Scam Pandemic: How Attackers Exploit Public Fear through Phishing

Marzieh Bitaab*, Haehyun Cho*, Adam Oest*†, Penghui Zhang*, Zhibo Sun*, Rana Pourmohamad*, Doowon Kim‡, Tiffany Bao*, Ruoyu Wang*, Yan Shoshitaishvili*, Adam Doupé* and Gail-Joon Ahn*§

*Arizona State University, †PayPal, Inc. ‡University of Tennessee, Knoxville, §Samsung Research
*{mbitaab, haehyun, aoest, penghui.zhang, zhibo.sun, rpourmoh, tbao, fishw, yans, doupe, gahn}@asu.edu
†doowon@utk.edu

***Abstract*—As the COVID-19 pandemic started triggering widespread lockdowns across the globe, cybercriminals did not hesitate to take advantage of users' increased usage of the Internet and their reliance on it. In this paper, we carry out a comprehensive measurement study of online social engineering attacks in the early months of the pandemic. By collecting, synthesizing, and analyzing DNS records, TLS certificates, phishing URLs, phishing website source code, phishing emails, web traffic to phishing websites, news articles, and government announcements, we track trends of phishing activity between January and May 2020 and seek to understand the key implications of the underlying trends.**

We find that phishing attack traffic in March and April 2020 skyrocketed up to 220% of its pre-COVID-19 rate, far exceeding typical seasonal spikes. Attackers exploited victims' uncertainty and fear related to the pandemic through a variety of highly targeted scams, including emerging scam types against which current defenses are not sufficient as well as traditional phishing which outpaced the ecosystem's collective response.

I. Introduction

The COVID-19 pandemic upended daily life across the globe and has led to unprecedented changes from two perspectives. First, the ensuing widespread lockdowns, travel restrictions, and telecommuting arrangements (working from home) have significantly increased users' reliance on online services. Second, continuous updates from news outlets and social media caused panic about the rapid spread and dangers of the disease [15]. Unfortunately, this increased usage of the Internet and the unstable emotions of its users has left the users vulnerable to online social engineering attacks such as scams and phishing [4]. For instance, attackers exploit users' fear to trick them into *acting now* instead of making an informed decision: For example, one COVID-19 phishing email exploits Internet users' fear by stating that *"this is the last set of test kits."* available for purchase. Besides fear, attackers also capitalize on their victims' generosity: Because many people desire to help others during major tragedies, scammers create fake donation campaigns as a lure to mount attacks.

Abundant news reports and government alerts about phishing attacks underscore the significance of anti-phishing systems [31]. However, such reports are generally anecdotal, and comprehensive studies on phishing (and other cybercrime) related to the pandemic are needed to inform society to better respond to these threats.

This need, combined with a lack of studies on the relationship between large societal shifts (such as the pandemic) and phishing attacks, motivates us to investigate the effect of COVID-19 on phishing trends, the effects of these changing trends on phishing victims, and possible defenses that can be implemented or enhanced to protect users in this dangerous online landscape. Specifically, in this paper, we seek to answer the following research questions:

- How has the COVID-19 situation affected trends in phishing attacks?
- How many victims have visited phishing websites related to the pandemic?
- What are the attackers exploiting?
- How can we improve anti-phishing systems to protect users and organizations from phishing threats that leverage massive global situations like COVID-19?

To answer the research questions, we collected a variety of datasets in the course of conducting our research: (1) we collected news articles and government announcements about phishing attacks related to the COVID-19 pandemic; (2) we gathered and monitored DNS records, TLS certificate transparency logs, and phishing website reports to measure how the pandemic has affected the Internet infrastructure; (3) we crawled the source code of COVID-19-themed phishing websites from among the reported URLs to explore novel types of phishing content; (4) by collaborating with a major financial services organization, we used a specialized network monitor to analyze trends in victim traffic to phishing websites and the volume of phishing reports by users of the organization. This gives us an unparalleled view from the organization's perspective; and (5) we collected COVID-19-related discussions from two large underground forums to understand cybercriminals' objectives and activities related to the pandemic.

We performed a multi-faceted analysis of the collected datasets. Through our analysis, we made several interesting findings about the early months of COVID-19:

- **Record-breaking attack volume.** We observed that traffic to phishing websites reached record levels in March

978-0-7381-3261-7/20/$31.00

and April 2020, with up to 2.2 times more users falling victim to phishing than in the preceding months. Cybersecurity warnings from governments and major industry organizations lagged behind these attacks.

- **Social engineering strategies.** During COVID-19, attackers exploited both users' altruism and self-interest. For example, we found attacks that impersonated the Centers for Disease Control and Prevention (CDC) and harvested user credentials and identities while making users believe they were making a donation. Conversely, myriad fraudulent storefronts pretended to sell personal protective equipment (PPE) or attempted to sell counterfeit goods such as fake COVID-19 testing kits.
- **Current defenses.** Traditional anti-phishing systems are primarily reactive in nature and, thus, they struggle to quickly protect users, at scale, in the face of novel types of phishing attacks. In addition, ecosystem defenses against non-phishing scams have a lesser degree of maturity.

Much to our surprise, despite historically being the most prevalent browser-based threat [14], phishing was not the most common threat among all COVID-19-related online attacks. In the first four months of 2020, we identified 467,323 COVID-19-related domain registrations; a curated whitelist indicated that just 0.16% (774) of these domains were benign [25]. Among all the registered domains, we found out that only 0.22% (1,047) of them appeared on phishing blacklists. Therefore, we concluded that phishing websites only represented *a small fraction* of malicious COVID-19 domains. As such, defenses against other types of scams are as important as anti-phishing defenses. To this end, we provide in our paper a taxonomy of other types of scams, such as fake storefronts or deceptive donation pages. We also recommend new ecosystem defenses to identify these scam websites and protect users from them as future work. The contributions of this paper are thus as follows:

- Our study clearly shows that attackers move quickly to develop novel types of attacks to exploit users' increased vulnerability during a crisis.
- This work is the first step in comprehensively investigating phishing attack trends as a result of COVID-19 and motivates standardized approaches to not only keep up with the agility of attackers but also help guide timely mitigations to protect users from sophisticated online scam threats.

II. Background

In social engineering attacks, attackers lure victims to disclose sensitive information. Phishing is a common type of social engineering attack in which criminals masquerade as trustworthy entities to take advantage of targeted victims. Attackers typically manipulate the victims to submit their credentials by exploiting their fear, curiosity, charitable spirit, or apprehension [40, 41]. As more routine tasks become digital, people increasingly rely on the Internet. The migration of tasks from traditional (e.g., paper-based) formats to online services has provided opportunities for cybercriminals to lure victims [17].

Cybercrime typically has three main components: (1) a victim who is the target of a cyber-attack, (2) a motive which is the criminal's incentive for committing the attack, and (3) a vulnerability or opportunity that enables the crime to take place [34]. When the online presence of users increases, the first two conditions will be met. Different principles affect the third factor for a successful attack, such as distraction, time pressure, compassion, and need [35]. When the aforementioned conditions are met, attackers usually ramp up their activities to maximize their success rate [38]. In the past, attackers have seen natural disasters as a prime opportunity to carry out social engineering. For example:

Ebola Virus Outbreak. The largest Ebola outbreak occurred in 2014 and lasted two years in west Africa. Although the Ebola virus did not spread worldwide, attackers targeted affected groups of people with phishing and scams. Barracuda Networks reported that 200,000 spam emails with Ebola news updates attempted to make people open malicious links, and 700,000 scam emails solicited donations to fictitious organizations [22].

Australia's Bushfire. During the Australian bushfire that happened in late 2019, attackers claimed to be from large organizations, the government, or popular charities to deceive people into donating money or providing sensitive information [12].

Unlike the disasters above, the COVID-19 pandemic has caused worldwide panic and, thus, miscreants have been exploiting the empathy and fear of people on a larger scale, in part by deploying scams and phishing websites with COVID-19-themed content [6].

III. Dataset

In this section, we discuss the datasets that we collected to enable us to investigate social engineering scams related to the pandemic in the remainder of the paper.Our datasets cover news reports and government announcements, domain names, TLS certificates, reported phishing URLs, reported phishing emails, and posts from underground forums. Table I shows an overview of these datasets.

A. Terminology

For the sake of brevity, throughout the paper, we use the term *corona-related* to refer to any data related to the COVID-19 pandemic.

To find corona-related news and domains in our datasets, we used keywords that include: *covid*, *covid-19*, and *coronavirus*, along with their permutations. Using regular expressions, we also considered keywords related to COVID-19 that use special characters or numbers, such as *c0-vi-d-19*.

B. Summary of The Dataset

We collected the following data for this study:

Data Content	Data Source	Date Range	Number of Samples
News, Government, and Companies announcements	Google News	01/01/20 – 05/12/20	756
Domain names	RiskIQ, Domaintools, WhoisDS	01/01/20 – 04/30/20	467,323
TLS certificates	Google Rocketeer CT log	01/01/20 – 04/30/20	33,596,126
Reported phishing URLs	APWG, OpenPhish	06/01/19 – 04/30/20	3,191,012
Source code of phishing websites	APWG	01/01/20 – 04/30/20	49,306
Reported phishing emails	Financial services provider	01/01/20 – 04/30/20	387,251
Victim web traffic to phishing websites	Financial services provider	01/01/20 – 04/30/20	Not disclosed
Posts from underground forums	Nulled.to, Cracked.to	01/25/20 – 05/06/20	3,530

TABLE I: An overview of our datasets.

News and government announcements. From media outlets, governments, and private companies, we automatically collected news and announcements about social engineering attacks related to the pandemic. We then filtered those that are relevant to phishing attacks that reference the pandemic. We searched Google News with the keywords {*corona, covid-19, scam, phishing*} to gather news from both government and non-government websites.

Domain names. To investigate changes in domain registration trends, we collected DNS records from three different sources: Domaintools [20], Whois Domain Search [3], and RiskIQ [31] to find corona-related domain names registered daily. Both Domaintools and RiskIQ provide filtered lists of registered corona-related domains, whereas Whois Domain Search provides all daily domain registrations which we then retrieved and scanned for corona-related domains.

TLS certificates. To find certificates issued to web sites with corona-related domain names and phishing websites related to the pandemic, we collected 144,590,199 TLS certificates using the Google Rocketeer CT log [13].

Reported phishing URLs. We collected phishing URLs submitted to OpenPhish [1] and the Anti-Phishing Working Group (APWG) [2].

Source code of phishing websites. We crawled the source code (i.e., page content) of 49,306 phishing websites between January 2020 and April 2020 (using the APWG URLs) to investigate corona-related phishing content and techniques.

Phishing emails and traffic to phishing websites. Between January and April 2020, we analyzed 387,251 phishing emails reported by users, and signals based on victim traffic to phishing websites, by collaborating with an organization commonly targeted by phishing.

Underground forums. To understand shifts in criminals' activities amid the COVID-19 pandemic, we crawled corona-related discussions in two popular underground forums: *Nulled.to* and *Cracked.to*, which have more than 2.8 million and 1.1 million registered members, respectively.

IV. Measurement Results

In this section, we first discuss overall measurement results based on the combination of our datasets, and we then present detailed findings from each individual dataset. In Section IV-B, we show trends in corona-related DNS records and reported phishing websites. Next, in Section IV-C, we study news from both private and government news outlets to understand the perceived importance of COVID-19 themed phishing and scams. We follow with an analysis of real victim traffic to phishing websites in Section IV-D. Then, we categorize and explain various types of COVID-19-themed phishing websites in Section IV-E. In Section IV-F, we use the underground forum data to characterize the corona-related topics discussed by criminals.

A. General Findings

a) Record-breaking victims: Even though in our dataset, phishing attacks leveraging the pandemic seem to be negligible compared to traditional attacks, there was a record-breaking number of overall phishing victims during this period (described fully in Section IV-D). We believe this is because phishing lures effectively exploited their victims' pandemic-related concerns, and that attackers caught victims off guard with high-quality phishing websites (Section IV-E). We observed that the number of news reports and government announcements regarding corona-related phishing attacks increased rapidly from March 2020 (Section IV-C). However, there were still many victims after that, which implies that typical anti-phishing systems' reactive mitigation strategy is not enough to protect users from a surge of novel phishing attacks.

b) No ecosystem defenses for non-phishing scams: We found 467,323 new corona-related domain names and 17,699 new certificates issued to corona-related domains from our observation period. Only 0.22% (1,047) of these new corona-related domains were reported to phishing blacklists (Section IV-B and Section IV-D). Also, one curated list suggests that only 774 of the domains hosted legitimate websites [25]. Furthermore, a recent FTC report showed that 54,813 corona-related scams reported across the U.S. from January led to $40.13M in loss due to fraud [43]. Based on this report, online shopping is the most commonly reported scam. This also implies that phishing is just one type of many corona-related attacks. Other corona-related domains can be used for different fraudulent purposes, such as non-phishing scam websites or e-mail spam. Cybercriminals exploit the ecosystem's lack of defenses against such non-phishing scams.

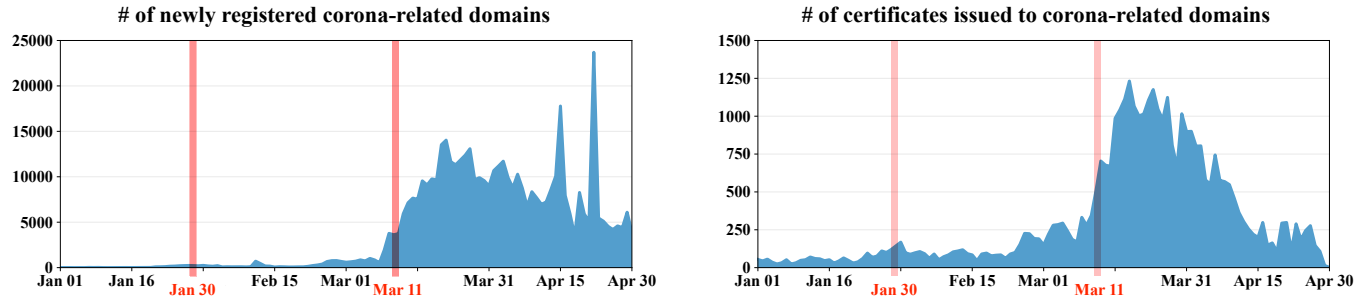


Fig. 1: The number of newly issued certificates used for corona-related domains and the number of newly registered corona-related domains.

B. Domain Names and Certificates

We provide an overview of corona-related websites and phishing sites based on domain names and URLs.

As Figure 1 illustrates, we found that the number of corona-related domain names started increasing significantly from early March 2020. On average, 155 corona-related domain names were registered every day before March 2020; however, this number increased to 7,453 in March and April. With the increase in new corona-related websites, it is difficult to distinguish between legitimate and malicious websites. As shown in Figure 1, the number of certificates issued to corona-related domain names increases starting in February 2020 and peaks in March 2020.

We further analyzed 72 certificates of corona-related HTTPS phishing websites as in Table II. Table III shows that CAs which issued certificates to corona-related phishing websites with the number of certificates and revoked certificates. Except for GoDaddy and Sectigo, the other CAs use the Automatic Certificate Management Environment protocol, allowing attackers to obtain TLS certificates easily. Through May, only GoDaddy and cPanel revoked certificates issued to phishing websites (five and two certificates, respectively).

We queried the collected DNS records against the APWG, OpenPhish, GSB, and RiskIQ blacklists to estimate how long it takes for a corona-related phishing URL to be detected after launch. Table IV shows the number of intersections between collected corona-related DNS records and each phishing blacklist. We calculate the *average gap* between the registration date and date of each report. From $467,323$ DNS records we collected, 110 domains were reported to APWG, on average 3.6 days after registration, and 72 domains were reported to OpenPhish 4.8 days after registration. As GSB and RiskIQ do not provide the date that blacklisted domains were reported, we were unable to calculate the average gap. Also, we found that 110 domains reported to the APWG were newly registered in 2020; however the other 44% of reported domains existed before January 1^{st} 2020. Table II shows that 198 domains were reported to the APWG from January 2020 to April 2020.

Month	# of Reported URLs	# of HTTPS Domains
January 2020	0	0
February 2020	5	1
March 2020	171	37
April 2020	140	34
Total	316	72

# of unique domains among the reported URLs = 198.

TABLE II: The number of corona-related URLs reported to the APWG per month and the number of HTTPS domains among the reported URLs.

CAs	ACME	# of Certs.	# of Revoked Certs.
Let's Encrypt	✓	31	0
cPanel	✓	22	2
Go Daddy	✗	8	5
Cloudflare	✓	6	0
Sectigo	✗	6	0

TABLE III: CAs, the number of certificates, and the number of revoked certificates that were used for phishing.

C. Public Phishing Guidance

To study the effect of the COVID-19 pandemic on corona-related social engineering attacks, we examined how news outlets, governments, and large companies have provided guidance or warnings against phishing and scam attacks. To this end, we collected daily news reports and government announcements starting from January 2020.

Figure 2 shows the number of corona-related news reports and government and company announcements. The news reports began to reference corona-related scams on January 30^{th}, stating that several phishing campaigns were sending corona-related emails containing malware. The first official U.S. government announcement about corona-related scams was made on February 4^{th} by the U.S. Securities and Exchange Commission (SEC), which warned people about cybercriminals trying to leverage the COVID-19 situation.

After the original SEC announcement, the number of news reports related to corona-related scams increased rapidly. How-

Blacklists	# Intersections	Avg. Gap (days)
APWG	110	3.6
OpenPhish	72	4.8
RiskIQ-BL	316	N/A
GSB	833	N/A

# of unique domains among the blacklisted domains = 1,047.

TABLE IV: Number of intersections and average gap for anti-phishing entities' blacklists

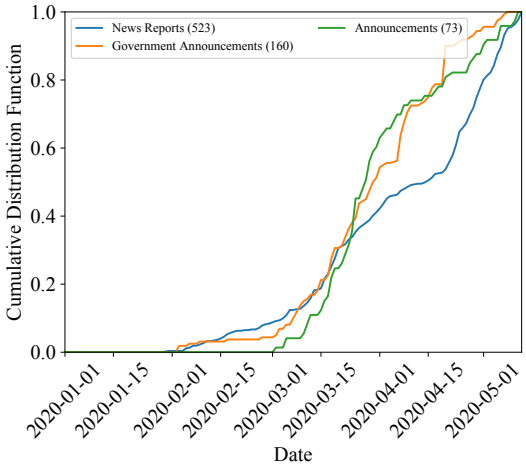


Fig. 2: The number of news and government announcements about corona-related phishing attacks.

ever, the U.S. government did not make many announcements about corona-related phishing attacks until the beginning of March 2020, when several additional government departments posted alerts. As shown in Figure 2, companies started to directly address corona-related scams at the beginning of March, with an increasing trend thereafter. The rapid growth of news from different sources motivates us to further investigate these scams.

As illustrated in Figure 3, most of the U.S. government announcements are from FTC, followed by the Department of Justice, while states only published one warning on their official websites (shown as "other"). The announcements warn people about scams and the threat of the theft of sensitive information such as the Social Security Number (SSN), and typically include detailed guidance for how members of the public can protect themselves from fraud.

D. Phishing Trends

As shown in Figure 4, the number of phishing hostnames reported to two major clearinghouses of phishing URLs did *not* increase significantly during the COVID-19 outbreak. However, hostname counts alone fail to accurately reflect the damage caused by phishing attacks, as certain high-impact websites may receive substantially more traffic than others as a result of their increased spamming activity or the ability to evade defenses [29].

To deepen our insight into phishing trends during the crisis, we collaborated with a financial service organization that is

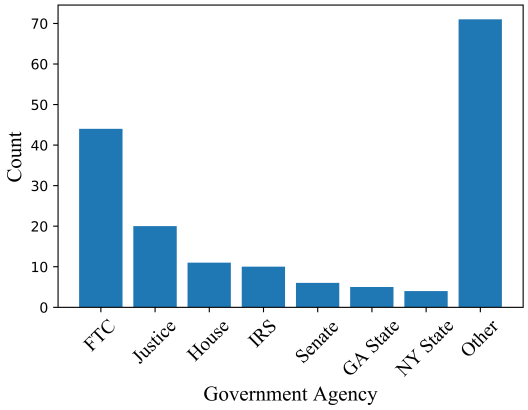


Fig. 3: Government announcements regarding corona-related phishing and scams, grouped by government agency.

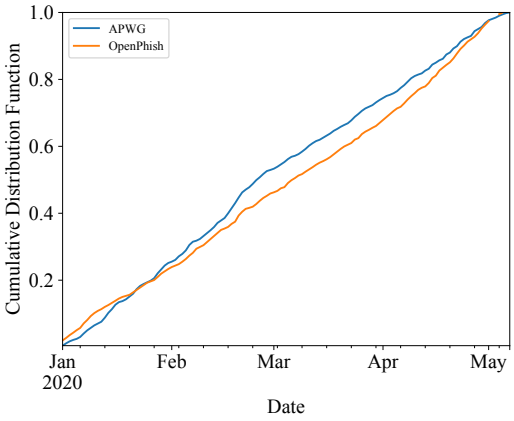
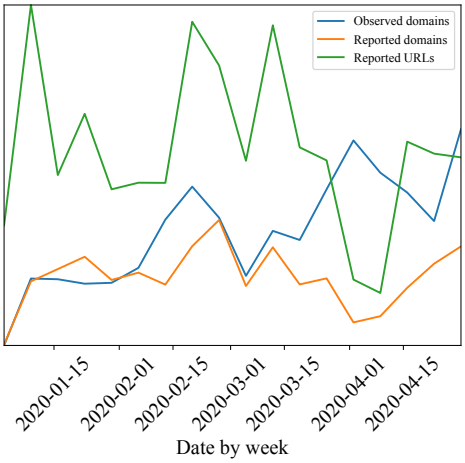


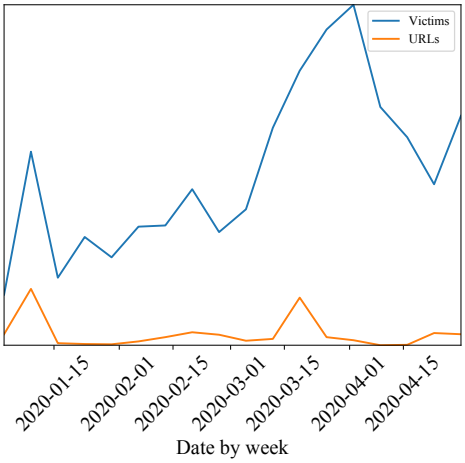
Fig. 4: CDF of unique phishing hostnames reported to two major anti-phishing entities.

commonly targeted by phishers and analyzed two additional datasets: (1) traffic to phishing websites by targeted victims and (2) phishing emails reported by users. The phishing website traffic was collected using a recently proposed network monitoring approach [29] which passively measures victim visits to live phishing websites based on signals (referrer headers as well as third-party resources embedded on phishing websites) detectable by the organization. Specifically, it first analyzes web traffic logs to find events of interest. Then, after filtering out benign events, it looks for events correlated with known phishing URLs. Further analysis of the events enables us to determine how many victims have fallen for corona-related phishing websites[1]. As shown in Figure 5, the network monitor recorded a surge in phishing victims from late March; attack volume remained elevated throughout April. Overall, the total number of observed phishing victims in March and April was 2.207 and 1.706 times higher than in February, respectively, and also 2.165 and 1.674 times higher than in

[1]The network events recorded by this approach have a high probability of being linked to victims successfully fooled by phishers, and have been de-duplicated to reflect individual sessions.



(a) Number of observed/reported phishing domains and URLs.



(b) Victim visits to phishing websites.

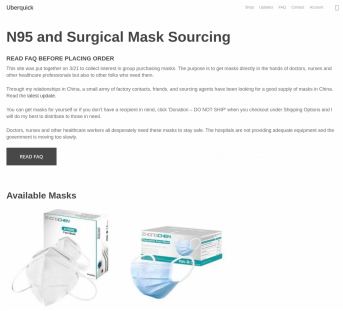
Fig. 5: Number of potential phishing victims identified by our network monitor, compared to the number of corresponding unique phishing URLs. Victim counts increased substantially in March and April despite a lack of a significant increase in URLs.

January, when the organization typically sees elevated phishing volume due to holiday shopping.

Phishing reports that users sent to the organization validate our observed increase in victim traffic, as a sustained rise in reporting is directly linked to an increase in spamming [23]. 1.06 times more emails were reported to the organization in March, and 2.64 times more in April, compared to the number of reports in February.

Interestingly, within this dataset, only 0.51% of phishing websites had corona-related content. Similarly, 0.02% of emails had COVID-19 keywords in the title or body, while 0.43% had such keywords in the sending email address.

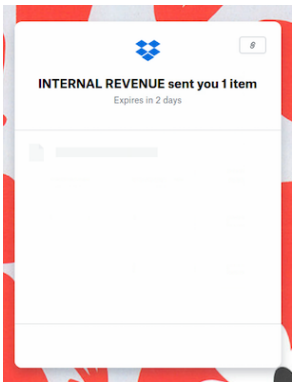
The World Health Organization (WHO) declared a global pandemic on March 11^{th}. Even though there were not many corona-related phishing websites, the number of victims increased dramatically after the WHO's pandemic announcement, as shown in Figure 5b. Moreover, the government announcements increased *after* the victim counts reached their peak in March. We suspect that not many corona-related phishing websites could take advantage of many victims within a short period after March 11^{th}.



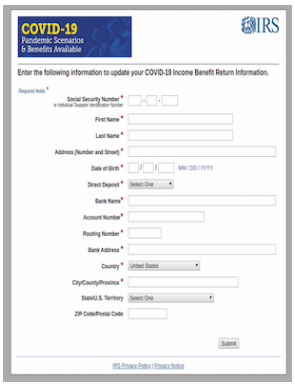
(a) Fraudulent PPE store.

(b) Free shipping fraudulent.

Fig. 6: Fraudulent website selling PPE and free shipping fraudulent store.



(a) Dropbox themed.



(b) IRS themed.

Fig. 7: Phishing websites exploiting IRS.

We conclude that through an increase in spamming activity against a larger attack surface, the pandemic led to record numbers of phishing victims. However, in the case of this organization, phishing attacks that leveraged COVID-19 as a lure were negligible compared to traditional attacks that simply impersonate the brand. We note that the latter trend may be skewed by the brand's industry sector, however.

E. COVID-19 Themed Phishing

We crawled and retrieved the source code of 49,306 phishing URLs from the APWG feed. We further considered corona-related scam websites reported to *scam.directory*. By analyzing their source code and page content, we found 255 unique phishing and scam websites with corona-related content.

Donation-Themed Phishing. Some phishing websites steal sensitive information while making users think they are making a donation to a charity. For example, we identified one prominent phishing website which copied the look-and-feel of a donation portal run by a major organization. The phishing

website deceptively informed visitors that they were making a donation to the "CDC Response to CoronaVirus" through a registered charity, and provided detailed information about this (legitimate but unrelated) charity. Victims successfully fooled by the attack—which we observed across several distinct domains—would think they were making a small donation to support the charity. However, the phishing website would instead steal the victims' account credentials and credit card numbers, rather than processing actual donation payments.

PPE Sale. Personal protective equipment (PPE), such as face masks and gloves, is in high demand when the public and healthcare workers try to protect themselves from communicable diseases. In the early months of the pandemic, such equipment was also in short supply (and/or excessively priced) at major online retailers such as Amazon, eBay, or Walmart. Therefore, people in urgent need of PPE may turn to other, less trustworthy sources. To exploit the high demand, attackers designed fake shopping websites that sell PPE; an example is shown in Figure 6a.

Several such websites embedded up-to-date corona-related information (e.g., COVID-19 statistics) in an effort to appear more legitimate. They lure people who require such equipment and then either steal their credentials, steal their money (without shipping any items), or sell them low-quality PPE.

Fraudulent Online Shopping Websites. Fraudulent websites try to keep up with the look and feel of legitimate websites. As more and more legitimate organizations started to inform their customers about pandemic-related matters such as policy updates, new features, and COVID-19 statistics, attackers also included such information on their websites. Similarly, some attackers advertised misleading "free shipping" offers on their fake shopping websites. Free shipping offers increase online sales and help attract visitors [32]. Figure 6b shows an example of such website.

Exploiting Corona-related Events. Phishers not only generate corona-related phishing websites, but they also exploit other events related to the pandemic. For example, to help address widespread financial hardship, the U.S. government offered stimulus funds by either direct deposit or a paper check. However, different groups of people received payments at different times. When people who received a check and shared this on social media, others might start to worry about if and when they also could receive their funds. Phishers were quick to disguise themselves as the IRS to steal the personal information of people looking for the status of their stimulus payments. For example, in Figure 7, IRS phishing websites steal Dropbox credentials or SSNs from users amid the pandemic. The text "Expires in 2 days" in the phishing website from Figure 7a conveys urgency so that visitors are more willing to open it. Figure 7b acquires users' SSN by declaring that they need such Personally Identifiable Information (PII) to process stimulus payments.

F. COVID-19 in Underground Forums

As underground forums are a key rendezvous point for cybercriminals [36, 37], we studied data from such forums to answer two questions: (1) Are corona-related topics popular in underground forums? (2) What do members discuss about corona-related topics?

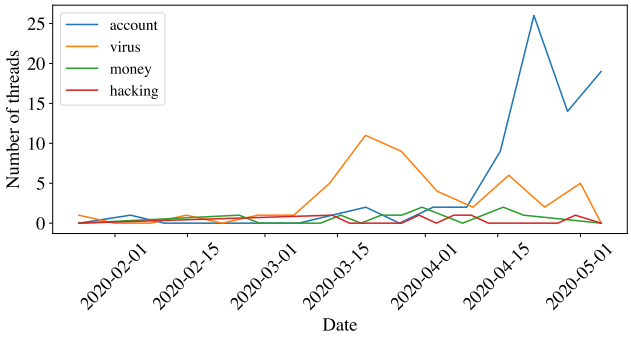


Fig. 8: Topics discussed in underground forums.

Content Category	Sub-category	% of discussions
Virus	General Discussion	31.9%
	Virus Protection	2.1%
Account	Game Accounts	20.6%
	VPN Accounts	13.5%
	Video Site Accounts	9.9%
	Music Site Accounts	5.7%
	Porn Site Accounts	5.0%
Money	General Discussion	3.5%
	Making Money via E-whoring	4.3%
Hacking	Cracking	1.4%
	Account Checkers	1.4%
	Phishing	0.7%

TABLE V: Coronavirus-related discussion in underground forums.

First, we measured the number of threads and posts within a one-week sliding window, as these are indicative of new discussion topics and members' overall levels of activity, respectively. We found 2,913 members engaging in 144 coronavirus-related threads among 3,530 posts. Then, we manually analyzed all of these threads and categorized them into four topics: (1) account (54.7%, e.g., selling compromised accounts); (2) virus (34%, e.g., what the coronavirus is); (3) money (7.8%, e.g., how to make money during the pandemic); (4) hacking (3.5%, e.g., setting up COVID-19 related phishing sites). Members in underground forums predominantly discussed compromised accounts and the coronavirus itself, as shown in Table V.

Figure 8 shows the number of different topics discussed in underground forums over time. At the beginning of the pandemic, people mainly discussed and shared information on the coronavirus; discussions shifted to compromised accounts from April 1^{st}, and there is a significant peak at April 15^{th}. 58.3% of the new threads about selling compromised accounts were started during this period. Because the total number of observed phishing victims peaked between March 1^{st} and April 15^{th} (as shown in Figure 5), we suspect that the attackers posted about stolen accounts after they successfully launched their attacks.

V. Discussion

Our dataset and analysis represent a current snapshot of specific corona-related cybercrime and shed light on how attackers exploited the COVID-19 pandemic in the early months of the outbreak. The pandemic upended daily lives across the globe and, consequently, resulted in unprecedentedly rapid changes across the digital world. These changes open up a new yet important set of challenges for website owners, users, governments, and security researchers to be able to adapt accordingly.

Our results demonstrate that attackers remain several steps ahead of typical modern anti-phishing defenses and will take advantage of a global crisis to directly harm online users. It is of great importance to collaboratively deploy anti-phishing systems that can better adapt to changes in the ecosystem, to narrow the attack window available to phishers and perhaps go further to offer proactive defenses.

From the corona-related domain perspective, 0.16% of the domains in our dataset were known to be benign, and 0.22% of the domains were known to be malicious. We discovered that many of the other unknown domains are used for non-phishing scams such as fake storefronts, which can harm users, yet remain out of the reach of traditional anti-phishing defenses.

From our deep dive into the content of corona-related phishing websites, we noted the importance of human factors and how attackers exploit individuals' pandemic-driven wants and needs. Hence, it is critical to raise awareness so that users better understand social engineering attacks and have access to the appropriate resources to protect themselves when technical mitigations fail to do so.

This paper is the first step toward investigating phishing attacks, trends, and consequences amid a global pandemic through multi-faceted measurements. We will continue our measurements to observe the effect on phishing and scams once the crisis subsides. To enable early characterization and detection of emerging types of phishing attacks, future research should also focus on developing real-time monitoring approaches to reliably conduct comprehensive and holistic measurements of phishing in an automated way.

VI. Related work

A. Mitigating Phishing

Phishing attacks, the most prevalent web-based threat, have caused substantial damage to victims [18, 39]. To detect and mitigate phishing attacks, much research effort focused on analyzing phishing URLs [9, 10, 19, 21] and website content [8, 11, 44, 46, 48].

Sahingoz et al. [33] proposed a method to detect phishing websites based on the URL. They extract Natural Language Processing (NLP) based features such as word counts, word length, and TLD to train a random forest classifier capable of detecting phishing URLs. While they show their proposed method outperforms previous models, adversaries can bypass URL classification algorithms [5].

As content-based approaches are proved to have a better performance than URL based methods [28], most of the new methods focus on analyzing the page content and search engine metadata [45]. Ardi et al. proposed a content-based method for detecting phishing websites on demand. Their method leverages the Document Object Model (DOM) of a webpage to detect phish. This method breaks the DOM tree into chunks and computes the hash of each chunk. If the number of chunks that matches the hashed blacklist content is greater than a threshold, it flags the webpage as phishing. This method provides good performance and zero false positive rate [7]. However, an attacker can simply use homographs (look-alike characters) or replace the content with images to bypass the detection method [7].

Google Safe Browsing [42] and Microsoft SmartScreen [24] are currently deployed mitigation systems across major web browsers for protecting users from phishing attacks. They detect a phishing website based on a URL blacklist or a heuristic classifier. As the only mitigation against phishing attacks is the blacklists, if blacklists do not offer adequate protection, users will be exposed to phishing threats without any protection [26].

B. Limitations of Current Anti-phishing Systems

Several research works revealed the limitations of blacklist-based anti-phishing approaches [16, 27, 28, 30]. Han et al. [16] monitored the lifecycle of phishing websites from the creation of them by using a honeypot web server. Oest et al. [27] conducted an empirical study on the blacklisting coverage and response time. In this work they propose the PhishFarm framework. PhishFarm first deploys different phishing websites, then it reports the deployed websites and waits for the anti-phishing entities to blacklist the reported websites. By using this framework, they test the resilience of anti-phishing entities. Peng et al. [30] measured the performance of the VirusTotal and its third-party vendors with their own phishing sites. They use a similar method as PhishFarm to study the reliance and robustness of VirusTotal and its 68 third-party vendors. In this work, they set up their own phishing websites while monitoring the incoming traffic and the VirusTotal labeling process. All of them suggested a significantly faster blacklist response time for protecting users more effectively. In addition, cybercriminals continue to use evasion techniques to make phishing websites remain online so that it is accessible to victim users for a long time [28, 29, 47]. These studies imply that, as far as the standard anti-phishing defense is operated in a reactive manner, phishing attacks will still remain a significant threat to Internet users.

VII. Conclusion

Amid widespread panic and uncertainty, the increased usage of online services during the COVID-19 pandemic resulted in an early spike of online social engineering attacks. To gain insight into how the pandemic changed trends in phishing and scams and how attackers took advantage of this situation, we synthesized multiple sources of web-related data. Our analysis revealed the potential for new ecosystem defenses and enhanced collaboration among entities to support a more

timely and effective ecosystem strategy to combat surges in phishing volume and sudden shifts in the nature of attacks.

References

[1] https://openphish.com, OpenPhish.

[2] https://apwg.org/ecx/, The APWG eCrime Exchange (eCX).

[3] https://whoisds.com/, Whois Domain Search.

[4] Sherly Abraham and InduShobha Chengalur-Smith. An overview of social engineering malware: Trends, tactics, and implications. *Technology in Society*, 32(3):183–196, 2010.

[5] Ahmed AlEroud and George Karabatis. Bypassing detection of url-based phishing attacks using generative adversarial deep neural networks. In *Proceedings of the Sixth International Workshop on Security and Privacy Analytics*, pages 53–60, 2020.

[6] Tony Anscombe. Beware scams exploiting coronavirus fears, 2020. welivesecurity.com/2020/03/13/beware-scams-exploiting-coronavirus-fears/.

[7] Calvin Ardi and John Heidemann. Auntietuna: Personalized content-based phishing detection. In *NDSS Usable Security Workshop (USEC)*, 2016.

[8] Leyla Bilge, Engin Kirda, Christopher Kruegel, and Marco Balduzzi. Exposure: Finding malicious domains using passive dns analysis. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2011.

[9] Sun Bin, Wen Qiaoyan, and Liang Xiaoying. A dns based anti-phishing approach. In *2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing*, volume 2, pages 262–265. IEEE, 2010.

[10] Aaron Blum, Brad Wardman, Thamar Solorio, and Gary Warner. Lexical feature based phishing url detection using online learning. In *Proceedings of the 3rd ACM Workshop on Artificial Intelligence and Security*, pages 54–60. ACM, 2010.

[11] Davide Canali, Davide Balzarotti, and Aurélien Francillon. The role of web hosting providers in detecting compromised websites. In *Proceedings of the 22nd international conference on World Wide Web*, pages 177–188. ACM, 2013.

[12] Australian Competition and Consumer Commission (ACCC). Bushfires and scams, 2020. https://www.scamwatch.gov.au/news-alerts/bushfires-and-scams-0.

[13] Graham Edgecombe. Google rocketeer certificate-transparency, 2020. https://ct.grahamedgecombe.com/logs/3.

[14] Google. Google safe browsing transparency report. https://safebrowsing.google.com/.

[15] Jeff Gorter. Impact of the coronavirus on eaps: Managing the fear of communicable disease. *Journal of Employee Assistance*, 2020.

[16] Xiao Han, Nizar Kheir, and Davide Balzarotti. Phisheye: Live monitoring of sandboxed phishing kits. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 1402–1413. ACM, 2016.

[17] Hiscox. The hiscox cyber readiness report 2019, 2019. https://www.hiscox.co.uk/cyberreadiness.

[18] Grant Ho, Asaf Cidon, Lior Gavish, Marco Schweighauser, Vern Paxson, Stefan Savage, Geoffrey M Voelker, and David Wagner. Detecting and characterizing lateral phishing at scale. In *Proceedings of the 28th USENIX Security Symposium*, pages 1273–1290, 2019.

[19] Huajun Huang, Liang Qian, and Yaojun Wang. A svm-based technique to detect phishing urls. *Information Technology Journal*, 11(7):921–925, 2012.

[20] Jackie Abrams. Free COVID-19 Threat List - Domain Risk Assessments for Coronavirus Threats. https://www.domaintools.com/resources/blog/free-covid-19-threat-list-domain-risk-assessments-for-coronavirus-threats, 2020.

[21] Mahmoud Khonji, Andrew Jones, and Youssef Iraqi. A novel phishing classification based on url features. In *2011 IEEE GCC conference and exhibition (GCC)*, pages 221–224. IEEE, 2011.

[22] Maria Korolov. Scammers move from ebola phishing to fundraising, 2014. https://www.csoonline.com/article/2848481/scammers-move-from-ebola-phishing-to-fundraising.html.

[23] Neil Kumaran. Protecting businesses against cyber threats during covid-19 and beyond, 2020. https://cloud.google.com/blog/products/identity-security/protecting-against-cyber-threats-during-covid-19-and-beyond.

[24] Windows defender smartscreen. 2019. https://github.com/MicrosoftDocs/windows-itpro-docs/blob/public/windows/security/threat-protection/windows-defender-smartscreen/windows-defender-smartscreen-overview.md.

[25] MISP. Open source threat intelligence platform & open standards for threat information sharing, 2020. https://github.com/MISP/misp-warninglists, https://github.com/krassi/covid19-related.

[26] NSS Labs. Nss labs conducts first cross-platform test of leading web browsers, Oct 2017. https://www.nsslabs.com/company/news/press-releases/nss-labs-conducts-first-cross-platform-test-of-leading-web-browsers/.

[27] Adam Oest, Yeganeh Safaei, Adam Doupé, Gail-Joon Ahn, Brad Wardman, and Kevin 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)*, pages 1344–1361. IEEE, 2019.

[28] Adam Oest, Yeganeh Safei, Adam Doupé, Gail-Joon Ahn, Brad Wardman, and Gary Warner. Inside a phisher's mind: Understanding the anti-phishing ecosystem through phishing kit analysis. In *2018 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2018.

[29] Adam Oest, Penghui Zhang, Brad Wardman, Eric Nunes, Jakub Burgis, Ali Zand, Kurt Thomas, Adam Doupé, and Gail-Joon Ahn. Sunrise to sunset: Analyzing the end-to-end life cycle and effectiveness of phishing attacks at scale. In *29th USENIX Security Symposium (USENIX Security 20)*, 2020.

[30] Peng Peng, Limin Yang, Linhai Song, and Gang Wang. Opening the blackbox of virustotal: Analyzing online phishing scan engines. In *Proceedings of the 2019 conference on Internet measurement (IMC)*. ACM, 2019.

[31] Team RiskIQ. The (many) benefits of offering free shipping. 2020. https://www.riskiq.com/blog/analyst/covid19-cybercrime-update/.

[32] ARMANDO ROGGIO. The (many) benefits of offering free shipping. 2015. https://www.practicalecommerce.com/The-Many-Benefits-of-Offering-Free-Shipping.

[33] Ozgur Koray Sahingoz, Ebubekir Buber, Onder Demir, and Banu Diri. Machine learning based phishing detection from urls. *Expert Systems with Applications*, 117:345–357, 2019.

[34] Debra Littlejohn Shinder and Michael Cross. *Scene of the Cybercrime*. Elsevier, 2008.

[35] Frank Stajano and Paul Wilson. Understanding scam victims: seven principles for systems security. *Communications of the ACM*, 54(3):70–75, 2011.

[36] Zhibo Sun, Adam Oest, Penghui Zhang, Carlos E Rubio-Medrano, Tiffany Bao, Ruoyu Wang, Ziming Zhao, Yan Shoshitaishvili, Adam Doupé, and Gail-Joon Ahn. Having your cake and eating it: An analysis of concession-abuse-as-a-service. In *Proceedings of the 30th USENIX Security Symposium (USENIX)*, 2021.

[37] Zhibo Sun, Carlos E. Rubio-Medrano, Ziming Zhao, Tiffany Bao, Adam Doupé, and Gail-Joon Ahn. Understanding and Detecting Private Interactions in Underground Forums. In *Proceedings of the Ninth ACM Conference on Data and Application Security and Privacy (CODASPY)*. ACM, 2019.

[38] Ken Tysiac. How cybercriminals prey on victims of natural disasters, 2018. https://www.journalofaccountancy.com/news/2018/sep/cyber-criminals-prey-on-natural-disaster-victims-201819720.html.

[39] Amber van der Heijden and Luca Allodi. Cognitive triaging of phishing attacks. In *Proceedings of the 28th USENIX Security Symposium*, 2019.

[40] Javier Vargas, Alejandro Correa Bahnsen, Sergio Villegas, and Daniel Ingevaldson. Knowing your enemies: Leveraging data analysis to expose phishing patterns against a major us financial institution. In *2016 APWG Symposium on Electronic Crime Research (eCrime)*, pages 1–10. IEEE, 2016.

[41] Arun Vishwanath, Tejaswini Herath, Rui Chen, Jingguo Wang, and H Raghav Rao. Why do people get phished? testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems*, 51(3):576–586, 2011.

[42] Colin Whittaker, Brian Ryner, and Marria Nazif. Large-scale automatic classification of phishing pages. In *Proceedings of the 28th Network and Distributed System Security Symposium (NDSS)*, 2010.

[43] Paul Witt. Covid-19 scam reports, by the numbers, 2020. https://www.ftc.gov/system/files/attachments/coronavirus-covid-19-consumer-complaint-data/covid-19-daily-public-complaints.pdf.

[44] Min Wu, Robert C Miller, and Greg Little. Web wallet: preventing phishing attacks by revealing user intentions. In *Proceedings of the*

second symposium on Usable privacy and security, pages 102–113. ACM, 2006.

[45] Guang Xiang, Jason Hong, Carolyn P Rose, and Lorrie Cranor. Cantina+ a feature-rich machine learning framework for detecting phishing web sites. *ACM Transactions on Information and System Security (TISSEC)*, 14(2):1–28, 2011.

[46] Haijun Zhang, Gang Liu, Tommy WS Chow, and Wenyin Liu. Textual and visual content-based anti-phishing: a bayesian approach. *IEEE Transactions on Neural Networks*, 22(10):1532–1546, 2011.

[47] Penghui Zhang, Adam Oest, Haehyun Cho, Zhibo Sun, RC Johnson, Brad Wardman, Shaown Sarker, Alexandros Kpravelos, Tiffany Bao, Ruoyu Wang, Yan Shoshitaishvili, Adam Doupé, and Gail-Joon Ahn. CrawlPhish: Large-scale Analysis of Client-side Cloaking Techniques in Phishing. In *Proceedings of the 42nd IEEE Symposium on Security and Privacy (Oakland)*, San Francisco, CA, May 2021.

[48] Yue Zhang, Jason I Hong, and Lorrie F Cranor. Cantina: a content-based approach to detecting phishing web sites. In *Proceedings of the 16th international conference on World Wide Web*, pages 639–648. ACM, 2007.