# Phishing Landscape 2020

## A Study of the Scope and Distribution of Phishing

*by*

Greg Aaron

Lyman Chapin

David Piscitello

Dr. Colin Strutt

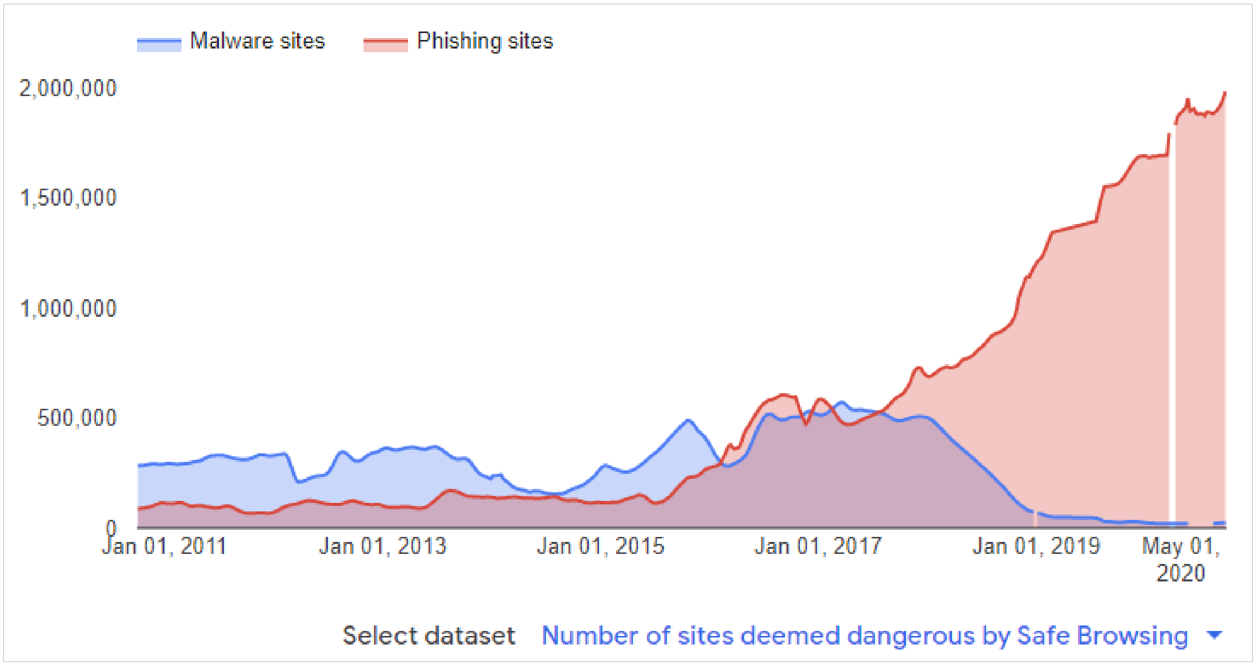
Interisle Consulting Group, LLC

*13 October 2020*

## 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:



*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.**
2. **Phishers themselves register more than half of the domain names on which phishing occurs**. We call these "malicious registrations."
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.
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.**
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.
6. **Sixty-five percent of maliciously registered domain names are used for phishing within five days of registration.**
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.**
8. **About 9% of phishing occurs at a small set of providers that offer subdomain services.**

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.
- **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.
- Phishing occurred in the IP spaces **of 2,169 different Autonomous Systems** (AS).
- 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.