Phishing Landscape 2020

A Study of the Scope and Distribution of Phishing

by

Greg Aaron

Lyman Chapin

David Piscitello

Dr. Colin Strutt

Interisle Consulting Group, LLC

13 October 2020

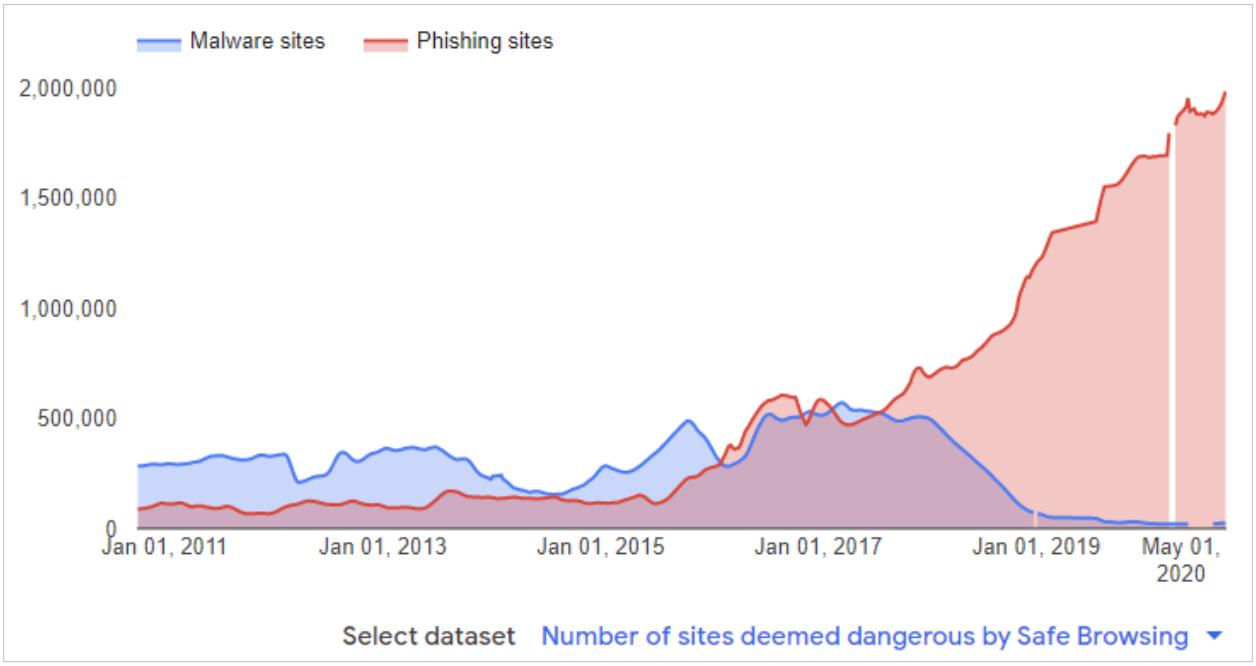
Interisle Consulting Group

Table of Contents

Executive Summary

Phishing is a significant threat to millions of Internet users. Phishing attacks lure victims to a website purportedly run by a trusted entity, such as a bank or other service the victim uses, and the victim is fooled into entering sensitive information. These bogus websites are actually run by criminals, and they steal extensive financial and personal information from the victims, leading to large aggregate financial losses and identity theft. At the same time, phishing inflicts financial costs and reputational damage to the targets, which are companies, government entities such as tax authorities, and universities. Phishing also inflicts damage on the systems of compromised web hosts, on the email providers who must defend against phishing spam, and on responders charged with protecting users and networks.

The amount of phishing being found continues to increase. Google's Safe Browsing program offers an excellent measurement of verified phishing activity over an extended period. Google Safe Browsing has logged a significant increase in phishing sites over the past four years:[1]



Phishing sites detected by Google Safe Browsing, Sept. 2010 to Sept. 2020
Source: https://transparencyreport.google.com/safe-browsing/overview

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 looked at when phishers launch attacks, to determine when attacks occur and how quickly phishers act. We studied where phishers are getting the resources they need to perpetrate their crimes — 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 URLs, domain names, IP addresses, and other data about phishing attacks from four widely used and respected threat data providers: the Anti-Phishing Working Group (APWG), OpenPhish, PhishTank, and Spamhaus. Over a three-month collection period, we learned about more than 100,000 newly discovered phishing sites.

Major Findings

Based on the data, our major findings and conclusions are:

1. **Most phishing is concentrated at small numbers of domain registrars, domain registries, and hosting providers.** These concentrations may be due to business decisions that these providers make. **These providers can make a significant impact on phishing if they implement better anti-abuse programs.** *(See pages 11-25.)*
2. **Phishers themselves register more than half of the domain names on which phishing occurs**. We call these "malicious registrations." *(See pages 17-21.)*
3. **Domain name registrars and registry operators can prevent and mitigate large amounts of phishing, by finding and suspending maliciously registered domains.** It is possible to identify malicious registrations with a high degree of reliability. Registries and registrars possess dispositive data about their customers that no one else has, and that data provides additional opportunities to identify risky registrations. *(See pages 17-21 and 33-34.)*
4. **Registries, registrars, and hosting providers should focus on both mitigation and *prevention*. Some anti-abuse programs are purely reactive, and address phishing only after it begins. Such programs can create incremental reductions of damage and losses, but may do nothing to prevent ongoing cycles of phishing and sustained abuse over time.** *(See pages 10 and 20-21.)*
5. **The problem of phishing is bigger than is reported, and the exact size of the problem is unknown.** This is due to gaps in detection and in data sharing. The over-redaction of contact data in WHOIS is contributing to the under-detection problem. *(See pages 27-28.)*
6. **Sixty-five percent of maliciously registered domain names are used for phishing within five days of registration.** *(See pages 8-10.)*
7. **New top-level domains introduced since 2014 account for 9% of all registered domain names, but 18% of the domain names used for phishing. Of the domains used for phishing in the new gTLDs, 81% were maliciously registered by phishers. Most of those were concentrated in a small number of new gTLDs.** *(See pages 12-13, 19.)*
8. **About 9% of phishing occurs at a small set of providers that offer subdomain services.** *(See pages 30-31.)*

The methodologies and data from this study will be used to conduct additional longitudinal surveys of phishing over time, and to investigate other forms of cybercrime and DNS abuse.

Key Statistics

We collected data over a three-month study period that ran from 1 May 2020 through 31 July 2020. In the data we found:

- **298,012 phishing reports**. This is the number of URLs and domains that were *added* to the four feeds during the study period — in other words, reports of newly found (reported) phishing incidents. Some URLs were duplicates, reported separately by one or more of the sources.
- **122,092 phishing attacks**. The reports told about a smaller number of attacks. An *attack* is defined as a phishing site (a web location) that targets a specific brand or entity. Some phishers point many different URLs to one phishing site, using redirection techniques. A phishing site usually contains multiple pages, more than one of which is reported. A single domain name can host several discrete phishing attacks (sites), each targeting a different company.
- **99,412 unique domain names**. We found how many unique domain names the reports contained. These are second-level domain names, and third-level domain names where the relevant registry offers third-level registrations (such as domain.co.uk).
- The domain names used for phishing were in **439 top-level domains**.
- **414 registrars** sponsored gTLD domains that were used for phishing.
- In addition, there were 619 attacks on URLs that contained IPv4 addresses and no domain name. (For example: *http://95.142.44.203/sparkasse.html*).
- **60,935 maliciously registered domain names.** Of the 99,412 domains used for phishing, we identified 60,935 that we believe were registered maliciously, by phishers. The rest were "compromised domains," owned by innocent parties on vulnerable hosting. *(See pages 16-20)*
- **684 targeted brands**. The phishing sites emulated 684 different entities. These including banks, social media companies, webmail providers, games, national tax services to which citizens pay taxes, universities, and cryptocurrency exchanges. *(See page 29.)*
- Phishing occurred in the IP spaces **of 2,169 different Autonomous Systems** (AS). *(See page 22).*
- In our data set we saw phishing on **219 IDN domain names**, used in 232 attacks. That was just 0.2% of the domains used for phishing. About 50 of those domains could be classified as homographic attacks. *(See page 32.)*

Explanations of the collection and counting methodologies are detailed in the appendices.

Introduction

The goal of 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 looked at when phishers launch attacks, to determine when attacks occur and how quickly phishers act. We studied where phishers are getting the resources they need to perpetrate their crimes — 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.

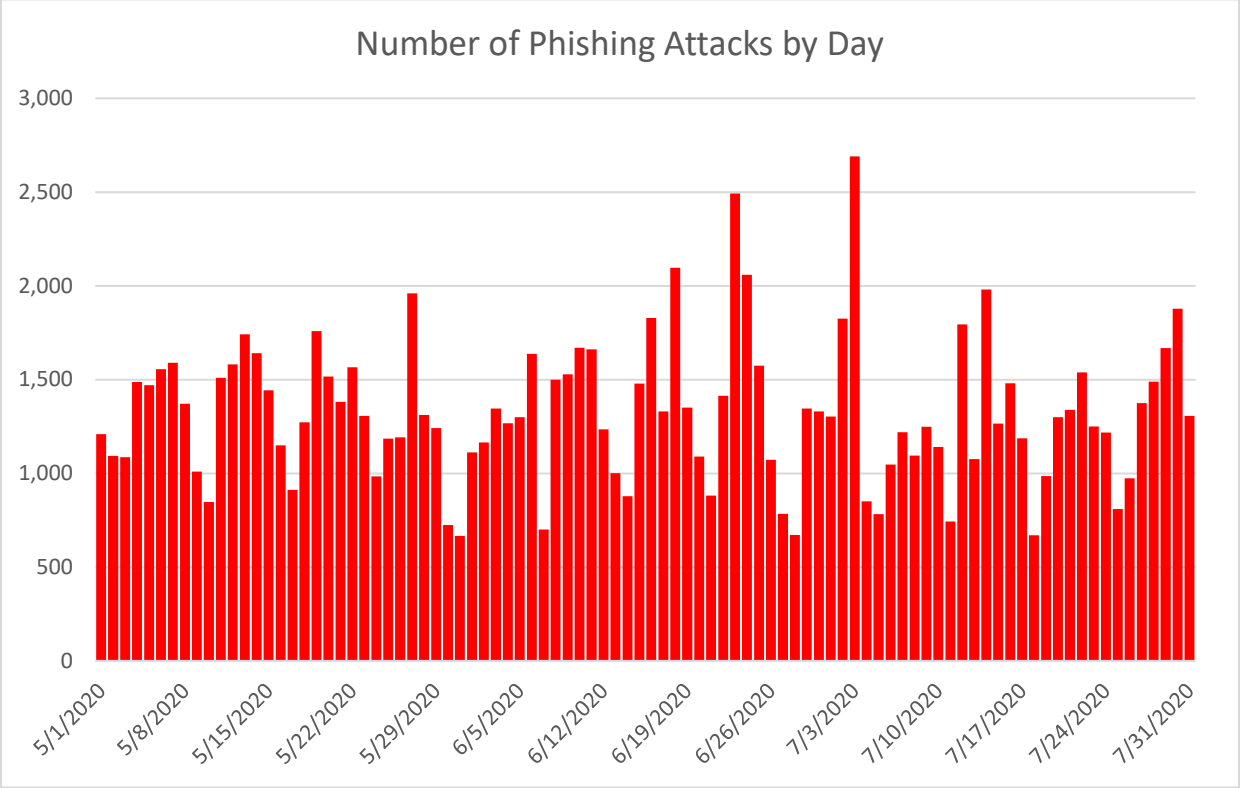
The first section of this study report describes the characteristics of phishing attacks. We looked at attack activity by day to determine whether or not phishers find particular days of the week more opportunistic than others. We also looked at the time elapsed between the registration of a domain for a phishing attack to the time when that maliciously registered domain was used for phishing, to further understand how phishers prepare for attacks.

The second section describes where phishing attacks are concentrated among Top-Level Domain registries, registrars, and hosting providers. We distinguished phishing attacks in which malicious registrations were used from attacks hosted on compromised domains or web sites, to better understand where (*e.g.*, registry, registrar, hosting provider) additional phishing detection or mitigation efforts could be applied most effectively. We also report on the wide range of brands targeted by phishers, their use of subdomain servers, and the use of International Domain Names in phishing attacks.

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.

Attack Activity by Day

Our data confirms a pattern that has held steady for many years. Phishing is lowest on the weekends when potential victims are away from their email. Phishing then ramps up early in the week as phishers send email lures, when the attention of potential victims is highest:



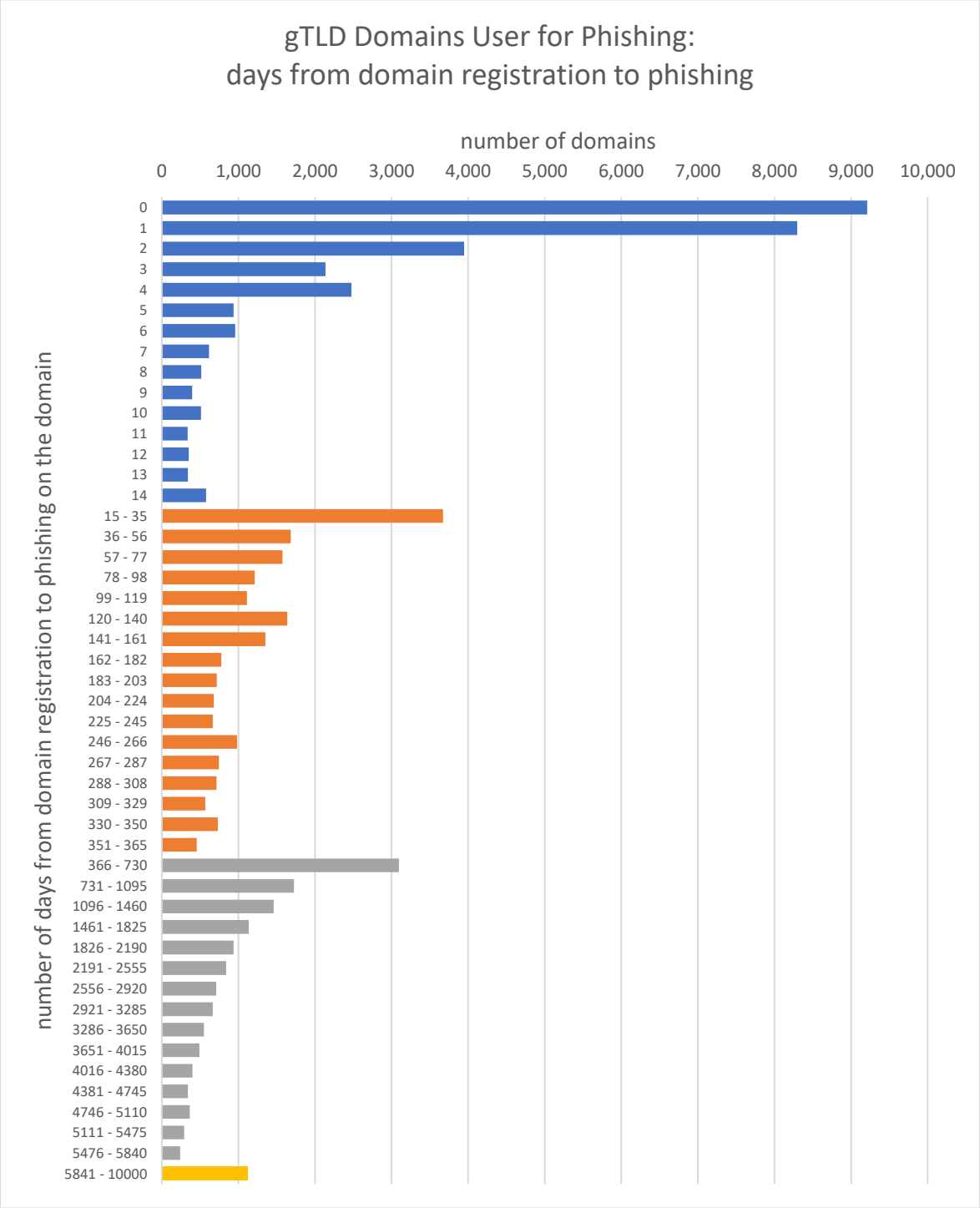
In the graph, blocklistings usually peak around Wednesdays. We note that there is a delay between when an attack begins and when it is blocklisted, meaning that attacks actually peak a bit earlier.

A team of researchers at PayPal, Google, and Arizona State University recently published a study called "Sunrise to Sunset: Analyzing the End-to-end Life Cycle and Effectiveness of Phishing Attacks at Scale."[2] This group leveraged some unique capabilities of Google and PayPal to measure visits to phishing sites. The study confirmed previous findings that the average phishing attack is short, and that the majority of the victimization occurs before the phishing is listed on blocklists. That study notes:

1. **The practical lifetime of a phishing attack is only 21 hours**. That is the average time from the first visit by a victim to the last visit by a victim.
2. **The detection of each attack by anti-phishing entities occurs, on average, 8 hours and 44 minutes *after* the victims start visiting**.[3] The majority of victimization (63%) occurs in this period *before the phishing attack is discovered and blocklisted.*
3. The "accounts of 7.42% of distinct Known Visitors subsequently suffered a fraudulent transaction; we believe this represents a lower bound on success rates and subsequent damage from phishing." Those victims tended to suffer a fraudulent transaction within 5 days of the phishing, on average.

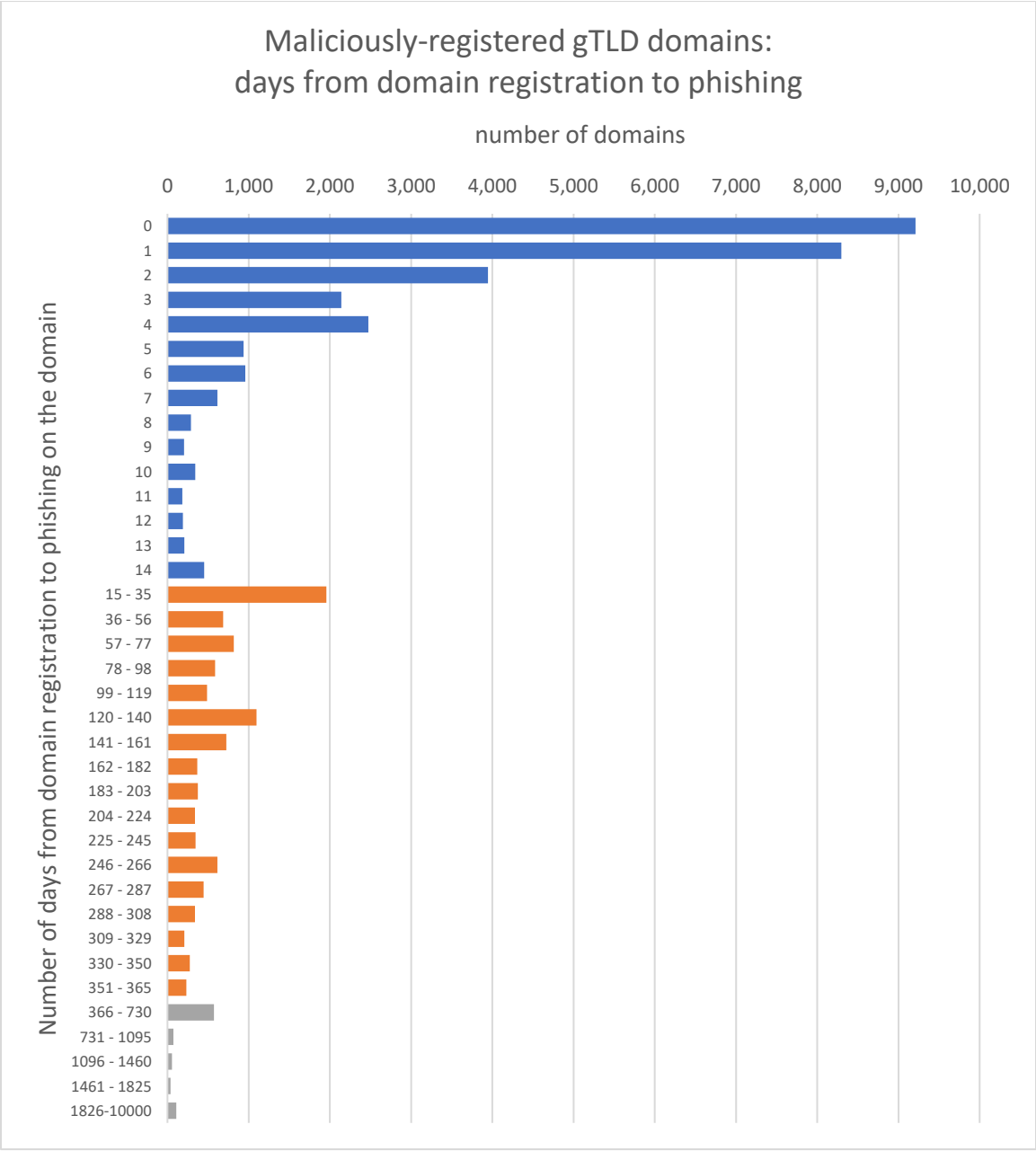
Time Elapsed between Domain Registration and Phishing

We analyzed how much time elapsed between when a domain name was registered and when that domain was first flagged for phishing by one of the phishing data feeds. This data set contained 65,255 gTLD domains for which we were able to obtain a registration date.



The chart above shows that 45% of the domains (31,610 out of 65,255) were used for phishing within 14 days of registration. This reinforces 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.[4, 5] Almost 78% of the domains were flagged for phishing within the first year of registration. The remaining 22% (14,360) of domains used for phishing were more than a year old.

We saw 41,210 gTLD domains which we classified as malicious registrations and for which we had registration dates. Malicious registrations are domain names we believe were registered by phishers to perpetrate phishing. (For more, see the section "Malicious Domain Name Registration," below.) We see that **57% of malicious domain registrations are used for phishing within the first three days**:



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

One implication is that phishers are either paying for their domain names with legitimate means, and/or that the payment processors used by the registrars are not recognizing many suspicious or fraudulent transactions at the time of transaction or in the days thereafter. If a domain purchase transaction is flagged as fraudulent, the domain registrar will usually 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.

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 like those incorporated into the COMAR system.) For more about this, see Appendix A: Identifying Malicious vs. Compromised Domains.

The data shows that 17% of the maliciously registered domains were not used until more than 90 days after they were registered. We call this “aging,” and it is a phenomenon that has been observed with batches of domains used for spamming. Recently registered domains receive low reputation scores from security and anti-spam companies, and some phishers apparently wait for their domains to move out of “very new domain” status.

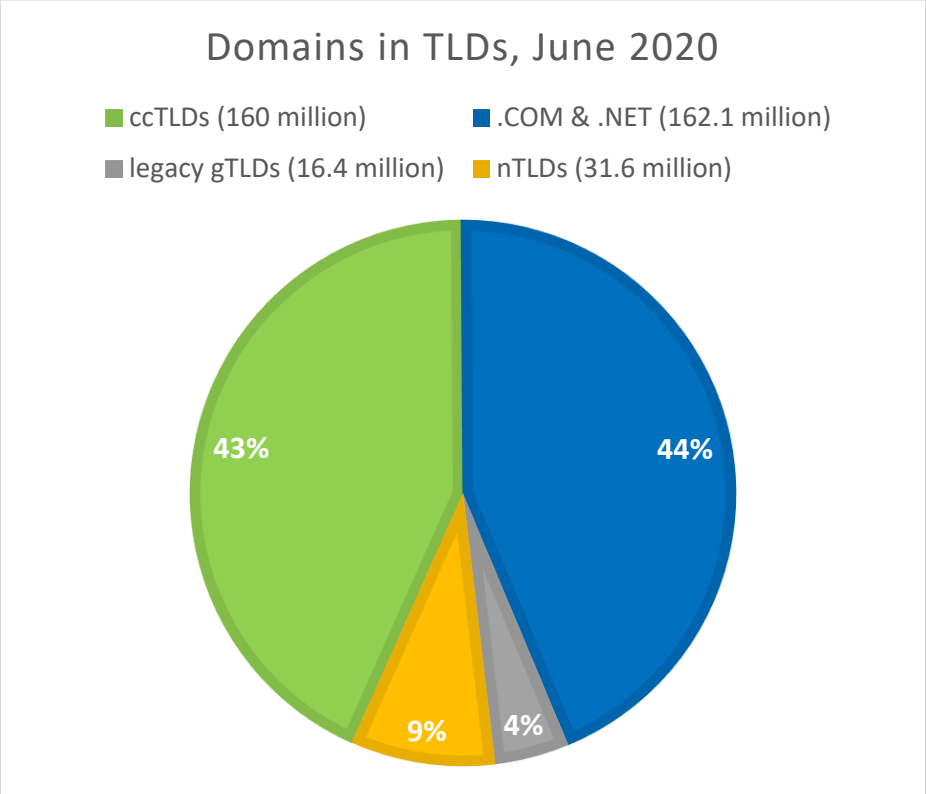
Our data also shows a small set of about 800 domains that appear to be maliciously registered but were flagged for phishing more than a year after the first year of registration, and sometimes several years after registration. These included domain names such as: microsftoffice365.com, signinsupport.com, com-verify.space, microsoftoutlookoffice.com, and customer-account-support-uk.com. These domains were either kept for a year or more by the phisher who used them (and therefore paid domain renewal fees) or phishers picked up these domains on the secondary market, such as through auctions.

The two charts above demonstrate a population of domains that get compromised by phishers within a year of registration. The data is roughly consistent with findings from the COMAR project, which found that among compromised domains used for phishing, 12% of the domains get compromised within three months of their registration.[6]

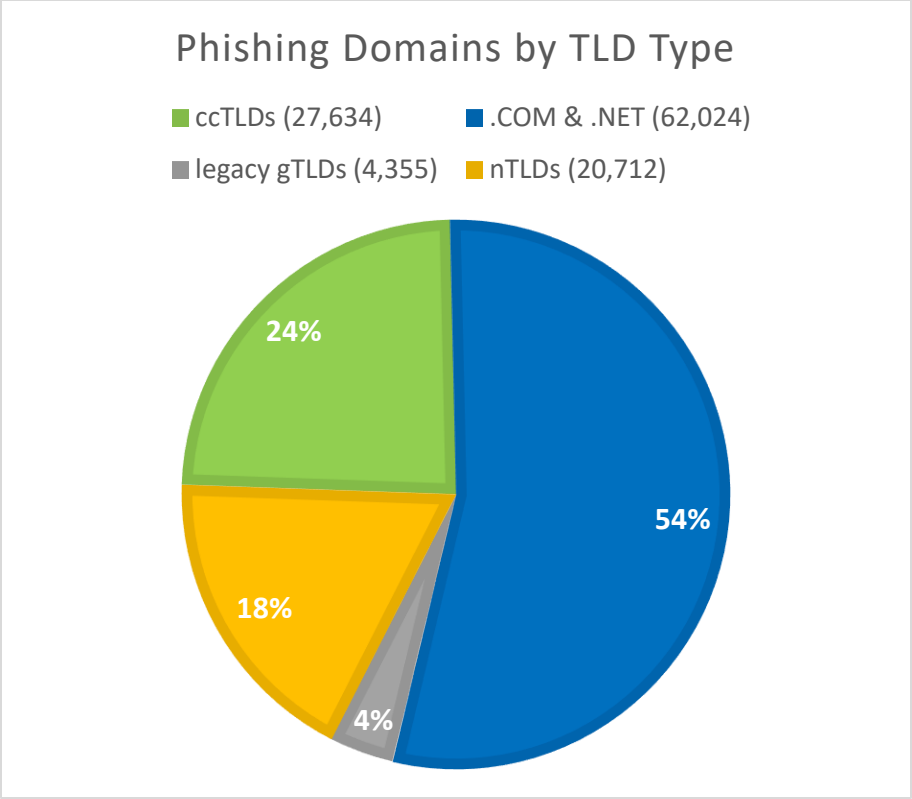
Phishing by Top-Level Domain (TLD)

We analyzed the phishing domains and attacks to see how they were distributed across the top-level domains. Most phishing continues to be concentrated in just a few namespaces, with some TLDs attracting many more problems (and/or more prevalent problems) than others. During our three-month study period, we observed phishing in 439 different top-level domains.

As of June 2020, there were 370.1 million registered domain names in the world's registries.[7] The domain name space can be divided into four categories: the .COM and .NET registries are operated by Verisign and represented 44% of the domains in the world; country-code domains (ccTLDs) represented another 43%; the legacy generic TLDs introduced before 2013 (.ORG, .BIZ, .INFO, etc.) represented 4%; and the new gTLDs (nTLDs) introduced from 2014 to the present were the remaining 9%:



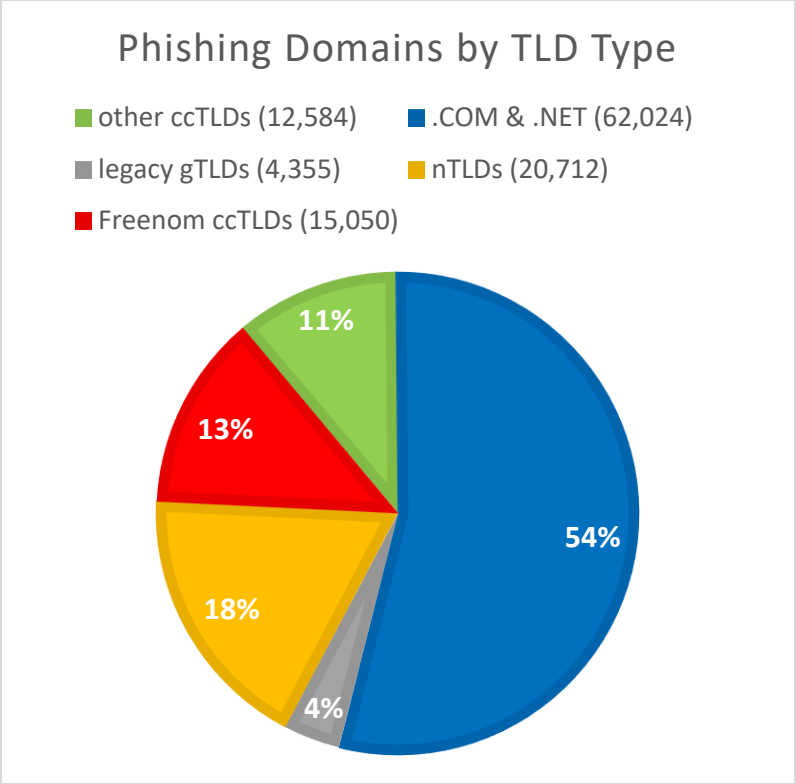
However, the distribution of domains used for phishing is different:



54% of the domains used for phishing were in .COM and .NET. This happened for two reasons: there are many web sites in .COM that were compromised and had phish placed on them, and phishers also registered large numbers of .COM domains.

While the new gTLDs were only 9% of domains in registries, 18% of the domains used for phishing were in the new gTLDs — twice that of the new gTLDs' market share. Of the domains used for phishing in the new gTLDs, 81% were maliciously registered by phishers. (See "Malicious Domain Name Registrations," below.)

While ccTLDs are 43% of the domain name market, only 24% of domains used for phishing were in ccTLDs. It is important to note that the ccTLD category was artificially swollen by large numbers of phishing domains in five "commercialized" ccTLDs that offer free domain name registrations. Those were .TK, .GA, .ML, .CF, and .GQ, operated by Freenom, a company in the Netherlands. The Freenom ccTLDs represented more than half of all phishing domains in ccTLDs, and 13% of all domains used for phishing. This leaves all the other ccTLDs with only 11% of the domain names used for phishing:



The TLDs with the highest numbers of new phishing domains discovered in our May through July 2020 study period were:

Rank	TLD	TLD type	Domains in TLD	Phishing Domains	Phishing attacks
1	com	legacy gTLD	151,931,301	43,753	58,685
2	xyz	new gTLD	3,136,553	4,059	4,271
3	tk	ccTLD	25,644,936	3,798	3,829
4	top	new gTLD	3,748,802	3,003	3,064
5	buzz	new gTLD	604,706	2,704	2,716
6	ga	ccTLD	5,057,226	2,574	2,599
7	ml	ccTLD	4,162,031	2,559	2,582
8	net	legacy gTLD	13,705,756	2,319	3,339
9	info	legacy gTLD	4,787,440	2,316	2,449
10	cf	ccTLD	4,453,018	1,915	1,927
11	gq	ccTLD	3,692,011	1,738	1,749
12	org	legacy gTLD	10,648,071	1,639	2,182
13	icu	new gTLD	6,611,658	1,589	1,651
14	wang	new gTLD	1,392,249	1,385	1,386
15	ru	ccTLD	4,867,074	1,281	2,069
16	cn	ccTLD	15,961,895	1,216	1,262
17	online	new gTLD	1,586,898	1,173	1,187
18	live	new gTLD	719,372	1,116	1,121
19	br	ccTLD	4,442,239	1,103	1,241
20	in	ccTLD	2,284,123	926	1,033

The gross numbers of phishing domains above 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.*

It is also interesting to compare whether a TLD has a higher or lower incidence of phishing relative to others. To measure the prevalence of phishing in each TLD, we use "Phishing Domains per 10,000." 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. "Phishing Attacks per 10,000" can highlight where high-volume phishers place multiple phish on one domain. (See "Abuse of Subdomain Providers," below.) These kinds of metrics, with variations, have been used by observers to score cybercrime.[8, 9, 10, 11, 12, 13, 14]

In our 2020 data set:

- The median phishing-domains-per-10,000 score was 2.5.
- .COM, the world's largest and most ubiquitous TLD, had a domains-per-10,000 score of 2.9.
- Among TLDs with 10,000 or more domains in them, the average score was 4.7.

We therefore suggest that domains-per-10,000 scores between 2.5 and 4.7 occupy the middle ground, with scores above 4.7 indicating TLDs with increasingly prevalent phishing.

The top TLDs by phishing domain score (considering those with at least 30,000 domains in the TLD and at least 25 phishing domains) are:

Rank	TLD	TLD type	Domains in TLD	Phishing Domains	Phishing Domain Score	Phishing attacks	Attack Score
1	host	new gTLD	97,718	667	68.3	754	77.2
2	buzz	new gTLD	604,706	2,704	44.7	2,716	44.9
3	best	new gTLD	113,614	433	38.1	436	38.4
4	casa	new gTLD	30,000	83	27.7	92	30.7
5	services	new gTLD	53,454	146	27.3	147	27.5
6	ph	ccTLD	107,421	185	17.2	196	18.2
7	monster	new gTLD	104,126	176	16.9	186	17.9
8	live	new gTLD	719,372	1,116	15.5	1,121	15.6
9	xyz	new gTLD	3,136,553	4,059	12.9	4,271	13.6
10	ve	ccTLD	31,788	37	11.6	52	16.4
11	pk	ccTLD	89,707	103	11.5	114	12.7
12	id	ccTLD	344,198	389	11.3	449	13.0
13	business	new gTLD	41,500	44	10.6	45	10.8
14	wang	new gTLD	1,392,249	1,385	9.9	1,386	10.0
15	ke	ccTLD	80,960	80	9.9	82	10.1
16	pe	ccTLD	109,174	93	8.5	101	9.3
17	top	new gTLD	3,748,802	3,003	8.0	3,064	8.2
18	center	new gTLD	41,437	31	7.5	32	7.7
19	digital	new gTLD	58,002	43	7.4	44	7.6
20	online	new gTLD	1,586,898	1,173	7.4	1,187	7.5

In 2016, the top ten TLDs by score were all ccTLDs.[15] In our 2020 set, eight of the top ten TLDs are new gTLDs.

Phishing score is not a great indicator of whether a person is more or less likely to encounter a dangerous domain while surfing the Web, because users only encounter phishing domains when lured and they click on phishing links.

For more data about top-level domains, see "Malicious Domain Name Registrations," below.

Prevalence of Phishing by gTLD Registrar

To perpetrate phishing, phishers need domain names which they advertise in spam email. Phishers register domain names for their own uses and also break into the domain management and hosting accounts of innocent domain name owners. In this section and the next section ("Malicious Domain Name Registrations"), we look at where gTLD domain names were purchased and managed, and where larger-than-usual concentrations of phishing occur in registrars' domain portfolios.

A total of 414 registrars sponsored the domains used for phishing that appeared in our three-month study period. Of those, 202 registrars had only one or two domains with phishing on them. The median phishing-domains-per-10,000 score was 2.2. The median score for registrars with at least 10,000 domains under management was 1.2. GoDaddy, the world's largest registrar with 30% of the gTLD market, had a domains-per-10,000 score of 1.7.

The registrars with the highest numbers of gTLD phishing domains newly discovered in our May-July 2020 study period were:

Rank	Registrar Name	IANA ID	gTLD domains under management	Phishing Domains	Phishing Domains Score
1	GoDaddy.com, LLC	146	63,168,934	10,822	1.7
2	NameCheap, Inc.	1068	10,323,962	7,816	7.6
3	NameSilo, LLC	1479	3,430,974	6,823	19.9
4	PDR Ltd. d/b/a PublicDomainRegistry.com	303	4,845,099	6,068	12.5
5	Tucows Domains Inc.	69	10,194,582	2,852	2.8
6	Google LLC	895	4,894,266	2,736	5.6
7	ALIBABA.COM SINGAPORE E-COMMERCE PRIVATE LIMITED	3775	830,808	2,489	30.0
8	Wild West Domains, LLC	440	2,755,518	1,686	6.1
9	Chengdu West Dimension Digital Technology Co., Ltd.	1556	3,536,841	1,521	4.3
10	eNom, LLC	48	5,314,291	1,500	2.8
11	Name.com, Inc.	625	2,077,924	1,151	5.5
12	Hosting Concepts B.V. d/b/a Openprovider	1647	812,193	1,072	13.2
13	Shinjiru Technology Sdn Bhd	1741	24,309	1,020	419.6

Rank	Registrar Name	IANA ID	gTLD domains under management	Phishing Domains	Phishing Domains Score
14	Jiangsu Bangning Science & technology Co. Ltd.	1469	641,974	961	15.0
15	Registrar of Domain Names REG.RU LLC	1606	917,444	920	10.0
16	GMO Internet, Inc. d/b/a Onamae.com	49	5,325,923	856	1.6
17	Web Commerce Communications Limited dba WebNic.cc	460	2,043,479	780	3.8
18	Alibaba Cloud Computing Ltd. d/b/a HiChina (www.net.cn)	1599	5,444,550	759	1.4
19	Wix.com Ltd.	3817	924,058	723	7.8
20	FastDomain Inc.	1154	2,334,868	689	3.0

The top registrars by score (considering registrars with at least 20,000 gTLD domains under management, and at least 25 phishing domains) were:

Rank	Registrar Name	IANA ID	gTLD domains under management	Phishing Domains	Phishing Domains Score
1	Shinjiru Technology Sdn Bhd	1741	24,309	1,020	419.60
2	ALIBABA.COM SINGAPORE E-COMMERCE PRIVATE LIMITED	3775	830,808	2,489	29.96
3	Center of Ukrainian Internet Names (UKRNAMES)	1436	16,224	48	29.59
4	NameSilo, LLC	1479	3,430,974	6,823	19.89
5	DomainPeople, Inc.	65	226,319	416	18.38
6	NICENIC INTERNATIONAL GROUP CO., LIMITED	3765	23,973	40	16.69
7	CNOBIN INFORMATION TECHNOLOGY LIMITED	3254	21,593	34	15.75
8	Jiangsu Bangning Science & technology Co. Ltd.	1469	641,974	961	14.97
9	TLD Registrar Solutions Ltd.	1564	82,239	116	14.11
10	Hosting Concepts B.V. d/b/a Openprovider	1647	812,193	1072	13.20
11	Innovadeus Pvt. Ltd.	3812	29,557	39	13.19
12	PDR Ltd. d/b/a PublicDomainRegistry.com	303	4,845,099	6,068	12.52
13	Domainshype.com, LLC	1660	33,214	39	11.74
14	NETIM SARL	1519	35,052	41	11.70
15	Sav.com, LLC	609	135,595	154	11.36
16	OnlineNIC, Inc.	82	635,435	646	10.17

Rank	Registrar Name	IANA ID	gTLD domains under management	Phishing Domains	Phishing Domains Score
17	BigRock Solutions Ltd.	1495	276,454	280	10.13
18	Registrar of Domain Names REG.RU LLC	1606	917,444	920	10.03
19	Internet Domain Service BS Corp	2487	352,086	351	9.97
20	CommuniGal Communication Ltd.	418	49,971	48	9.61

Shinjiru Technology Sdn Bhd is a small registrar and hosting provider in Malaysia. It had more than 1,020 phishing domains newly appear in the study period, representing more than 4% of the registrar's total gTLD domain portfolio. As we will see below, it appears that 983 of those domains were registered by phishers.

For more data about domain registrars, see "Malicious Domain Name Registrations," below.

Malicious Domain Name Registrations

We analyzed how many domain names were registered by phishers to operate phishing sites — which we refer to as "malicious registrations." Malicious domains can be suspended by a registrar or registry operator, without causing any collateral damage. These are in contrast to phishing that appeared on compromised (hacked) domains and their vulnerable hosting. Compromised domains are owned by innocent parties. Suspending such domains will stop the phishing, but it will also prevent the legitimate services on the domain name from functioning. Compromised domain names should not be suspended; instead, the phishing pages can be taken offline by the web hosting provider.

Of the 99,412 domains used for phishing in the study period, we identified 60,935 that we believe were registered maliciously, by phishers. That represented 61% of the domains, with the other 39% classified as compromised. This year a separate study, using a similar data set and designed by researchers at ccTLD operators SIDN and AFNIC, found that 58% of phishing domains (in all TLDs) are maliciously registered, and 42% are compromised.[16]

Domain name registrars and registry operators are therefore in an excellent position to find and prevent the majority of phishing, which takes place on maliciously registered domains. It is possible for registrars and registry operators to identify maliciously registered phishing domains with a high degree of accuracy, often at the time of registration. For example, many domains registered by phishers also have telltale characteristics that can be used to identify them quickly and with low false-positive rates. For more about the methods and tools available, see Appendix A: Identifying Malicious vs. Compromised Domains. Registrars also possess dispositive information that no one else does: the registrant's identity (contact information, now mostly redacted in public WHOIS as allowed by a recent change in ICANN policy), the registrant's payment information, the registrant's IP address, and the registrant's purchase history. Registry operators also have access to the registrant contact information.

These factors are highly useful to determine whether a registration is risky, and whether the registrant customer has been honest about its contact information.

The larger the number of malicious phishing registrations in a portfolio controlled by one company, the greater the need for that company to identify and suspend malicious phishing registrations early.

We flagged a domain as malicious if it was reported for phishing within seven days of being registered or if it contained a brand name or misleading string and was reported for phishing. Quick usage of the domain indicates that it is not compromised, and a misleading domain name is a sign of bad intent. For a fuller description of the methodology, and how it is congruent to methods used by other researchers, see Appendix B.

Malicious registrations appeared in 283 TLDs. **More than 88% of the malicious domains in our data set occurred in just these 20 TLDs**:

Rank	TLD	TLD type	Domains in TLD	Phishing Domains	Malicious Phishing Domains	% malicious
1	com	legacy gTLD	151,931,301	43,753	24,925	57.0%
2	tk	Freenom ccTLD	25,644,936	3,798	3,362	88.5%
3	buzz	new gTLD	604,706	2,704	2,674	98.9%
4	xyz	new gTLD	3,136,553	4,059	2,669	65.8%
5	top	new gTLD	3,748,802	3,003	2,441	81.3%
6	ga	Freenom ccTLD	5,057,226	2,574	2,186	84.9%
7	ml	Freenom ccTLD	4,162,031	2,559	2,046	80.0%
8	info	legacy gTLD	4,787,440	2,316	1,960	84.6%
9	cf	Freenom ccTLD	4,453,018	1,915	1,636	85.4%
10	gq	Freenom ccTLD	3,692,011	1,738	1,522	87.6%
11	icu	new gTLD	6,611,658	1,589	1,456	91.6%
12	wang	new gTLD	1,392,249	1,385	1,367	98.7%
13	live	new gTLD	719,372	1,116	1,019	91.3%
14	net	legacy gTLD	13,705,756	2,319	1,013	43.7%
15	cn	ccTLD	15,961,895	1,216	910	74.8%
16	online	new gTLD	1,586,898	1,173	730	62.2%
17	host	new gTLD	97,718	667	657	98.5%
18	org	legacy gTLD	10,648,071	1,639	540	32.9%
19	us	ccTLD	1,668,953	709	490	69.1%
20	ru	ccTLD	4,867,074	1,281	446	34.8%

Most of the above TLDs had high percentages of malicious registrations. The exceptions were .NET, .ORG, and .RU, where the majority of domains used for phishing were compromised.

Of the domains used for phishing in all the new gTLDs, 81% were maliciously registered by phishers. Most of those were concentrated in a small number of new gTLDs.

Many of the maliciously registered domains in .BUZZ targeted users of Microsoft services. The domains were registered in various batches, using scripts to create the domain names. For example one large batch used the name of a U.S. town appended with a state abbreviation (homesteadpa.buzz, bingerok.buzz, knoxborony.buzz, etc.). Another set consisted of long, random domains, such as:

dgds54h1b65rf41ehjn6t8e74j8rt74j8t5er47j8rtj747-56ej4tr4j46t4g7.buzz

and were used to attack users of Amazon, on URLs of format:

http://amazon.co.jp.dgds54h1b65rf41ehjn6t8e74j8rt74j8t5er47j8rtj747-56ej4tr4j46t4g7.buzz/1

To understand which registrars are experiencing more than their expected share of malicious domain registrations, we looked at the ratio of malicious vs. compromised gTLD domains handled by that registrar. The raw numbers of maliciously registered domains are important — they indicate where phishers were able to purchase domains:

Rank	Registrar Name	IANA ID	gTLD domains under management	gTLD Phishing domains	Malicious phishing domains	% malicious
1	NameCheap, Inc.	1068	10,323,962	7,816	6,052	77%
2	NameSilo, LLC	1479	3,430,974	6,823	5,999	88%
3	GoDaddy.com, LLC	146	63,168,934	10,822	4,322	40%
4	PDR Ltd. d/b/a PublicDomainRegistry.com	303	4,845,099	6,068	3,855	64%
5	Tucows Domains Inc.	69	10,194,582	2,852	1,869	66%
6	Google LLC	895	4,894,266	2,736	1,860	68%
7	Chengdu West Dimension Digital Technology Co., Ltd.	1556	3,536,841	1,521	1,342	88%
8	ALIBABA.COM SINGAPORE E-COMMERCE PRIVATE LIMITED	3775	830,808	2,489	1,207	48%
9	Wild West Domains, LLC	440	2,755,518	1,686	1,096	65%
10	Shinjiru Technology Sdn Bhd	1741	24,309	1,020	983	96%
11	Hosting Concepts B.V. d/b/a Openprovider	1647	812,193	1,072	972	91%
12	Jiangsu Bangning Science & technology Co. Ltd.	1469	641,974	961	855	89%
13	Name.com, Inc.	625	2,077,924	1,151	833	72%
14	Registrar of Domain Names REG.RU LLC	1606	917,444	920	812	88%
15	eNom, LLC	48	5,314,291	1,500	752	50%
16	Wix.com Ltd.	3817	924,058	723	709	98%
17	GMO Internet, Inc. d/b/a Onamae.com	49	5,325,923	856	666	78%

Rank	Registrar Name	IANA ID	gTLD domains under management	gTLD Phishing domains	Malicious phishing domains	% malicious
18	Web Commerce Communications Limited dba WebNic.cc	460	2,043,479	780	656	84%
19	OnlineNIC, Inc.	82	635,435	646	519	80%
20	Register.com, Inc.	9	1,708,466	539	428	79%

This means that at least 47% of all maliciously registered domains were registered at just the top ten gTLD registrars above. In addition, Freenom acts as the sole registrar for domains in its TLDs, which contained 17.6% of the maliciously registered domains in the world.

Addressing Malicious Registrations

Multiple TLDs and registrars had both high numbers of maliciously registered domains and high percentages of malicious registrations. What happened, and how can these problems be solved?

These concentrations of abuse can be caused by one or a combination of 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.[17]
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.[18]
4) Ineffective anti-abuse programs, which do nothing to deter constant abuse.

An example of a TLD with high numbers of malicious registrations was .BUZZ. In July 2019 .BUZZ was a small TLD, with only 17,771 domains in it. .BUZZ then entered a period of aggressive expansion, and grew to about 620,000 domains in July 2020. The expansion was accompanied by an increase in abuse, and we recorded 2,610 maliciously registered domains in .BUZZ in our three-month study period. All but ten of those phishing domains were registered at the U.S. registrar NameSilo. NameSilo only sponsored 70,900 .BUZZ domains total in May 2020,. But the largest .BUZZ registrar, Eranet, sponsored 447,000 .BUZZ domains and had no phishing domains in .BUZZ at all during out study period. Eranet may have brought growth to .BUZZ without the side effect of abuse, while NameSilo brought less growth and brought some highly problematic customers that used .BUZZ domains to perpetrate cybercrime. The registry operator of .BUZZ may not have known about the phishing, and/or may not have addressed it in a timely fashion.

As previously noted, the Freenom TLDs (.TK, .GA, .ML, .CF, and .GQ) offer domain names for free, and contained 13% of the phishing domains in our entire data set. Those domains are registered at Freenom directly, rather than through registrars, and at least 80% appear to be maliciously registered. Freenom's TLDs contained 10,752 maliciously registered domains, representing 17.6% of the maliciously registered domains reported in our data set.

Mitigation is taking steps to stop a phishing attack once it is underway. Mitigation is *reactive*. As noted above, phishing attacks do half of their damage before they are discovered and blocklisted, if they are discovered and blocked at all. Freenom offers an API that allows vetted security companies to suspend Freenom domain names involved in cybercrime.[19] While this is a useful feature, it has not reduced the number of Freenom domain names used for phishing. Indeed, the number of phishing domains in Freenom's TLDs has grown. A research survey found 2,758 phishing domains in .TK during *all* of 2016.[20] In 2020 we saw 3,362 maliciously registered phishing domains appear in .TK over the course of just *three months*. Some of those domains are suspended only after the phishing attacks start, and therefore after much of the damage has taken place**.** While there is a mitigation program at the Freenom TLDs, it has not prevented constant phishing.

The .XYZ registry has had a phishing mitigation program in place for more than five years, including a "sophisticated abuse monitoring tool" and three full-time employees devoted to handling abuse.[21, 22] But during our study period, .XYZ had 2,669 malicious registrations blocklisted for phishing — the fourth-largest number in any TLD. While the uptime of those phishing attacks might be reduced by the mitigation program, they were likely mitigated only after most of the damage had been done. In the meantime, phishers have been finding .XYZ to be a suitable place to buy large numbers of domains and use them for attacks — suggesting that the anti-abuse program has not impacted the phishers in a meaningful way.

All gTLD registrars are contractually required by ICANN to have mitigation programs.[23] They must:

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

However, the data about malicious registrations shows that phishers are able to register and use large numbers of domains at specific registries and registrars, again and again over time.

Mitigation programs can produce incremental reductions of damage and losses if implemented well. But they may also allow constant cycles of new phishing, leading to no overall improvement of Internet safety. Mitigation and *prevention* are two very different things, and both are needed.

Notably, only 40% of the domains used for phishing under GoDaddy's management were maliciously registered, a much lower percentage than other registrars on the above list. The number may mean that GoDaddy is better at deterrence or is suspending some phishing domains before they are actually used and blocklisted.

Phishing by Autonomous System and Hosting Provider

We looked to see where phishing sites were being hosted. We collected the IP addresses (A records) that phishing attacks were resolving to; those IPs were provided by OpenPhish, APWG, and PhishTank as part of their phishing reports and were captured at the time of the phishing. We then looked up what autonomous system (AS) each IP was in. This provides insight into the entities hosting the domains.

An AS is a collection of the IP addresses (routing prefixes) controlled by a common network administrator. That administrator may be a business, a university, an Internet Service Provider, or a network operator providing service to several of those types of entities. It is common for larger hosting providers and infrastructure providers to have several AS numbers. Two or more hosting providers may be allocated space within a single AS. For more about our methodology, please see Appendix B.

The largest numbers of phishing domains were hosted in the following ASes (with a minimum of 768 addresses, or three /24 blocks):

Rank	AS Name	AS Number	# routed IPv4 Addresses	Phishing Domains	Domains Score
1	CLOUDFLARENET	13335	1,570,560	1,769	11.3
2	NAMECHEAP-NET	22612	35,072	1,674	477.3
3	AS-26496-GO-DADDY-COM-LLC	26496	935,168	1,041	11.1
4	UNIFIEDLAYER-AS-1 [Endurance.com]	46606	1,373,952	671	4.9
5	MICROSOFT-CORP-MSN	8075	37,570,816	545	0.1
6	GOOGLE	15169	10,280,448	529	0.5
7	HOST4GEEKS-LLC	393960	5,120	404	789.1
8	AMAZON-02	16509	50,380,544	261	0.1
9	PIHL-AS - Private Internet Hosting LTD	213058	768	229	2,981.8
10	BCPL-SG BGPNET Global ASN	64050	216,576	226	10.4
11	SHINJIRU-MY-AS-AP Shinjiru Technology Sdn Bhd	45839	21,248	223	105.0
12	OVH - OVH SAS	16276	3,485,440	217	0.6
13	CONTABO - Contabo GmbH	51167	218,368	192	8.8
14	AS-HOSTINGER - Hostinger International Limited	47583	70,912	190	26.8
15	GODADDY - Host Europe GmbH	20773	162,816	181	11.1
16	MULTA-ASN1	35916	3,441,664	165	0.5
17	DIGITALOCEAN-ASN	14061	2,329,088	162	0.7
18	BEON-AS-ID PT. Beon Intermedia	55688	2,048	162	791.0
19	CNNIC-ALIBABA-US-NET-AP Alibaba (US) Technology Co.	45102	10,199,296	151	0.1
20	ASN-QUADRANET-GLOBAL	8100	506,112	130	2.6

At #1, with the most phishing domains, was Cloudflare. Technically, Cloudflare does not host domains on its servers. Rather, Cloudflare is a reverse proxy, and acts as an intermediary. DNS requests for domains using Cloudflare's service go to Cloudflare's servers, which then send the traffic on to the actual hosts, and back. The domains resolving to Cloudflare were ultimately hosted somewhere else, and the actual hosting locations are known only to Cloudflare.

At #2, #3, and #4 are companies that are accredited as ICANN registrars which also offer hosting services as part of their product mixes. Such companies hosted significant numbers of phishing domains; in the case of malicious registrations, phishers purchased the domain name and also hosting. These companies include Namecheap at #2, GoDaddy at #3 and #15, and Endurance International at #4. (Unified Layer is part of the Endurance International Group, which is the parent of more than 80 hosting, domain name, and Internet services companies, including Domain.com, Dotster, FastDomain, PublicDomainRegistry, Constant Contact, ResellerClub, BlueHost, and HostGator.[24] A number of those utilize AS46606.)[25]

At 5, #6, and #8 were the cloud hosting providers Microsoft, Google, and Amazon. These services are popular and host large numbers of domains, and control tens of millions of IP addresses each.

At #7 was Host4Geeks, a small hosting provider registered in California, according to WHOIS records at ARIN. However, that address is merely a virtual office, and the company is actually operated from India. Over the course of our three-month study period, phishers registered domain names at various registrars and then hosted the domains at Host4Geeks. It appears that one phisher may have registered more than 200 domains at registrar Tucows during that time span, repeatedly targeting users of Amazon and Square. In these cases, phishing was reported on the same day as the domain registration. Sometimes Tucows suspended the domains on the same day; but other times it took Tucows up to 16 days to suspend the domains.[26] This is an example where the registrar did not catch onto the repeat problems, abuse continued over time, and so mitigation after the attacks was often too late to prevent victimization.

At #9 was PIHL-AS - Private Internet Hosting LTD. It is a small hosting provider registered in Belize City, Belize, but has connections to Russia; its only upstream connection to the Internet is AS35196 Ihor Hosting LLC in Moscow.[27] On 23 September 2020, Spamhaus declared a block of IP addresses at Private Internet Hosting LTD to be "cybercriminal bulletproof hosting" for supporting both phishing hosting servers and botnet controllers.[28] "Bulletproof hosting" companies are lenient about what they host, are resistant to requests to take down illegal activities, and are often located in "offshore" countries in order to gain immunity from legal process.[29]

The highest scores for malicious phishing domains were at the following Autonomous Systems (considering those with a minimum of 768 IP addresses (three /24 blocks), and at least 25 phishing domains):

Rank	AS Name	AS Number	# routed IPv4 Addresses	Phishing Domains	Malicious Phishing Domains	Malicious Domains Score
1	PIHL-AS - Private Internet Hosting LTD	213058	768	229	198	2578.1
2	BEON-AS-ID PT. Beon Intermedia	55688	2,048	162	157	766.6
3	HOST4GEEKS-LLC	393960	5,120	404	389	759.8
4	NAMECHEAP-NET	22612	35,072	1,674	1,508	430.0
5	SIMPLECARRER2 - Simple Carrier LLC	34888	768	30	19	247.4
6	VERDINA - Verdina Ltd.	201133	3,328	41	36	108.2
7	SHINJIRU-MY-AS-AP Shinjiru Technology Sdn Bhd	45839	21,248	223	171	80.5
8	DDOS-GUARD - DDOS-GUARD LTD	57724	8,448	68	47	55.6
9	INTERNET-IT - INTERNET IT COMPANY INC	200313	9,728	64	51	52.4
10	HOSTKEY-AS - HOSTKEY B.V.	57043	12,288	48	45	36.6
11	NCONNECT-AS - LLC "Server v arendy"	49335	20,736	68	65	31.3
12	AS-HOSTINGER - Hostinger International Limited	47583	70,912	190	178	25.1
13	GODADDY-AMS - Host Europe GmbH	21501	43,008	115	100	23.3
14	CONTABO	40021	12,288	45	26	21.2
15	A2HOSTING	55293	86,528	115	110	12.7
16	IHOR-AS - Ihor Hosting LLC	35196	29,714	42	37	12.5
17	AS-26496-GO-DADDY-COM-LLC	26496	935,168	1,041	962	10.3
18	GODADDY - Host Europe GmbH	20773	162,816	181	167	10.3
19	CLOUDFLARENET	13335	1,570,560	1,769	1,552	9.9
20	BCPL-SG BGPNET Global ASN	64050	216,576	226	213	9.8

At #2 was PT Beon Intermedia, a hosting provider in Indonesia. The phishing there was carried out with at least 635 free .TK domain names, registered in batches, and easily identifiable once the pattern is recognized. Most targeted users of Facebook, for example:

fb-recovery-10000076857-it.tk
fb-recovery-10000076858-it.tk
fb-recovery-10000076859-it.tk
fb-recovery-10000076860-it.tk
fb-payment-update10182802.tk
fb-payment-update10182803.tk
fb-payment-update10182804.tk
fb-payment-update10182805.tk

and also users of Checkpoint:

checkpoint-bussiness-manager-corporation-012.tk
checkpoint-bussiness-manager-corporation-013.tk
checkpoint-bussiness-manager-corporation-014.tk
checkpoint-bussiness-manager-corporation-016.tk

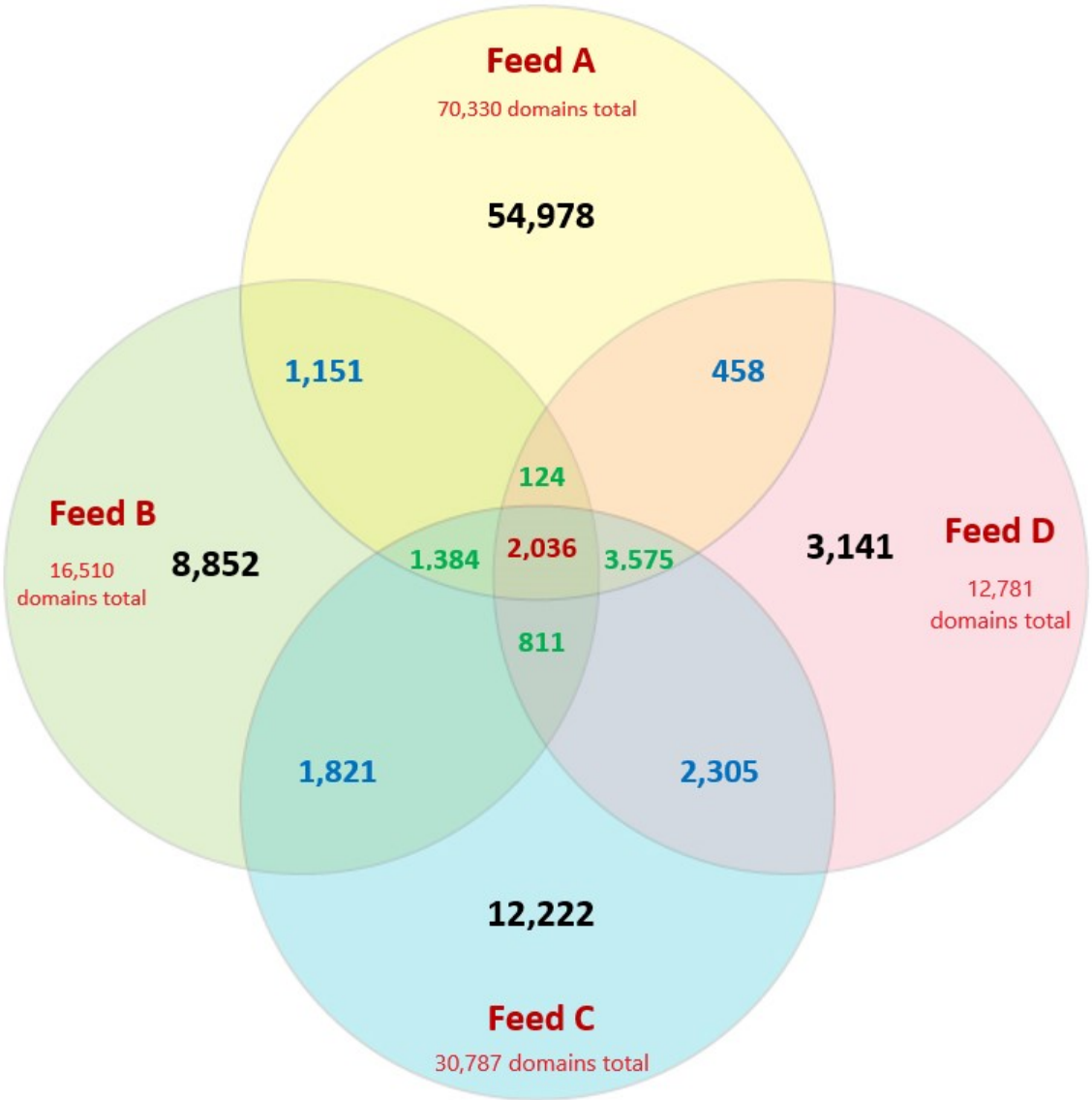
The phishers also used 69 .ID domains (the ccTLD of Indonesia) to target Facebook users, such as:

social-network-correction-identify-us1.my.id
social-network-correction-identify-us2.my.id
social-network-correction-identify-us3.my.id
social-network-correction-identify-us4.my.id
activation-center-social-media-us2020.my.id
advanced-pages-certificated-valid-2020.my.id
center-activation-social-media-2020.my.id

List Coverage: The Phish That Get Away

By collecting data from multiple sources, we confirmed that there is low overlap between anti-phishing blocklists. This emphasizes that **phishing is a much larger problem than is reported, and that even the best detection systems are finding only a percentage of the phishing attacks on the Internet. An ominous problem is: 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?

The four sources we looked at discovered a total of 99,412 unique domain names listed for phishing (either URLs on those domains, or the domain itself) during the three-month study period. Most of the domains were listed by one source and one source only. Only 2,036 of the domains were identified by all four sources. The detection overlap of domain names used for phishing was: [30]



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.[31, 32, 33, 34, 35, 36, 37]

What explains the low overlap? Some are problems 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.[38, 39, 40] 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.[41, 42]
4. The sharing of data is uneven and is not always timely. Some phishing targets do not share data about the phishing that affect them, for fear that it will reflect negatively on their brands. Some anti-phishing vendors do not share their data due to competitive concerns.[43]
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. **This over-redaction of WHOIS data is leading to the under-identification of phishing domains**.[44, 45, 46]

When faced with these factors, even the most professional and experienced observers can only find a portion of the phishing that occurs and are challenged to do so in a timely fashion.

Another implication is that a party who uses only one blocklist to protect itself will leave its users exposed to a significant number of phishing attacks over time. This is not to disparage the value of blocklists — they are essential tools for cybersecurity, they prevent enormous damage, and all organizations should take advantage of them directly or through their service providers. It is simply an acknowledgement that no solution provides complete protection, and that phishers place defenders at a disadvantage.

Regional Phishing and the Effect of Data Sharing

Our 2020 data set seems to significantly under-represent the phishing that takes place in certain regions. The data contains only a handful of phishing attacks against popular, online Chinese targets such as Alibaba, Made-in-China, and WeChat, and no attacks against Chinese banks and major providers such as Baidu and JD.com. We suspect this under-reporting is the result of both an under-detection and an under-sharing of data. 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.[47] Those studies obtained data from 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 kind of in-country data is absent from the four sources we observed in 2020, 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, and, if they are, they may not be parsing Chinese-language emails effectively.

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. Notably, our data showed that several Chinese registrars had significant numbers of malicious domain name registrations, but the attacks on those domains targeted non-Chinese brands, notably Microsoft and Japanese companies.

Our data set also seems to under-represent phishing against Russian brands. It contains only three attacks against leading search site Yandex, a few against mail provider Mail.RU, and none against Wildberries, the largest Russian online retailer. The set did contain almost 300 attacks against Russian social media site Vkontakte.

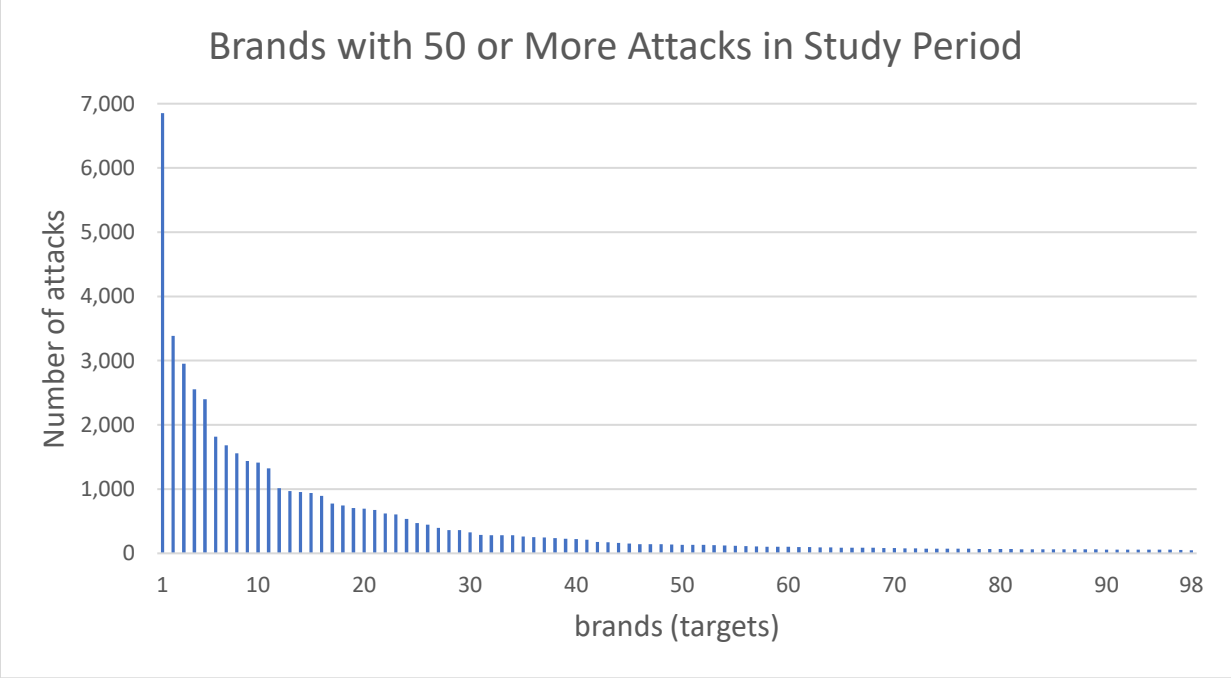
In contrast, our data set contains more than 1,700 attacks against brands in South America, including the leading Brazilian retailer Magazine Luisa, and banks across Latin America. The data also contains more than 1,800 attacks against brands in Japan, including against Rakuten's Japanese-language site. Many of these attacks against Latin American and Japanese targets were reported by members of the Anti-Phishing Working Group operating in those regions. *Here, data sharing provided visibility, better blocklisting, and better protection.*

Target Distribution

The data sources reported 684 different brands that were targeted by phishing. These were a wide variety of targets, including banks, social media companies, webmail, and games; national tax services; universities; and cryptocurrency exchanges.

A specific target was identified for 52,048 of the 122,092 attacks in our data set, or 43% of the attacks. Because target data was not confirmed for the majority of attacks, we decline to provide per-target numbers. The most-attacked targets identified by our data sources were, in alphabetical order: Amazon, Apple, AT&T, Chase, Facebook, LinkedIn, Microsoft, Outlook (owned by Microsoft), PayPal, and WhatsApp. The brunt of phishing was borne by those top ten targets, which suffered 50% of the identified phishing attacks.

But more than 300 brands were attacked at least five times each during our study period — a drumbeat of damaging crime that the targets had to deal with.



Unfortunately, a brand can become a phishing target at any time. Phishers are looking for companies that have potentially lucrative user information, are newly popular, and/or are not ready to respond to phishing. Phishing attacks against Zoom were not reported to the APWG or PhishTank in early 2020, but Zoom became a regular target of phishing attacks in April 2020, as companies and schools came to rely on Zoom when the COVID-19 pandemic forced in-person interaction online. That was a reflection of Zoom's popularity, rather than a reflection on the Zoom service's inherent security. Phishing is a social engineering attack that ultimately targets users and exploits their confusion or naivete, and the attack creates problems for brand owners, who want to protect their users and reputations. If a company takes in personal data, phishers may take steps to exploit it.

Abuse of Subdomain Service Providers

Our analysis reveals that 9% of all phishing attacks took place using resources at subdomain service providers. Subdomain services give customers services on a domain name that the provider owns. This gives users their own DNS space, on 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. Some 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 11,330 phishing attacks using subdomains provider services, 9% of the 122,090 total attacks in our data set. They sat on just 330 second-level domain names. Of those 11,330 attacks, 89% of them (10,031) occurred on resources 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	Domain	Phishing attacks
1	Hostinger	000webhostapp.com	3,626
2	Weebly	weebly.com	2,510
3	ChangeIP	multiple	1,062
4	Google	appspot.com and web.app	861
5	No-IP	multiple	552
6	GoDaddy	godaddysites.com	540
7	Yola	yolasite.com	220
8	Blogger (Google)	blogspot.com	232
9	Duck DNS	duckdns.org	203
10	Miarroba Networks	webcindario.com	157

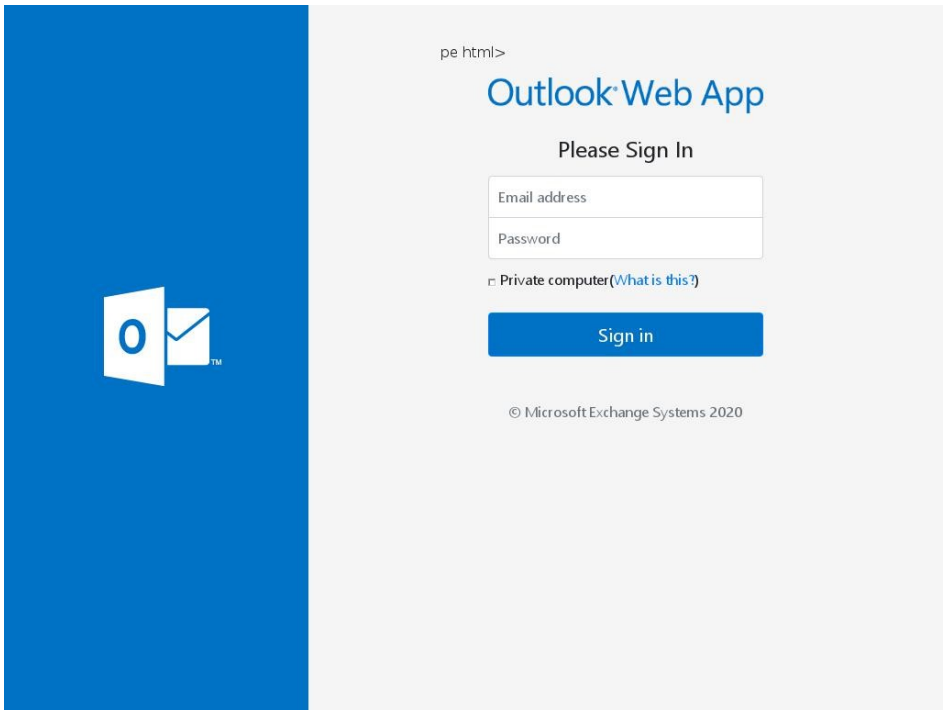
#1 Hostinger is a web hosting provider headquartered in Lithuania. It offers free web hosting via its 000WebHost arm. Phishers took advantage of this free resource, using it to launch thousands of different phishing sites, on thousands of different third-level domains on one domain: 000webhostapp.com. Those phish attacked at least 113 brands around the world. Based on that diversity of targets, it appears that the Hostinger service was used by many different phishers.

#2 Weebly, #6 GoDaddy, and #7 Yola provide website-building tools, and make third-level domains available to their customers.

#3 ChangeIP is an American company that offers free dynamic DNS. It provides subdomains that can be pointed to an IP address of the user's choice. The majority of the phishing that took place on

subdomains provided by ChangeIP pointed to phishing hosted at Contabo GmBH, a hosting provider in Germany.

#4 Google offers subdomains on Appspot.com,a cloud computing platform for developing and hosting web applications in Google-managed data centers; and on web.app, a mobile platform used for building mobile apps hosted by Firebase, which is Google's mobile app platform.



Phishing page on appspot.com, 22 June 2020.
https://login-microsoft-online.el.r.appspot.com

#5 No-IP is a free dynamic DNS provider. It points subdomains to the IP address of the user's choice. These phishing attacks occurred on multiple domains operated by No-IP, including ddns.net, bounceme.net, hopto.org, and viewdns.net. These subdomains resolved to phish on a wide variety of hosting providers around the world.

#8 Blogger provides blog-building tools and hosting and makes third-level domains available to its customers. Blogger is owned by Google and its blogs are hosted by Google.

#9 Duck DNS is a free dynamic DNS provider. It points subdomains to the IP address of the user's choice. These subdomains resolved to phish on a wide variety of hosting providers around the world.

#10 was Spanish hosting provider Miarroba Networks.

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 fifteen 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 domain name. These look-alike domains can be displayed in browser address bars if IDN display is enabled.

In our data set we saw 219 IDN domain names, used in 232 attacks. That was just 0.2% of the domains used for phishing.

- 190 domains were on non-IDN TLDs, such as: xn--blockchin-c2d.com. Of those, 116 were in 10 gTLDs, and 74 were in 20 (two-letter) ccTLDs.
- 29 domains were in six IDN TLDs, such as xn--e1afilellcz.xn--p1ai (усполитех.рф)

We classified 50 of the domains as true homographic attacks, for example:

xn--santnder-l8a.com → santąnder.com

and

xn--verzonwreless-yibe.com → verízonwíreless.com

Other 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--reperation-iphone-gteborg-hsc.com → reperation-iphone-göteborg.com

In studies from January 2007 to December 2014, APWG studies found nine true homographic phishing attacks. So, while more homographic attacks showed up in our three-month study set, domains that leverage the unique characteristics of IDNs for phishing remain a numerically small problem.

Appendix A: Identifying Malicious vs. 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. 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 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.

For this study, we considered a domain to be maliciously registered if it appeared on a blacklist within seven days of being registered, or if it had a famous brand name or misleading string in the domain name. When the above criteria identified domains, we then used clear evidence of common control and usage as an indicator to flag additional domains in a batch.

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).[48] 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

excellent 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 did, we looked for misspellings of brand names. COMAR 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, and from them created a list of keyword strings distinctive enough to avoid false-positives.[49] (For example we decided that "Uber" is not distinctive enough, since it is a common word in German.) We then compared that list to the domains used for phishing. COMAR used dnstwister and Levenshtein distance (with distance = 1) to find misspellings of brand names. 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. Similarly 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 facially designed to fool users and were used for phishing within a week of registration.

Appendix B: 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.[50, 51, 52, 53, 54] 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.[55, 56, 57]

- **Anti-Phishing Working Group eCrime eXchange (eCX) phishing feed**.[58] 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.**[59] 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)**.[60] 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)**.[61] 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 from 1 May through 31 July 2020. We collected and analyzed only newly found phishing incidents reported during that time. (Some of the sources also offer historical data, and as of 1 May 2020 the Spamhaus list contained domains added to its list before 1 May 2020. We did not include any of those historical entries in our data set.) 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. So 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 "Phishing Data Sources" 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 and separately stored the A record we found, determined the AS by using Team Cymru's IP to ASN mapping service[62]. We relied upon RIPE-NCC's WHOIS[63] 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[64]. We used the Public Suffix List[65] 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[66] for normalization. Some gTLD registries rate-limited[67] 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.[68] 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.[69]

Identifying Phishing Attacks

An *attack* is defined as a phishing site that targets a specific brand or entity. Many URLs can point to one phishing site, due to wildcarding and redirection. A single domain name can host several discrete phishing attacks against different companies. A phishing site can have more than one page (multiple URLs). To identify unique attacks, we designed scripts to compare URLs and meta-data. At a high level:

- We de-duplicated URLs; some URLs were reported by more than one source.
- We then compared URLs. One of the basic rules was: if the hostname (for example hmrc.gov.check-details.com) in the URL was the same, and if the abuse report dates for those reports were within 7 days of each other, and if the target across those URL reports was the same, then those URLs were considered to be involved in one attack.
- Phishers use a wide variety of URL construction methods, and we formulated additional rules to group URLs into attacks based on observed cases. The identification of attacks required a final round of manual examination to find additional batches of related URLs. For example, some phishers clearly generated multiple subdomains, programmatically, as part of one attack. 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.
- Due to the many use cases, other observers may independently arrive at slightly different numbers.

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 uses slightly different nomenclature and brand identification conventions. We normalized simple variations such as "PayPal" and "PayPal Inc." by combining them into a single brand, but did not collapse reports that identified targets discretely despite readily recognizable relationships between them. For example, WhatsApp is owned by Facebook; some reports might identify the target of an attack as "WhatsApp," while others identify the target as "Facebook." In those cases we counted the reports as referring to two different 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.[70, 71, 72, 73] 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*, 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 thirty five 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.
- Peter Cassidy and Foy Shiver of the Anti-Phishing Working Group.
- 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] Google Safe Browsing Transparency Report. https://transparencyreport.google.com/safe-browsing/overview

[2] A. Oest, P. Zhang, B. Wardman, E. Nunes, et al: "Sunrise to Sunset: Analyzing the End-to-end Life Cycle and Effectiveness of Phishing Attacks at Scale." Proceedings of the 29th USENIX Security Symposium, August 12–14, 2020. https://www.usenix.org/system/files/sec20-oest-sunrise.pdf

[3] This gap is almost exactly the same as a gap measured in 2014, when research found that once discovered, phishing sites stayed up for a median time of 8 hours and 42 minutes to 10 hours 6 minutes. G. Aaron and R. Rasmussen. Anti-Phishing Working Group: Global Phishing Survey 2H2014. https://docs.apwg.org//reports/APWG_Global_Phishing_Report_2H_2014.pdf

[4] 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

[5] 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/

[6] 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/

[7] Verisign Domain Name Industry Brief, 2Q2020. Volume 13, issue 3, August 2020. https://www.verisign.com/assets/domain-name-report-Q22020.pdf

[8] Netcraft: Cybercrime on Top-Level Domains. https://trends.netcraft.com/cybercrime/tlds

[9] G. Aaron and R. Rasmussen. Anti-Phishing Working Group: Global Phishing Survey series, 2008 to 2016. https://apwg.org/globalphishingsurvey

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

[11] Comparitech: "Which Countries Have the Worst (and best) Cybersecurity?" https://www.comparitech.com/blog/vpn-privacy/cybersecurity-by-country/

[12] Cisco Talos IP & Domain Reputation Center. https://talosintelligence.com/reputation_center/

[13] The Spamhaus Project. "The World's Most Abused TLDs." https://www.spamhaus.org/statistics/tlds/

[14] C. Larsen, Symantec: "The "Top 20: Shady Top-Level Domains". https://symantec-enterprise-blogs.security.com/blogs/feature-stories/top-20-shady-top-level-domains

[15] G. Aaron and R. Rasmussen, "Global Phishing Survey: Trends and Domain Name Use in 2016." Anti-Phishing Working Group. https://docs.apwg.org/reports/APWG

[16] 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/

[17] 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.

[18] 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

[19] Freenom Anti Abuse API. https://www.freenom.com/en/antiabuse_api.html

[20] G. Aaron and R. Rasmussen. Anti-Phishing Working Group: Global Phishing Survey series, 2008 to 2016. https://apwg.org/globalphishingsurvey

[21] "New gTLD phishing still tiny, but .XYZ sees most of it." DomainIncite.com, 27 May 2015. http://domainincite.com/18592-new-gtld-phishing-still-tiny-but-xyz-sees-most-of-it

[22] .XYZ blog: "XYZ says NO to abuse." 7 June 2016. https://gen.xyz/blog/antiabuse

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

[24] https://www.endurance.com/our-customers

[25] https://bgp.he.net/AS46606#_prefixes

[26] For example, see https://www.whois.com/whois/anazonbilling.com which was registered on 20 July 2020 and suspended by Tucows the same day, and https://www.whois.com/whois/womromsnd.com which was registered on 11 July 2020 and suspended by Tucows on 27 July 2020.

[27] https://bgp.he.net/AS213058#_asinfo

[28] The Spamhaus Project. SBL364207: 178.159.36.0/24, listed 23 September 2020; https://www.spamhaus.org/sbl/query/SBL364207 and SBL495124: 178.159.36.0/25, listed 23 September 2020; https://www.spamhaus.org/sbl/query/SBL495124

[29] See https://us.norton.com/internetsecurity-emerging-threats-what-is-bulletproof-hosting.html and https://www.spamhaus.org/news/article/792/bulletproof-hosting-theres-a-new-kid-in-town

[30] 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.

[31] 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. Available: 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)

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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

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

[38] 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

[39] 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

[40] 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/

[41] 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

[42] 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

[43] 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

[44] 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

[45] 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

[46] 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

[47] 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

48 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/

[49] 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.

[50] 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

[51] 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

[52] 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

[53] 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/

[54] 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

[55] 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

[56] 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

[57] 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

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

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

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

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

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

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

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

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

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

[67] 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

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

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

[70] 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

[71] 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

[72] 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

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