Phishing Landscape 2021

An Annual Study of the Scope and Distribution of Phishing

by

Greg Aaron

Lyman Chapin

David Piscitello

Dr. Colin Strutt

Interisle Consulting Group, LLC

22 September 2021

Interisle Consulting Group

Table of Contents

Table of Figures

Table of Tables

Executive Summary

Phishing is a significant threat to millions of Internet users. Phishing attacks lure victims to a web site that appears to be operated by a trusted entity, such as a bank, a merchant, or other service. The web site, however, is a deception, a fake, and the site's fake content is designed to persuade a victim to provide sensitive information.

Our goal in this study was to capture and analyze a large set of information about phishing attacks, to better understand how much phishing is taking place and where it is taking place, and to see if the data suggests better ways to fight phishing. To do so we determined when attacks occur and how quickly phishers act. We studied where phishers get the resources that they need to perpetrate their crimes — such as where they obtain domain names, and what web hosting is used. This analysis can identify where additional phishing detection and mitigation efforts are needed and can identify vulnerable providers. We also report on the wide range of brands targeted by phishers, and how often they take advantage of the unique properties of internationalized domain names (IDNs).

To assemble a deep and reliable set of data, we collected 1,487,914 phishing reports from 1 May 2020 to 30 April 2021 from four widely used and respected threat intelligence providers: the Anti-Phishing Working Group (APWG), OpenPhish, PhishTank, and Spamhaus. From that data we identified 695,823 unique phishing attacks.

Our major findings are:

The number of phishing attacks and domain names reported for phishing trended up over the yearly period. The number of reports we ingested from phishing feeds increased by 46% from the first quarter to the last quarter. We observed a 70% increase in phishing attacks and a 69% increase in unique domains reported for phishing.

When phishers register domains, they tend to use them quickly. 57% of domains reported for phishing were used within 14 days following registration and more than half of those were used within 48 hours. 89% of these *maliciously registered* phishing domain names were reported for phishing within 14 days following registration, and 98% of maliciously registered domain names were reported for phishing within the first year of registration.

Most phishing is concentrated at small numbers of domain registrars, domain registries, and hosting providers. We identified 497,949 unique domains used for phishing across the whole year. These domains were registered in 623 Top-level Domains (TLDs) and registered through 997 gTLD registrars. 69% of the domains used for phishing were in 10 TLDs; 69% were registered through 10 registrars.

Phishing attacks are disproportionately concentrated in new gTLDs (nTLDs). In June 2020, nTLDs represented 9% of domain names in the world but 18% of domains used for phishing. The new TLDs' market share decreased during our yearly reporting period (to 6% in March 2021), but phishing reported in the new TLDs increased to 21% during our yearly period.

Phishing domain registrations in some TLDs are overwhelmingly dominated by a small number of registrars. In some TLDs, 90% or more of the malicious domains were registered through one gTLD registrar.

Most phishing occurs on domains purposely (maliciously) registered for phishing attacks. 65% of domains associated with phishing attacks were maliciously registered. In the new TLD space,

70% of phishing domains reported in new TLDs were malicious. Twenty gTLD registrars accounted for 83% of all reported maliciously registered domains. Of these, the top four gTLD registrars (NameCheap, NameSilo, GoDaddy, and Public Domain Registry) account for 53%.

Ten hosting networks accounted for 41% of all phishing attacks. We identified 4,110 hosting networks (ASNs) where phishing web sites were reported; of these, four hosting networks (NameCheap, Cloudflare, Unified Layer, and Google) accounted for 28% of all phishing attacks.

11% of all phishing attacks took place using resources at subdomain service providers. Ten providers accounted for 90% of the phishing attacks hosted at subdomain service providers.

Phishers targeted 1,804 businesses or organizations during the 1 May 2020 to 30 April 2021 period. The top 10 brands targeted over the course of our annual period account for 46% of the reported phishing attacks.

We observed only a small overlap of phishing reports among the four feeds we monitored. This suggests that organizations could benefit from incorporating multiple sources of threat intelligence in their phishing defenses.

We under-report the phishing that takes place in certain regions. We suspect that this is the result of our limited access to phishing reports from these regions and challenges associated with data sharing from region to region.

Our data suggest that there may be opportunities for registry operators and registrars to identify maliciously registered phishing domains with a high degree of accuracy, often at the time of registration.

1. gTLD registrars and TLD operators are in an excellent position to identify and suspend malicious domain name registrations early, in some cases before they can be used victimize users and brands.
2. gTLD registrars and TLD operators possess key information – contact data and billing data – that no one else does. This data is highly useful for identifying malicious customers at the time of registration.
3. Domain name registrars or registry operators all have terms of service that allow them to suspend domains for malicious and illegal activity. Opportunities exist for registrars and registry operators to monitor for such activity, and to suspend domains for malicious purposes.

Introduction

For this study we analyzed nearly 1.5 million phishing reports representing about 700,000 phishing attacks. The study revealed that phishing increased by nearly 70% across the period 1 May 2020 through 30 April 2021 and thus continues to pose a significant threat to millions of Internet users.

Figure 1 Monthly Number of Phishing Attacks, 1 May 2020 to 30 April 2021

We began our study by examining phishing activity over a one-year period. We looked at attack activity, daily and weekly. We determined whether phishers found certain days of the week more opportunistic than others. We next compared the dates when our study set of domain names were registered to the dates when the domains were reported for phishing, to understand how phishers prepare for attacks.

We examined where phishing attacks occurred among Top-Level Domain registries, gTLD registrars, and hosting providers (Autonomous Systems, ASNs). We ranked these operators according to raw counts and metrics (*e.g.,* phishing scores).

For this 2021 Landscape study, we improved on our method for distinguishing phishing attacks where *maliciously registered domain names* were used from phishing attacks that were hosted on *compromised domains* or web sites. This distinction is important because it suggests where additional phishing detection or mitigation efforts could be applied most effectively, and importantly, which operator (registry, registrar, hosting provider) is best positioned to implement these.

We completed this study by reporting on the wide range of brands targeted by phishers and how phishers have added cryptocurrencies to their financial fraud target lists.

The statistics that we present in this report include both absolute metrics (*e.g.,* the number of domain names registered in a particular TLD that appear on a blocklist) and relative metrics (*e.g.,* the number of those domain names as a percentage of the total number of domains registered in that TLD). Attention to this distinction is critical to understanding and properly interpreting our analysis and findings.

Key Statistics

To assemble a deep and reliable set of data, we collected phishing reports for a one-year period, from 1 May 2020 through 30 April 2021 from four widely used and respected threat data providers: the Anti-Phishing Working Group (APWG), OpenPhish, PhishTank, and Spamhaus, (see the section *Phishing Data Sources* on page 57). In Table 1 we highlight key statistics for this period of phishing activity.

Measurement	May 2020 to April 2021
Total number of phishing reports	1,487,914
Total number of distinct phishing attacks	695,823
Unique domain names reported for phishing	497,949
Top-level domains where we observed phishing	623
Malicious domain registrations	322,145
Registrars that had domains under management reported for phishing	997
Hosting Networks (ASNs) where phishing web sites were reported	4,110
Brands targeted in phishing attacks	1,804

Table 1 Key Statistics for the Yearly Period of Phishing Activity, 1 May 2020 to 30 April 2021

The total number of phishing reports is the sum of reports we ingested from phishing feeds from 1 May 2020 to 30 April 2021.

The total number of phishing attacks is a sum of the attacks that we identified using the methodology we describe in *Appendix A: Identification of Phishing Attacks*.

Unique domain names reported for phishing is based on our determination of "the first occurrence of a domain name in a phishing report". We use this number to account for domains which recurred in multiple quarters during the yearly period.

The numbers of TLDs, gTLD registrars, and Hosting Networks where we observed phishing were obtained by counting each operator that appeared in the yearly study data.

The number of brands targeted in phishing attacks, as reported by the phishing feeds, was calculated during our processing of phishing attacks. See the section *Targeted Brands* on page 40. (Note: the study of attacks against cryptocurrencies were processed independently from the studies of targeted brands).

Table 2 shows the quarterly figures reported at the Cybercrime Information Center [1] for these key statistics. Note that the sums of quarterly numbers for total number of phishing reports and phishing attacks in Table 2 will equal the yearly figures. Other quarterly numbers, for example TLDs or gTLD registrars, require de-duplication, so the sum of the four quarterly numbers will not be the same the numbers in Table 1.

Measurement (by Quarter)	May 2020 – July 2020	August 2020 – October 2020	November 2020 – January 2021	February 2021 – April 2021
Total number of phishing reports	298,012	303,922	450,333	435,647
Phishing attacks	126,439	160,876	193,058	215,451
Top-level Domains (TLDs) where we observed phishing	439	454	493	516
Registrars with domains under management reported for phishing	414	552	481	553
Hosting Networks (ASNs) where phishing web sites were reported	2,169	2,321	2,335	2,421
Unique domain names reported for phishing	99,203	117,001	146,724	168,395
Malicious Phishing Domain Registrations	56,191	75,114	93,228	106,820
Phishing Attacks associated with malicious domain registrations	61,334	80,220	93,228	106,820
Brands targeted in phishing attacks	695	883	1034	945

Table 2 Quarterly Key Statistics, 1 May 2020 to 30 April 2021

Trends of Key Statistics

Figure 2 illustrates that phishing domains reported, phishing attacks, and unique domains reported for phishing all trended up over the yearly period.

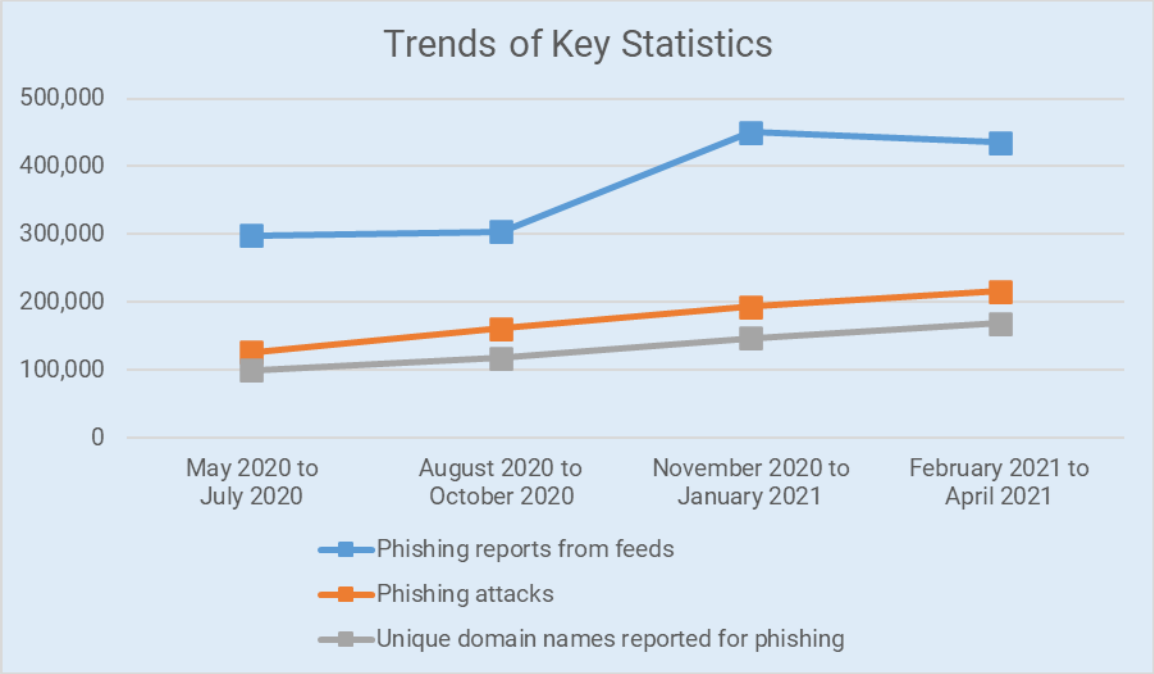

Figure 2 Phishing Domains, Attacks, and Unique Domains Reported for Phishing Trended Up

Phishing Activity

We define a phishing attack as a phishing site that targets a specific brand or entity. In *Appendix A: Identification of Phishing Attacks*, we describe how we determined if multiple phishing reports refer to the same phishing site. When they did, we eliminated duplicates to yield a count of distinct phishing attacks.

During the study period, phishing peaked in November and December 2020. Historically, phishing has usually risen during the holiday season, when consumers are making online purchases and online fraud checks are more permissive [2], so the late 2020 rise was not surprising. Following a dip after Christmas 2020, phishing then peaked again in late January through February 2021, and remained at a relatively elevated level through April. The phishing in early 2021 occurred at a variety of providers, and attacked a wide variety of targets, representing an across-the-board increase.

In Figure 3 we show the daily counts of phishing attacks for the study period, to illustrate that phishing activity is a generally chronic (persistent) problem, with periods of acute daily activity that is increasing over time.

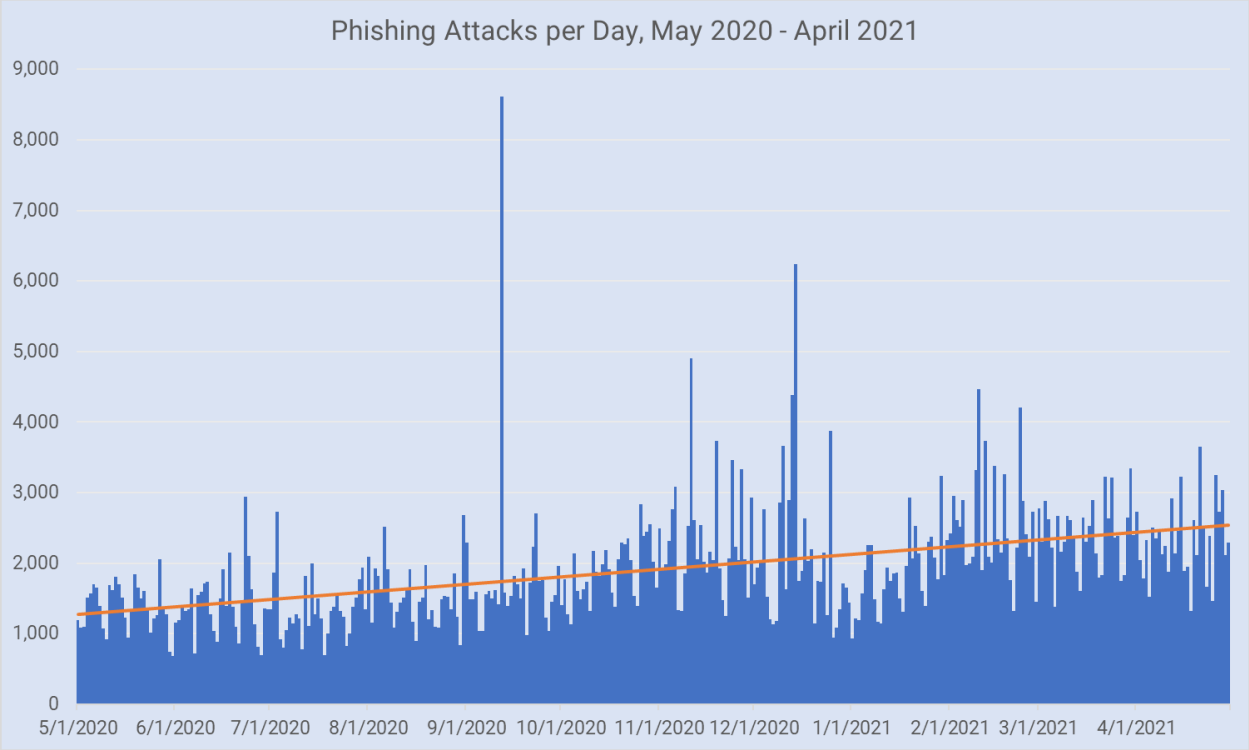

Figure 3 Daily Number of Phishing Attacks, 1 May 2020 to 30 April 2021

Figure 4, depicts when phishing attacks occurred, by day of the week.

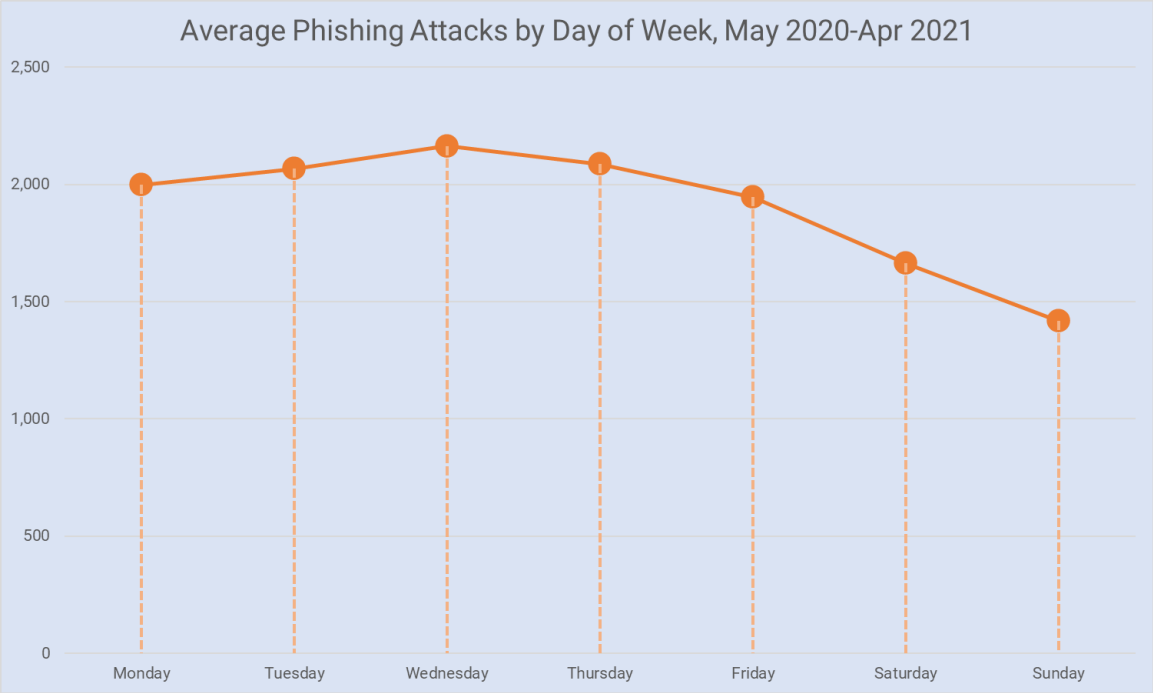

Figure 4 Phishing Attacks by Day of Week, 1 May 2020 to 30 April 2021

In the graph in Figure 4, we observe that blocklistings usually peak around Wednesdays. We note that there is a delay between when an attack begins and when it is blocklisted, meaning that attacks actually peak a bit earlier.

Historically, phishing activity has been highest in the Monday through Wednesday period. Phishers advertise their attacks via spam mail at what they believe to be an optimal time, *i.e.,* when people check their work and personal email on returning to work or after the weekend is over.

Time Elapsed between Domain Registration and Phishing

We analyzed how much time elapsed between when a domain name was registered and when that domain was associated with a phishing attack by one of the phishing data feeds. For this analysis, we only included domain names for which we were able to obtain a creation date from domain registration data.

Figure 5 shows that, during the yearly study period, 57% of domains reported for phishing were used within 14 days following registration and that the majority of these were reported within 48 hours. 84% of domain names associated with a phishing attack were reported within the first year of registration.

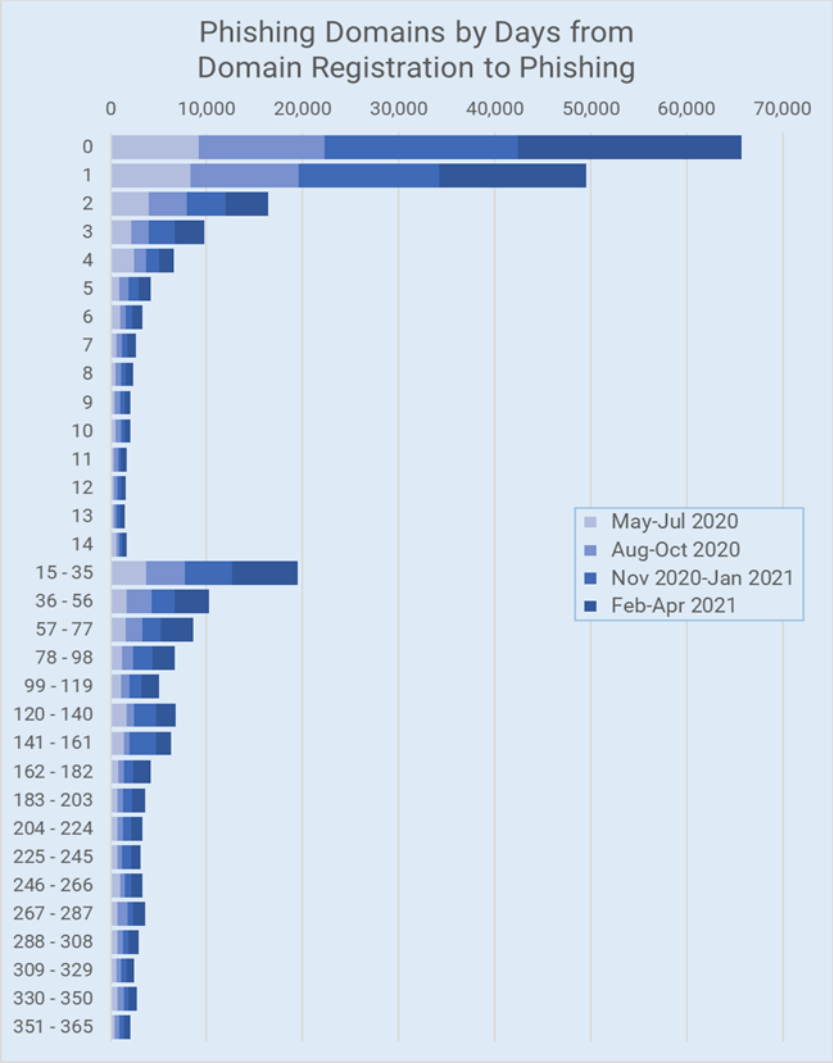

Figure 5 Phishing Domains: Days from Domain Registration to Reported for Phishing

The Cybercrime Information Center [1] reports quarterly numbers of phishing domains. Here, we show the total number of phishing domains using a different shade of color to represent the different quarters. Our findings continue to reinforce the conventional wisdom that when phishers register domains, they tend to use them quickly to avoid detection. This is consistent with research concerning the risk associated with newly registered domain names [3, 4].

In the section *Malicious Domain Name Registrations* on page 23, we explain how we classified phishing domains as maliciously registered (*a domain name registered by a criminal to carry out phishing*) or compromised (*a domain name that was registered for legitimate purposes but had its web hosting broken into by a phisher*). Figure 6 shows that, for the yearly study period, 89% of domain names that we classified as malicious were reported for phishing within 14 days following registration and 98% of domain names that we classified as malicious were reported for phishing within the first year of registration.

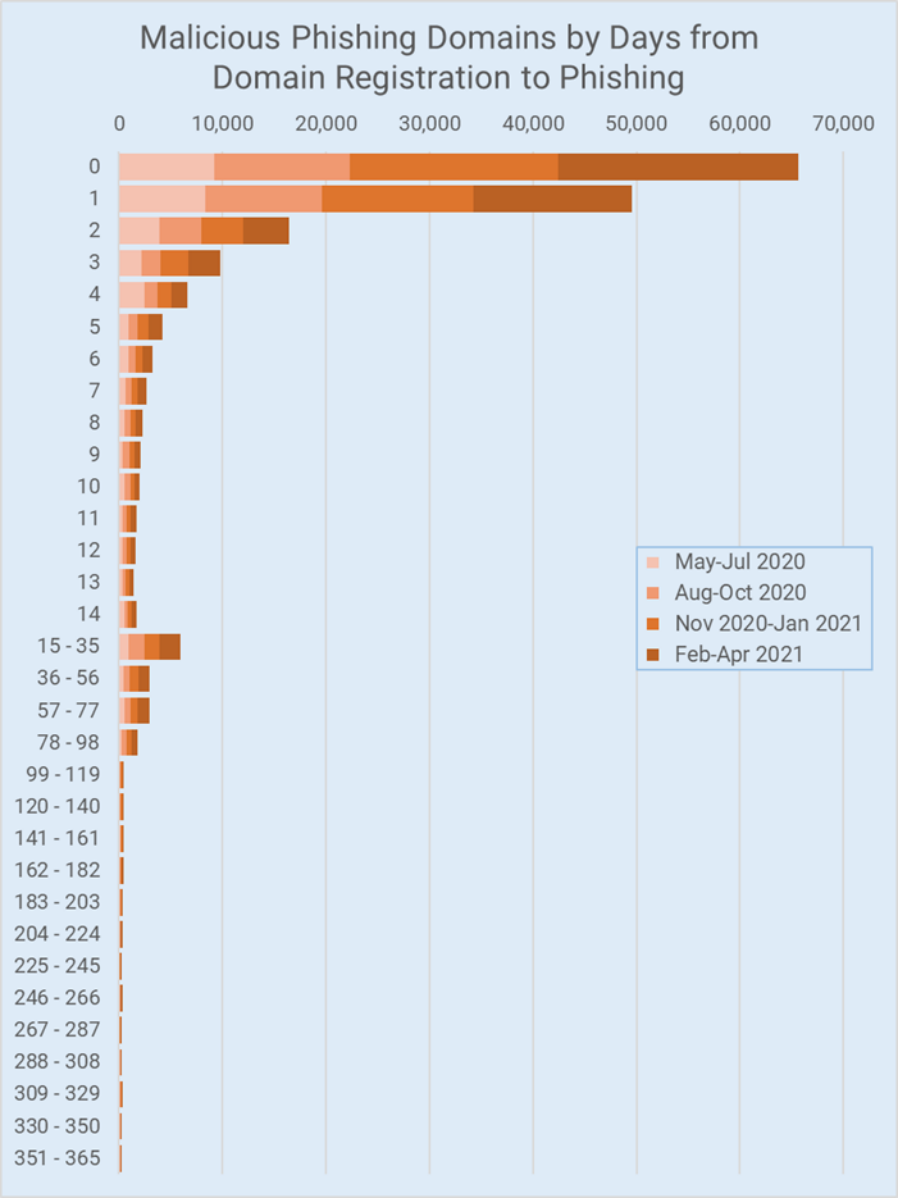

Figure 6 Malicious Phishing Domains: Days from Domain Registration to Reported for Phishing

The data also indicates that many malicious domain registrations remained undetected for days (and sometimes months) by the registrars and registry operators, allowing the phishers to carry out their attacks**.**

If a domain registration were flagged as fraudulent, for example purchased with a stolen credit card, the transaction could either be rejected at the time of the transaction or be flagged by the registrar's payment processor, usually within a few days after the transaction. Where fraud is detected after a domain sale, the registrar could suspend the domain names involved, which makes phishing on them impossible. In the United States, credit card holders can dispute a fraudulent charge for up to 60 days after the transaction date. Payment processors do not rely just on complaints from their customers; they also run anti-fraud algorithms of their own. The yearly study data reinforce our suspicion that phishers are either paying for their domain names with legitimate means, or that the payment processors and the registrars are not recognizing many suspicious or fraudulent transactions at the time of transaction or in the days thereafter.

It also appears that domain registrars are not taking advantage of tools that will allow them to recognize maliciously registered domains in a short time immediately after registration. (These include checks for inaccurate contact data and checks that can identify – or label as suspicious – a domain that was purposely registered for phishing from a domain from a domain that was registered for a legitimate purpose. See *Appendix B: Distinguishing Maliciously Registered Domain Names from Compromised* .)

We observed little "aging" of domains purposely registered for phishing. The yearly study data shows that only 2% of the maliciously registered domains were not used until more than 90 days after they were registered. This is consistent with our finding that 89% of domain names that we classify as malicious are reported for phishing within 14 days following registration. Phishers do not appear to be waiting for their domains to move out of *"*very new domain*"* status.

Our data also show that a small number of domains *appear* to be maliciously registered but that they were flagged for phishing well past the first year of registration, in some cases several years after registration.

Prevalence of Phishing by Top-Level Domain (TLD)

The Q1 2021 Verisign Domain Name Industry Brief[5] reported that there were 363.5 million registered domain names in the world's registries. The overall domain name space can be divided into four categories and is illustrated in Figure 7:

- the .COM and .NET registries are operated by Verisign and represented 46% of the domains in the world,
- country-code domains (ccTLDs) represented 43% of the domains,
- the legacy generic TLDs (those other than .COM and .NET and introduced before 2013, *e.g.,* .ORG, .BIZ, .INFO) represented 5% of the domains, and
- the new gTLDs (nTLDs) introduced from 2014 to the present (*e.g.,* .ONLINE, .XYZ, .ICU) were the remaining 6% of the domains.

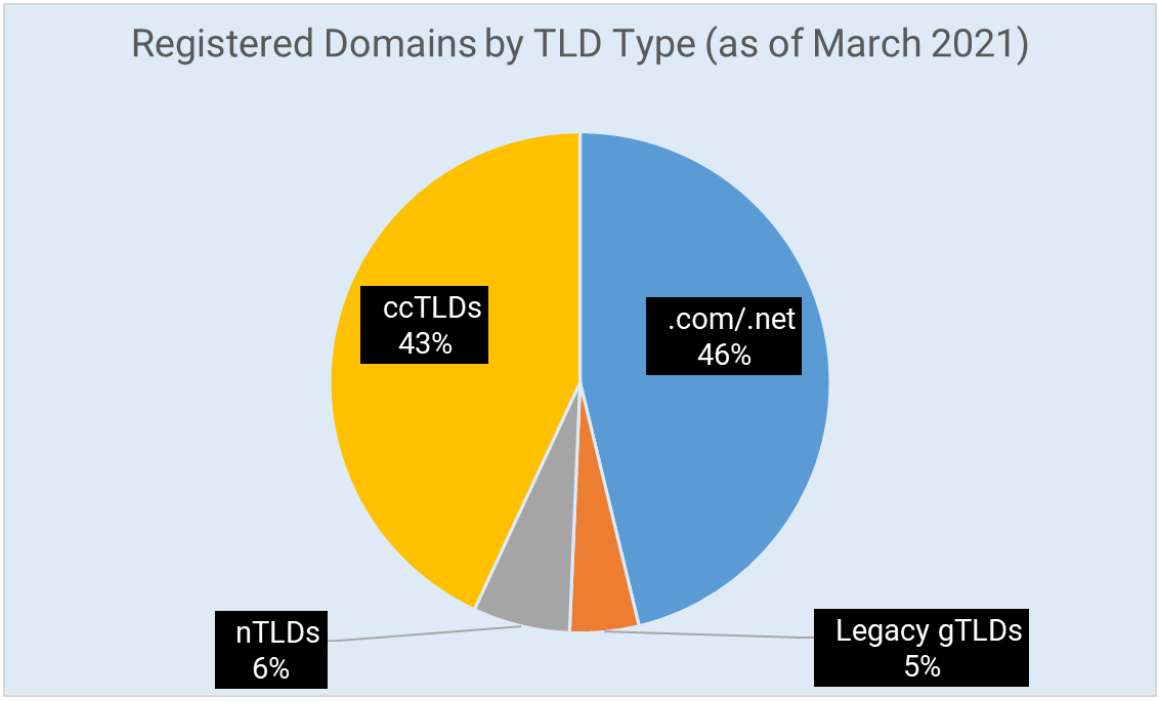

Figure 7 Registered Domain Names in the World's Registries, per Verisign by TLD Type, March 2021

We analyzed the phishing domains and attacks to see how they were distributed across the top-level domains. While we observed phishing in 623 TLDs during the yearly study period, we note that phishing activity continues to be concentrated in just a few namespaces.

Some TLDs attract many more problems (and/or more persistent problems) than others. Figure 8 shows that the distribution of domains used for phishing by TLD differs from market share.

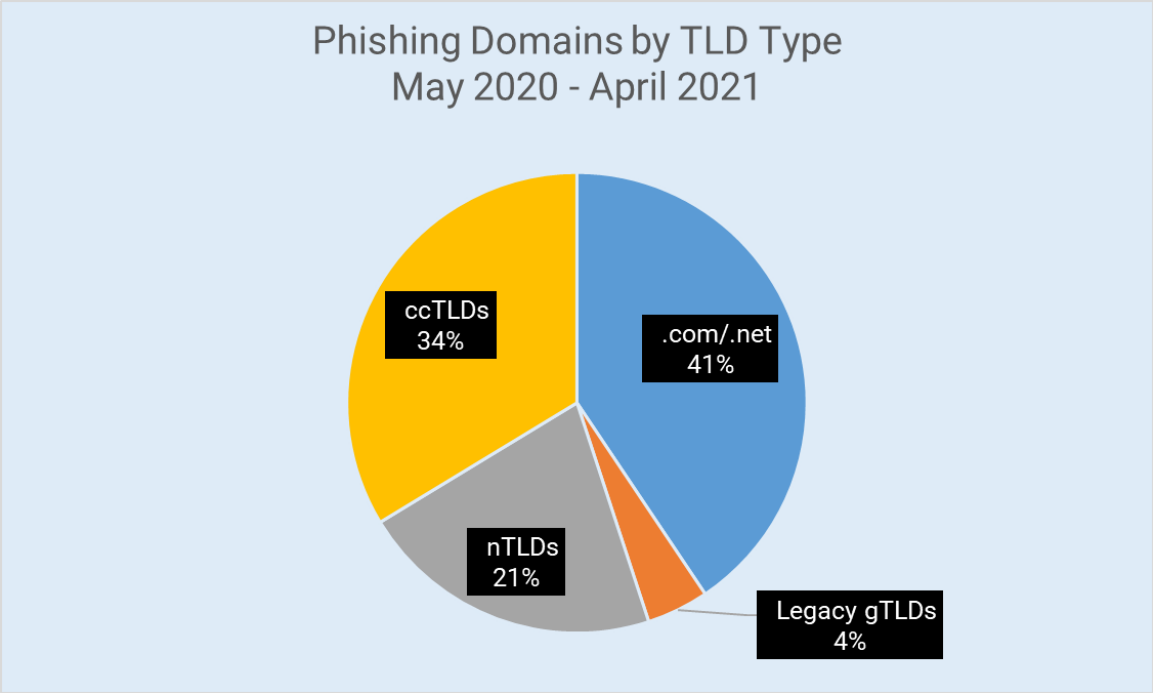

Figure 8 Phishing Domains by TLD Type, 1 May 2020 to 30 April 2021

41% of all domains reported for phishing were in .COM and .NET. This percentage is smaller than the combined market share (46%) of these TLDs.

21% of phishing was in the new TLDs. This is 3.5 times the new TLDs' market share of 6%, indicating that domains in the new gTLDs were used disproportionately for phishing. In June 2020, nTLDs were 9% of the market, with 18% of phishing domains [6]. While the nTLDs' market share has decreased since then, phishing in this category increased through April 2021.

34% of domains used for phishing were in ccTLDs. This is smaller than the 43% of the domain name market share represented by ccTLDs.

Phishing in the ccTLD category was artificially swollen by 97,380 phishing domains reported in five "commercialized" ccTLDs run by Freenom (.TK, .ML, .GA, .CF, .GQ) which offers free domain name registrations. This number represented 58% of all ccTLD phishing domains reported and 20% of phishing domains reported in all TLDs. Setting aside phishing on the Freenom domains, the other ccTLDs suffered far less phishing than might be expected based on market share.

The remaining 4% of phishing was in the legacy TLDs other than .COM and .NET, roughly in line with their market share.

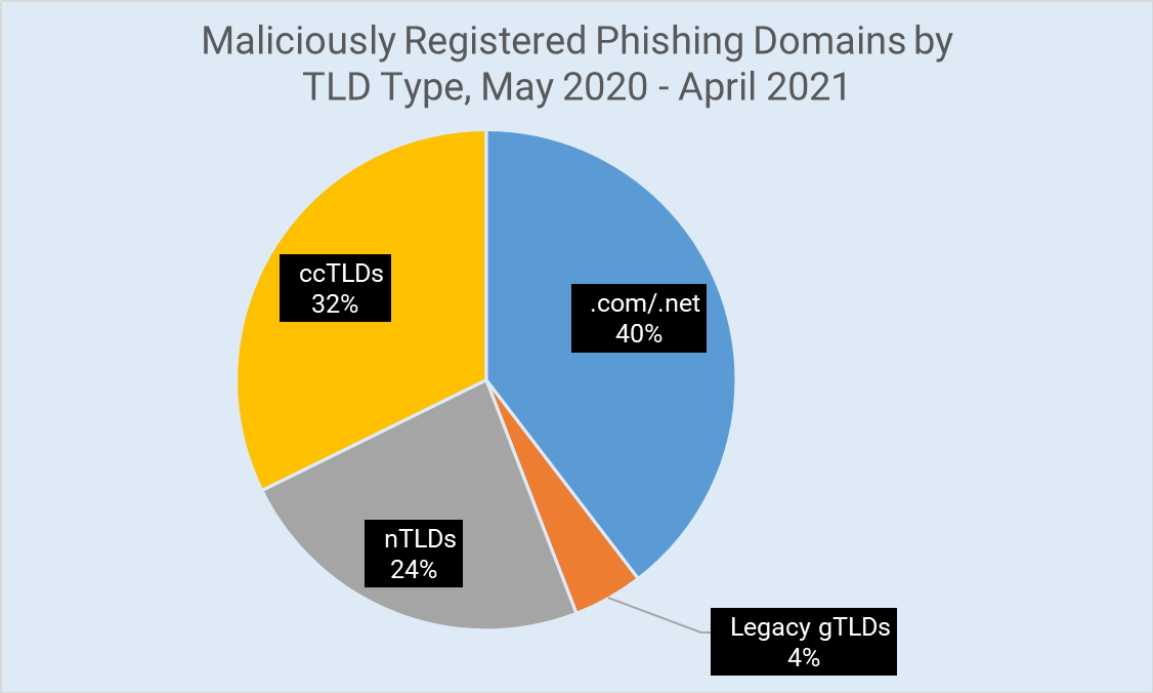

Figure 9 Maliciously Registered Phishing Domains by TLD Type, 1 May 2020 to 30 April 2021

Phishers did register domains purposely for phishing in .COM and .NET, but the percentage of phishing domains that we classified as malicious registrations is smaller than the combined market share of .COM and .NET. Approximately 132,086 domains in these TLDs had their web hosting compromised by phishers, who placed phishing pages on the sites without the owners knowledge. We estimate that 60% of the domains used for phishing in .COM and .NET were maliciously registered.

New TLDs continue to present attractive registration opportunities for phishers: the percentage of malicious registrations in the new TLDs was 3.75 times the segment's market share. **We estimate that 70% of the nTLD domains used for phishing were maliciously registered.**

Note that Figure 9 combines malicious domain registrations for .COM and .NET. In the section *Malicious Domain Name Registrations and TLDs* on page 25 we examine these and other TLDs with large numbers of malicious domain name registrations individually.

Ranking of TLDs by Phishing Domains Reported

Table 3 shows the TLDs with the highest numbers of new phishing domains reported in our yearly study period.

Rank	TLD	Registry Operator	Domains in TLD	Cumulative Phishing Domains Reported ▼
1	com	Verisign	151,618,533	**260,636**
2	tk	Freenom	19,987,952	**40,002**
3	xyz	XYZ.COM	2,978,332	**27,532**
4	ml	Freenom	3,816,199	**27,284**
5	ga	Freenom	4,661,469	**21,657**

Rank	TLD	Registry Operator	Domains in TLD	Cumulative Phishing Domains Reported ▼
6	cf	Freenom	4,179,760	**19,187**
7	gq	Freenom	3,375,388	**16,168**
8	cn	CNNIC	13,708,468	**16,052**
9	top	Jiangsu Bangning	2,306,018	**15,129**
10	net	Verisign	13,407,660	**14,398**

Table 3 TLDs with Most New Phishing Domains, 1 May 2020 to 30 April 2021

For the period, six ccTLDs were ranked among the top ten. Five of them (.TK, .ML, .GA, .CF, and .GQ) are operated by Freenom, a company in the Netherlands that offers free domain registrations in these ccTLDs. The Freenom ccTLDs appeared repeatedly in the Top 10 quarterly TLD phishing activity reports published at the Cybercrime Information Center [7].

It is worth noting that while the legacy gTLD .INFO and new TLDs .SHOP and .BUZZ did not make the yearly Top 10 ranking, these TLDs made appearances in the Top 10 in quarterly phishing activities reported at the Cybercrime Information Center.

Ranking of TLDs by Scoring Metrics

The gross numbers of phishing domains are significant because more phishing domains means more damage and victimization. The larger the number of phishing domains in a space or portfolio controlled by one company, the greater the opportunity (and need) for that company to take effective anti-abuse measures — including measures to find and suspend malicious phishing registrations early.

Scoring metrics allow for comparisons between TLDs of different sizes. In the quarterly phishing activity published at the Cybercrime Information Center, the metric *"*Phishing Domains per 10,000*"* is used to compare whether a TLD has a higher or lower incidence of phishing relative to others. This is a ratio of the number of domain names used for phishing in a TLD to the number of registered domain names in that TLD. We call this metric **TLD Phishing Score**:

TLD Phishing Score =
*(number of phishing domains / total number of domains under management in the TLD) * 10,000*

Here, we use a similar metric to measure the prevalence of phishing in each TLD for the period beginning 1 May 2020 and ending 30 April 2021 (365 days). We take the sum of the four quarters of unique phishing domains reported and divide by the average of the domains under management per TLD for each of the four quarters. We call this metric **Yearly TLD Phishing Score**:

Yearly TLD Phishing Score =
(sum of the four quarters of unique phishing domains reported /
*average of the domains under management per TLD for each of the four quarters) * 10,000*

This method considers the fact that a TLD can grow and shrink, sometime appreciably, over the course of a year of phishing activity.

Table 4 shows the ranking of Top-level Domains by annual TLD phishing score:

Rank	TLD	Registry Operator	Domains in TLD	Yearly Phishing Domain Score ▼
1	cyou	ShortDot	54,848	**269.5**
2	bar	Punto 2012 S.A.P.I.	128,154	**233.3**
3	best	BestTLD	100,508	**206.0**
4	casa	Minds + Machines	40,570	**199.7**
5	buzz	DOTSTRATEGY	467,985	**190.5**
6	services	Donuts	52,649	**181.0**
7	live	Donuts	526,119	**160.6**
8	monster	XYZ.COM	109,300	**159.7**
9	link	UNR	132,529	**157.5**
10	host	Radix FZC	68,101	**123.6**

Table 4 Ranking of TLDs by Yearly Phishing Score, 1 May 2020 to 30 April 2021

A person is more likely to encounter a dangerous domain when they click on a hyperlink in an email message or visit a web site address that contains a domain name registered in a TLD with a high phishing score.

The phishing score of an individual TLD cannot be used to predict the likelihood that a person will encounter a dangerous domain when clicking on an arbitrary domain name (that is, without regard to which TLD it belongs) in an email message or on a web site, because phishing score applies to one specific TLD, not to the distribution of phishing domains across the entire domain name space.

Prevalence of Phishing by gTLD Registrar

Of the 695,823 phishing attack reports that we collected during the 1 May 2020 to 30 April 2021 period, 3,135 contained IP addresses, and of these, 2,478 were unique addresses. The remainder of phishing reports that we processed contained domain names.

Phishers acquire domain names by registering names purposely for phishing. They also break into the domain name management accounts or the hosting accounts of domain name owners. Table 5 shows that phishers purchase and manage domain names through many gTLD registrars.

	May 2020 – July 2020	August 2020 – October 2020	November 2020 – January 2021	February 2021 – April 2021
Registrars with domains under management reported for phishing	414	552	481	553

Table 5 Quarterly Counts of gTLD Registrars with Domains Under Management Reported for Phishing

The table shows quarterly gTLD registrar counts as reported at the Cybercrime Information Center. Some gTLD registrars had phishing domains reported against their domains under management in more than one quarter. Overall, 1,009 gTLD registrars appeared at least once during the period covered by this report.

Some gTLD registrars appear to be more attractive to phishers than others. We consider this phenomenon in the section *Malicious Domain Name Registrations and gTLD Registrars* on page 27.

Ranking of gTLD Registrars by Phishing Domains Reported

Table 6 shows where larger-than-usual concentrations of phishing occur in registrars' domain portfolios. The registrars with 5,000 or more gTLD domains reported for phishing from 1 May 2020 to 30 April 2021 period were:

Rank	Registrar	Registrar IANA_ID	gTLD Domains under Management	Phishing Domains Reported ▼
1	NameCheap	1068	11,045,487	**79,118**
2	NameSilo	1479	3,501,471	**37,067**
3	GoDaddy.com	146	63,844,325	**35,150**
4	PublicDomainRegistry.com (PDR)	303	4,996,592	**19,065**
5	Tucows Domains	69	10,389,339	**9,972**
6	Wild West Domains	440	2,812,669	**8,582**
7	Google LLC	895	5,360,500	**8,413**
8	ALIBABA.COM SINGAPORE E-COMMERCE PRIVATE LIMITED	3775	969,502	**7,883**
9	GMO Internet (Onamae.com)	49	5,000,613	**7,276**
10	eNom,	48	5,171,823	**6,754**
11	Alibaba Cloud Computing d/b/a HiChina (www.net.cn)	1599	4,700,511	**6,368**

Rank	Registrar	Registrar IANA_ID	gTLD Domains under Management	Phishing Domains Reported ▼
12	Web Commerce Communications Limited dba WebNic.cc	460	1,446,221	**6,318**
13	Name.com	625	2,143,807	**5,812**
14	Registrar of Domain Names REG.RU LLC	1606	868,229	**5,124**

Table 6 gTLD Registrars with at Least 5,000 Reported Phishing Domains, 1 May 2020 to 30 April 2021

Ranking of gTLD Registrars by Scoring Metrics

Gross numbers influence how one compares a set of operators that have size diversity (numbers bias). In the quarterly phishing activity published at the Cybercrime Information Center, the metric *"*Phishing Domains per 10,000*"* is used to compare whether a gTLD registrar has a higher or lower incidence of phishing relative to others. This is a ratio of the number of domain names used for phishing to the number of registered domain names under management at that gTLD registrar. We call this metric **gTLD Registrar Phishing Score**:

gTLD Registrar Phishing Score =
*(number of phishing domains / domains under management at gTLD Registrar) * 10,000*

For this report, as we did for TLDs, we use a similar metric to measure the prevalence of phishing in each gTLD registrar for the period 1 May 2020 to 30 April 2021. Here, we take the sum of the four quarters of unique phishing domains reported and divide by the average of the domains under management per gTLD registrar for each of the four quarters. We call this metric **Yearly gTLD registrar Phishing Score**:

Yearly gTLD registrar Phishing Score =
(sum of the four quarters of unique phishing domains reported /
*average of the domains under management at gTLD Registrar) * 10,000*

Note that the calculation of these two metrics yields different results (simply put, we use different inputs for the numerators and denominators in the division); in particular, one cannot draw any conclusion by comparing the scores from a quarterly phishing score against an annual phishing score. Instead, we encourage comparisons of quarterly phishing scores over time, as well as annual phishing scores.

Table 7 shows the ranking of gTLD registrars by annual gTLD registrar phishing score:

Rank	Registrar	Registrar IANA_ID	gTLD Domains under Management	Phishing Domains	Yearly Phishing Domain Score ▼
1	TLD Registrar Solutions	1564	87,996	1,033	**117.4**
2	NameSilo	1479	3,501,471	37,607	**105.9**
3	ALIBABA.COM SINGAPORE	3775	969,502	7,883	**81.3**

Rank	Registrar	Registrar IANA_ID	gTLD Domains under Management	Phishing Domains	Yearly Phishing Domain Score ▼
4	Jiangsu Bangning Science & technology Co.	1469	632,725	4,697	**74.2**
5	BigRock Solutions	1495	271,419	2,001	**73.7**
6	NameCheap	1068	11,045,487	79,118	**71.6**
7	Key-Systems	1345	472,577	3,329	**70.4**
8	NETIM SARL	1519	43,107	257	**59.6**
9	Registrar of Domain Names REG.RU	1606	868,229	5,124	**59.0**
10	Internet Domain Service BS Corp	2487	379,984	2,085	**54.9**

Table 7 Ranking of gTLD Registrars by Yearly Phishing Domain Score, 1 May 2020 to 30 April 2021

High gTLD registrar phishing scores may indicate that phishers find that the registrar's processes, pricing, or services are attractive or favorable for registering domain names for phishing. To explore this proposition further, we next consider domain names that have been purposely registered for phishing attacks.

Malicious Domain Name Registrations

We define a maliciously registered domain as a *domain registered by a criminal to carry out a malicious or criminal act.* For our studies, we distinguish maliciously registered domains from compromised domains, which we define as *domain names that were registered for legitimate purposes but co-opted by criminals* through some form of compromise.

For example, an attacker may hijack a legitimate user's domain registrar account, alter the corresponding DNS entry to resolve a name or URL to a host that the attacker controls; here, the domain and DNS are compromised. An attacker may also exploit a vulnerability at a legitimate web hosting site, upload fake or malicious content to a web site, and create a phishing URL that points to the malicious content at the legitimate web site; in this case, the web server is compromised.

This distinction is important because it often identifies where investigators should go for assistance with mitigation of the criminal activity:

- If the domain is maliciously registered, an investigator will seek assistance from a domain name registrar, a TLD operator, or the operator that provides DNS for the malicious domain to suspend the domain name registration or name resolution.
- For a compromised domain, suspension would further victimize a legitimate party already victimized by the compromise, so investigators will contact the administrator of the compromised host to have the malicious content removed.

Note that parties that discover phishing pages will do their best to blocklist URLs that identify malicious content to avoid further victimization, whereas they may block maliciously registered domain names (and thus all hostnames and URLs created using this name) to contain the pervasive malicious activity.

For this measurement, we consider:

- **The age of the domain name — the number of days elapsed between domain registration and the use of the domain for a malicious purpose**. In general, the older the domain name, the higher the likelihood it will legitimate. Miscreants tend to use their domains within the first year of registration, before they must pay for renewal. The shorter the time between registration and use for phishing, the more likely the domain was maliciously registered.
- **The content of the domain name.** We apply rules to determine whether the composition of the name contains indicators of misuse or harmful intent, for example, the presence of a famous brand, a misspelled brand or a string intended to resemble a brand.

When the above criteria identify domains, we then look for clear evidence of common control and usage as an indicator to flag additional domains in a batch.

Prevalence of Maliciously Registered Phishing Domains in TLDs

Of the 497,949 domains reported for phishing in the study period, we identified 322,145 that we believe were registered maliciously, by phishers. This represents 65% of the domains**,** with the other 35% classified as compromised domains. This percentage is consistent with findings from our October 2020 Phishing Landscape study (where we found that 61% were maliciously registered) and a separate study by researchers at ccTLD operators SIDN and AFNIC, who found that 58% of phishing domains (in all TLDs) are maliciously registered, and 42% are compromised [8].

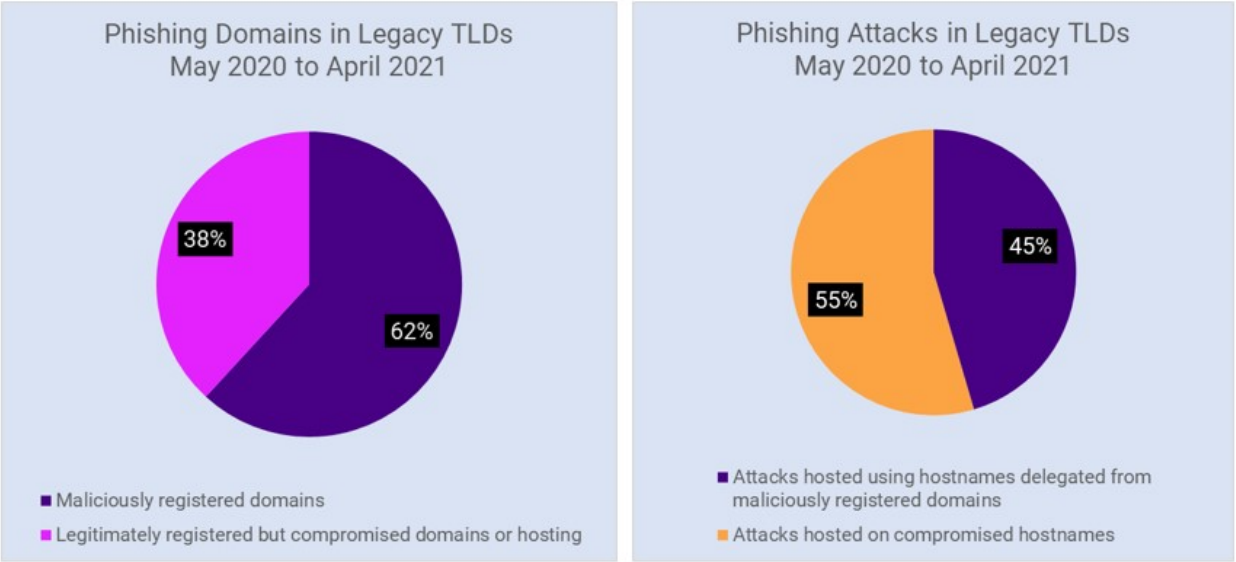

Figure 10 Phishing Domains and Phishing Attacks in Legacy TLDs, 1 May 2020 to 30 April 2021

In Figure 10 (left-hand chart), maliciously registered phishing domains account for a higher percentage of reported phishing domains (a 62:38 ratio), but in Figure 10 (right-hand chart), we observe that phishers host more attacks on compromised domains (a 55:45 ratio). This is consistent with theory that phishers find compromised hostnames attractive because they are harder to take down. In Figure 11 we see that this behavior is the nearly the same across ccTLDs (59:41 ratio compared to 51:49).

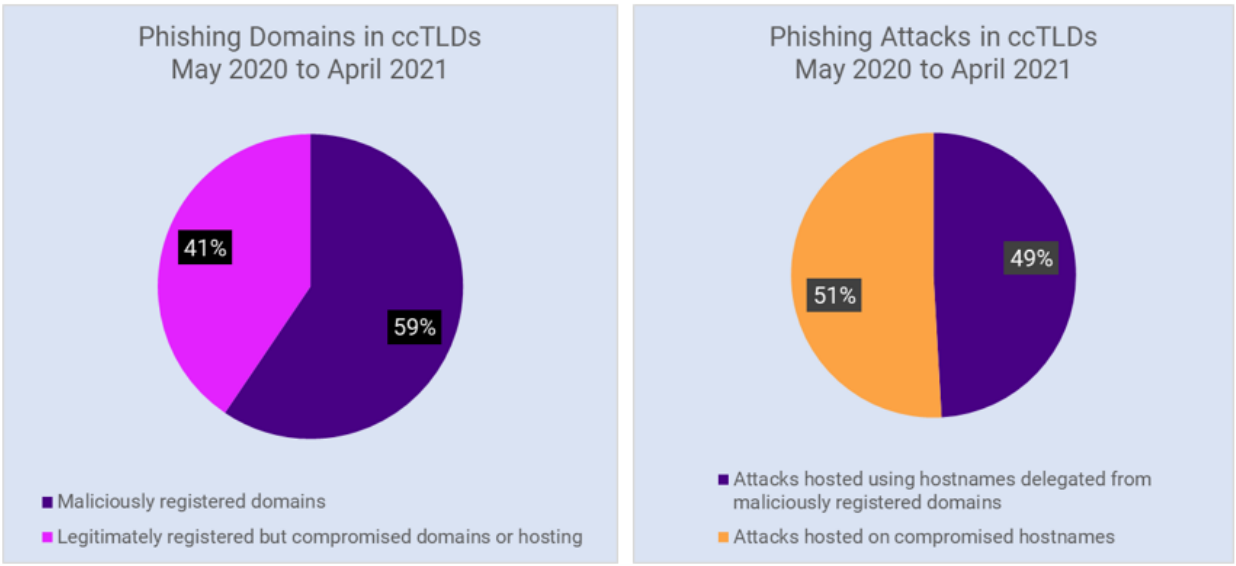

Figure 11 Phishing Domains and Phishing Attacks in ccTLDs, 1 May 2020 to 30 April 2021

In the new TLDs (Figure 12) we see nearly the same ratios (70:30 compared to 75:25). Many of the new TLDs where we observe phishing have extraordinarily high percentages of malicious registrations. In some cases, and for some periods of time, the percentages are so extreme that organizations have elected to blocklist the entire TLD.

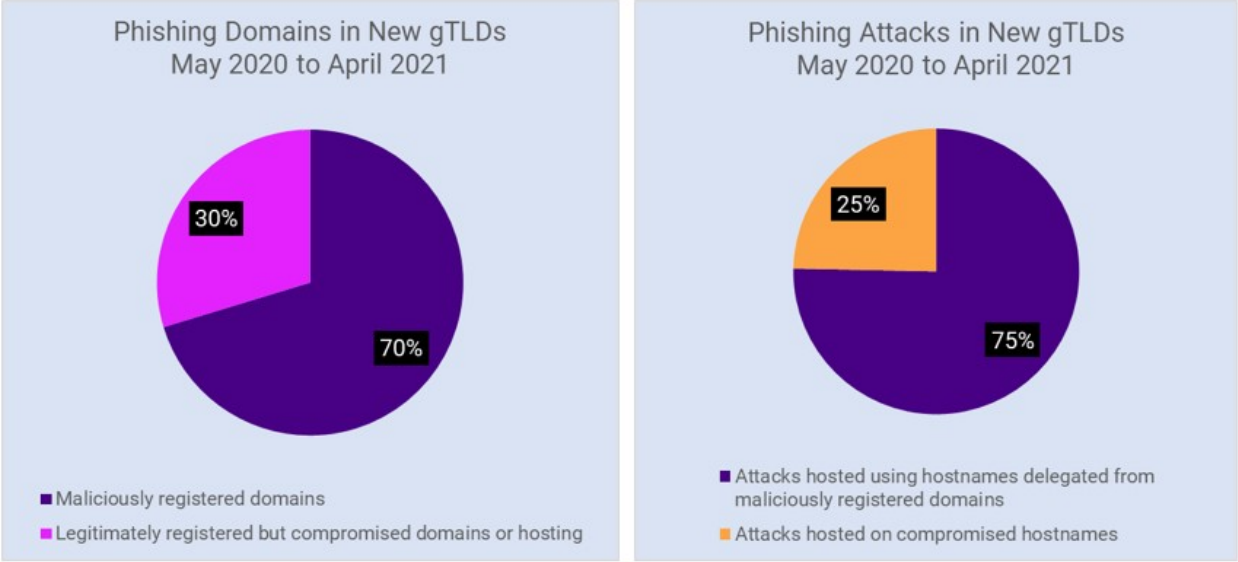

Figure 12 Phishing Domains and Phishing Attacks in New gTLDs, 1 May 2020 to 30 April 2021

We next look at the effects that malicious domain name registrations have on the levels of phishing activity in TLDs.

Malicious Domain Name Registrations and TLDs

Table 8 shows the Top 20 TLDs with cumulative malicious phishing domain registrations from 1 May 2020 to 30 April 2021.

Rank	TLD	Registry Operator	Domains in TLD	Malicious Phishing Domain Registrations (May 2020 – April 2021) ▼
1	com	Verisign	151,618,533	**160,896**
2	tk	Freenom	19,987,952	**40,002**
3	ml	Freenom	3,816,199	**27,284**
4	ga	Freenom	4,661,469	**21,657**
5	xyz	XYZ.COM	2,978,332	**20,039**
6	cf	Freenom	4,179,760	**19,187**
7	gq	Freenom	3,375,388	**16,168**
8	info	Afilias	4,278,926	**11,398**
9	top	Jiangsu Bangning	2,306,018	**10,744**
10	net	Verisign	13,407,660	**8,122**
11	buzz	DOTSTRATEGY	467,985	**7,733**
12	live	Dog Beach	526,119	**6,923**
13	online	Radix FZC	1,648,332	**5,546**
14	icu	ShortDot	4,158,251	**5,189**

Rank	TLD	Registry Operator	Domains in TLD	Malicious Phishing Domain Registrations (May 2020 – April 2021) ▼
15	org	Public Interest Registry	10,405,909	**4,891**
16	shop	GMO Registry	742,095	**3,965**
17	club	Registry Services	1,164,540	**3,301**
18	site	Radix FZC	1,549,543	**2,882**
19	best	BestTLD	100,508	**1,942**
20	link	UNR	132,529	**1,751**

Table 8 Malicious Phishing Domain Registrations, by TLD, 1 May 2020 to 30 April 2021

Counts of phishing domains help us to identify where domain names reported for phishing were registered, but further analysis is needed to understand what acts of prevention or mitigation are appropriate for individual TLDs.

By discriminating maliciously registered phishing domains from compromised domains (web sites), we identify the parties that are best positioned to combat phishing. Maliciously registered phishing domains can be suspended by the registrar or registry operator; this stops the attacks and will not cause any damage or inconvenience to anyone except the phisher. Registries with high numbers of maliciously registered domain names can collaborate with their registrars to adopt phishing identification and prevention measures. For compromised phishing domains, hosting network operators are best suited to mitigate vulnerabilities. They are also able to deploy measures to detect compromises and to recommend security content management practices that can reduce their customers' web vulnerability attack surfaces.

Figure 13 compares the number of maliciously registered domain names to the number of compromised domain names reported for phishing by TLD:

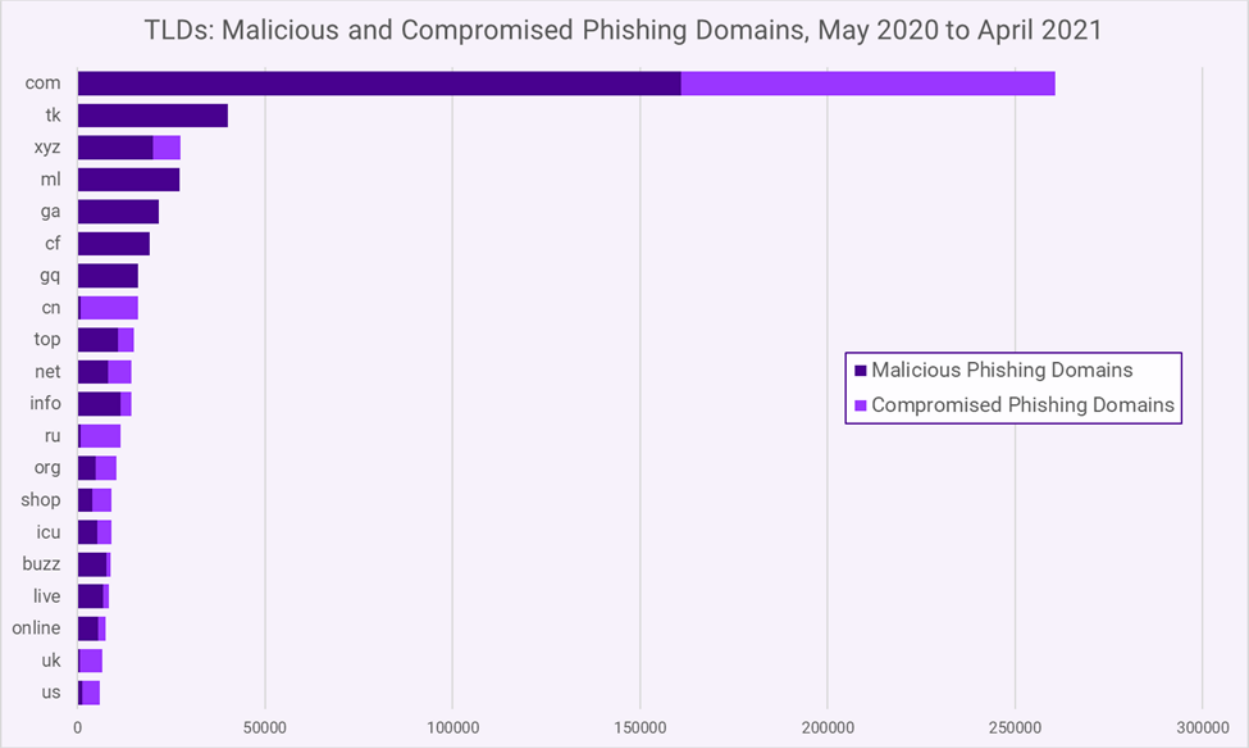

Figure 13 Malicious and Compromised Phishing Domains, by TLD, 1 May 2020 to 30 April 2021

In some TLDs, malicious phishing domain registrations dominate the count of phishing domains for the yearly period. This is particularly the case for the Freenom TLDs (.TK, .ML, .GN, .CF, and .GQ) but also the case for .INFO, .ICU, .BUZZ, .TOP, and .LIVE. The larger the number of malicious phishing registrations in a portfolio controlled by one TLD, the greater the need for that TLD to identify and suspend malicious phishing registrations early.

The number of phishing domains in other TLDs – notably, .COM, .CN, .NET, .RU, .UK, and .US – is less influenced by malicious domain registrations; here, compromised domains (hostnames) dominate the total. **When phishing occurs on compromised hosting, hosting providers are best positioned to take appropriate mitigation efforts**. While administrators of web sites can remove the phishing pages from the hosting server, phishers are highly unlikely to do so. The responsibility to remove fraudulent phishing content, disable an unauthorized web server, or suspend accounts of subscribers who are perpetrating phishing falls upon hosting operators. Typically, these are violations of the operator's own acceptable use policy.

Malicious Domain Name Registrations and gTLD Registrars

Counts of phishing domains help us to identify where domain names reported for phishing were registered. Further analysis is needed to understand what acts of prevention or mitigation are appropriate for gTLD registrars. By identifying characteristics of maliciously registered domain names and distinguishing these from compromised domains, we can identify which parties are best positioned to act to prevent phishing.

The classification *compromised domains* represents the set of domains where the domain name owner who operates a legitimate web site may be a victim. Here, investigators should seek out hosting

providers to mitigate phishing attacks (*e.g.,* by having the phishing page and related content removed from the compromised web site).

The classification *maliciously registered phishing domains* represents the set of domains that were purposely registered for phishing, by an actor with criminal intent (to commit fraud). Here, a gTLD registrar is often well positioned to (proactively) identify a domain as *"*intended for phishing*"*; for example, *only* a gTLD registrar has the means to:

- examine a domain name such as amazongjgasb14sjh21saknx.icu, appleidsupport-us.com, or customersupport-netflix.com during registration,
- detect a trademark or brand within the domain name (Amazon, Apple, Netflix), and
- suspend the registration while it reviews the registrant's contact data to assess the legitimacy of the registration.

The maliciously registered classification also represents the types of domains where investigators should seek the assistance of gTLD registrars to mitigate phishing attacks (*e.g.,* by suspending the domain name or registrant account). For example, when a phishing investigator determines that a phishing campaign is using dozens or more domain names containing random patterns, *only* a gTLD registrar can determine during the early hours of a phishing attack whether the contact data for a set of verified phishing domains is the same (an historically reliable indicator of a phisher). The gTLD registrar should review the evidence of phishing presented by a phishing investigator quickly and accommodate requests to reveal the contact data of a registrant once they verify the evidence.

Table 9 shows gTLD registrars with more than 1,000 cumulative malicious phishing domain registrations under management from 1 May 2020 to 30 April 2021.

Rank	Registrar	Malicious Domain Registrations May 2020 to April 2021 ▼
1	NameCheap	**60,629**
2	NameSilo	**28,105**
3	GoDaddy.com	**12,122**
4	PublicDomainRegistry.com (PDR)	**8,200**
5	Tucows Domains	**6,359**
6	Wild West Domains	**5,978**
7	Google	**5,679**
8	GMO Internet, Inc. (Onamae.com)	**5,394**
9	Name.com	**4,498**
10	Web Commerce Communications Limited (WebNic.cc)	**4,343**
11	Wix.com	**4,222**
12	Registrar of Domain Names (REG.RU)	**3,811**
13	eNom	**3,658**

Rank	Registrar	Malicious Domain Registrations May 2020 to April 2021 ▼
14	Jiangsu Bangning Science & technology Co.	**3,031**
15	Register.com, Inc.	**2,938**
16	Hosting Concepts B.V. (Registrar.eu)	**1,682**
17	ALIBABA.COM SINGAPORE E-COMMERCE PRIVATE LIMITED	**1,602**
18	Hosting Concepts B.V. (Openprovider)	**1,475**
19	BigRock Solutions	**1,449**
20	Eranet International Limited	**1,299**
21	Internet Domain Service BS Corp	**1,273**
22	Porkbun	**1,052**
23	Hostinger, UAB	**1,035**

Table 9 Registrars with at Least 1,000 Unique Malicious Domain Registrations, 1 May 2020 to 30 April 2021

We next compared malicious phishing domain registrations to compromised domains, by gTLD registrars. The raw numbers of maliciously registered domains are important — they indicate where phishers were able to purchase domains. Figure 14 uses the gTLD registrars from Table 9 to calculate the percentage of maliciously registered phishing domains vs. all phishing domains for each gTLD registrar. The gTLD registrars are sorted high to low by that percentage.

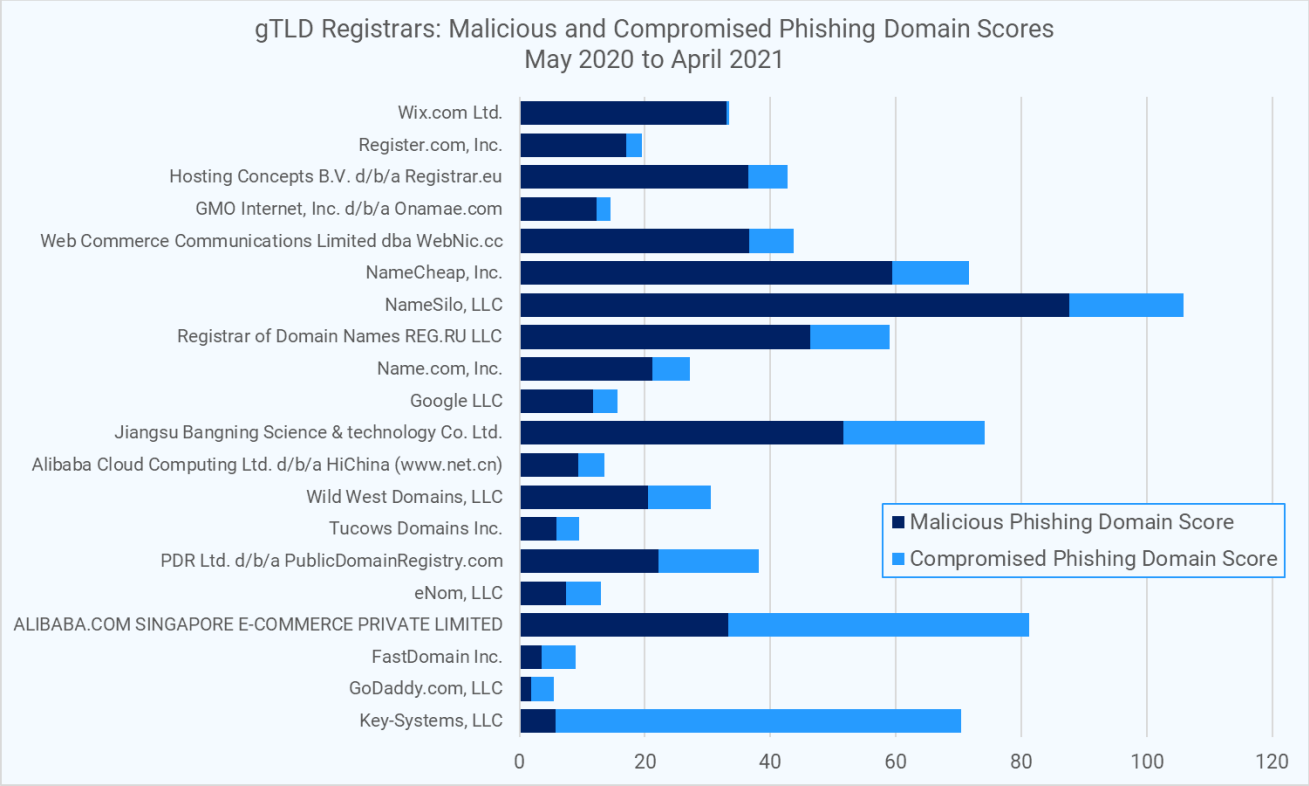

Figure 14 Malicious and Compromised Phishing Domains, by gTLD Registrar, 1 May 2020 to 30 April 2021

Malicious registrations directly influence the reputations of the gTLD registrars that are most targeted by phishers when they register domains purposely for phishing. In some cases, a gTLD registrar's malicious domain registration can also have a disastrous effect on the phishing score of a Top-level Domain and consequently on that TLD's reputation. For some TLDs, one gTLD registrar adversely influences a TLD's reported phishing domain counts month after month.

Which gTLD registrars have an adverse effect on which TLDs, and to what degree? In Table 10 we identify gTLD registrars that had an extraordinary or disproportionate influence on a particular TLD's phishing domain count over the entire period from 1 May 2020 to 30 April 2021.

Registrar	TLD	Registrar's share of all phishing domains reported for that TLD ▼
Key-Systems	bar	99%
GMO Internet, Inc. d/b/a Onamae.com	tokyo	99%
NameSilo	buzz	98%
NameSilo	date	98%
NameCheap	casa	96%
NameSilo	monster	88%
Chengdu West Dimension Digital Technology Co., Ltd.	wang	88%

Registrar	TLD	Registrar's share of all phishing domains reported for that TLD ▼
NameCheap	host	85%
NameCheap	services	80%
NameCheap	digital	78%
NameCheap	link	74%
Alibaba Cloud Computing Ltd. d/b/a HiChina	shop	67%
NameCheap	website	55%
NameCheap	club	55%
NameCheap	pro	55%

Table 10 Registrars with Highest Percentage of a TLD's Phishing Domains, 1 May 2020 to 30 April 2021

The new TLDs receive the most negative attention in policy communities and domain industry reports where the topics of DNS abuse or the criminal misuse of domain names and the DNS are discussed. Legacy gTLD registrars receive less attention but analyses of malicious domain registrations make it clear that *something* attracts criminals to certain registrars, and that there is consequential harm to the reputation of individual TLDs as well as to the new TLD program overall. Organizations and even individual Internet users have and continue to implement measures at firewalls or DNS resolvers to blocklist entire TLDs[9]. It is more difficult to blocklist *"*all the domains under the management of a gTLD registrar*"* but such a measure may be practical and necessary.

Opportunities to Prevent or Mitigate Malicious Registration Activity

Most phishing responses are reactive. Domains that will be used to lure users to phishing sites have already been registered, or they are obtained through some form of compromise attack. The phishing content has been composed and hosted. Spam emails or other means of presenting lures to Internet users have been transmitted. Victims have been harmed.

All gTLD registrars are contractually required by ICANN to have mitigation programs[10]. They must:

- maintain an abuse contact to receive reports of abuse and illegal activity, and publish the abuse contact address,
- publish on their website a description of their procedures for the receipt, handling, and tracking of abuse reports,
- document their receipt of and response to all such reports, and
- *"*take reasonable and prompt steps to investigate and respond appropriately to any reports of abuse.*"*[10]

Mitigation programs can reduce harms or losses, but our findings regarding malicious registrations show that phishers can and do register and use large numbers of domains at specific registries and registrars, again and again over time.

These levels of phishing activity might be caused by one or more of the following factors:

1) Low pricing, offered as part of a registrar and/or a registry operator's sales strategy. In general, phishers tend to be attracted to low prices [11].
2) Inattention to abuse problems by the registrar and/or the registry operator. This allows phishers to buy and use domains over time.
3) Features at the registrar that facilitate phishing, such as APIs that allow registrations in bulk, or payment methods that offer anonymity or have weak fraud detection. Cybercriminals take advantage of bulk registration services to "weaponize" large numbers of domain names [12].

Our purpose for defining a method to distinguish phishing domains as maliciously registered or as hosted on compromised assets is to make clear how phishers acquired resources. If phishers use malicious registrations more frequently than compromised assets, then prevention programs that deal with these registrations more proactively would be most helpful.

There may be opportunities for registry operators and registrars to use the methods that phishing investigators apply when phishing is first seen to suspend domains for malicious or illegal activity before they can victimize users and brands.

gTLD registrars and TLD operators are in an excellent position to identify and suspend malicious domain name registrations with a high degree of accuracy, often at the time of registration, and often by using the same methods that phishing investigators apply when phishing is first seen in the wild. For example, many domains registered by phishers also have telltale characteristics – name composition, common creation dates, similarities in contact data – that an operator can use to identify malicious registrations quickly and with low false-positive rates.

gTLD registrars and TLD operators possess key information – contact data and billing data – that no one else does. This data is highly useful for identifying malicious customers at the time of registration. Access to contact information – the registrant's identity, payment information, IP address, and purchase history – can be essential in a phishing investigation. Traditionally, phishing investigators would use WHOIS contact data to find other domains with similar contact data elements, and thus owned by the same cyber criminals. Only by identifying virtually *all* of a phisher's domain names can investigators hope to fully mitigate a phishing campaign.

gTLD registrars and TLD operators all have terms of service that allow them to suspend domains for malicious and illegal activity. Opportunities exist for registrars and registry operators to monitor for such activity, and to suspend domains for malicious purposes. Many operators have AUPs. Phishing is a recognized manifestation of fraud in arguably every jurisdiction in which registrars and TLDs operate. Stringently (and uniformly) enforcing a prohibition against phishing should result in a reduction in maliciously registered domains.

Phishing Attacks by Hosting Networks (Autonomous Systems)

An Autonomous System (AS) is a collection of the IP addresses (routing prefixes) controlled by a common network administrator. That administrator may be a hosting provider, a business, a university, an Internet Service Provider, or a network operator providing service to several of those types of entities. Each Autonomous System is assigned a unique AS number (ASN) for routing and identification purposes. It is common for larger hosting providers and infrastructure providers to have several AS numbers. Business and operational practices may cause an Autonomous System (and number) to be transferred from one hosting or infrastructure provider to another (*e.g.,* following an acquisition or divestiture). An AS and its number may be re-allocated as a result of other events (*e.g.,* bankruptcy or business closure). In light of this churn, we report on individual hosting networks (ASNs) rather than named hosting organizations.

We studied where phishing sites were being hosted, to determine if any hosting providers have outsized phishing problems. We collected the IP addresses (A records) that phishing attacks were resolving to. We then looked up what autonomous system (AS) each IP address was in. This provides insight into the entities that hosted the phishing attacks.

We are not seeing phishing on IPv6 addresses; the following sections are about IPv4 addresses only.

Ranking of Hosting Networks (ASNs) by Phishing Attacks Reported

Table 11 shows where larger-than-usual numbers of phishing occurred.

Rank	AS Name	AS number	# Routed IPv4 Addresses	Phishing Attacks ▼
1	NAMECHEAP-NET	22612	62,208	**55,903**
2	CLOUDFLARENET	13335	2,249,408	**52,011**
3	UNIFIEDLAYER-AS-1	46606	1,385,856	**35,363**
4	GOOGLE	15169	15,953,280	**32,330**
5	DIGITALOCEAN-ASN	14061	2,379,072	**15,794**
6	AWEX - Hostinger International Limited	204915	768	**13,186**
7	OVH - OVH SAS	16276	3,627,968	**12,604**
8	WEEBLY	27647	2,112	**10,701**
9	CONTABO - Contabo GmbH	51167	219,008	**10,635**
10	AMAZON-02	16509	41,090,304	**10,257**
11	AS-26496-GO-DADDY-COM-LLC	26496	1,385,280	**9,190**
12	MICROSOFT-CORP-MSN-AS-BLOCK	8075	38,122,752	**8,817**
13	HETZNER-AS - Hetzner Online GmbH	24940	1,977,408	**7,593**
14	CNNIC-ALIBABA-US-NET-AP Alibaba (US) Technology Co.	45102	9,553,984	**6,078**
15	DYNDNS	33517	65,537	**4,742**

Rank	AS Name	AS number	# Routed IPv4 Addresses	Phishing Attacks ▼
16	AS-COLOCROSSING	36352	788,544	**4,731**
17	ASN-QUADRANET-GLOBAL	8100	624,000	**4,534**
18	PUBLIC-DOMAIN-REGISTRY	394695	35,712	**4,383**
19	AMAZON-AES	14618	16,259,712	**4,319**
20	DDOS-GUARD CORP.	262254	12,608	**4,043**

Table 11 Ranking of Hosting Networks (ASNs) by Phishing Attacks Reported, 1 May 2020 to 30 April 2021

A few notes about the top results:

- #1 Namecheap is a domain name registrar that also offers hosting for its customers. Namecheap has a modest allocation of IPv4 addresses, but the largest number of phishing attacks.
- #2 Cloudflare provides a DNS redirection service that protects its customers from denial-of-service attacks. Cloudflare's service also prohibits observers from seeing the real hosting locations behind this defense network, and phishers take advantage of this to hide the hosting locations of phishing pages.
- #6 AWEX/Hostinger provides a free service that allows people to register subdomains and hosting. This service is abused heavily by phishers and had the highest ratio of phishing attacks to IPv4 addresses allocated to an ASN (17:1).
- #8 WEEBLY, another free website operator, had the second highest ratio of phishing attacks to IPv4 address allocation (5:1).

Ranking of Hosting Networks (ASNs) by Scoring Metrics

The gross numbers of phishing attacks reported are significant. Here, as with TLDs and gTLD registrars, more phishing attacks means more damage and victimization. A heavily abused ASN can enable many attacks. If it makes improvements to its anti-abuse efforts, it can reduce victimization and make things harder for phishers.

Gross numbers influence how one compares operators who have more of less IP addresses than each other (numbers bias). In the quarterly phishing activity published at the Cybercrime Information Center, the metric *"*Phishing Attacks per 10,000*"* is used to compare whether a hosting network (AS) has a higher or lower *incidence* of phishing relative to others. This is a ratio of the number of phishing attacks hosted in an Autonomous System to the IPv4 addresses routed by that hosting network (AS). We call this metric **hosting network (AS) Phishing Attack Score**:

hosting networks (AS) Phishing Attack Score =
*(number of phishing attacks/IPv4 addresses routed by AS) * 10,000*

For this report, and as we did for TLDs, we measured this prevalence of phishing in each hosting networks (ASNs) for the period 1 May 2020 to 30 April 2021. Here, we take the sum of the four quarters of unique phishing attacks reported and divide by the average of the IP addresses routed by each hosting network (ASN) during each of the four quarters. We call this metric **Yearly hosting network (ASNs) Phishing Attack Score**:

Yearly hosting networks (AS) Phishing Score =
(sum of the four quarters of unique phishing attacks reported /
*average of the IP addresses routed by AS) * 10,000*

Note that the calculation of these two metrics yields different results (we use different inputs for the numerators and denominators in the division); in particular, one cannot draw any conclusion by comparing the scores from a quarterly phishing score against an annual phishing score. Instead, we encourage comparisons of quarterly phishing scores over time, as well as annual phishing scores.

Table 12 shows the ranking of autonomous systems by phishing score:

Rank	AS Name	AS number	# Routed IPv4 Addresses	Phishing attacks	Phishing Attack Score ▼
1	AWEX - Hostinger International Limited	204915	768	13,186	**171,692.7**
2	WEEBLY	27647	2,112	10,701	**50,667.6**
3	PIHL-AS - Private Internet Hosting LTD	213058	704	1,714	**24,346.6**
4	IDNIC-JALANET-AS-ID PT. Jupiter Jala Arta	131775	2,304	3,446	**14,956.6**
5	WIX_COM - Wix.com Ltd.	58182	1,024	1,219	**11,904.3**
6	BEON-AS-ID PT. Beon Intermedia	55688	2,560	2,589	**10,113.3**
7	NAMECHEAP-NET	22612	62,208	55,903	**8,986.5**
8	ELITETEAM-PEERING-AZ1 - 1337TEAM LIMITED	51381	256	208	**8,125.0**
9	SEDO-AS - SEDO GmbH	47846	896	693	**7,734.4**
10	TRELLIAN-AS-AP Trellian Pty. Limited	133618	1,024	653	**6,377.0**
11	FLOWSPEC-AS - FLOWSPEC LTD	210138	256	155	**6,054.7**
12	IDNIC-JETORBIT-AS-ID PT Jetorbit Teknologi Indonesia	141584	256	149	**5,820.3**
13	LANDGARD-AS - Landgard Management Inc	44015	1,280	732	**5,718.8**
14	BODIS-NJ	395082	512	284	**5,546.9**
15	SKB-ENTERPRISE - SKB Enterprise B.V.	64425	1,963	912	**4,646.7**
16	AIRNET-AS - AIRNET llc	212860	256	118	**4,609.4**
17	HOST4GEEKS-LLC	393960	5,184	2,253	**4,346.1**
18	SNTHOSTINGS-AS-AP SnTHostings	140947	512	205	**4,003.9**
19	VFMNL-AS - Verotel International B.V.	31624	4,352	1,651	**3,793.7**
20	DDOS-GUARD CORP.	262254	12,608	4,043	**3,206.7**

Table 12 Ranking of Hosting Networks (ASNs) by Phishing Attack Score, 1 May 2020 to 30 April 2021

Examination of the phishing at the worst-ranked providers reveals that phishers took advantage of free services. These hosting providers did not effectively prevent or mitigate the phishing, allowing large-scale phishing often over an extended period.

#1 Hostinger is a Cyprus-based company that offers free web hosting and free subdomains on the domain 000webhostapp.com. This service has been abused heavily by phishers for the last several years. During our year-long study period, Hostinger's free service was used to impersonate at least 215 targets, especially WhatsApp, Google, and Facebook.

#2 Weebly also offers a free website-building service. Its users get free subdomains on the domain weebly.com. During our year-long study period, Weebly's free service was used to attack at least 100 targets, mainly AT&T, Verizon, and Yahoo!.

#3 PIHL-AS (Private Internet Hosting LTD) is nominally headquartered in Belize. The same company operates AS43350 in Russia. Half of the attacks launched on PIHL-AS's hosting were set up on free Freenom domains (.CF, .TK, .GA, .GQ, and .ML).

#4 IDNIC-JALANET-AS-ID PT. (Jupiter Jala Arta) is a provider in Indonesia. It appears that one phisher registered 2,898 free .TK domains and placed them on Jupiter Jala Arta's hosting. (The domains contained consecutive numbers, such as service-update10000000005674565288121-bg.tk, service-update10000000005674565288122-bg.tk, etc.) The attacks targeted Facebook.

#5 Wix offers a popular no-credit-card-required website building service. Of the 1,219 attacks, 507 appeared on the domain Wixsite.com, where Wix gives its customers free subdomains. Wix is also a domain name registrar. Phishers also took advantage of this aspect of Wix's business, registering hundreds of malicious domain names such as usersupport-alert.com and youraccounthasaproblem.com, and then hosting these domains on Wix’s servers.

List Coverage: The Phish That Get Away

By collecting data from multiple sources, we confirmed that there is low overlap between anti-phishing blocklists. For our 1 May 2020 to 30 April 2021 study period, we identified a total of 497,949 unique domain names listed for phishing (either URLs on those domains, or the domain itself). The Venn diagram (Figure 15) illustrates that most of the domains were reported via a single feed [13]. Only 1% of the domains – 6,231 out of 497,949 – were reported by all four feeds.

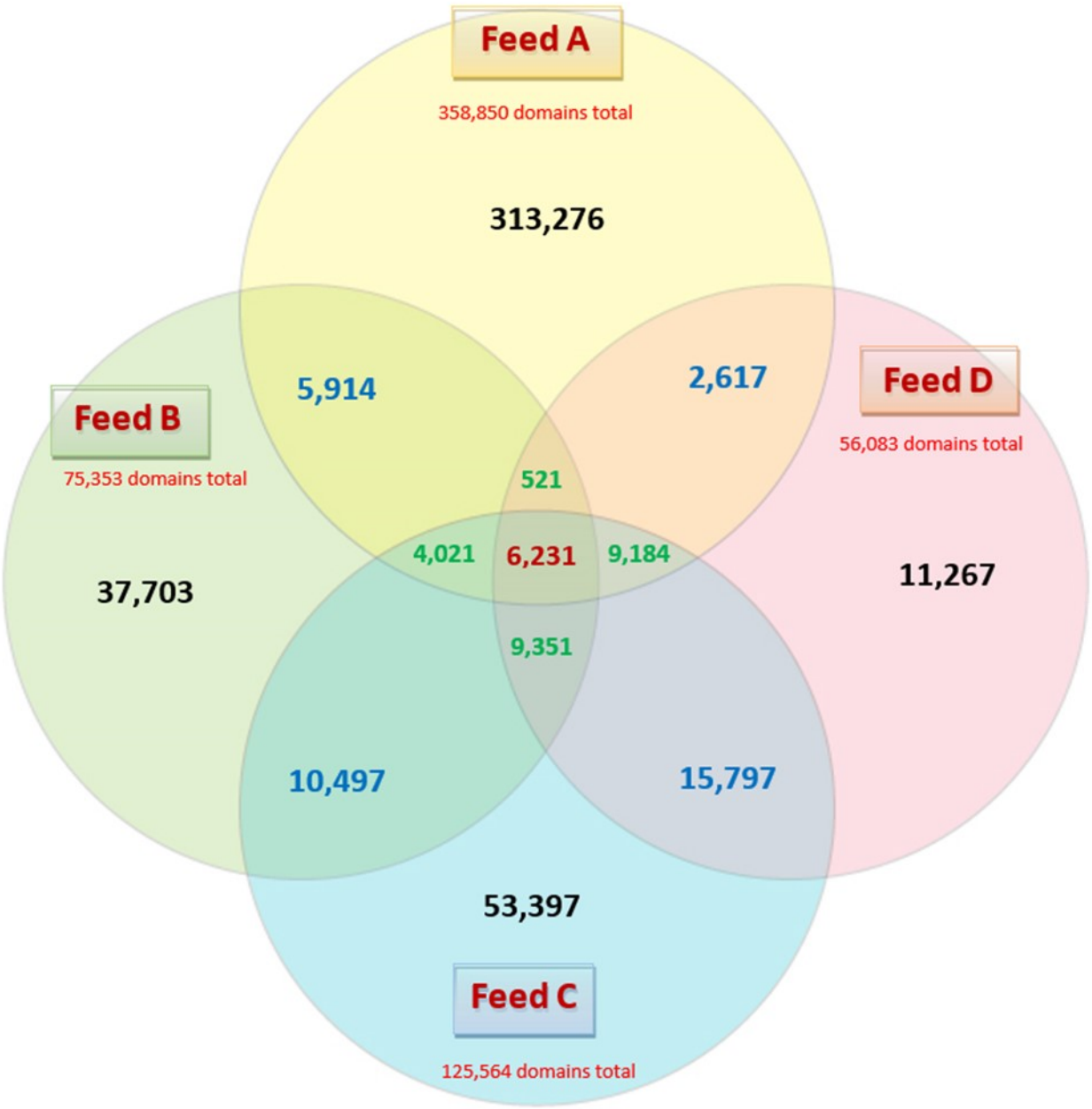

Figure 15 Overlap Across Feeds Providing Phishing Reports, 1 May 2020 to 30 April 2021

The existence of this coverage problem has been confirmed in a series of studies, which have found similar gaps for cybercrime data generally and for specific types of abuse including phishing [14, 15, 16, 17, 18, 19, 20].

What factors contribute to the low overlap? Some factors are common to detecting cybercrime generally, and some are especially relevant to phishing:

1. The Internet is a big place, and each blocklist provider only has a certain window of visibility into it. For example, a provider will have access to only a certain amount of email spam that it can scan for phishing lures.
2. The limited duration of phishing attacks provides only a small period in which observers can confirm the presence of a phishing site.
3. Phishers employ a variety of evasive techniques that complicate the confirmation of phishing attacks [21, 22, 23]. One called *"cloaking"* notably decreases the likelihood that a phishing site will be blacklisted, and if a URL does get blacklisted, the cloaking substantially delays blocking in browsers [24, 25].
4. The sharing of data is uneven and is not always timely. Some phishing targets do not share data about the phishing that affects them, for fear that it will reflect negatively on their brands. Some anti-phishing vendors do not share their data due to competitive concerns [26].
5. ICANN policy now allows gTLD domain registrars to redact all domain contact data from publication in WHOIS, even those records not covered by a privacy law such as GDPR. That contact data is a key tool for identifying malicious registrations and differentiating them from compromised domains. As we discussed in the section Why WHOIS is Important on page 51, **over-redaction of WHOIS data continues to contribute to the under-identification of phishing domains** [27, 28, 29].

How much phishing is not being detected at all? What is the number of "unknown unknown" attacks, and what is the total size (upper boundary) of the phishing problem? No one knows for sure. **Phishing is a much larger problem space than is reported.** The factors we list above inhibit even the best detection systems from finding much of the phishing attacks that occur and even the most professional and experienced observers can find only a portion of the phishing that occurs and are challenged to do so in a timely fashion**.**

Blocklists are essential tools for cybersecurity: they prevent enormous damage, and all organizations should take advantage of them directly or through their service providers. Organizations should further consider whether they are well served with one blocklist, or whether they would benefit from incorporating multiple sources of threat intelligence in their phishing defenses.

Regional Phishing and the Effect of Data Sharing

Our data set seems to significantly under-represent the phishing that takes place in certain regions. We suspect this under-reporting is the result of both an under-detection and an under-sharing of data.

For example, the data seems to under-represent the amount of phishing occurring in China, and against Chinese brands. Three of our four sources report target data. The data sources reported:

- No phishing attacks against China's four largest banks: the Industrial & Commercial Bank of China (ICBC), the China Construction Bank (CCB), the Bank of China, and the Agricultural Bank of China (ABC).

- No phishing attacks against major Chinese ecommerce companies Taobao, Baidu, JD.com, Pinduoduo, and Suning.
- 2,053 attacks against Tencent, plus an additional 17 attacks against Tencent's WeChat service. These were all reported by just one of our sources.
- 738 attacks against ecommerce giant Alibaba. Of those, 695 were reported by the same source that reported the phishing against Tencent.

Clearly, only one source is getting some data about phishing in China, and that data is apparently limited to certain brands, and limited in scope.

APWG studies in 2015 and 2016 found that a significant amount of phishing takes place in China, against the types of targets noted above, but that such phishing was not being discovered or reported by sources outside of China [30]. Those studies included data contributed by the Anti-Phishing Alliance of China (APAC), which works with phishing targets inside of China. The 2016 APWG study found that more than half of malicious gTLD registrations worldwide were being made by Chinese phishers, and that six of the top ten registrars of malicious phishing domains were located in China and had primarily Chinese customers. That volume of phishing may have changed in the last five years – but we cannot tell based on the current inputs. That kind of in-country data was largely absent from the data sources we observed in 2020-2021, and its absence is obvious. Observers outside of China are not making detections of those kinds of phishing attacks because they are not receiving Chinese-language email and SMS lures, may not be parsing Chinese-language emails effectively, and because sources in China are not sharing data. There are commercial forces at work as well — anti-phishing and blocklist providers outside of China may not have customers inside of China and therefore do not have an incentive to find phishing that affects Chinese targets and victims. Our data showed that over 4% of phishing attacks were on domains registered at Chinese registrars, but the attacks targeted non-Chinese brands almost exclusively, notably Microsoft and Japanese companies.

In contrast, our data set includes hundreds of attacks against Russian phishing targets, because an APWG member in Russia is contributing data to the APWG phishing feed. This provides information about targets such as Столото, also known as Stoloto.ru, which distributes state lotteries in the Russian Federation. Much of the phishing that targeted Russian companies occurred on .RU and .SU domains. Still, our data set contained only 17 attacks against the popular Russian email and search provider Yandex, suggesting that Yandex or its anti-phishing vendors are not sharing data with the sources we monitor.

Our data set also contained 3,202 attacks against the leading Brazilian retailer Magazine Luisa, plus records of attacks against scores of banks across Central and South America. The data also contains more than 13,300 attacks against brands in Japan, including Rakuten's Japanese-language site and Japanese government sites. Most of these attacks against Latin American and Japanese targets were reported by Anti-Phishing Working Group members operating in those regions.

The coverage of phishing against Russian, Japanese and Latin American targets provided by APWG members is especially valuable. This data sharing provided visibility, better blocklisting, and better protection.

Targeted Brands

Phishers targeted 1,804 businesses or organizations during the 1 May 2020 to 30 April 2021 period, including banks, social media companies, webmail, and games, national tax services, universities, and cryptocurrency exchanges. Figure 16 depicts the top 10 brands targeted over the course of our annual period, accounting for 46% of the phishing attacks associated with specific brands.

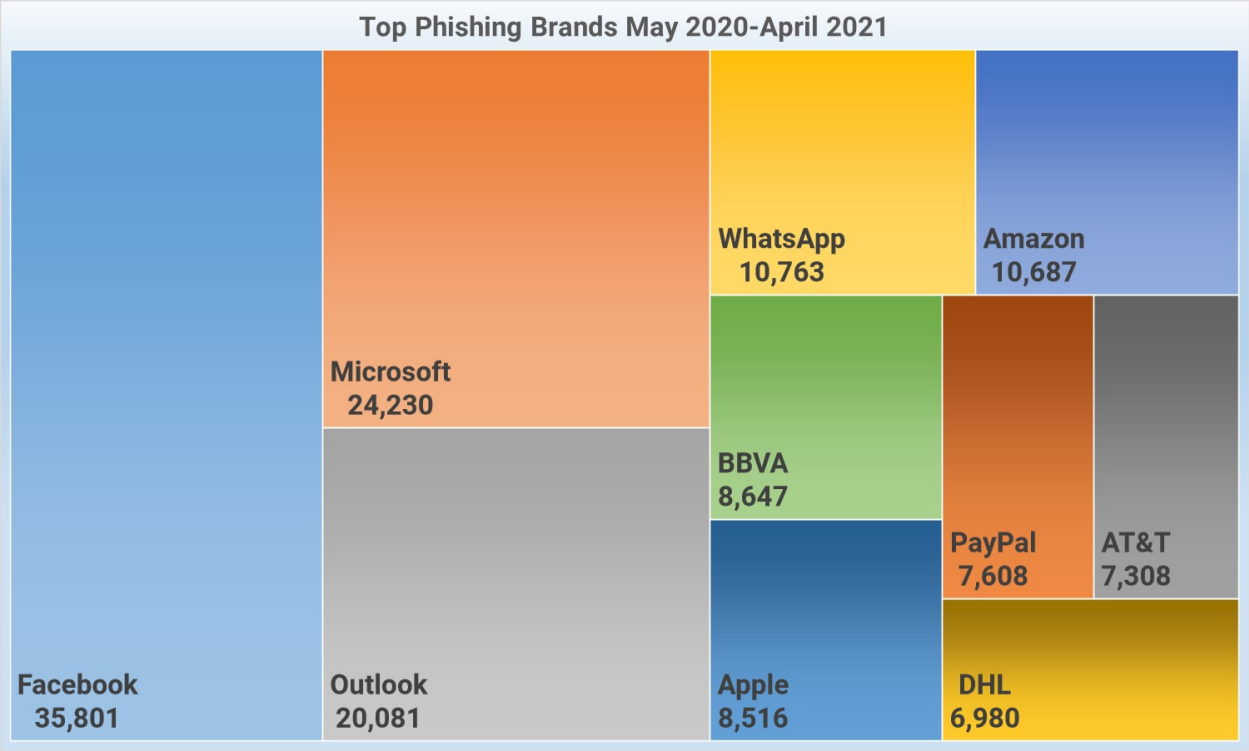

Figure 16 Top Phished Brands, 1 May 2020 to 30 April 2021

In Figure 17, we used quarterly reporting data published at the Cybercrime Information Center to show the number of brands identified as targets in phishing attacks quarter over quarter. The trendline shows that phishing has increased quarter over quarter. (Note that the reporting period 1 February 2021 to 30 April 2021 has the fewest number of days reported.)

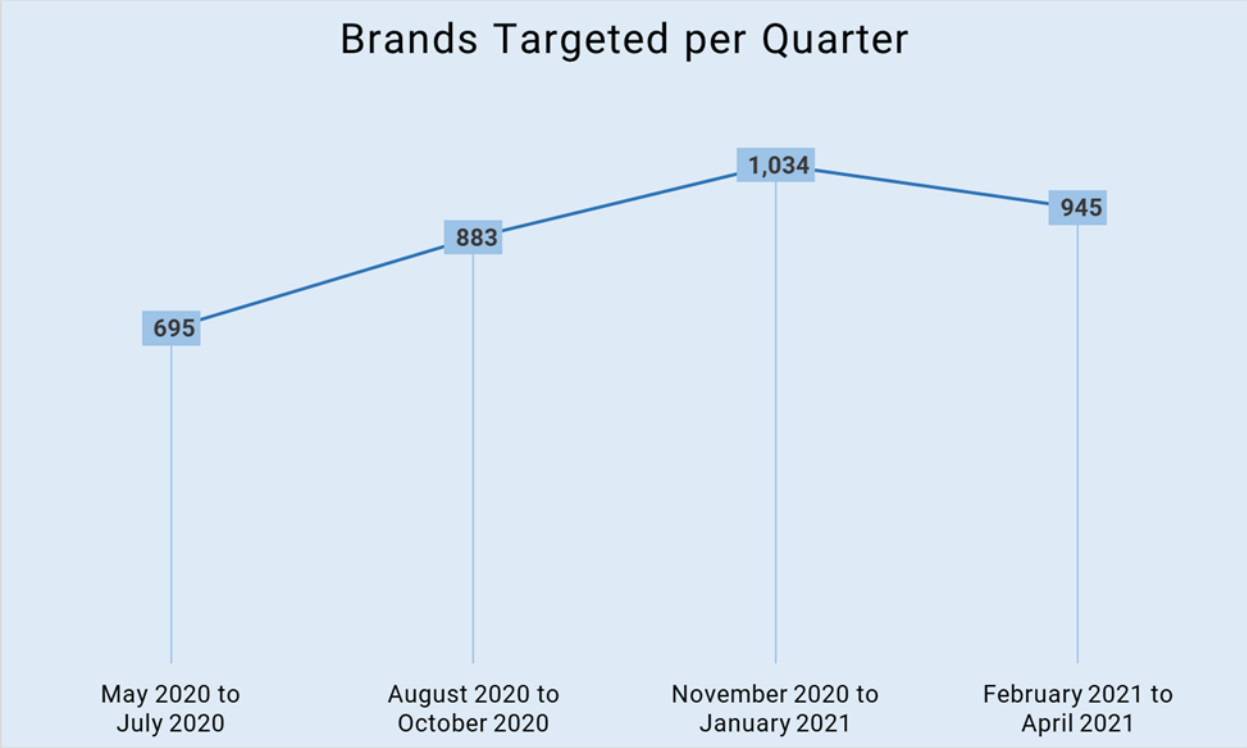

Figure 17 Brands Targeted each Quarter

To identify targeted brands, we used three URL blocklists that identify targets in the metadata included in phishing reports. Reports from each phishing feed we consume vary slightly in granularity and nomenclature. We compiled lists of these variations and normalized spelling as part of our curation; for example, if one feed uses "PayPal" while another uses "PayPal Inc.", we treated these as one target and normalized our data to a common form of the company name so that we could analyze brand data.

Some feeds pose classification challenges. For example, WhatsApp is owned by Facebook. Some sources report WhatsApp as a separate brand, but other sources report the same WhatsApp phishing URLs as attacks against Facebook. We used the target reported by each feed, with the granularity (discrimination) that feed offers.

In some cases, one source may positively identify a URL as a phish against a specific target, but another source may only report the same URL as a phishing attack against "unknown" or "generic" brand. In these cases, we used the most detailed information available and attributed that attack to the specific brand. In the cases where an attack's target is not determined by any feed, we set those attacks aside when analyzing brand data.

Table 13 identifies the most targeted brands, by annual ranking and then quarterly, from May 2020 through April 2021:

Annual Ranking	Brand	Attempted Phishing Attacks	Quarterly Ranking			
			May-July 2020	August-October 2020	November 2020-February 2021	February-April 2021
1	Facebook	35,801	2	3	1	1
2	Microsoft	24,230	1	2	4	3
3	Outlook	20,081	8	1	3	6
4	WhatsApp	10,763	6	8	9	2
5	Amazon	10,687	4	4	8	7
6	BBVA	8,647	314	266	2	97
7	Apple	8,516	11	12	7	5
8	PayPal	7,608	3	7	13	14
9	AT&T	7,308	5	5	11	15
10	DHL	6,980	26	24	10	4
11	Lloyds	6,209	94	40	6	9
12	Halifax	5,465	206	9	5	30
13	eBay	4,945	13	6	22	29
14	Instagram	4,721	29	19	14	8
15	webmail	4,488	12	13	19	20
16	Chase	4,409	10	16	21	18
17	Netflix	4,031	16	11	23	22
18	Rakuten	3,961	20	10	15	26
19	Magazine Luiza	3,202	17	17	28	28
20	Intesa Sanpaolo	3,202	111	35	12	17

Table 13 Most Targeted Brands, 1 May 2020 to 30 April 2021

A brand can become a phishing target at any time. Phishers constantly look for companies that have potentially lucrative user information, are newly popular, or are not ready to respond to phishing. Phishers also use a variety of ploys to lure Internet users to their phishing pages including,

- a new product announcement,
- a critical software update to obtain,
- an issue with a social media account or financial account,
- a problem with a merchant transaction or subscription,

- an inquiry regarding a criminal matter or tax violation,
- a newsworthy or catastrophic event, such as the COVID-19 pandemic, or
- an emerging technology or service such as cryptocurrency.

Cryptocurrency Phishing

The cryptocurrency market topped two trillion dollars in April 2021 [31]. The bullish interest in the market, which was up over 180% from April 2020 to April 2021, also attracted phishers.

Cryptocurrency phishing objectives are the same as bank phishing: steal money, credentials, and personal identifying information. Many cryptocurrency phishing schemes involve attacks on wallets – a mobile app, browser extension, or hardware device that stores cryptocurrency keys and allows users to buy, sell, and store cryptocurrency. Figure 18 illustrates how one cryptocurrency phishing attack used the threat of having their wallet account closed to lure users to a bogus MyEtherWallet web site, where victims disclosed information used to access wallets[32].

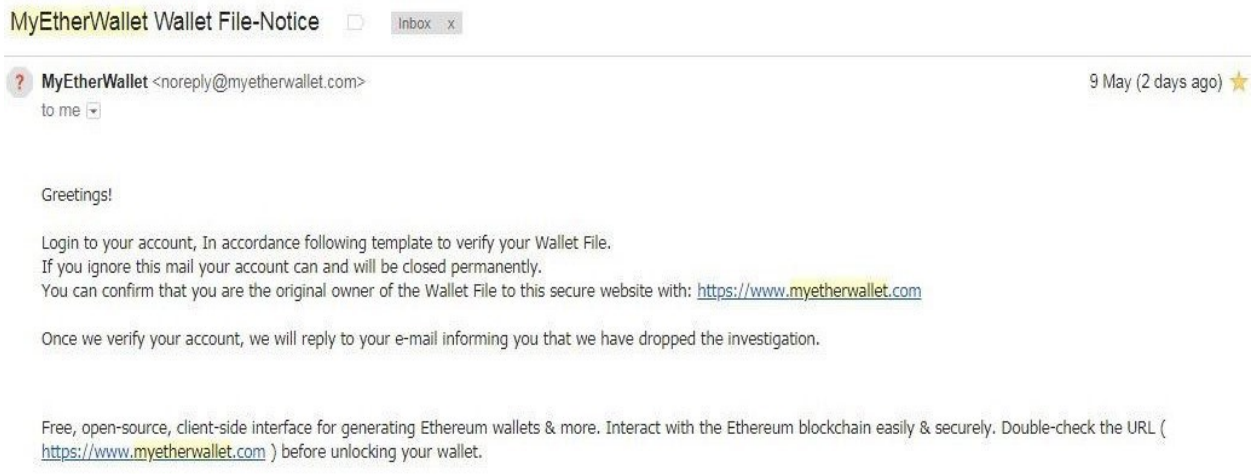

Figure 18 MyEtherWallet Phishing Email

A separate phishing attack told notified users that they needed to resynchronize their MyEtherWallet [33]. In this campaign, the phishers used a homoglyph attack: they added an underdot to the first letter *"e"* in the brand "Ledger" and composed a phishing URL from the domain ļedger.com to lure Ledger customers to a fake site [34].

Neither of these attack methods is new. Cryptocurrency phishers simply apply methods that they have successfully employed in prior campaigns. They are practiced at all aspects of these attacks as well; for example, a screenshot of a fake Trezor wallet website [35] (Figure 19) shows that phishers produce the same credible fakes to deceive cryptocurrency adopters as they have used to deceive banking customers.

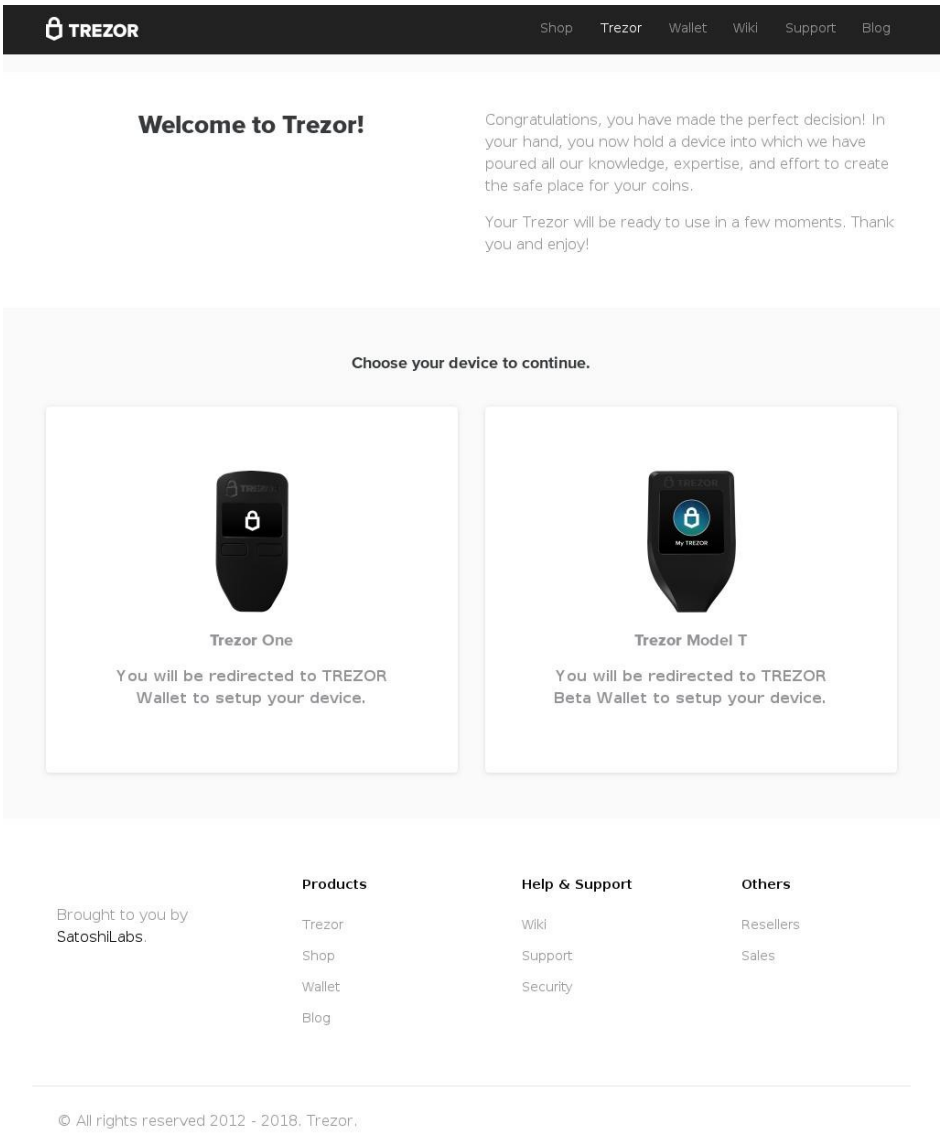

Figure 19 Fake Trezor Wallet Website, Courtesy of PhishTank

Phishers also attacks cryptocurrency trading platforms and traders. They have lured traders with offers of free trading bots to a fake site where the traders disclose their cryptocurrency keys (also known as API keys) [36].

To study phishing activity in cryptocurrencies, we constructed a list of keywords from names or trade symbols of cryptocurrencies and complemented this list with names of cryptocurrency end-uses, e.g., traders, exchanges, or payments. We also included companies that provided cryptocurrency related components – software apps, miners, wallets, or hardware (wallets). We next grouped the keywords and brands into target "classes": Virtual wallet companies, Cryptocurrencies, Trading exchanges, Mining and Mixing.

Using these, we identified 5,442 phishing attacks in our 1 May 2020 to 30 April 2021 data with a cryptocurrency connection. Keywords associated with digital wallet companies or their virtual wallet apps were the most frequently encountered among the attacks that we associated with cryptocurrency phishing (Figure 20). This is consistent with user reports [37, 38] and cryptocurrency attack tracking by Kaspersky, who reported that fake cryptocurrency exchanges, fake sales of crypto mining hardware and phishing pages designed to steal [virtual] wallet private keys were the most encountered forms of attack [39].

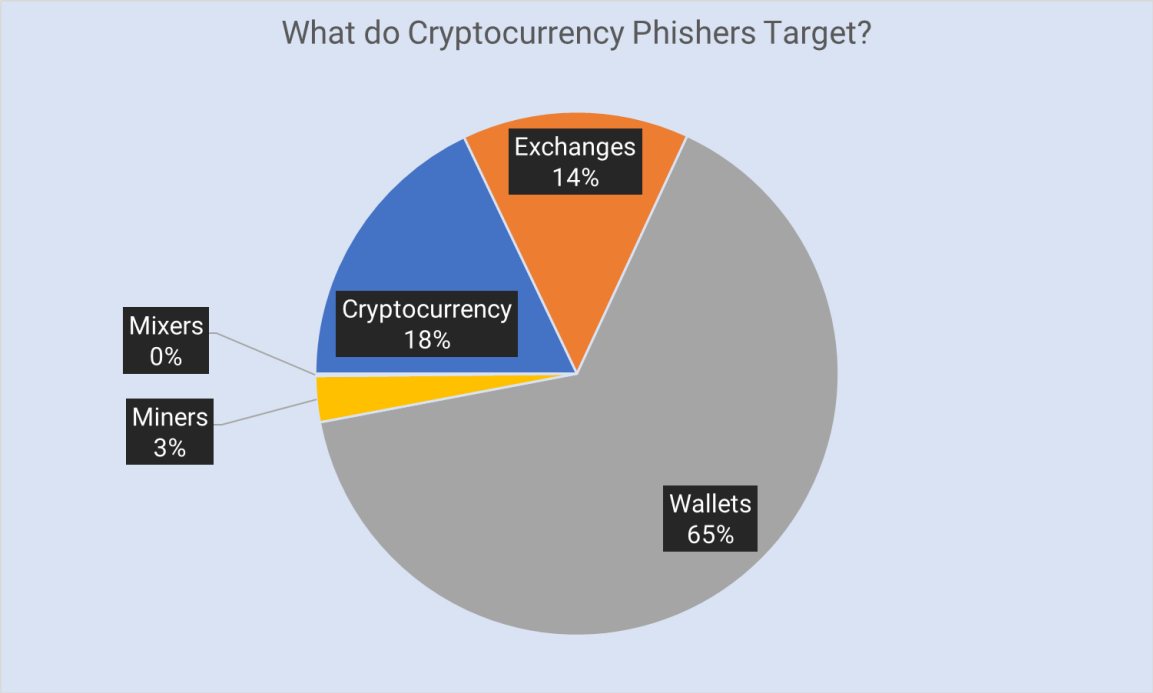

Figure 20 Most Often Encountered Target Classes of Cryptocurrency Phishing, 1 May 2020 to 30 April 2021

Bitcoin and Ethereum were the most frequently encountered "brand" keywords in the cryptocurrency phishing URLs that we collected (Figure 21).

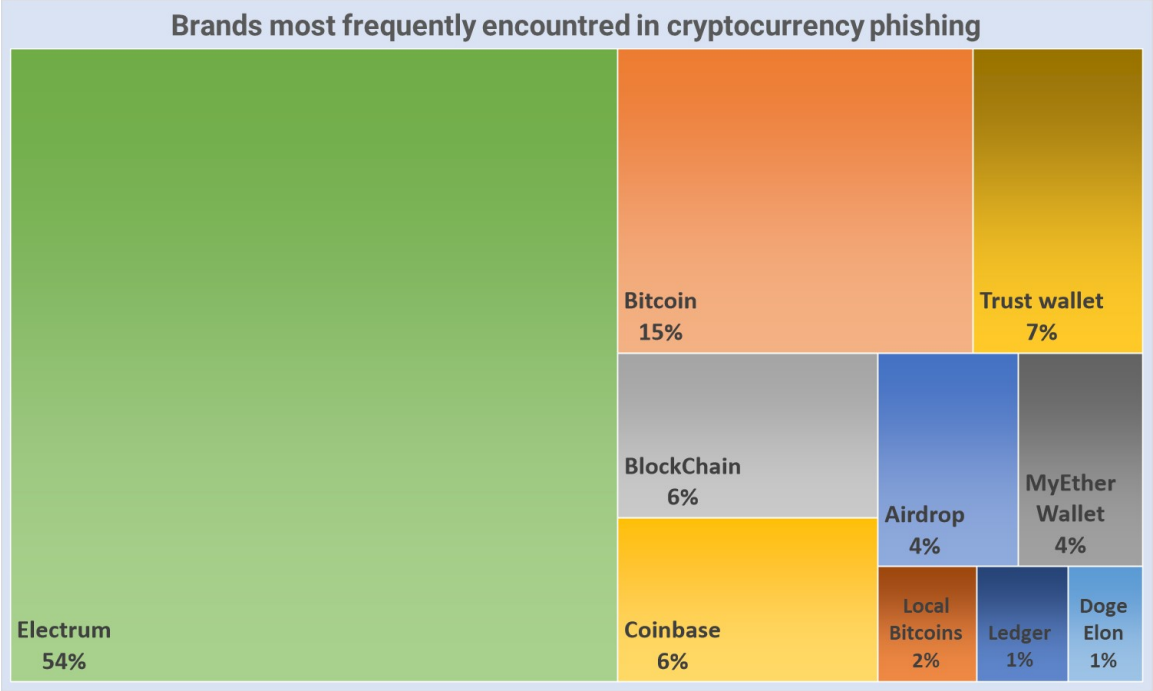

Figure 21 Brands Most Frequently Encountered in Cryptocurrency Phishing, 1 May 2020 to 30 April 2021

We were able to associate 5,023 unique domain names to these attacks. We identified 194 TLDs with at least one cryptocurrency phishing domain reported. Of these, 33% of the cryptocurrency phishing domains were registered in the .COM TLD. This is a much smaller percent than .COM's market share and suggests that phishers may cast their net wider for cryptocurrencies. While no other TLD had more than a 4% share of cryptocurrency phishing domains reported, we found that:

- 129 new gTLDs had at least one cryptocurrency phishing domain reported,
- 58 ccTLDs had at least one cryptocurrency phishing domain reported,
- 47% of the cryptocurrency phishing domains were registered in the new gTLDs,
- 13% of the cryptocurrency phishing domains were registered in ccTLDs,
- 7% of the cryptocurrency phishing domains were registered in legacy TLDs other than .COM, and
- 33% of the cryptocurrency phishing domains were registered in .COM.

Figure 22 shows the Top 10 TLDs, by number of cryptocurrency phishing domains reported.

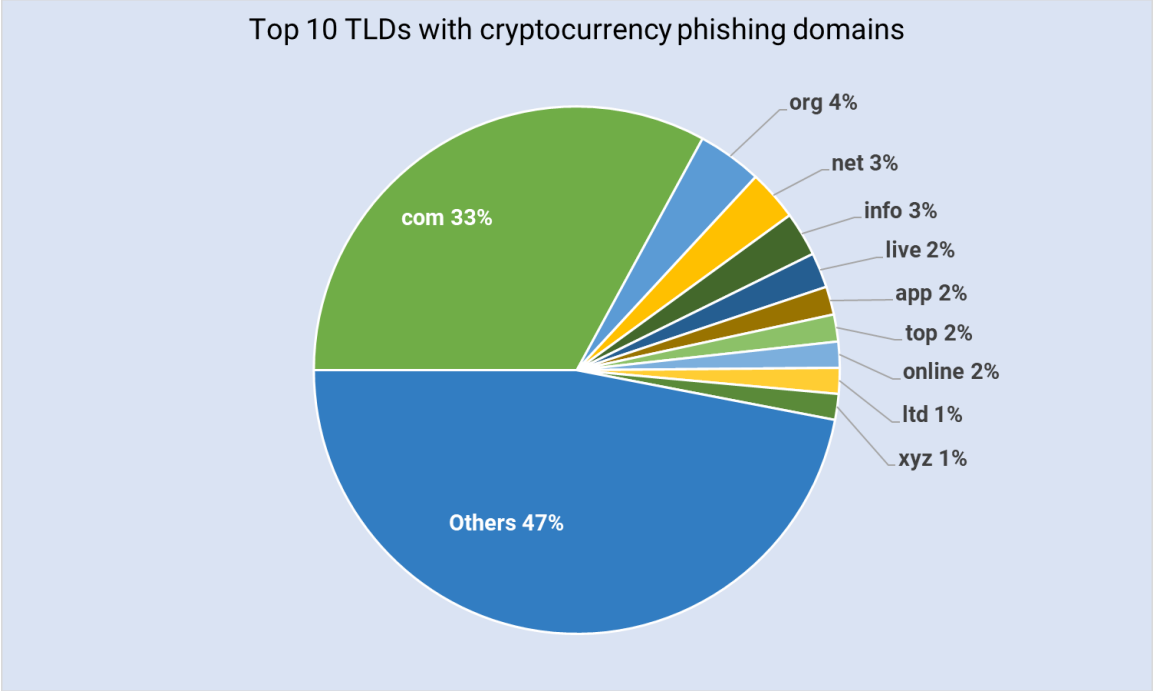

Figure 22 Top 10 TLDs, by Number of Cryptocurrency Phishing Domains, 1 May 2020 to 30 April 2021

One hundred and twenty-three gTLD registrars had at least one cryptocurrency phishing domain under management. Among gTLD registrars, NameCheap and NameSilo had the highest number of cryptocurrency phishing domains.

Figure 23 shows the Top 10 gTLD registrars, by number of cryptocurrency phishing domains under management.

gTLD registrar	Number of cryptocurrency phishing domains under management
NameSilo	1,965
NameCheap	1,157
GoDaddy	183
PublicDomainRegistry PDR	175
Hostinger, UAB	148
Web Commerce Communications (WebNic.cc)	82
Registrar of Domain Names REG.RU	81
Danesco Trading	67
OwnRegistrar	65
Tucows Domains Inc.	44

Figure 23 gTLD Registrars with the Most Cryptocurrency Phishing Domains, 1 May 2020 to 30 April 2021

We observed that 257 hosting networks (ASNs) had at least one IPv4 address that hosted a cryptocurrency phishing attack.

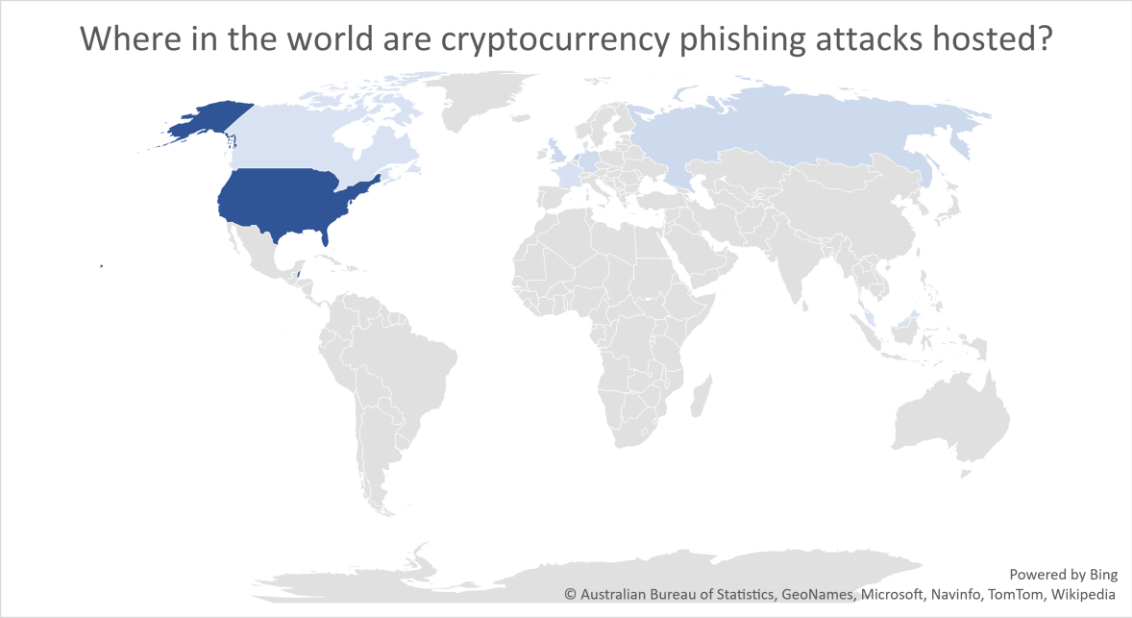

Figure 24 Where in the World are Cryptocurrency Attacks Hosted, 1 May 2020 to 30 April 2021

The United States (1,625), Great Britain (1,484), Belize (1,480), Netherlands (283), and Russia (135) are the countries where the highest numbers of cryptocurrency phishing attacks were hosted.

Abuse of Subdomain Service Providers

Our analysis reveals that 10.7% of all phishing attacks took place using resources at subdomain service providers. This was up from 9% in our previous report [6]. Subdomain services give customers services on a domain name that the provider owns. This gives users their own DNS space, on a third-level domain, of format:

subdomain.domainname.tld

Some of these providers are web hosts; some offer just the third-level domain with free DNS management so the domain owner can point it to other hosting. Phishers use the domains and hosting offered by these providers to build and maintain phishing sites.

This use of subdomain services is a challenge for several reasons. Many of these companies offer the services for free. Some offer anonymous registration, with little to no identify validation. Finally, only the subdomain service providers can effectively mitigate these phishing attacks. Some providers apparently lack proactive measures to keep criminals from abusing their services.

We identified 74,315 phishing attacks using subdomains provider services. They sat on just 583 second-level domain names. Of those 74,315 attacks, 90% of them (67,412) occurred on domains operated by just ten providers. **This emphasizes how a service of this type can be used to perpetrate significant amounts of damage, and how important it is for such providers to have proactive and quick anti-abuse monitoring and takedown capabilities.** Those providers were:

Rank	Provider	Domains	Phishing Attacks
1	Google	appspot.com, Blogspot domains	21,587
2	Hostinger	000webhostapp.com	12,763
3	Weebly	Weebly.com	10,155
4	DynDNS	multiple	9,580
5	ChangeIP	multiple	4,507
6	DuckDNS	duckdns.org	2,499
7	No-IP	multiple	2,391
8	GoDaddy	godaddysites.com	1,731
9	yolasite.com	yolasite.com	1,481
10	webcindario.com	webcindario.com	824
11	Wix	wixsite.com	771
12	ngrok	ngrok.io	638
13	SpaceWeb	swtest.ru	486
14	CentralNIC	uk.com, ru.com. gb.net, com.de, us.com, etc.	438
15	MailJet	mjt.lu	426
16	WebNodeAG	webnode.com	300
17	Moonfruit	moonfruit.com	293
18	Typeform	typeform.com	290

Rank	Provider	Domains	Phishing Attacks
19	Microsoft	azurewebsites.net, cloudapp.net	231
20	Digital Ocean	digitaloceanspaces.com	192

Table 14 Phishing Attacks via Subdomain Service Providers, 1 May 2020 to 30 April 2021

#1 Google had tens of thousands of phishing attacks hosted on Appspot.com, a Google cloud computing platform for developing and hosting web applications in Google-managed data centers, and on its Blogger product.

#2 Hostinger is a Cyprus-based hosting provider that offers free hosting on its 000webhostapp.com domain. This free service has been used extensively by phishers for several years.

#3 Weebly offers a free website builder service, which is used frequently by phishers. Weebly is a subsidiary of Square, the payment processing company.

#4 DynDNS, #5 ChangeIP, #6 DuckDNS, and #7 No-IP offer dynamic DNS services, which allow one to access devices from the Internet via a simple-to-remember domain name. These services are often free, and can obscure the real location of hosting, and thus are attractive to phishers.

Why WHOIS is Important

Domain name WHOIS data is essential for identifying and investigating phishing. WHOIS data allows investigators to determine when domain names were registered, at what registrar, and details about the technical infrastructure used to support the domain name. WHOIS data has traditionally included contact data for the party who registered the domain (the "registrant") and is responsible for it. This contact data is valuable for identifying malefactors, for identifying risky domains for blocklisting, and for alerting innocent registrants who have had their web hosting compromised by phishers.

Changes to WHOIS data availability since May 2018 have seriously undermined its usability for fighting online crime. Contact data is now mostly redacted in public WHOIS, as allowed by new ICANN policies designed to allow compliance with Europe's General Data Protection Regulation (GDPR). While there are mechanisms for qualified parties to request hidden contact data for legitimate purposes, these procedures are not uniformly implemented, the domain registrars who hold the data sometimes do not respond or reject legitimate requests, and when data is provided it is usually long after it would be useful for mitigating phishing attacks.

Two recent independent studies have found that the redaction of WHOIS contact data is excessive:

1. In February 2021, academic researchers presented a large-scale, systematic measurement study of how WHOIS data providers have been complying with ICANN's new GDPR policies [40]. Using a collection of 1.2 billion WHOIS records spanning two years and software they developed to automate compliance checking, the authors concluded that ***"the scope of privacy protection is usually excessive in practice, causing a global impact on the WHOIS system."***
2. In the January 2021 Interisle report *WHOIS Contact Data Availability and Registrant Classification Study* [41], we determined that **86.5% of registrants can no longer be identified via WHOIS – up from just 24% before the ICANN policy went into effect**. Our study explains how ICANN policy has **allowed registrars and registry operators to withhold much more contact data than is required by GDPR—perhaps five times as much as is necessary.**

In the spring of 2021, the two major professional organizations that fight online fraud and cybercrime polled their members about the utility of WHOIS. Conducted by the Anti-Phishing Working Group (APWG) and the Messaging Malware Mobile Anti-Abuse Working Group M3AAWG), the study found [42]:

- Two-thirds of respondents indicate that their ability to detect malicious domains has decreased,
- 94% report that redaction impairs their ability to investigate relationships between domains and actors, and
- 70% of respondents report that time to mitigate or respond exceeds an acceptable threat threshold due to the changes to WHOIS access, and
- 65% of cybersecurity experts indicated that they need a full response for non-public registration data within one day when addressing malware, phishing and botnet/command and control incidents.

The survey respondents also indicated that requests to access non-public WHOIS by legitimate investigators for legitimate purposes are not fulfilled effectively by the registrars and registry operators. Requests to disclose redacted WHOIS data are regularly ignored or denied, and "revealed" data are often not actionable (not provided in timely manner). Less than a quarter of respondents now make data requests, and more than a third do not bother because they deem it "too laborious, not worth it."

These complaints are consistent with a status report by AppDetex on requests made through the WHOIS Request System [43] during the course of legitimate trademark infringement cases:

> *"During the period from September 1, 2020 through February 28, 2021, we submitted 4,575 requests to 182 ICANN-accredited registrars. Of those individual requests, only 10.1% resulted in responses that included registrant data. Of the 182 registrars to whom we made requests, 121 registrars provided registrant data. Sixty-one registrars were completely unresponsive to our requests for registrant data. While the majority of registrars acknowledge requests for data, they provide NO data."*

In the meantime, the most recent effort at ICANN to improve the situation is concluding and will not implement any meaningful changes. ICANN plans to build a standardized access/disclosure system to provide access to qualified parties, but the system has not been designed, remains years from implementation, and parties representing date requestors have expressed doubt that the system will provide meaningful benefits [44, 45, 46].

As contact data is effectively accessible only to TLD registrars and some registry operators, it becomes incumbent on these parties to take a more prominent role in early phishing intervention and prevention.

Use of Internationalized Domain Names (IDNs) for Phishing

Data continues to show that the unique characteristics of Internationalized Domain Names (IDNs) are not being used to facilitate phishing in meaningful numbers.

IDNs are domain names that contain one or more non-ASCII characters. Such domain names can contain letters with diacritical marks such as ă and ü or be composed of characters from non-Latin scripts such as Arabic, Chinese, or Cyrillic. Over the past sixteen years, IDNs have been available at the second and third levels in many domain name registries. IDN TLDs allow the entire domain name to be in non-Latin characters, including the TLD extension.

The IDN homographic attack is a means by which a phisher seeks to deceive Internet users by exploiting the fact that characters in different language scripts may be nearly (or wholly) indistinguishable, thereby allowing the phisher to spoof a brand name. These look-alike domains can be displayed in browser address bars if IDN display is enabled.

In our data set we saw 1,222 IDN domain names, used in 1,469 attacks. That was just 0.25% of the domains used for phishing.

- 1,100 domains were on non-IDN TLDs, such as: .xn--blockchin-c2d.com
- 171 domains were in six IDN TLDs, mostly .xn--kprw13d (the Chinese "台湾") and .xn--p1ai (the Russian "рф")

We classified 125 true homographic attacks, for example:

xn--locabitcoin-j4b.com → localʼbitcoin.com

and

xn--santandrbnk-xrb6939g.com → santandęrbąnk.com

Some domains had strings that were misleading, but the domain did not feature a brand name. Yet others had the brand name in plain ASCII characters, and added IDN characters elsewhere in the domain, such as:

xn--laga-iphone-gteborg-26b.com → laga-iphone-göteborg.com

Yet others contained brand names, but were not closely spoofing famous brands, such as:

xn--ycka3kocybf0193c65r3s7adt2f.tokyo → 乳酸ジンジャー・楽天.tokyo

which translates roughly to "Lactic acid ginger / Rakuten".

We also found 48 URLs where the internationalized portion was in a subdomain only, such as:

xn--znsgrab11-c2a4h8cuc3e.aqp.red → zönâsëgürabñ11.aqp.red

In summary, the number of true homographic attacks was very small, just 125 or so out of the 363.5 million registered domain names in the world. Domains that leverage the unique characteristics of IDNs for phishing remain a numerically small problem.

Appendix A: Identification of Phishing Attacks

Phishers commonly point many URLs to one phishing site and use wildcarding [47] and redirection [48] techniques to hide the location of the phishing site from investigators. They may use a single domain name to host several discrete phishing attacks against different companies or may use multiple URLs for any given phishing site to host multiple pages.

To identify unique attacks from this diverse environment of domains, hostnames, and URLs, we examined URLs and metadata associated with URLs. We applied a set of rules to compare URLs for similarities; for example, if the hostname in two or more URLs is the same, and if the report dates for those URLs fall within 7 days of each other, and if the target across those URL reports was the same, then we treated this set of URLs as involved in one phishing attack.

Phishers use a wide variety of URL construction methods, so we formulated additional rules to group URLs into attacks based on observed cases. When we prepare our reports, we perform a final round of manual examination to find additional batches of related URL. For example, some phishers generate multiple subdomains as part of one attack. In some cases, phishers register large numbers of pseudo-randomly generated domain names (see Automating Detection of "Random-looking" Algorithmic Domain Names [49]). In such cases, if the date of the abuse report and the target (brand) were the same, and the reporting feed was the same, then we grouped all those URLs as part of one attack.

Our methodology may result in underreporting the number of attacks. Others who apply a similar methodology may independently arrive at slightly different (higher) numbers; for example, if one were to use the report date window of 30 days from the research paper, COMAR: Classification of Compromised versus Maliciously Registered Domains [50], but in all other respects apply our rules, the results might identify more attacks.

Appendix B: Distinguishing Maliciously Registered Domain Names from Compromised Domains

A maliciously registered domain is defined as a domain registered by a criminal to carry out a malicious act — in this case phishing. Compromised domains are domains registered by innocent parties; an attacker leverages a vulnerability, usually in the web hosting setup, to upload a phishing page on the domain. Because they are dedicated to abuse, maliciously registered domains can be blocklisted in their entirety, and can be suspended by the domain name's registrar or registry operator. Compromised domains generally should not be approached the same way — domain suspension would affect the legitimate services on the domain. When compromised domains appear on blocklists, it is usually a specific URL that is listed, so that URL only can be blocked and prevent collateral damage to legitimate uses of the domain.

To differentiate between compromised and maliciously registered domains, operational security professionals and researchers have relied primarily on two factors:

1. The content of the domain string.
2. The age of the domain name — the number of days elapsed between domain registration and the use of the domain for a malicious purpose. In general, the older the domain name, the higher the likelihood it will legitimate. Miscreants tend to use their domains within a short time after registration in order to avoid detection of their registrations, and almost always within the first year of registration, before they must pay for renewal.

For this study, we refined the algorithm we used on our 2020 study. In the present study, we considered a domain to be maliciously registered if it appeared on a blacklist within fourteen days of being registered, or if a blocklisted domain had a famous brand name or misleading string in it, subject to certain time limits. We also applied additional rules that indicated common control and risk.

Our approach was at its core similar to the COMAR methodology, which was designed by researchers at two security-minded ccTLD operators, SIDN (.NL) and AFNIC (.FR) [51]. COMAR's inputs are *"*public data,*"* in that it is freely available or can be purchased commercially and does not contain personally private data, such as registrant data. Our data shared those characteristics.

In one way our method is more conservative that the COMAR method, which considers a domain to be maliciously registered if it appeared on a blacklist within *three months* of its registration time, or if it has a famous brand name/misleading string in the domain name. COMAR found that among compromised domains used for phishing, only 12% of the domains get compromised within three months of their registration. The implication is that a new domain name is unlikely to be compromised; it usually takes some time for a phisher to discover new domains on vulnerable hosting.

COMAR uses additional criteria to ferret out compromised domains, such as the number of web pages on a suspicious site, the use of SPF records, and a TLD maliciousness score. These additional checks help to find more maliciously registered domains than our fewer criteria; they also refine out border cases. For its phishing data, COMAR used OpenPhish, APWG, and PhishTank — three of the four sources we used.

Neither we nor the COMAR program had access to one of the most useful pieces of data available: domain name contact data, *i.e.,* information about who registered the domain name. Recent changes in ICANN policy allow registrars to redact contact data at will. Falsified contact information is an excellent indicator of bad faith on the part of the registrant, and there are ways for registrars and registry

operators to validate accuracy to various degrees of rigorousness. Also, registrars possess additional detailed data that can help them detect suspicious registrations: the registrant's payment information, the registrant's IP address, and the registrant's purchase history. These are highly useful factors to determine whether a registration is risky, and whether the registrant customer has been honest about its contact information.

Like the COMAR project, we looked for misspellings of brand names. COMAR used dnstwister and Levenshtein distance (with distance = 1) to find misspellings of brand names. They identified 231 brand names mostly targeted by attackers in phishing attacks (e.g., PayPal, Amazon, Yahoo, or Gmail), and looked to see if those strings were contained in the domain name.

We created a list of more than 500 brand names that were targeted in phishing attacks. We used these as the basis for creating a list of a list of misspellings that were distinctive enough to avoid false positives [52]. For example, we decided that "Uber" is not distinctive enough, since it is a common word in German. We complemented this list with misspellings that we encountered in our phishing URLs. We then compared that list to the domains used for phishing. We also looked for variations contained within the domain name, and this identified domains such as feddexx.com, facebaak.gq, and faceb00k-seecuurity-dept.com. Similar to COMAR, we also looked for a short list of misleading words within the domain name designed to fool victims, such as "verification" and "login".

We then performed an examination of remaining domain names. Here we relied on some additional evidence:

- We found *evidence of common control and intent*. The tests above sometime led us to *batches* of domains that were registered, used for phishing, and hosted together, indicating *common control and intent*. Examples were: rebate-tax.uk, return-calculation.uk, and secure-rebate.uk (attacking Her Majesty's Revenue & Customs, the U.K. tax authority), and independent-social-network-000005.my.id, independent-social-network-000006.my.id, and independent-social-network-000007.my.id (used to attack Facebook). This also pointed to long strings of random and meaningless characters, whereas most domains intended for a useful purpose signify some sort of meaning.
- The Spamhaus DBL phishing feed contains a *"*return code*"* indicating whether Spamhaus considers a domain compromised (127.0.1.104, *"*abused legit*"*) or a domain that may be malicious (127.0.1.4).

Our methodology and the more involved COMAR methodology created generally comparable results. One reason is that many malicious registrations are simply *"*beyond the pale*"* — they are designed to fool users and were used for phishing within a week of registration.

Appendix C: Data Sources and Methodologies

Phishing Data Sources

The use of DNS blocklists as a way to track and measure Internet abuse has a long history, and collating data reported by multiple sources is a standard procedure in academic and professional cybercrime studies [53, 54, 55, 56, 57]. To find phishing attacks, blocklist operators use several techniques, including capturing spam email lures, reports from user, and heuristics that examine a variety of data and signals.

The following sources of phishing-specific data were chosen because they are used by a wide variety of organizations to protect users, have low false-positive rates, and have meta-data that is useful for studies such as ours 58, 59, 60.

- **Anti-Phishing Working Group eCrime eXchange (eCX) phishing feed** [61]. The eCX phishing feed is a repository of URLs reported to the APWG by APWG members, who are companies and government and academic investigators. Metadata associated with each uniquely identified URL includes the discovered date, targeted organization (brand) if identified, a confidence level, status (active, inactive), the discovered date, and the date of the last modification of the record.
- **OpenPhish Phishing Intelligence, premium level** [62]**.** The OpenPhish feed is a commercial source that contains phishing URLs discovered by OpenPhish or reported to OpenPhish directly and then verified. Metadata associated with each uniquely identified URL includes the IP address where phish was hosted, targeted brand, discovered timestamp, name of the ASN operator from which the IP address is delegated, hostname of the phish, country where the IP address is geo-located, and Top-level domain (TLD) from which the domain name in the URL was delegated.
- **PhishTank (API)** [63]. PhishTank is operated by OpenDNS, and publishes phishing URLs discovered by and confirmed by PhishTank community contributors. Metadata associated with each uniquely identified URL includes submission time (discovered), verification data (verified, yes/no, and verification time), status (online, yes/no), and details including IP address(es), IP network/prefix, ASN, RIR that delegated the ASN and IP allocations, and country.
- **Spamhaus Domain Block List (DBL)** [64]. The DBL is an rsync feed of registered domain names that have been associated with a malicious or criminal activity. For this study, we used only DBL-listed domains that were associated with two return codes: phish domain (127.0.1.4) and abused legit phish domain (127.0.1.104). We used as the discovery date the timestamp of each rsync access.

We collected data covering the period 1 May 2020 to 30 April 2021. We collected and analyzed only newly found phishing incidents reported during that time. We downloaded updated data from PhishTank and Spamhaus three times a day, and APWG and OpenPhish once a day. The APWG, OpenPhish, and PhishTank feeds allow the downloading of historical listings, and contain timestamps of when the listing was created. Thus we did not miss any listings that appeared between the daily downloads and did not have to worry about a delay of hours between the time the blocklist provider add an entry to its list and when we downloaded those blocklist updates. The Spamhaus DBL is stateful and does not offer *"*time-of-listing*"* time stamps, and it is possible that we missed some short-lived listings there.

These sources provide data about attacks that targeted the general public; they do not quantify *"spear-phishing"* attacks, which are directed at a few specific individuals and are therefore difficult to detect and count reliably.

Confidence Levels

We used only high-confidence reports in our collected data set.

- OpenPhish reports only URLs that are verified to support phishing attacks.
- The PhishTank API feed contains only phishing URLs that have been verified as supporting phish. It does not contain URLs that were reported to PhishTank but had not been verified.
- The APWG feed contains a confidence level provided by the reporting APWG member company. We used only APWG reports at the 90% level (verified by heuristics) and 100% level (verified by a human).
- The Spamhaus phishing feed does not offer confidence ratings. We consider them to be of high confidence because the Spamhaus Domain Blocklist is maintained as a *"*near-zero false positive list,*"* only containing domains that Spamhaus recommends be blocked in their entirety because they are considered dangerous. See the previous section for more about Spamhaus return codes.

Data Normalization and DNS Data

We collected reports from each feed at least once per day to find new entries. This *collected data set* then required curation to allow data from different sources to be stored together and compared. Each time a URL (or plain domain) was reported, we stored that as a separate report. Some URLs were reported by more than one source.

It was necessary to normalize certain metadata such as target (brand). For example, different sources reported slight variations of target names ("Microsoft" vs. "Microsoft Corp" vs. "Microsoft Corporation"). We normalized such examples to a common form of the company name.

UTC time is the time convention used by the four data sources, and in all gTLD registry and registrar systems including WHOIS. We used UTC.

Some sources provided IP (A record) data and AS data. For every domain reported, we also queried DNS and separately stored the A record we found and determined the AS by using Team Cymru's IP to ASN mapping service [65.] We relied upon RIPE-NCC's WHOIS [66] to find ASN name, organization, and IP prefix. When we list the number of IPv4 addresses in an AS, that is a count of routed addresses.

To identify TLDs we used the IANA root zone list [67]. We used the Public Suffix List [68] to identify registered domain names (zones in which registries offer third level registration, such as example.co.uk).

The "legacy generic TLDs" introduced before 2013 (other than .COM and .NET) are: .AERO, .ASIA, .BIZ, .CAT, .COOP, .INFO, .JOBS, .MOBI, .MUSEUM, .NAME, .ORG, .POST, .PRO, .TEL, .TRAVEL, and .XXX.

For gTLD domain names we obtained registry WHOIS to identify the sponsoring registrar, along with the registrar's IANA ID [69] for normalization. Some gTLD registries severely rate-limited [70] our queries and made it impossible to obtain basic data about their domain names, including the domain registration date and the identity of the domain's sponsoring registrar. For this reason, some gTLD domain names were not attributable to registrars and do not appear in the phishing-by-registrar tables and could not be included in the analysis of registration-to-phishing times. We did not obtain WHOIS for ccTLD domains due to limited access and non-uniformity of WHOIS output. Also, ccTLD registrars are not

identified via a uniform identifier across ccTLD registries, making the compilation of by-registrar statistics difficult.

In the tables, the number of domains in each gTLD, and the number of gTLD domains sponsored by each registrar, are from the monthly ICANN reports for May 2020, the latest month available when we began writing the report [71]. Reference to DUM are also made to NTLDSTATS.com and ICANN July 2020 reports. ICANN ccTLD domain counts are from the web sites of the registry operators and from DomainTools [72].

Target Identification

The APWG, OpenPhish, and PhishTank feeds identify target brand for each report; the Spamhaus DBL does not provide target information but classifies the domains according to the type of threat the domain is used to perpetrate. The sources determine target by either heuristics (which parses the content of the email phishing lure, and /or identifies the logos and wording on the phishing site), or by manual verification.

Each feed varies slightly in its granularity and nomenclature. We normalized variations in spelling — for example one feed used *"*PayPal*"* while another called it *"*PayPal Inc.*"* and so we consolidated those. Some feeds present classification differences. For example, WhatsApp is owned by Facebook. Some sources report WhatsApp as a separate brand, but another source reported the same WhatsApp phishing URLs as attacks against Facebook. In that case we accepted both and counted those as two brands.

In some cases, a source would positively identify a URL as a phish against a specific target. Another source would then report the same URL as an attack against an unknown or *"*generic*"* brand. In such cases we attributed that attack to the specific brand. In the cases where an attack's target was still unknown, we set those attacks aside when analyzing brand data.

AS Rankings

We took into consideration previous work done to develop security reputation metrics for hosting providers [73, 74, 75, 76]. That work notes that rankings are one way of unifying the scales on which normalized abuse is measured and allows cross comparisons, and that normalized abuse is an indicator of security performance by itself. Per the work of Noroozian *et al* [73], our work has some useful features, namely that our approach considered second-level domain-IP pairs as a unit of abuse, and that normalized abuse is abuse-type specific (because we considered phishing only).

In an AS, there may be multiple organizations which use a part of the IP space, and in the future, we wish to refine approaches to that issue. In the end we believe that our initial effort points to interesting concentrations of abuse in IP spaces under common control and are useful indicators for additional study.

About the Authors

Greg Aaron is an internationally recognized authority on the use of domain names for cybercrime, and is an expert on domain name registry operations, DNS policy, and related intellectual property issues. Mr. Aaron is Senior Research Fellow for the Anti-Phishing Working Group. As a member of ICANN's Security and Stability Advisory Committee (SSAC), he advises the international community regarding the domain name and numbering system that makes the Internet function. He works with industry, researchers, and law enforcement to investigate and mitigate cybercrime, and is also a licensed private detective. He was the Chair of ICANN's Registration Abuse Policy Working Group (RAPWG) and has been a member of ICANN's EPDP Working Group, which has been creating registration data access policies. He was the senior industry expert on a team that evaluated the policy and technical qualifications of more than one thousand new TLD applications to ICANN in 2012-2013. He has created products and services used by organizations to discover and track Internet-based threats, and has managed large top-level domains around the world, including .INFO, .ME, and .IN. He is President of Illumintel, Inc., a consulting company. Mr. Aaron is a *magna cum laude* graduate of the University of Pennsylvania.

Lyman Chapin has contributed to the development of technologies, standards, and policy for the Internet since 1977, and is widely recognized and respected as a leader in the networking industry and the Internet community. Mr. Chapin is a Life Fellow of the IEEE, and has chaired the Internet Architecture Board (IAB), the ACM Special Interest Group on Data Communication (SIGCOMM), and the ANSI and ISO standards groups responsible for Network and Transport layer standards. Mr. Chapin was a founding trustee of the Internet Society and a Director of the Internet Corporation for Assigned Names and Numbers (ICANN). He currently chairs ICANN's Registry Services Technical Evaluation Panel (RSTEP), which is responsible for assessing the impact of new Domain Name System (DNS) registry services on the security and stability of the Internet, and the DNS Stability Panel, which evaluates proposals for new Internationalized Domain Names (IDNs) as country code top-level domains (ccTLDs). He is also a member of ICANN's Security and Stability Advisory Committee (SSAC). He has written many other papers and articles over the past 40 years, including the original specification of the Internet standards process operated by the IETF. Mr. Chapin holds a B.A. in Mathematics from Cornell University.

David Piscitello has been involved in Internet technology and security for more than 40 years. Until July 2018, Mr. Piscitello was Vice President for Security and ICT Coordination at ICANN, where he participated in global collaborative efforts by security, operations, and law enforcement communities to mitigate Domain Name System abuse. He also coordinated ICANN's security capacity-building programs and was an invited participant in the Organisation for Economic Co-operation and Development (OECD) Security Expert Group. Dave is an Associate Fellow of the Geneva Centre for Security Policy. He served on the Boards of Directors at the Anti-Phishing Working Group (APWG) and Consumers Against Unsolicited Commercial Email (CAUCE). He is the recipient of M3AAWG's 2019 Mary Litynski Award, which recognizes the lifetime achievements of individuals who have significantly contributed to making the Internet safer.

Dr. Colin Strutt has published and spoken extensively on networking technology, name collisions, enterprise management, eBusiness, and scenario planning, and has represented the interests of Digital Equipment, Compaq, and the Financial Services Technology Consortium in national and international industry standards bodies. He holds six patents on enterprise management technology and brings more than forty years of direct experience with information technology, as a developer, architect, and consultant, with recent work including design and operation of a regional public safety network, providing technical expertise relating to patents, and analysis of world-wide Internet use. Dr. Strutt holds a B.A. (with First Class Honours) and Ph.D. in Computer Science from Essex University (UK).

About Interisle Consulting Group, LLC

Interisle's principal consultants are experienced practitioners with extensive track records in industry and academia and world-class expertise in business and technology strategy, Internet technologies and governance, financial industry applications, and software design. For more about Interisle, please visit: www.interisle.net

Acknowledgments

The authors extend thanks to:

- Spamhaus and OpenPhish, for their kind contribution of data for this study.
- The PhishTank and the APWG eCrime Exchange communities, for their collaborative efforts to identify phishing.
- Domain Tools, for access to historical and bulk parsed WHOIS.
- Saeed Abu-Nimeh for access to the Seclytics Predictive Threat Intelligence platform.
- Rod Rasmussen, who co-created the Global Phishing Survey with Greg Aaron, published via the Anti-Phishing Working Group from 2007 to 2017.
- All the security personnel who fight phishing.

End Notes

[1] Cybercrime Information Center.
https://cybercrimeinfocenter.org/phishing-activity

[2] See the Anti-Phishing Working Group's Phishing Activity Trends Reports (2004-2021, at https://apwg.org/trendsreports/), and the APWG's Global Phishing Survey series of papers (2007-2017, at https://apwg.org/globalphishingsurvey/).

[3] P. Foremski and P. Vixie. "Modality of Mortality in Domain Names: An In-depth Study of Domain Lifetimes." 2018.
https://www.farsightsecurity.com/assets/media/download/VB2018-study.pdf

[4] Palo Alto Networks' cybersecurity research team studies large numbers of newly registered domain names found in zone files and concluded that 70% of them are "malicious" or "suspicious" or "not safe for work." Palo Alto Networks, Unit 42. "Newly Registered Domains: Malicious Abuse by Bad Actors." 20 August 2019.
https://unit42.paloaltonetworks.com/newly-registered-domains-malicious-abuse-by-bad-actors/

[5] Verisign Domain Name Industry Brief, Q12021. Volume 13, issue 3, August 2020.
https://www.verisign.com/en_US/domain-names/dnib/index.xhtml

[6] G. Aaron, L. Chapin, D. Piscitello, and C. Strutt, Interisle Consulting. *Phishing Landscape 2020: A Study of the Scope and Distribution of Phishing.* 13 October 2020.
http://www.interisle.net/PhishingLandscape2020.html

[7] Cybercrime Information Center, https://cybercrimeinfocenter.org,
http://www.interisle.net/ContactStudy2021.html

[8] S. Maroofi, M. Korczynski, C. Hesselman, B. Ampeau, A. Dud, "COMAR: Classification of Compromised versus Maliciously Registered Domains." 2020 IEEE European Symposium on Security and Privacy (EuroS&P).
http://mkorczynski.com/COMAR_2020_IEEEEuroSP.pdf and
https://comar-project.univ-grenoble-alpes.fr/

[9] How Far Will Email Operators Take Blocklisting to Prevent Spam?
https://www.securityskeptic.com/2017/11/how-far-will-email-operators-take-blocklisting-to-prevent-spam-.html

[10] ICANN. 2013 Registrar Accreditation Agreement. Section 3.18.
https://www.icann.org/resources/pages/approved-with-specs-2013-09-17-en

[11] For examples of pricing-related issues, see the APWG Global Phishing Survey series of papers, 2007-2017, at https://apwg.org/globalphishingsurvey/. Further study of the effect of pricing on domain name abuse would be welcome. Such studies are hard to carry out because the sales specials and rebate programs offered by registry operators, and the bulk sale prices offered by some registrars, make it difficult to establish point-in-time wholesale and retail prices for specific domain names.

[12] D. Piscitello and C. Strutt. "Criminal Abuse of Domain Names: Bulk Registration and Contact Information Access." 17 October 2019.
http://interisle.net/sub/CriminalDomainAbuse.pdf

[13] Note that a two-dimensional Venn diagram cannot show the sets that were shared by just A and C, and by B and D. For this reason, if one adds up the seven numbers contained in a circle, the result will be less than the total number of domains found by that source.

[14] A. Pitsillidis, C. Kanich, G.M. Voelker, K. Levchenko, S. Savage. "Taster's Choice: A Comparative Analysis of Spam Feeds." Proceedings of the 2012 Internet Measurement Conference, 427-440.
https://cseweb.ucsd.edu/~apitsill/papers/imc12.pdf
This paper compares the contents of ten distinct feeds of spam-advertised domain names (which includes phishing URLs).

[15] L. Metcalf and J. Spring, "Everything You Wanted to Know About Blacklists But Were Afraid to Ask." Publication CERTCC-2013-39.
https://resources.sei.cmu.edu/asset_files/WhitePaper/2013_019_001_83445.pdf.
This study compares the contents of 25 different common public-internet blacklists in order to discover any patterns in the shared entries.

[16] L. Metcalf, J. Spring, "Blacklist Ecosystem Analysis: Spanning Jan 2012 to Jun 2014." CERT Division, Software Engineering Institute, Carnegie-Mellon University. Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security, 12 October 2015, 13-22.
https://discovery.ucl.ac.uk/id/eprint/10037798/
This study compared the contents of 86 Internet blacklists, include how many lists an indicator is unique to, list sizes, expanded list characterization and intersection, etc.

[17] L. Metcalf, E. Hatleback, J. Spring. "Blacklist Ecosystem Analysis: 2016 Update." Software Engineering Institute, CERT Coordination Center, Pittsburgh, PA. March 2016.
https://resources.sei.cmu.edu/asset_files/WhitePaper/2016_019_001_466029.pdf
This follow-study confirms that blacklists generally fail to overlap substantially with each other, suggesting that available blacklists present an incomplete and fragmented picture of the malicious infrastructure on the Internet.

[18] V. Guo Li, M. Dunn, P. Pearce, D. McCoy, G. Voelker, S. Savage, K. Levchenko. "Reading the Tea Leaves: A Comparative Analysis of Threat Intelligence." Proceedings of the 28th USENIX Security Symposium. August 14–16, 2019, Santa Clara, CA, USA.
https://www.usenix.org/system/files/sec19-li-vector_guo.pdf
This recent study systematically characterizes a broad range of public and commercial sources of threat intelligence. Although it concentrates on IP address and file hashes, the findings seem generally applicable to domain name-based threat intelligence feeds.

[19] H.Griffioen, T. Booij and C Doerr. "Quality Evaluation of Cyber Threat Intelligence Feeds." International Conference on Applied Cryptography and Network Security (ACNS), May 2020.
https://www.researchgate.net/publication/341385656_Quality_Evaluation_of_Cyber_Threat_Intelligence_Feeds

[20] FireHOL.
http://iplists.firehol.org/#comparison
This is a comparison of IP blocklists, noting overlaps.

[21] C. Kanich, N. Chachra, D. McCoy, C. Grier, D.Y. Wang et al. "No Plan Survives Contact: Experience with Cybercrime Measurement." CSET, 2011.
http://damonmccoy.com/papers/cset11kanich.pdf

[22] A. Oest, Y. Safaei, A. Doupé, G. Ahn, B. Wardman, and G. Warner. "Inside a Phisher's Mind: Understanding the Anti-Phishing Ecosystem Through Phishing Kit Analysis". In Proceedings of the 2018 APWG Symposium on Electronic Crime Research (eCrime), pages 1–12, May 2018.
https://docs.apwg.org/ecrimeresearch/2018/5349207.pdf

[23] Palo Alto Networks, Unit 42. "Newly Registered Domains: Malicious Abuse by Bad Actors." 20 August 2019.
https://unit42.paloaltonetworks.com/newly-registered-domains-malicious-abuse-by-bad-actors/

[24] A. Oest, Y. Safaei, A. Doupé, G. Ahn, B. Wardman, and K. Tyers. "PhishFarm: A Scalable Framework for Measuring the Effectiveness of Evasion Techniques Against Browser Phishing Blacklists." In: 2019 IEEE Symposium on Security and Privacy (SP), 19-23 May 2019.
https://ieeexplore.ieee.org/document/8835369

[25] A. Oest, Y. Safaei, P. Zhang, B. Wardman, et al: "PhishTime: Continuous Longitudinal Measurement of the Effectiveness of Anti-phishing Blacklists." Proceedings of the 29th USENIX Security Symposium, August 12–14, 2020.
https://www.usenix.org/system/files/sec20-oest-phishtime.pdf

[26] T. Moore and R. Clayton. "How Hard Can it Be to Measure Phishing?" Mapping and Measuring Cybercrime, 2010.
https://www.cl.cam.ac.uk/~rnc1/cyberbias.pdf

[27] Europol European Cybercrime Centre (EC3). "The Indispensable Role of WHOIS for Global Cybersecurity: Statement by the EC3 Advisory Group on Internet Security." 25 January 2018.
https://www.icann.org/en/system/files/files/gdpr-statement-ec3-europol-icann-proposed-compliance-models-25jan18-en.pdf

[28] Messaging, Malware and Mobile Anti-Abuse Working Group (M3AAWG) and Anti-Phishing Working Group. "ICANN GDPR WHOIS Policy Eliminates Pre-emptive Protection of Internet Infrastructure Abuse; Obstructs Routine Forensics to Cybercriminals' Advantage." 24 October 2018.
https://www.m3aawg.org/rel-WHOISSurvey2018-10

[29] D. Piscitello. "Facts & Figures: WHOIS Policy Changes Impair Blocklisting Defenses." 8 March 2019.
https://www.securityskeptic.com/2019/03/facts-figures-whois-policy-changes-impair-blacklisting-defenses.html

[30] G. Aaron and R. Rasmussen, "Global Phishing Survey: Trends and Domain Name Use in 2016." Anti-Phishing Working Group. Pages 17, 19.
https://docs.apwg.org/reports/APWG Global Phishing Report 2015-2016.pdf

[31] Cryptocurrency market value tops $2 trillion for the first time as Ethereum hits record high,
https://www.cnbc.com/2021/04/06/cryptocurrency-market-cap-tops-2-trillion-for-the-first-time.html

[32] Twitter post by @MyEthernWallet warns of phishing attack,
https://twitter.com/myetherwallet/status/995043917878902784

[33] Analysing a Cryptocurrency phishing attack that earns $15K in two hours
https://www.kpn.com/zakelijk/blog/analysing-cryptocurrency-phishing-attack.htm

[34] Ledger owners lose 1.1 million XRP to scam site
https://cointelegraph.com/news/ledger-owners-lose-1-1-million-xrp-to-scam-site

[35] [PSA] Phishing Alert: Fake Trezor Wallet Website
https://blog.trezor.io/psa-phishing-alert-fake-trezor-wallet-website-3bcfdfc3eced

[36] Cybercriminals Exploiting API Keys to Steal Cryptocurrency
https://cyware.com/news/cybercriminals-exploiting-api-keys-to-steal-cryptocurrency-6e52b50b

[37] Top 5 ways how criminals steal crypto in 2020
https://hacken.io/research/top-5-ways-how-criminals-steal-crypto-in-2020/

[38] Users of Crypto Wallets Electrum and MyEtherWallet Face Phishing Attacks
https://cointelegraph.com/news/users-of-crypto-wallets-electrum-and-myetherwallet-face-phishing-attacks

[39] The top 3 cryptocurrency scams of 2021
https://www.techrepublic.com/article/the-top-3-cryptocurrency-scams-of-2021/

[40] C. Lu, B. Liu et al. "Measurement Study of Domain Registration Privacy under the GDPR." Network and Distributed Systems Security (NDSS) Symposium 2021, 21-25 February 2021, virtual.
https://www.ndss-symposium.org/wp-content/uploads/ndss2021_2A-2_23134_paper.pdf

[41] WHOIS Contact Data Availability and Registrant Classification Study
http://www.interisle.net/ContactStudy2021.html

[42] L. Weissinger, D. Piscitello, and B. Wilson. "ICANN, GDPR, and the WHOIS: A Users Survey – Three Years Later." June 2021.
https://www.m3aawg.org/sites/default/files/m3aawg_apwg_whois_user_survey_report_2021.pdf

[43] Appdetex WHOIS Requestor System (AWRS) | Appdetex
https://www.appdetex.com/appdetex-whois-requestor-system-awrs-3/

[44] ICANN Generic Names Supporting Organization (GNSO). *Final Report of the Temporary Specification for gTLD Registration Data Phase 2 Expedited Policy Development Process.* 31 July 2021.
https://gnso.icann.org/sites/default/files/file/field-file-attach/epdp-phase-2-temp-spec-gtld-registration-data-31jul20-en.pdf

[45] ICANN Security and Stability Advisory Committee. *SAC118: SSAC Comments on Initial Report of the Expedited Policy Development Process (EPDP) on the Temporary Specification for gTLD Registration Data Team – PHASE 2A*. 15 July 2021.
https://www.icann.org/en/system/files/files/sac-118-en.pdf

[46] ICANN Security and Stability Advisory Committee. *SAC112: Minority Statement on the Final Report of the Temporary Specification for gTLD Registration Data Phase 2 Expedited Policy Development Process (EPDP).* 20 August 2020.
https://www.icann.org/en/system/files/files/sac-112-en.pdf

[47] Definition of Wildcarding
https://tools.ietf.org/html/rfc4592

[48] DNS hijacking and redirection
https://www.imperva.com/learn/application-security/dns-hijacking-redirection/

[49] Automating Detection of "Random-looking" Algorithmic Domain Names
https://www.farsightsecurity.com/blog/txt-record/automatingdetection-20190517/

[50] COMAR: Classification of Compromised versus Maliciously Registered Domains
http://mkorczynski.com/COMAR_2020_IEEEEuroSP.pdf

51 Maroofi, M. Korczynski, C. Hesselman, B. Ampeau, A. Dud, "COMAR: Classification of Compromised versus Maliciously Registered Domains." 2020 IEEE European Symposium on Security and Privacy

(EuroS&P)
http://mkorczynski.com/COMAR_2020_IEEEEuroSP.pdf and
https://comar-project.univ-grenoble-alpes.fr/

[52] W. Wang and K. Shirley, "Breaking Bad: Detecting Malicious Domains Using Word Segmentation," In Proc. 9th Workshop on Web 2.0 Security and Privacy, 2015.

[53] A. Oest, Y. Safaei, A. Doupé, G. Ahn, B. Wardman, and K. Tyers. "PhishFarm: A Scalable Framework for Measuring the Effectiveness of Evasion Techniques Against Browser Phishing Blacklists." In: 2019 IEEE Symposium on Security and Privacy (SP), 19-23 May 2019.
https://ieeexplore.ieee.org/document/8835369

[54] D. Piscitello, G. Aaron. "Domain Abuse Activity Reporting (DAAR) System Methodology". Internet Corporation for Assigned Names and Numbers (ICANN). November 2017.
https://www.icann.org/en/system/files/files/daar-methodology-paper-30nov17-en.pdf

[55] Dietrich C.J., Rossow C. (2009) Empirical research of IP blacklists. In: Pohlmann N., Reimer H., Schneider W. (eds) ISSE 2008 Securing Electronic Business Processes. Vieweg+Teubner.
https://doi.org/10.1007/978-3-8348-9283-6_17

[56] S. Maroofi, M. Korczynski, C. Hesselman, B. Ampeau, A. Dud, "COMAR: Classification of Compromised versus Maliciously Registered Domains." 2020 IEEE European Symposium on Security and Privacy (EuroS&P). http://mkorczynski.com/COMAR_2020_IEEEEuroSP.pdf and
https://comar-project.univ-grenoble-alpes.fr/

[57] Pitsillidis, C. Kanich, G.M. Voelker, K. Levchenko, S. Savage. "Taster's Choice: A Comparative Analysis of Spam Feeds." Proceedings of the 2012 Internet Measurement Conference, 427-440.
https://cseweb.ucsd.edu/~apitsill/papers/imc12.pdf

[58] D. Piscitello. "Reputation Block Lists: Protecting Users Everywhere." 1 November 2017. Internet Corporation for Names and Numbers (ICANN).
https://www.icann.org/news/blog/reputation-block-lists-protecting-users-everywhere

[59] B. Greene. "What Makes a Good 'DNS Blacklist"?"
https://blogs.akamai.com/2017/08/what-makes-a-good-dns-blacklist.html and
https://www.akamai.com/us/en/products/security/enterprise-threat-protector.jsp

[60] G. Aaron, D. Piscitello. "Domain Abuse Activity Reporting (DAAR) System Methodology". Internet Corporation for Assigned Names and Numbers (ICANN). November 2017.
https://www.icann.org/en/system/files/files/daar-methodology-paper-30nov17-en.pdf

[61] Anti-Phishing Working Group eCrime Exchange.
https://apwg.org/ecx/

[62] OpenPhish.
https://openphish.com

[63] PhishTank.
https://www.phishtank.com/

[64] The Spamhaus Project.
https://www.spamhaus.org/

[65] Team Cymru. IP to ASN Mapping Service.
https://team-cymru.com/community-services/ip-asn-mapping/

[66] RIPE-NCC.
https://stat.ripe.net/ and
https://www.ripe.net/manage-ips-and-asns/db/tools

[67] IANA root zone list.
https://www.iana.org/domains/root/db

[68] Public Suffix List.
https://publicsuffix.org/

[69] IANA Registrar IDs.
https://www.iana.org/assignments/registrar-ids/registrar-ids.xhtml

[70] ICANN Security and Stability Advisory Committee (SSAC): *SAC101v2: SSAC Advisory Regarding Access to Domain Name Registration Data*. 12 December 2018.
https://www.icann.org/en/system/files/files/sac-101-v2-en.pdf

[71] ICANN Monthly Registry Reports
https://www.icann.org/resources/pages/registry-reports

[72] DomainTools, Domain Count Statistics for TLDs.
https://research.domaintools.com/statistics/tld-counts/

[73] A. Noroozian, M. Korczynski, S. Tajalizadehkhoob, and M. van Eeten, "Developing Security Reputation Metrics for Hosting Providers," in Proc. Workshop on Cyber Security Experimentation and Test (CSET), 2015.
http://mkorczynski.com/UsenixCSETNoroozian.pdf

[74] M. Korczynski, S. Tajalizadehkhoob, A. Noroozian, M. Wullink, C. Hesselman, and M. v. Eeten, "Reputation Metrics Design to Improve Intermediary Incentives for Security of TLDs," in IEEE EuroS&P, 2017, pp. 579–594.
http://mkorczynski.com/SPEurope2017Korczynski.pdf

[75] B. Stone-Gross, C Kruegel, K. Almeroth, A. Moser, and E. Kirda. "FIRE: FInding Rogue nEtworks". In: ACSAC. 2009, pp. 231–240.
https://sites.cs.ucsb.edu/~chris/research/doc/acsac09_fire.pdf

[76] J. Armin, editor. "World Hosts Report 2014".
http://hostexploit.com/?p=whr-201403