870	49	BIZ	190	1
BIZ	588	5	PRO	46	1	COM	624,852	49	ORG	1,153	1
ORG	6,347	5	COM	15,795	1	PRO	1,740	42	NET	2,082	1
INFO	2,369	4	COOP	0	0	ORG	30,227	28	PRO	43	1
NET	7,123	4	JOBS	0	0	MOBI	849	13	COM	22,226	1
COOP	3	3	NAME	8	0	NAME	79	5	JOBS	0	0
ASIA	59	2	MUSEUM	0	0	COOP	1	1	NAME	7	0

SURBL ph			SURBL mw			SURBL ws			SURBL jp		
TLD	**# Domains**	**Score**	**TLD**	**# Domains**	**Score**	**TLD**	**# Domains**	**Score**	**TLD**	**# Domains**	**Score**
INFO	2,336	4	INFO	31,494	58	BIZ	20,765	91	BIZ	125,286	550
COM	41,522	3	ORG	30,298	28	ASIA	1,440	64	INFO	63,062	117
ASIA	57	2	COOP	2	2	PRO	1,912	46	NET	152,163	100
ORG	2,368	2	ASIA	34	1	NET	60,552	39	ORG	65,365	62
NET	3,259	2	NET	2,037	1	COM	299,537	23	COM	342,603	27
NAME	21	1	JOBS	0	0	NAME	331	21	NAME	335	21
CAT	16	1	NAME	9	0	ORG	22,635	21	MOBI	970	15
BIZ	391	1	MUSEUM	0	0	INFO	10,602	19	ASIA	208	9
PRO	78	1	AERO	0	0	MOBI	1,160	18	PRO	328	7
COOP	0	0	XXX	0	0	COOP	0	0	COOP	2	2

CleanMX ph			CleanMX mw			CleanMX pt		
TLD	**# Domains**	**Score**	**TLD**	**# Domains**	**Score**	**TLD**	**# Domains**	**Score**
CAT	26	2	COOP	2	2	INFO	850	1
INFO	1,037	1	TRAVEL	4	2	NAME	21	1
COOP	1	1	INFO	1,076	1	CAT	13	1
ORG	1,596	1	CAT	17	1	TRAVEL	3	1
NET	1,800	1	MOBI	85	1	ORG	1,743	1
COM	21,759	1	BIZ	252	1	NET	2,222	1
JOBS	1	0	ORG	1,718	1	PRO	62	1
NAME	10	0	NET	2,320	1	COM	18,960	1
MUSEUM	0	0	COM	22,934	1	COOP	0	0
AERO	0	0	JOBS	0	0	JOBS	0	0