Sunrise to Sunset: Analyzing the End-to-end Life Cycle and Effectiveness of Phishing Attacks at Scale

Adam Oest and Penghui Zhang, *Arizona State University;* Brad Wardman, Eric Nunes, and Jakub Burgis, *PayPal;* Ali Zand and Kurt Thomas, *Google;* Adam Doupé, *Arizona State University;* Gail-Joon Ahn, *Arizona State University, Samsung Research*

https://www.usenix.org/conference/usenixsecurity20/presentation/oest-sunrise

This paper is included in the Proceedings of the 29th USENIX Security Symposium.

August 12–14, 2020

978-1-939133-17-5

Open access to the Proceedings of the 29th USENIX Security Symposium is sponsored by USENIX.

Sunrise to Sunset:
Analyzing the End-to-end Life Cycle and Effectiveness of Phishing Attacks at Scale

Adam Oest[*], Penghui Zhang[*], Brad Wardman[†],
Eric Nunes[†], Jakub Burgis[†], Ali Zand[‡], Kurt Thomas[‡], Adam Doupé[*], and Gail-Joon Ahn[*,§]

[*]Arizona State University, [†]PayPal, Inc., [‡]Google, Inc., [§]Samsung Research
[*]{aoest, pzhang57, doupe, gahn}@asu.edu, [†]{bwardman, enunes, jburgis}@paypal.com,
[‡]{zand, kurtthomas}@google.com

Abstract

Despite an extensive anti-phishing ecosystem, phishing attacks continue to capitalize on gaps in detection to reach a significant volume of daily victims. In this paper, we isolate and identify these detection gaps by measuring the end-to-end life cycle of large-scale phishing attacks. We develop a unique framework—*Golden Hour*—that allows us to passively measure victim traffic to phishing pages while proactively protecting tens of thousands of accounts in the process. Over a one year period, our network monitor recorded 4.8 million victims who visited phishing pages, excluding crawler traffic. We use these events and related data sources to dissect phishing campaigns: from the time they first come online, to email distribution, to visitor traffic, to ecosystem detection, and finally to account compromise. We find the average campaign from start to the last victim takes just 21 hours. At least 7.42% of visitors supply their credentials and ultimately experience a compromise and subsequent fraudulent transaction. Furthermore, a small collection of highly successful campaigns are responsible for 89.13% of victims. Based on our findings, we outline potential opportunities to respond to these sophisticated attacks.

1 Introduction

Phishing attacks target millions of Internet users each year, resulting in sensitive data exposures, financial fraud, and identity theft [43, 62]. These attacks also harm the reputation of targeted brands as well as incur collateral damage to the broader ecosystem and user trust. Modern phishing attacks fall into two general categories: *spearphishing*, where attackers target specific high-value individuals or groups [30, 32], and *large scale* attacks, where attackers target a broad range of potential victims to profit through volume [57]. In this work, we focus on the latter.

Prior research has shown that large scale phishing lures have a low click-through rate (5-8%) [58] and that the likelihood that targeted users will hand over credentials to attackers is similarly low (9%) [29]. Yet, the volume of observed phishing attacks shows no signs of subsiding [3, 4]. Furthermore, social engineering techniques such as phishing remain one of the primary stepping stones to even more harmful scams [21].

In an adversarial race—fueled in part by the underground economy [59]—phishers collectively seek to stay one step ahead of the security community through a myriad of evasion techniques [52]. Recent work has shown how cloaking and related strategies significantly delay browser-based phishing detection and warnings—a defense layer adopted by every major browser [51]. However, the implications of such delays on the success of each attack are not yet well-understood, nor is the precise window of opportunity available to attackers between the launch and detection of their phishing websites.

In this paper, we present a longitudinal, end-to-end analysis of the progression of modern phishing attacks, from the time of deployment to the time a victim's account is compromised. Our study relies on a key observation: despite cloaking and related evasive efforts, a substantial proportion of phishing pages make requests for web resources (e.g., images and scripts) hosted by third-parties, including the websites that attackers impersonate. Based on this insight, we collaborate with one of the most-targeted financial services brands in the current ecosystem [64] to develop and deploy a re-usable framework to meaningfully analyze victim traffic to live phishing pages.

We start by analyzing 404,628 distinct phishing URLs in our traffic dataset to gain an understanding of the aggregate volume and timing of key events within the life cycle of phishing attacks. Next, we correlate the traffic with the original phishing email lures to map the distribution phase of the attacks. Finally, we investigate the timing and success rates of attackers' monetization efforts based on the subsequent account compromise and fraudulent transactions—of the same victims—at a major financial services provider. We show that the data sampled by our approach provides *visibility* into 39.1% of all known phishing hostnames which targeted the same brand during our observation period.

We find that the average phishing attack spans 21 hours between the first and last victim visit, and that the detection of each attack by anti-phishing entities occurs on average nine hours after the first victim visit. Once detected, a further

seven hours elapse prior to peak mitigation by browser-based warnings. This gap constitutes the "golden hours" during which attackers achieve a significant return-on-investment from their attacks that might otherwise be mitigated. Alarmingly, 37.73% of all victim traffic within our dataset took place *after* attack detection, and at least 7.42% of all targeted victims suffer subsequent fraud.

Our approach allows us to identify characteristics of particularly successful phishing attacks. We found that the top 10% largest attacks in our dataset accounted for 89.13% of targeted victims and that these attacks proved capable of effectively defeating the ecosystem's mitigations in the long term. Phishing campaigns would remain online for as long as nine months while deceiving tens of thousands of victims in the process—using sophisticated, but off-the-shelf phishing kits on a single compromised domain name. As a result, we propose a practical methodology for organizations targeted by phishing to proactively mitigate similar attacks, and we deploy our approach to secure the affected victims' accounts in our study.

Our work motivates the expansion of collaborative, defense-in-depth anti-phishing approaches as a means to cope with phishers' evasion techniques and increasing sophistication. It underscores the importance of not only making improvements to existing ecosystem defenses such as browser-based detection, but also more widely adapting proactive mitigations like those that we propose. The contributions of our work are as follows:

- A longitudinal measurement study of the end-to-end life cycle of real phishing attacks representative of the modern anti-phishing ecosystem.
- A framework for the proactive detection and mitigation of phishing websites that embed external resources.
- Security recommendations to address the limitations within the current anti-phishing ecosystem based on our analysis of highly successful phishing attacks.

2 Background

Phishing is a type of attack through which malicious actors (i.e., phishers) leverage social engineering to trick victims into unknowingly disclosing sensitive information such as account credentials, personal data, or financial details [15]. Typically, victims are lured to a fraudulent website that impersonates a well-known brand solely for the purpose of harvesting such information. Attackers then use the stolen information for their own gain, either directly or through monetization services prevalent in underground marketplaces [59].

In recent years, criminals have shown no signs of slowing down their phishing attacks; the increased difficulty of drive-by downloads and exploits has given a resurgence to large attacks grounded in social engineering [69]. As such, phishing continues to evolve in sophistication to adapt to best practices within the broader Internet [4, 52]. When used by criminals, credentials obtained through phishing have proved to work the most reliably due to the broad range of other identifying information obtainable through phishing [16, 62]. Subsequent account compromise occurs at scale and accounts for substantial monetary damage to both users and businesses [21].

2.1 Phishing Websites and Phishing Kits

Large scale phishing attacks consist of three main phases. First, an attacker launches a deceptive website that spoofs the look-and-feel of a legitimate website (e.g., of a prominent brand). Thereafter, the attacker sends messages to potential victims (e.g., via spam email) with a link to the phishing website. Social engineering techniques often play an important role in these messages; they may lure users by conveying a sense of urgency and a need for action, such as correcting a billing error or securing a (fictitiously) compromised account [14]. If the lure successfully deceives the victim [68], the victim submits the sensitive information requested by the attacker. The phishing website may also record metadata about the user, such as the user's IP address or language from the HTTP request. Finally, the attacker downloads the stolen information from the phishing page or associated drop box [18].

Beyond email, attackers also rely on social media lures [1, 10], spoofed or exploited mobile or cloud applications [67], search engine listings, text messages, and phone calls [34, 65] in order to reach victims. However, email-based attacks remain dominant within the ecosystem [4]; such attacks enjoy greater scalability thanks in part to the help of underground services that simplify bulk messaging [61, 63]. Also, advanced features in phishing kits help phishers deceive their victims and bypass anti-phishing mitigations [51]; some can even intercept (and thus defeat) multi-factor authentication like SMS in real time [45, 66].

Furthermore, the deployment of phishing websites—even those which are technically sophisticated and laden with evasion techniques—has a low barrier to entry, likewise thanks to services in the underground economy. *Phishing kits* are readily available off-the-shelf packages that attackers can use to deploy phishing websites without the need for any technical knowledge [11]. Such kits are often bundled with exploits which can be used to install the kit on a compromised web server (thus minimizing any cost to the attacker).

2.2 Detecting and Mitigating Phishing

The risk of phishing attacks has given rise to an extensive anti-abuse ecosystem [52]. Multiple layers of defense include email spam classification filters [17, 35], crimeware and credential drop analytics [33], URL and content classification and blacklisting [39, 73, 75, 76], malware and vulnerability scanning by web hosts [5, 9], DNS, domain, and certificate intelligence [31, 50], user training [15], content take-down [2], and direct abuse reports [25].

Browser-based phishing detection [57]—like Google Safe Browsing or Microsoft Windows Defender—serve a particularly important role due to its scale and always-on nature.

All major web browsers have phishing detection built-in by default for both desktop and mobile platforms [51]. When a user visits a URL, their browser will make a call to a detection backend (e.g., a URL blacklist or a heuristic classifier) and show a prominent warning if the URL is deemed harmful. This represents a considerable improvement over early decentralized mitigations such as add-on toolbars [74].

2.3 Evasion Techniques

The longer that phishing websites remain online and are accessible to victims, the more attackers stand to profit. Therefore, modern phishing websites seek to maximize their own longevity through a variety of strategies to remain stealthy [51]. We provide an overview of key strategies below, and offer further insight, based on our findings, in Section 7.1.

Cloaking: In an effort to prevent security infrastructure from verifying malicious content, phishing websites with *cloaking* display benign content or an inconspicuous error message whenever they detect that a web request originates from an anti-phishing crawler [34]. Cloaking is typically implemented through client- or server-side code which applies filters using attributes such as IP address, geolocation, user agent, session cookies, or browser fingerprints. The presence of cloaking is the norm, rather than the exception, within modern phishing websites [52].

Redirection Links: URLs are the most direct indicators of phishing attacks and are therefore one of the primary data points used by anti-phishing systems. To evade detection, some attackers initially distribute URLs that might appear benign but redirect to different landing URLs which may contain deceptive keywords [10]. Such redirection not only hampers the use of URL heuristics as a detection strategy [75], but also makes it difficult to correlate URLs that are part of the same redirection chain in the wild [68]. Redirection links themselves may leverage cloaking or frequently change to evade detection, even when pointing to the same phishing landing page.

2.4 Measuring the Impact of Phishing

Meaningfully assessing long-term trends in the volume of phishing attacks has historically proved to be challenging due to a lack of transparency and consistency in the methodology applied [46]. Data sources that could be effectively used for such measurements are spread throughout the ecosystem and typically held closely by their owners. Other data, such as phishing URLs, is more readily available and suitable for classification or fingerprinting purposes, but not directly coupled with attack volume or impact [12].

Since 2004, the Anti-Phishing Working Group (APWG)—an industry-wide consortium of key anti-phishing entities—has regularly published summary reports of monthly phishing volume and ecosystem attack trends based on diverse partner data [4]. Although these reports have provided phishing volume figures for over a decade, changes in methodology and data sources over time prevent longitudinal analysis and only enable conclusions such as "phishing continues on a large scale". Research with deeper insight into the progression of phishing attacks has thus far been limited to smaller datasets or isolated scope [29, 48].

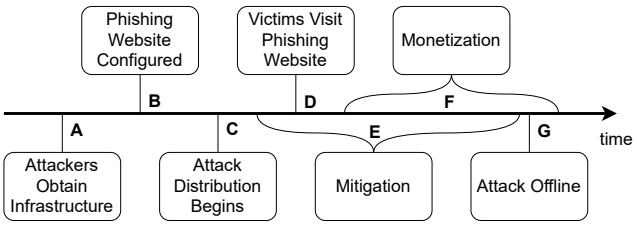


Figure 1: High-level stages of a typical phishing attack.

Obtaining data relating to the *damage* caused by phishing attacks (i.e., as a result of account compromise or credential theft) at specific organizations is even more challenging due to its sensitive nature in terms of both individual victims' and businesses' confidentiality. Additionally, victims themselves have shown a tendency to under-report cybercrime to authorities [20]. Aggregate summaries of such damage are thus often extrapolations based on certain assumptions [43].

3 Methodology

In this section, we discuss our approach to measuring the end-to-end life cycle of phishing websites, from the time of configuration to the time the attack goes offline.

3.1 Phishing Attack Stages

We show an overview of the stages of a typical phishing attack in Figure 1. Attackers first obtain infrastructure (*A*) and configure a phishing website on this infrastructure (*B*), often by installing a phishing kit. Once the website is operational, attackers begin distributing it to their victims (*C*) and victims start accessing it (*D*), as previously discussed in Section 2.1. After this point, the remaining stages are not necessarily consecutive.

Once detected by anti-phishing infrastructure, the attack will be mitigated by the ecosystem's defenses such as browser-based phishing warnings (*E*). In an optimal scenario, this mitigation would occur *before* time *D* and would prevent all future victim traffic. If these conditions are not satisfied, victim visits may continue for an extended period, and attackers will proceed to monetize the data stolen by the phishing website through various means (*F*) [63], which could entail testing stolen credentials against the corresponding platforms, or submitting fraudulent transactions using stolen financials [19, 71]. The original phishing website will eventually go offline, either as a result of take-down efforts [2] or deliberately by attackers (*G*). Note that malicious infrastructure configuration (*A*) is outside of the scope of our work, as we focus purely on phishing attacks themselves [5, 6].

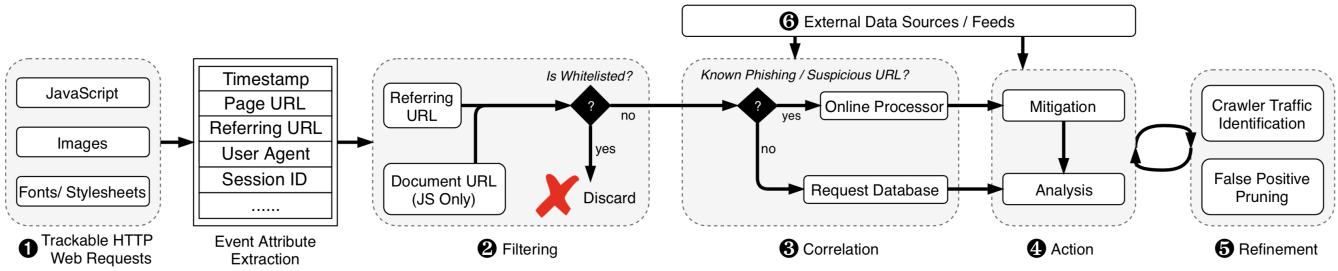


Figure 2: Golden Hour framework design.

3.2 Observations

As a preliminary step in our study, in June 2018, we manually inspected a sample of live phishing websites shortly after their URLs were submitted to PhishTank [55], OpenPhish [53], or the APWG eCrime Exchange [4] (large clearinghouses of phishing URLs). We made two key observations: first, that phishing websites routinely embed resources (e.g., images, fonts, or JavaScript) hosted on third-party domains, including domains which belong to the organizations being impersonated; and second, that some phishing websites redirect the victim back to the organization's legitimate website after the victim submits their information.

It thus follows that third parties—including the organizations being targeted and impersonated by phishers—could, with the right methodology, directly track visitor activity on certain phishing websites by inspecting HTTP/HTTPS requests for the aforementioned web resources within their *own* systems, and by identifying referrals [22] from suspicious sources. Such tracking could capture not only victim interaction with the phishing websites, but also visits from attackers themselves as they configure and test their attacks.

Moreover, the data could be used to proactively identify phishing URLs and propagate them through the anti-phishing ecosystem. Correlating the data with victim information (e.g., if a visitor's request for a resource on a phishing website has the same session identifier as a prior visit to the organization's legitimate website [7]) could help organizations better mitigate attacks by securing any accounts tied to the victims, while simultaneously measuring the effective damage likely caused by phishing. Lastly, correlating URLs in phishing lures (e.g., email messages) with victim traffic to phishing websites can paint a clear picture of the distribution phase of phishing attacks.

Recent work used a similar approach to identify characteristics of successful email lures and discover the corresponding URLs [68]. Our analysis of web event data instead focuses on mapping the overall progression of phishing attacks: consequently, we correlate the timing of key events within phishing attacks to a deeper extent, and on a larger scale, than previous studies [29, 48]. We also consider the success of phishing attacks, and we *directly* leverage the web event data as an anti-phishing mitigation.

3.3 Data Analysis Framework

The aforementioned analysis necessitates access to data only available to specific organizations (i.e., those commonly targeted by phishers or engaged in anti-phishing). We collaborated with one such organization—a major financial services provider—to develop and deploy a generic framework for processing the relevant data. The *Golden Hour* framework, shown in Figure 2, extracts web tracking events associated with phishing websites for analytical purposes or as a real-time proactive mitigation.

Our framework is brand-agnostic and could thus realistically be adapted for use by a broad range of organizations that have access to the appropriate data. We start by providing an abstract overview of the framework and then discuss our deployment in Section 4.1. In Section 5, we show that our framework enables insight into phishing attacks during their early *golden hours*, and that it can effectively disrupt attacks during or prior to this period.

In the *Golden Hour* framework, we first ingest web events of interest (❶), which can be obtained from raw web traffic logs (i.e., requests for images or style elements) or pre-processed data from web trackers or JavaScript web application code. We annotate each ingested event with a timestamp and extract further attributes, such as the IP address, user agent, session identifiers (i.e., from prior requests), referring URL, and the main page URL which was visited. We then take the latter two URL attributes and apply whitelist filtering (❷) to eliminate benign events which would normally be expected to be seen in this context, such as requests to the organization's legitimate website or requests with referrers on approved partner websites. Thereafter, we correlate (by substring matching) the URLs of the remaining events with a recent list of known phishing URLs from additional data sources (❸); this correlation enables the discovery of new phishing URLs which might only share a similar hostname or path with a previously-reported URL, but differ otherwise. It is also possible to apply phishing URL classification heuristics to identify previously-unknown URLs of interest [23].

The event correlation can take place in an online manner, or be deferred, in which case events are archived for later analysis (❹). In both cases, to allow for scalability, a chosen *observation window* defines a range of time (i.e., before and after a URL is reported) within which correlations for a

	Date Range	No. of Samples
Golden Hour web events	10/01/18 - 09/30/19	22,553,707
(distinct phishing URLs)		404,628
E-mail reports	09/01/19 - 09/30/19	68,502
APWG phishing URLs	10/01/18 - 09/30/19	52,116
Organization's phishing URLs	10/01/18 - 09/30/19	37,438
Fraudulent account transactions	10/01/18 - 09/30/19	Not disclosed
Compromised user accounts		

Table 1: Overview of the datasets analyzed.

given phishing URL are made. Successive reports of the same URL naturally extend the observation window; otherwise, correlations against unnecessary data can be avoided.

Events that are identified as phishing are additionally marked for immediate mitigation. Over time, we further refine the archived events (❺) by identifying false positive correlations, noise from automated (i.e., web crawler) traffic, and phishing URLs detected at a later point in time, with the use of statistical analysis and external data sources (❻).

To benefit from our framework's mitigation capabilities, it should ideally be deployed *online*, on a stream of live (or recent) data during the ingestion stage (❶). However, the framework can also process archived (i.e., historical) event data alone. In the long term, as the anti-phishing ecosystem builds ground truth (i.e., by having access to a vetted list of known phishing URLs), both approaches will enable the same level of analytical insight. Thus, the framework can accommodate different data ingestion strategies to support the infrastructure of the organization deploying it.

4 Dataset Overview

We deployed the *Golden Hour* framework to collect and analyze one year of phishing web traffic data between October 1, 2018 and September 30, 2019 (inclusive) at the same organization mentioned in the previous section—a major financial services provider and one of the most-targeted brands within the current ecosystem [13, 44]. We provide an overview of the scope of all of our datasets in Table 1. Note that this data was collected ethically and in compliance with user privacy laws within the originally-intended context (see Section 8.4).

4.1 Data Collection

We operated the *Golden Hour* framework in an *online* manner from July 1, 2019 through September 30, 2019 and additionally processed archived data from the preceding nine months. To efficiently query a data warehouse, we limited our *observation window* for web event data (as discussed in Section 3.3) to one week before and one week after the corresponding hostname appeared in a phishing feed. We found that this approach did not lead to the omission of any relevant events, as phishing URLs which remained live for longer periods would reappear in the feeds at a later date, and would thus also be extracted by our framework for analysis.

The resulting dataset initially contained a total of 22,553,707 web events resulting from traffic to phishing websites from victims, attackers, and security crawlers alike. Using the traffic data, we are able to gain detailed insight into stages *B*, *D*, *E*, and *G* within the life cycle of phishing attacks. For the framework's correlation (❸) and refinement (❺) steps, we programmatically queried additional data sources: phishing URLs for the same brand in the APWG eCrime Exchange feed, the organization's proprietary phishing URL feed, and the organization's proprietary automated (i.e., crawler) traffic detection system (❻).

During our deployment, we pruned 3,194,031 events by identifying traffic to legitimate websites (based on a whitelist and manual review) and false-positive URLs that were under-represented in the phishing feeds or flagged as such by the organization. Thus, our final dataset contained 19,359,676 total events. These events corresponded to 404,628 distinct phishing URLs—more than either phishing feed we considered, as our hostname correlation enables the identification of unreported variants of URLs similar to those which appeared in feeds. However, additional types of data are required to obtain timings of attack distribution (stage *C*) and monetization (stage *F*) and thus complete our end-to-end analysis, as these are not captured by the traffic dataset alone.

Phishing URL Distribution: To measure URL distribution to victims, we extracted metadata from phishing emails that users forwarded (i.e., as spam reports) to the organization. The timestamps within the original email lures allow us to calculate when phishers originally distributed their attacks. To correlate these timestamps with web events in our traffic dataset, we extracted URLs from each email and followed redirects (if any) to obtain the URL of the final phishing landing page. In cases when a redirect was followed, or if the phishing URL was no longer accessible, we would additionally query the organization's internal anti-phishing system to obtain any other landing page URLs known to be previously associated with the URL in the email. To complete the correlation, we search for events within the traffic dataset with the same hostname and a common path.

We were able to correlate 21,244 email reports with phishing URLs in our event dataset[1]. We found that 84.44% of these emails contained a timestamp detailed enough (i.e., date, time, and timezone) for our analysis. Determining final landing page URLs from links within the email proved integral, as only 3.99% of emails contained the same URL as the final phishing page (i.e., others made use of redirection).

Account Compromise and Monetization: To understand one way in which criminals exploit credentials from phishing victims, we analyzed session identifiers programmatically extracted from events in the traffic dataset (i.e., victim visits to a phishing website which had cookies from a prior interaction with the legitimate organization's website). The organization then mapped these identifiers to user

[1]The uncorrelated emails either were outside of the *visibility* of our approach, or had redirection chains which could not be reconstructed. We discuss the relatively small size of our email dataset in Section 8.3.

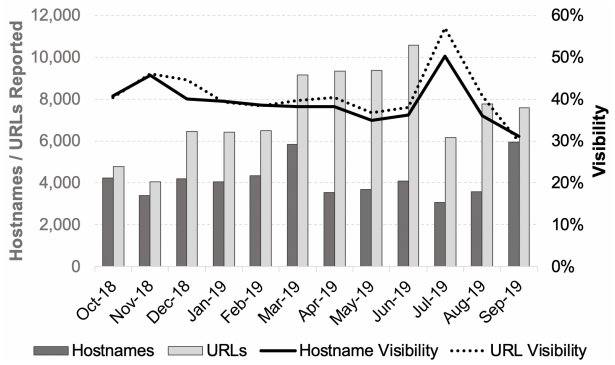


Figure 3: Visibility of phishing websites targeting the organization in our dataset.

accounts and provided timestamps of fraudulent transactions associated with these accounts, and timestamps of when corresponding credentials appeared in a public dump. We could then correlate these timestamps with the victims' original interaction with the phishing page per the traffic dataset.

Due to the sensitive nature of user information, we present our findings related to this data in aggregate form only (in Section 5). Note that no Personally Identifiable Information (PII) was given to us for the purposes of this analysis.

4.2 Level of Visibility

An immediate question that arises about our analysis concerns the level of *visibility* which can be provided by the *Golden Hour* framework. We define *visibility* as the proportion of phishing websites that target the organization in our study that can be analyzed through our approach. While we cannot provide a definite visibility measurement, as this would require knowledge of *all* phishing campaigns that target the organization, we estimate the visibility of our dataset by dividing the number of distinct hostnames with at least one associated web event by the total number of phishing hostnames, for the same brand, known to us from other data sources during the data collection period (i.e., in the APWG or organization's phishing URL feed). We also calculate the same ratio for full URLs. We note that it is easier for phishers to create multiple paths on a single domain compared to multiple subdomains; thus, the hostname ratio better represents unique attacks.

We found that our approach had visibility into an average of 39.1% of all hostnames and 40.9% of all URLs which were found in the aforementioned feeds and targeted the organization during our data collection period. We present the visibility ratios by month in Figure 3. Given the evasiveness of modern phishing attacks, we suspect that the list of phishing URLs known to us is an underestimate of phishing URLs in the wild [52, 62]; however, consequently, the same would apply to the URLs for which we have event data.

The degree of visibility for both hostnames and URLs remained fairly consistent throughout our data collection period, with the exception of July 2019. During this month, we observed a spike to 50.2% and 57.1% visibility, respectively, which coincided with the launch of numerous sophisticated, large-scale attacks that were detectable by our approach. We discuss these attacks in more detail in Section 7.

Per the APWG eCrime Exchange, the brand in our dataset accounted for 10.6% of all phishing hostnames (with known brands) during the same one-year period. This allows the extrapolation of the potential visibility of our approach into the population of phishing websites.

4.3 Event Distribution

We collected web events of two broad types: visits that occurred directly on phishing websites (*Page URLs*) and referral traffic from a phishing website back to the organization's legitimate website (*Referring URLs*). We show the monthly distribution of these events in our dataset in Figure 4. We observe that phishing attacks are not uniformly distributed; some months see substantially more traffic than others. Historically, phishing attacks have been associated with a certain seasonality, particularly near holidays. The spike in the final three months of our dataset is consistent with the Q3 2019 APWG report, which found this period to have the largest volume of phishing URLs in three years [4]. However, we expose a limitation of counting URLs alone as a measurement of overall phishing volume, as the spike in our traffic dataset is far more dramatic than the change in total URLs[2].

In Table 2, we further subdivide the events by the type of user. Events from *Known Visitors* are those which contain a session or device identifier previously known to the organization, and can thus be linked with certainty to a known account at the organization. *Crawler* events are those which we or the organization classified as automated traffic based on request attributes. The *Other* events fall into neither category but follow a similar distribution to *Known Visitors*, and thus represent potential victims which cannot be immediately traced back to an account at the organization.

To ensure consistency across our measurements in the following sections, we define the set of *Compromised Visitors* as those *Known Visitors* whose accounts were subsequently either accessed by an attacker or had at least one fraudulent transaction. We consider only these events in our analysis of monetization efforts, as the sequence of observations strongly suggests that a phishing attack succeeded against the corresponding victims. We do not disclose the total number of unique victims within these two sets for reasons discussed in Section 8.3.

5 Progression of Phishing Attacks

To create an end-to-end timeline of the progression of phishing attacks, we calculate the relative difference between the timestamp of each *Golden Hour* web event and the

[2] We believe that both of these spikes are associated with the effectiveness (and proliferation) of highly sophisticated phishing websites, which we characterize in Section 7.1.

	Known Visitor	Crawler	Other	Total
Page URL	2,968,735	2,934,976	7,982,475	13,886,186 (71.73%)
Referring URL	1,879,179	820,716	2,773,595	5,473,490 (28.27%)
Total	4,847,914 (25.04%)	3,755,692 (19.40%)	10,765,070 (55.56%)	19,359,676

Table 2: Breakdown of Golden Hour web events by type.

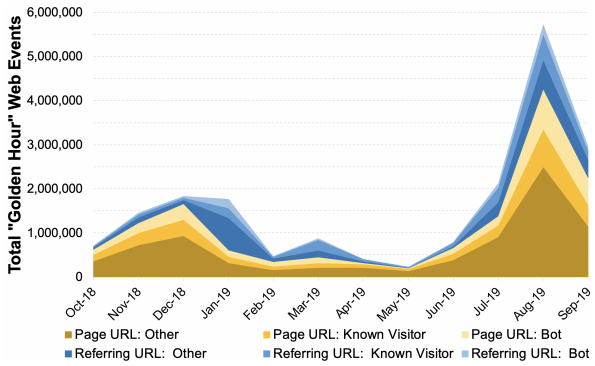


Figure 4: Distribution of Golden Hour web events by month.

original detection of the corresponding phishing URL within a feed, as correlated by our framework. We calculate similar timestamp differences for email lures, account compromise, and fraudulent account transactions. We can then plot a histogram of victim traffic relative to attack detection, alongside the average timestamps of key attack milestones. Note that the effect of outliers on these averages is inherently suppressed by our use of a fixed *observation window* for each phishing URL's web events.

In Figure 5, we show such a histogram for *Compromised Visitors*: in other words, every user represented in the figure was highly likely to have been successfully deceived by a phishing attack. We count multiple events from the same victim on the same phishing website only once. For brevity, we do not separate *Page URL* and *Referring URL* events in our figures, as these did not differ significantly except in the success rates of subsequent account compromise (discussed in Section 5.3).

We observe that phishers enjoy a large window of opportunity when carrying out their attacks. Nearly nine hours elapse on average between the first victim visit and detection by the ecosystem. By this time, the phishing websites have already lured 62.73% of victims. Moreover, victim visits continue at a slower pace for the next 12 hours. We show the Cumulative Density Function (CDF) of *Compromised Visitor* web events in Figure 6a. Despite the 21-hour time frame (-08:44 to +12:26) of a typical phishing attack illustrated in Figure 5, there exist some attacks with a longer overall duration.

5.1 Initial Traffic

The average first non-victim visit to each phishing website occurs 9 hours and 42 minutes prior to attack detection, as shown in Figure 5. We believe that such visits are representative of attackers' initial testing of each phishing website.

Country	Other Traffic	Country	Known Visitor Traffic
United States	32.84%	United States	65.48%
Morocco	9.17%	United Kingdom	6.15%
Indonesia	8.16%	Canada	4.26%
United Kingdom	6.08%	Italy	3.05%
Algeria	3.73%	Spain	2.78%
Canada	2.99%	Australia	2.58%
Germany	2.88%	Germany	2.29%
Brazil	2.35%	Mexico	1.46%
Tunisia	2.29%	France	0.93%
Italy	2.24%	Netherlands	0.79%
France	1.92%	Brazil	0.72%
Iraq	1.60%	Singapore	0.64%
Egypt	1.44%	Ireland	0.40%
Spain	1.39%	Belgium	0.40%
Nigeria	1.39%	Portugal	0.38%

Table 3: Geolocation of initial visits to live phishing websites, by traffic category.

We performed an unequal variance T-test [56] to compare the distribution of the relative timestamps of the *first* event (for each attack) within the *Other* category to the *first* event for *Known Visitors*. We find the means of the two distributions to be statistically significantly different, with a p value of 0.011. Furthermore, in Table 3, we show that top geolocations within the former set closely coincide with countries disproportionately associated with cybercrime [37] (and inconsistent with the organization's customer base).

5.2 Phishing Email Distribution

We show the CDF of phishing email distribution in Figure 6b. We note that prior to attack detection, the cumulative proportion of victim visits to phishing websites (in Figure 6a) grows at a faster rate than emails sent. In other words, traffic from phishing emails to phishing websites drops after attack detection, as should be expected following the intervention of spam filters. However, just one day after detection, the rate of victim visits once again starts outpacing the sending of emails. This suggests that victims will follow links in old emails, and, thus, attackers continue to profit without further intervention. Take-down can potentially assist with mitigating these long-lasting phishing attacks [2].

5.3 Progression of Monetization Efforts

In our dataset, 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, as our approach does not identify victimization of the *Other* traffic. After each victim's visit to a phishing website, we found that such a transaction would occur with an average delay of 5.19 days. However, as we show in Figure 6c, fraudulent transactions grow consistently over a 14-day period, with the earliest ones occurring less than one hour after a victim visit. Although about 3.99% of fraudulent transactions occur *after* this period, the increasing potential for mitigation encourages attackers to act quickly.

The credentials of 63.61% of these compromised victims would additionally appear in a public dump, with an average

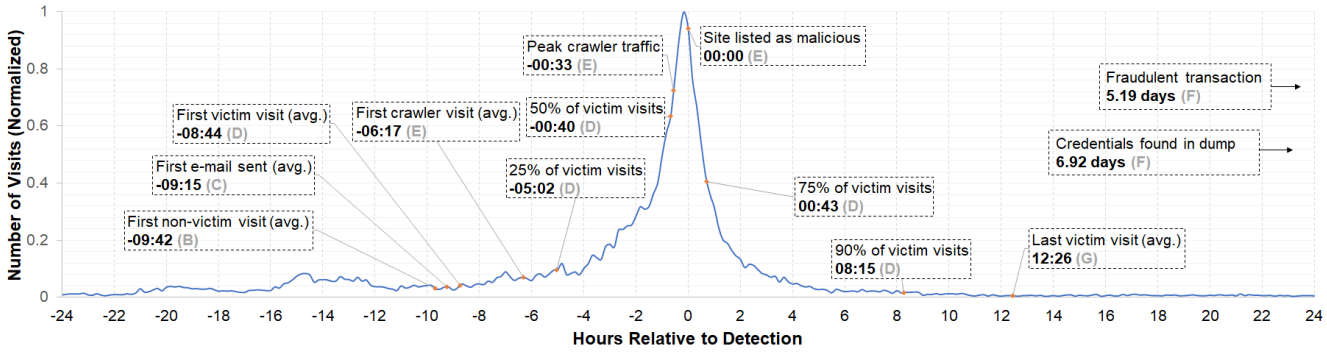
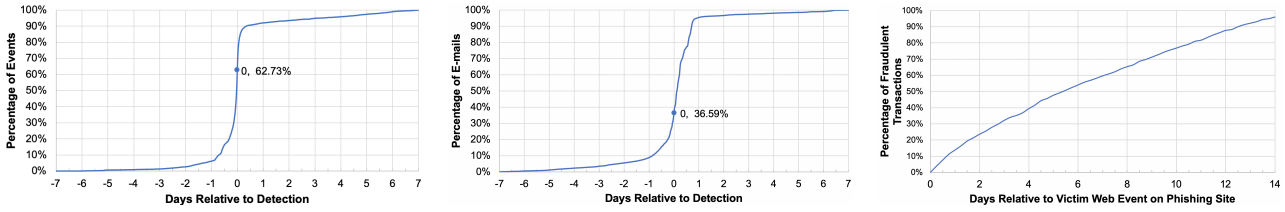


Figure 5: Histogram of Compromised Visitor traffic to phishing websites, annotated with attack stages.



(a) Compromised Visitor traffic. (b) Phishing email distribution. (c) Fraudulent transactions over time.

Figure 6: Cumulative Density Function (CDF) plots depicting key phishing attack stages.

delay of 6.92 days. This trend suggests that criminals tend to first monetize the accounts of their victims, and only later sell credentials within underground economies [8].

Our dataset does not provide insight into the monetization of each victim's stolen personal information beyond the organization's own systems. We find that the average victim makes 2.43 page loads during his or her interaction with a phishing website—enough to visit a landing page and submit credentials. Some victims, however, made substantially more visits during a single session. After inspecting the chain of phishing URLs visited in such sessions, we believe that such victims provide additional personal information to the phishing website (i.e., one with multiple data collection forms), and could thus suffer from identity theft or other financial fraud. Per our dataset, we observed that victims with an above-average number of page loads who also appeared in a *Referring URL* event (i.e., returned to the organization's website after presumably completing interaction with the phishing website) were 10.03 times more likely to later encounter a fraudulent withdrawal from their account.

5.4 Browser-based Detection Effectiveness

Given the ubiquity of browser-based detection and warnings, the role of these defenses in preventing phishing in the wild is a key measurement we seek to estimate. The mitigation from browser-based detection can be delayed for two main reasons: failure of backend systems to flag a given phishing URL or the lag between backend flagging and data propagation to clients (e.g., browsers) [26]. This lag period may vary between the same browser on different devices due to differences in cache state [51].

We can meaningfully estimate the overall impact of

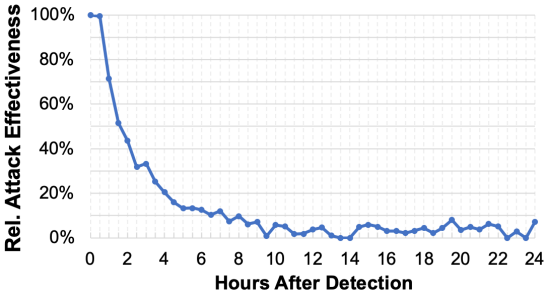


Figure 7: Impact of browser-based detection on phishing effectiveness.

browser-based warnings on phishing attack effectiveness by calculating the ratio of *Compromised Visitors* for browsers with native defenses (Google Safe Browsing, Windows Defender) and *Compromised Visitors* for *all* browsers, at regular time intervals after attack detection (i.e., after the midpoint of Figure 5), and subsequently comparing this ratio to a baseline ratio just prior to detection. This ratio is not sensitive to the decrease in absolute phishing traffic as it simply isolates the likelihood that the phishing attack will be successful (*Compromised Visitors* are visitors who likely submitted credentials to a phishing website).

As we show in Figure 7, browser-based warnings start to substantially reduce the relative effectiveness of phishing attacks within one hour after detection, at which point the ratio of *Compromised Visitors* drops to 71.51%. By the end of the second hour, the ratio drops further to 43.55%: at this point, attacks are less than half as effective as they were originally. The effectiveness continues declining more slowly until the seventh hour and thereafter stabilizes within the 0-10% range.

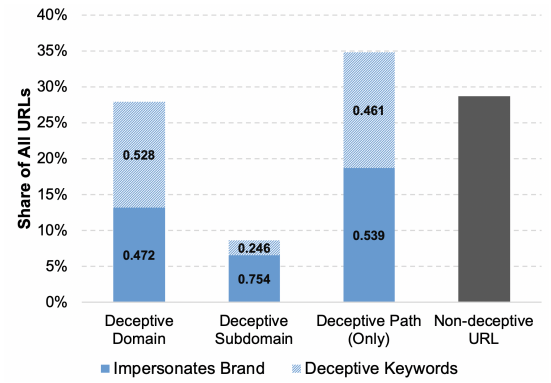


Figure 8: Classification of URLs in the Golden Hour dataset.

Browser Name	Traffic Share
Chrome Mobile	29.72%
Safari Mobile	22.38%
Chrome	21.56%
Samsung Browser	7.97%
Edge	5.53%
Safari	4.10%
Firefox	3.66%
Internet Explorer	3.21%
Other	1.87%

(a) By browser

Device	Traffic Share
Android	35.70%
Windows	28.13%
iOS	27.03%
OS X	8.35%
Other	0.79%

(b) By device

Table 4: Known Visitor traffic share by browser and device.

Browser-based phishing warnings are clearly an effective mitigation overall, but attackers can and do abuse their reactive delay, as we have demonstrated. In addition, certain evasion techniques, which we discuss in Section 7.1, can avoid triggering such warnings even after attack detection. Additional mitigations are required to thwart the trickle of *Compromised Visitor* visits which we observed many hours after detection.

6 Phishing Attack Characteristics

In this section, we analyze metadata related to the phishing websites in our dataset in an effort to better understand the characteristics of successful attacks. We consider all phishing URLs with at least one *Compromised Visitor* event.

6.1 Phishing URL Classification

Attackers have traditionally crafted phishing URLs to deceive victims by either mimicking the brand being impersonated by the phishing website (e.g., *www.brand-alerts.com*), or by including misleading keywords which convey a desire to help the victim (e.g., *secure-my-account.com*) [23].

We apply a previously-proposed classification scheme [52] to the URLs in our dataset and show the results in Figure 8. We observe that 28.70% of all URLs have no deceptive content whatsoever; 34.76% have non-deceptive domains with deceptive paths only. 8.64% use deceptive subdomains on a non-deceptive domain, and the remaining 27.90% have deceptive domains (0.52% with Punycode). The nature of deceptiveness is similarly split between brand names and misleading keywords, except in the case of subdomains, which favor brand names. Bare IP addresses were negligible in our dataset and thus are excluded from the figure.

The vast majority of phishing URLs (98.58%) were hosted on traditional, paid domain names. Only 0.79% of URLs leveraged subdomains from free hosting providers; 0.63% had domains with free TLDs [52]. However, compromised hosting infrastructure plays a key role, which we assess in Section 7.

With the increasing use of mobile devices to browse the Internet, the importance of URL content has diminished (i.e., because of limited screen real estate on such devices) [40, 72]. However, the heavy use of redirection in phishing lures allows attackers to continue using deceptive URLs (which would otherwise be easily detectable by text-based classifiers) on their landing pages.

6.2 Device and Browser Type

As shown in Table 4, mobile devices accounted for 62.73% of all victim traffic in our dataset. Browsers protected by *Google Safe Browsing*—*Chrome*, *Safari*, and *Firefox*—accounted for 81.42% of the traffic (roughly consistent with overall market share) [60]. The wide use of these browsers, in particular on mobile platforms, underscores the importance of the efficacy of the anti-phishing features which they natively include. The *Samsung Browser*, which does not currently include browser-based phishing detection to the best of our knowledge, and thus leaves users particularly vulnerable to phishing, had a disproportionate representation of 7.97% in our dataset. The behavior of individual browsers has previously been studied in detail and is thus outside of the scope of our analysis [51].

6.3 Use of HTTPS

The web has moved away from traditional HTTP in favor of encrypted communication over HTTPS; phishers started following this trend in 2017 [3], which has been simplified through the wide availability of free SSL certificates [38]. Within our entire dataset, 66.85% of distinct URLs used HTTPS. However, these URLs accounted for 85.77% of the *Compromised Visitors*. Phishing attacks with HTTPS thus proved about three times more successful than HTTP. Even though some successful phishing attacks still occur on unencrypted websites, this now represents a minority of attacks. Simultaneously, the potential impact of Certificate Authorities in helping prevent abuse—especially on attacker-controlled domains—has grown.

7 Phishing Attack Longevity

Prior research has stipulated that individual phishing attacks tend to be short-lived and that they capitalize on the narrow gap between deployment and detection [41]. Despite some caveats, we have made a similar observation in Section 5. However, these observations do not capture trends within broader phishing campaigns, which may entail a group of organized criminals involved in the successive deployment of persistent and sophisticated attacks.

Rank	First Seen Date	Last Seen Date	Campaign Duration (Days)	Known Visitor Events	Average Events Per Day	Distinct URLs Reported	URL Text Classification	Domain Type
1	01/06/2019	09/22/2019	259	145,306	560	41	Deceptive Path Only	Compromised
2	08/30/2019	09/26/2019	27	115,616	4,329	41	Deceptive Subdomain	Compromised
3	07/20/2019	09/14/2019	56	102,601	1,847	40	Non-deceptive	Free Subdomain
4	01/11/2019	01/15/2019	4	82,636	20,487	6	Deceptive Path Only	Regular Registration
5	06/14/2019	06/20/2019	6	71,478	11,681	56	Non-deceptive	Compromised
6	04/21/2019	05/27/2019	36	71,037	1,992	39	Deceptive Path Only	Regular Registration
7	08/11/2019	08/17/2019	5	59,911	11,296	40	Deceptive Subdomain	Free Domain
8	03/14/2019	04/22/2019	39	55,147	1,427	81	Deceptive Subdomain	Regular Registration
9	08/30/2019	09/26/2019	27	50,402	1,877	28	Deceptive Subdomain	Compromised
10	01/07/2019	01/07/2019	1	49,627	49,627	8	Deceptive Subdomain	Free Subdomain
11	12/22/2018	12/26/2018	4	44,502	10,806	45	Non-deceptive	Compromised
12	06/23/2019	06/28/2019	6	42,574	7,708	22	Deceptive Subdomain	Free Subdomain
13	09/24/2019	09/25/2019	2	42,406	21,203	29	Deceptive Domain	Regular Registration
14	12/12/2018	01/02/2019	21	38,484	1,814	16	Deceptive Path Only	Compromised
15	10/06/2018	02/22/2019	140	32,591	233	39	Deceptive Path Only	Compromised
16	12/11/2018	12/29/2018	18	30,983	1,768	63	Deceptive Subdomain	Regular Registration
17	10/31/2018	03/24/2019	145	30,853	213	90	Deceptive Path Only	Regular Registration
18	09/12/2019	09/22/2019	10	30,781	2,990	23	Deceptive Path Only	Compromised
19	03/19/2019	03/24/2019	4	23,552	5,399	21	Deceptive Path Only	Regular Registration
20	08/13/2019	08/15/2019	3	22,254	7,418	16	Deceptive Domain	Regular Registration

Table 5: Top phishing campaigns by number of Known Visitor events.

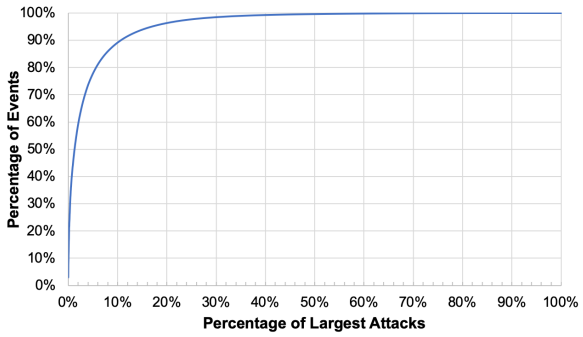


Figure 9: Share of Known Visitor events by top attacks.

To gain better insight into long-term phishing campaigns, we group phishing URLs from events in our dataset by domain (or hostname in the case of free subdomain hosting providers). We then sort the groups by the total number of unique *Known Visitor* events to capture variations in hostname or path for attacks which are likely related[3]. We define the *date range* of a campaign as the time between the first and last web event from a *Compromised Visitor*; we found the average *date range* to be 13.55 days.

We discovered that the top 5% of attacks accounted for 77.79% of *Known Visitor* events within our dataset, and the top 10% for 89.13%, as shown in the CDF in Figure 9. We then manually analyzed the top 20 campaigns (these alone accounted for 23.57% for *Known Visitor* events), some of which lasted several months each, as shown in Table 5. We also determined whether they were hosted on compromised domains (i.e., otherwise belonging to a legitimate website) or domains directly controlled by attackers.

[3]Some threat actors pivot across different infrastructure and might thus be underestimated by domain-based grouping of attacks. Confidently grouping attacks by other attributes, such as phishing kit signature or drop email, would require additional data. The same applies in case different threat actors were to leverage a single domain.

7.1 Sophistication and Evasion

To understand the success of the top phishing attacks, we manually inspected the content (and, when possible, phishing kits) of high-impact phishing URLs that were live during our *online* deployment between July 1 and September 30, 2019. We identified such URLs by spikes in the number of *Known Visitor* events associated with any individual hostname.

The characteristics we found contribute to the attacks' success not only by avoiding detection by anti-phishing infrastructure, but also by more effectively targeting human victims. We quantify our observations to the extent possible given our methodology; however, a more comprehensive measurement would be suitable for future work.

Broad Data Collection: The sophisticated phishing websites which we analyzed mark a clear departure away from single-page login forms, and thus venture far beyond mere theft of usernames and passwords [11]. Phishers fully match the page structure of the victim organization's website, complete with a homepage with links to (fraudulent) login pages and resources in case the victim was to navigate away from the initial landing page. Once the victim returns to the login page and starts interacting with the phishing website, it will seek to harvest extensive personal and financial information, identity documents, and even photographs (i.e., *selfies*) to steal and more effectively monetize victims' identities.

Automatic Translation: Five of the phishing websites in Table 5 used the visitor's geolocation to automatically translate their deceptive content. Manual analysis of the phishing kit used on one of these websites revealed a total of 14 language options that coincided with the targeted brand's major markets.

Human Verification: We observed that as part of a URL redirection chain, some attackers would show a reCAPTCHA challenge [70] prior to redirecting the victim to the phishing landing page. Also, one specific phishing kit showed a

CAPTCHA challenge directly on its landing page, prior to allowing the victim to input any credentials. Such challenges not only hamper the verification of phishing content by automated systems, but may also trick users into proceeding due to the use of CAPTCHA on legitimate websites.

Cloaking: All phishing websites which we analyzed leveraged server-side cloaking, a well-known technique that seeks to block traffic based on a blacklist (or a whitelist) of request attributes such as IP address or hostname [34, 52, 62]. Such cloaking intends to restrict access from security crawlers or other non-victim entities. Also, some phishing kits include an initial landing page that contains nothing but a simple piece of JavaScript code or an HTML Meta tag to redirect the victim to the true phishing page. Such code could defeat basic crawlers that look at static HTML only.

Victim-specific Paths: Eight of the campaigns in Table 5 had a landing page that automatically generates a sub-path unique to each visitor's IP address, and then immediately redirects to that path. The path is not visible to other IP addresses, and would thus evade crawlers visiting a previously-generated path rather than the attack's initial landing page.

Fake Suspension Notices: As a deterrent to take-down efforts [2], when a visitor fails *cloaking* checks, we observed that several phishing websites displayed a misleading page indicating that the domain has been suspended, rather than a generic HTTP 404 or 403 error message [22].

Man-in-the-Middle Proxies: Rather than a traditional phishing kit, two of the large phishing attacks we analyzed used a proxy that would make live requests to the legitimate organization's website and display the page to the user while intercepting all data submitted [28]. Such proxies can defeat most forms of two-factor authentication [66] and may require special care from the targeted organization—such as requiring security keys—to mitigate.

7.2 Attack Mitigation

While analyzing the sophisticated attacks in Table 5, we simultaneously manually reported them to anti-phishing entities and hosting providers. By the time the many original URLs were added to detection backends, the attackers redeployed subsequent attacks on different subdomains or paths, which would, in turn, necessitate another cycle of reporting or detection. In this manner, attackers can stay one step ahead of the ecosystem. When paired with bulletproof hosting (i.e., resistant to take-down from abuse reports) [36] or successive re-compromise of legitimate, albeit vulnerable, infrastructure, such attacks can remain effective for prolonged periods.

To help overcome the challenges faced by the ecosystem, we adapted the *Golden Hour* framework to perform proactive mitigation of attacks. We reported events corresponding to *Known Visitor* back to the victim organization, such that the organization could flag accounts to prevent successful compromise or re-secure accounts that had already been compromised. We reported tens of thousands of distinct events in this manner, which has motivated the permanent adoption of our framework by the organization.

Our framework can also be used to discover previously-unknown phishing URLs based on heuristics such as textual URL content (applied during correlation) or context. Such URLs can then be reported to detection backends and propagated through the ecosystem. Due to technical limitations, we did not automate this aspect of the framework during our deployment. In a retrospective analysis, we found that this would have potentially increased the number of web events in our dataset by 7.28%, which, if reported, could help narrow the gap in the detection of phishing attacks by the ecosystem.

8 Discussion

Although individual evasion techniques might not suffice to defeat the modern anti-phishing ecosystem, the increased degree of sophistication which arises from the combination of such techniques poses a key threat. We have shown that in terms of the number of victims compromised, sophisticated and persistent phishing attacks dominate, and should, therefore, be a priority for the ecosystem. At a more granular level, both the response time of browser-based warnings (which protect victims once a phishing attack is detected by the ecosystem) and speed of initial detection by backends (which closes attackers' window of opportunity) represent potential directions for improvement.

8.1 Data Sharing

The mere fact that so many of phishing websites in our dataset embed third-party resources shows that attackers do not fear being detected by certain organizations. Consequently, there is an opportunity for increased data sharing across the ecosystem to better detect threats based on proactive intelligence indicative of attacks: the web events from *Golden Hour* are just one example of such intelligence.

Reporting Phishing: Sophisticated phishing attacks currently exploit limitations within the detection ecosystem. In the case of cloaked phishing websites, simple URL-based reporting to anti-phishing backends—such as what is currently commonly carried out through automated systems and web submission forms [26]—fails to provide sufficient context for the backend to verify the phishing content. In particular, with only a URL in hand, the anti-phishing backend may not be able to determine the parameters required to defeat the cloaking or find new, but related threats. We experienced this phenomenon when manually reporting certain URLs from the sophisticated attacks in Table 5 to *Google Safe Browsing*; by the time such URLs were filtered, attackers would have shifted their websites to alternate paths or subdomains on the same web server. Enhanced phishing reporting protocols—potentially bolstered by trust between vetted entities within the ecosystem—could help anti-phishing entities share detailed attack data at scale. Similarly, with proper consent and privacy protections, sufficiently detailed

information could potentially be shared (e.g., request parameters, redirection chains, or a screenshot) [27]. Such measures might help close the gap between what users see and anti-phishing systems see, thus preventing cloaking.

Similarly, the ecosystem currently lacks *standardized* approaches for requesting malicious content to be taken-down [2]. Although major hosting providers may have well-documented avenues for removing phishing websites, attackers might migrate to bulletproof hosting [36] or small hosts with fewer resources for timely intervention, or by compromising infrastructure.

Phishing Links in Emails: Attackers make heavy use of redirection links in phishing email lures. As we have shown, such links complicate the correlation of phishing emails with live websites—and, in turn, hamper further mitigation efforts, such as browser-based detection. Additionally, we observed an average delay of 9.62 hours between the start of each phishing email campaign (i.e., the initial arrival of a phishing message) and the *first* report sent by a victim. Due to this delay, we believe that direct user reports should only serve as a secondary means for entities to discover new phishing attacks. As such, there is a potential for email providers to share known abusive URLs with the wider ecosystem [8].

In our dataset, at the granularity of individual phishing hostnames, email lures were sent in large spikes, similar to what has been previously observed [41]. If a message is initially classified as benign but the URL within it is later detected as phishing, additional measures are needed to ensure retroactive detection.

8.2 Third-party Resources

It may seem counter-intuitive for malicious websites to embed external web resources hosted by third parties, especially in light of our findings that these resources enable both analysis and mitigation of phishing attacks. However, we argue that phishing websites will nevertheless use external resources for a number of reasons.

Most importantly, anti-phishing systems use known filenames of scripts, images, favicons, and archives as one type of fingerprinting to identify malicious websites [24, 52]. Phishing pages which only link to external files can avoid such fingerprinting entirely; with the added use of cloaking on their landing pages, phishing websites can remain stealthy to avoid or delay detection by the ecosystem.

In some cases, attackers choose to use third-party *services* for their own benefit. Within our dataset, the use of reCAPTCHA is one such example. Additionally, we observed phishing websites hosted on single-page *pastebin* services [42]. In order for phishing pages to render correctly on such services (and thus successfully deceive victims), most images and scripts must be retrieved from external sources.

The use of external files can also ensure consistency between the look and feel of the legitimate website and a phishing page. Phishing kits can thus remain current without the need for frequent updates, which may be particularly desirable for phishers who do not want to invest money into sophisticated phishing kits. It is also easier for attackers to directly copy the source of the original page than to build a deceptive version from scratch.

Even if they do not embed third-party resources, phishing websites may link back to the legitimate organization's website and could thus be detected by our approach. The same applies if victims are redirected back to the legitimate website after being phished: a common strategy used by attackers to minimize victims' awareness of the attack.

8.3 Limitations

Our analysis should be considered alongside certain limitations. Despite a large sample size, our data is based on victim traffic to phishing websites that target a single organization, which may skew our findings. However, our *Golden Hour* framework is not tied to any one organization; thus, future analysis in other contexts could deepen the insight into the broader ecosystem.

Due to the nature of our agreements with the organization, we cannot disclose certain concrete findings from our analysis, such as the total profit secured by attackers. Also, the success of phishing attacks hinges on numerous factors—such as the content and type of the original lure, appearance of the landing page, or redirection services used—which we did not consider, but which could provide details about the ecosystem vulnerabilities being exploited.

Despite the incentives for phishers to use third-party resources, as discussed in Section 8.2, our approach does not guarantee the detectability of an arbitrary phishing website. Phishers could deliberately evade our approach by excluding any trackable third-party files and avoiding redirecting victims back to the organization's website.

The timeframe of our email dataset is shorter than that of our web event dataset. We originally intended to correlate the event data with phishing URLs sent to victim inboxes at a major email provider over the full data collection period. However, the prevalence of redirection links within phishing emails made such correlation difficult to scale.

Lastly, our web event correlation approach (stage ❸ of the framework) benefits from the ability to accurately classify URLs as suspicious from within a large stream of traffic data, or a reliable source of ground truth (i.e., known phishing URLs) to match events. We only did the latter during our deployment; however, we consider our data sources (in Section 4.1) to be of high quality: peaks in *Crawler* traffic in our event dataset coincided with detection times of URLs in the phishing feeds considered. Yet, recent research has shown that even reputable anti-phishing vendors fail to identify many phishing URLs reported to them [54]. Future deployments of our approach should maximize the number of data sources for correlation to further increase *visibility*.

8.4 Ethical Considerations

We took great care to ensure that user privacy was preserved throughout this research. Our analysis did not involve access to any PII, and processing which entailed datasets that could contain PII (such as user account information or email report content) was carried out in a purely programmatic fashion by existing, automated systems. During our analysis of user account compromise times, we only handled anonymized session or account identifiers which were interpreted and aggregated by the organization with which we collaborated.

Entities that become aware of compromised accounts within their systems—through internal or external data sources—should make reasonable efforts to re-secure such accounts [62]. During our research, we ensured that user accounts which we associated with phishing website traffic by *Golden Hour* were appropriately flagged by the organization. Furthermore, we recommended that user accounts that likely visited sophisticated phishing websites be investigated in an effort to identify and thwart the underlying threat actors.

9 Related Work

Because phishing attacks are by nature spread across diverse infrastructure, empirical measurements of the relationships between the different attack phases are difficult. Nevertheless, such measurements can deliver crucial insights that are not possible at a finer granularity. To the best of our knowledge, our work is the first to paint an end-to-end picture of phishing attacks at scale by correlating victim traffic to live phishing websites with attack distribution and monetization.

The work most similar to ours is that of Heijden and Allodi [68], who leveraged methodology similar to *Golden Hour* to correlate URLs in phishing emails (reported to an organization) with the timestamps of clicks by individual victims. The authors combined the click data with email content analysis to identify cognitive and technical factors that characterize successful phishing emails, which can help prioritize the mitigation of high-impact phishing URLs.

Han et al. monitored the life cycles of phishing kits installed on a honeypot [29]. Unlike our approach, the authors captured the credentials sent by each phishing kit and more closely analyzed attackers' interaction with the kit. However, honeypots are limited in scale and scope compared to our approach and do not offer insight into the damage caused by phishing, such as how stolen credentials are ultimately used.

Thomas et al. [62] analyzed a one-year dataset of data breaches, phished credentials, and keyloggers to study trends in the users victimized by such attacks, and the effectiveness of each type of attack. Although this work did not strictly focus on phishing attack anatomy, it underscores the effectiveness of large-scale, cross-organizational data analysis to capture the state of the ecosystem.

Ho et al. [32] analyzed over 113 million emails sent by employees of enterprise organizations to model lateral phishing attacks carried out via compromised email accounts. The authors revealed new types of attacks marked by both sophistication and effectiveness. Although this work does not focus on traditional phishing, it shows that important insight can be gained from analyzing attack data at scale.

Other prior work has scrutinized the time between phishing attack detection and blacklisting [49]. Oest et al. [52] conducted a controlled empirical analysis of the effectiveness of evasion techniques against the response time and coverage of blacklists. The study revealed weaknesses in blacklists and measured the gap between attack detection and mitigation under specific conditions.

In early measurements of the ecosystem, Moore and Clayton analyzed the temporal relationship between the sending of spam emails and the availability of phishing websites [48], and the latency between phishing deployment and detection by anti-phishing blacklists [47]. The authors cited a need for take-down due to the persistence of spam campaigns.

10 Conclusion

At their disposal, phishers have an array of sophisticated techniques that aim to circumvent existing anti-phishing defenses and increase the likelihood of compromising victims. With the addition of underground resources, such attacks are scalable, as has long been observed by the ecosystem [4]. However, the ecosystem itself is not powerless to fight back, as it has access to a wealth of data that can be used to analyze, detect, and prevent phishing. By correlating data from multiple ecosystem sources, we performed a longitudinal, end-to-end life cycle analysis of phishing attacks on a large scale: we not only gained insight into the timing of key events associated with modern phishing attacks, but also identified the gaps in defenses that phishers actively target.

Phishing remains a significant threat to Internet users in part because the *reactive* anti-phishing defenses that are standard throughout the ecosystem, such as browser-based detection and warnings, struggle to effectively address the agility and sophistication of attackers. Importantly, analysis such as that carried out in our research can inform anti-phishing entities of an appropriate response time threshold for specific mitigations, to ultimately narrow the window of opportunity available to phishers.

Our use of the *Golden Hour* framework to automatically secure the accounts of tens of thousands of phishing victims also motivates the continued expansion of *proactive* mitigations within the ecosystem. The framework could be practically adapted by any organization (commonly targeted by phishers) with access to its own phishing URL and web traffic data, and can help seal gaps in defenses by securing compromised user accounts and enabling earlier detection of phishing websites. Moreover, closer collaboration between anti-phishing entities, coupled with the development of enhanced and standardized mechanisms for sharing intelligence, would allow such mitigations to better scale to the ecosystem level.

Acknowledgments

The authors would like to thank the reviewers for their insightful feedback. This material is based upon work supported by the National Science Foundation (NSF) under Grant No. 1703644. This work was supported by the Global Research Laboratory (GRL) program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT, and Future Planning (NRF-2014K1A1A2043029), and by a grant from the Center for Cybersecurity and Digital Forensics (CDF) at Arizona State University.

References

[1] A. Aggarwal, A. Rajadesingan, and P. Kumaraguru. Phishari: Automatic realtime phishing detection on twitter. In *Proceedings of the 2012 APWG Symposium on Electronic Crime Research (eCrime)*, pages 1–12. IEEE, 2012.

[2] E. Alowaisheq, P. Wang, S. Alrwais, X. Liao, X. Wang, T. Alowaisheq, X. Mi, S. Tang, and B. Liu. Cracking the wall of confinement: Understanding and analyzing malicious domain take-downs. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2019.

[3] Anti-Phishing Working Group. APWG Trends Report Q1 2018. https://docs.apwg.org/reports/apwg_trends_report_q1_2018.pdf.

[4] Anti-Phishing Working Group. APWG Trends Report Q3 2019. https://docs.apwg.org/reports/apwg_trends_report_q3_2019.pdf.

[5] L. Bilge, E. Kirda, C. Kruegel, and M. Balduzzi. Exposure: Finding malicious domains using passive dns analysis. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2011.

[6] K. Borgolte, C. Kruegel, and G. Vigna. Meerkat: Detecting website defacements through image-based object recognition. In *Proceedings of the 24th USENIX Security Symposium*, pages 595–610, 2015.

[7] J. E. Brinskelle. Enforcement of same origin policy for sensitive data, Oct. 7 2014. US Patent 8,856,869.

[8] B. Butler, B. Wardman, and N. Pratt. Reaper: an automated, scalable solution for mass credential harvesting and osint. In *2016 APWG symposium on electronic crime research (eCrime)*, pages 1–10. IEEE, 2016.

[9] D. Canali, D. Balzarotti, and A. Francillon. The role of web hosting providers in detecting compromised websites. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13, pages 177–188, New York, NY, USA, 2013. ACM.

[10] S. Chhabra, A. Aggarwal, F. Benevenuto, and P. Kumaraguru. Phi.sh/$ocial: the phishing landscape through short urls. In *Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference*, pages 92–101. ACM, 2011.

[11] M. Cova, C. Kruegel, and G. Vigna. There is no free phish: An analysis of "free" and live phishing kits. In *Proceedings of the 2nd Conference on USENIX Workshop on Offensive Technologies*, WOOT, pages 4:1–4:8, Berkeley, CA, USA, 2008.

[12] Q. Cui, G.-V. Jourdan, G. V. Bochmann, R. Couturier, and I.-V. Onut. Tracking phishing attacks over time. In *Proceedings of the 26th International Conference on World Wide Web*, pages 667–676, 2017.

[13] F. C. Dalgic, A. S. Bozkir, and M. Aydos. Phish-iris: A new approach for vision based brand prediction of phishing web pages via compact visual descriptors. In *Proceedings of the 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, pages 1–8. IEEE, 2018.

[14] R. Dhamija, J. D. Tygar, and M. Hearst. Why phishing works. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI, pages 581–590, New York, NY, USA, 2006. ACM.

[15] R. C. Dodge Jr, C. Carver, and A. J. Ferguson. Phishing for user security awareness. *computers & security*, 26(1):73–80, 2007.

[16] P. Doerfler, K. Thomas, M. Marincenko, J. Ranieri, Y. Jiang, A. Moscicki, and D. McCoy. Evaluating login challenges as a defense against account takeover. In *Proceedings of the World Wide Web Conference*, 2019.

[17] S. Duman, K. Kalkan-Cakmakci, M. Egele, W. Robertson, and E. Kirda. Emailprofiler: Spearphishing filtering with header and stylometric features of emails. In *Proceedings of the Computer Software and Applications Conference (COMPSAC), 2016 IEEE 40th Annual*, volume 1, pages 408–416. IEEE, 2016.

[18] A. Emigh. ITTC report on online identity theft technology and countermeasures 1: Phishing technology, chokepoints and countermeasures. *Radix Labs*, Oct 2005.

[19] J. M. Esparza. Understanding the credential theft lifecycle. *Computer Fraud & Security*, 2019(2):6–9, 2019.

[20] S. Fafinski, W. H. Dutton, and H. Z. Margetts. Mapping and measuring cybercrime. OII Working Paper, 2010.

[21] Business e-mail compromise: The 12 billion dollar scam. https://www.ic3.gov/media/2018/180712.aspx, 2019.

[22] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee. Hypertext transfer protocol – HTTP/1.1, 1999.

[23] S. Garera, N. Provos, M. Chew, and A. D. Rubin. A framework for detection and measurement of phishing attacks. In *Proceedings of the 2007 ACM Workshop on Recurring Malcode*, WORM, pages 1–8, New York, NY, USA, 2007. ACM.

[24] G.-G. Geng, X.-D. Lee, W. Wang, and S.-S. Tseng. Favicon - a clue to phishing sites detection. In *Proceedings of the 2013 APWG Symposium on Electronic Crime Research (eCrime)*, pages 1–10. IEEE, 2013.

[25] Google. Report phishing page. https://safebrowsing.google.com/safebrowsing/report_phish/.

[26] Google. Safe browsing apis (v4). https://developers.google.com/safe-browsing/v4/.

[27] Google. Chrome suspicious site reporter. https://chrome.google.com/webstore/detail/suspicious-site-reporter/jknemblkbdhdcpllfgbfekkdciegfboi?hl=en-US, 2019.

[28] K. Gretzky. Evilginx - Advanced Phishing with Two-factor Authentication Bypass. https://breakdev.org/evilginx-advanced-phishing-with-two-factor-authentication-bypass/.

[29] X. Han, N. Kheir, and D. 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.

[30] Y. Han and Y. Shen. Accurate spear phishing campaign attribution and early detection. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, SAC '16, pages 2079–2086, New York, NY, USA, 2016. ACM.

[31] S. Hao, M. Thomas, V. Paxson, N. Feamster, C. Kreibich, C. Grier, and S. Hollenbeck. Understanding the domain registration behavior of spammers. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 63–76. ACM, 2013.

[32] G. Ho, A. Cidon, L. Gavish, M. Schweighauser, V. Paxson, S. Savage, G. M. Voelker, and D. Wagner. Detecting and characterizing lateral phishing at scale. In *Proceedings of the 28th USENIX Security Symposium*, pages 1273–1290, 2019.

[33] T. Holz, M. Engelberth, and F. Freiling. Learning more about the underground economy: A case-study of keyloggers and dropzones. *Computer Security–ESORICS 2009*, pages 1–18, 2009.

[34] L. Invernizzi, K. Thomas, A. Kapravelos, O. Comanescu, J.-M. Picod, and E. Bursztein. Cloak of visibility: Detecting when machines browse a different web. In *Proceedings of the 37th IEEE Symposium on Security and Privacy*, 2016.

[35] M. Khonji, Y. Iraqi, and A. Jones. Enhancing phishing e-mail classifiers: A lexical url analysis approach. *International Journal for Information Security Research (IJISR)*, 2(1/2):40, 2012.

[36] M. Konte, R. Perdisci, and N. Feamster. Aswatch: An as reputation system to expose bulletproof hosting ASes. *ACM SIGCOMM Computer Communication Review*, 45(4):625–638, 2015.

[37] N. Kshetri. *Cybercrime and cybersecurity in the global south*. Springer, 2013.

[38] Let's Encrypt. Let's Encrypt No Longer Checking Google Safe Browsing. https://community.letsencrypt.org/t/let-s-encrypt-no-longer-checking-google-safe-browsing/82168.

[39] B. Liang, M. Su, W. You, W. Shi, and G. Yang. Cracking classifiers for evasion: A case study on Google's phishing pages filter. In *Proceedings of the 25th International Conference on World Wide Web*, pages 345–356, 2016.

[40] M. Luo, O. Starov, N. Honarmand, and N. Nikiforakis. Hindsight: Understanding the evolution of ui vulnerabilities in mobile browsers. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 149–162. ACM, 2017.

[41] S. Marchal, J. François, R. State, and T. Engel. Phishstorm: Detecting phishing with streaming analytics. *IEEE Transactions on Network and Service Management*, 11(4):458–471, 2014.

[42] S. Matic, A. Fattori, D. Bruschi, and L. Cavallaro. Peering into the muddy waters of pastebin. *ERCIM News: Special Theme Cybercrime and Privacy Issues*, page 16, 2012.

[43] McAfee. Economic Impact of Cybercrime- No Slowing Down. https://www.mcafee.com/enterprise/en-us/assets/reports/restricted/rp-economic-impact-cybercrime.pdf. [Date last accessed 25-August-2019].

[44] N. Miramirkhani, T. Barron, M. Ferdman, and N. Nikiforakis. Panning for gold. com: Understanding the dynamics of domain dropcatching. In *Proceedings of the 2018 World Wide Web Conference*, pages 257–266, 2018.

[45] A. Mirian, J. DeBlasio, S. Savage, G. M. Voelker, and K. Thomas. Hack for hire: Exploring the emerging market for account hijacking. In *Proceedings of the World Wide Web Conference*, 2019.

[46] T. Moore and R. Clayton. How hard can it be to measure phishing? *Mapping and Measuring Cybercrime*, 2010.

[47] T. Moore and R. Clayton. Discovering phishing dropboxes using email metadata. In *Proceedings of the*

2012 APWG Symposium on Electronic Crime Research (eCrime), pages 1–9. IEEE, 2012.

[48] T. Moore, R. Clayton, and H. Stern. Temporal correlations between spam and phishing websites. In *LEET*, 2009.

[49] NSS Labs. Web browser security: Phishing protection test methodology v3.0. https://research.nsslabs.com/reports/free-90/files/TestMethodology_WebB/Page4, Jul 2016.

[50] C. Nykvist, L. Sjöström, J. Gustafsson, and N. Carlsson. Server-side adoption of certificate transparency. In *Proceedings of the International Conference on Passive and Active Network Measurement*, pages 186–199. Springer, 2018.

[51] 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 *Proceedings of the 2019 IEEE Symposium on Security and Privacy*, pages 764–781, May 2019.

[52] 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.

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

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

[55] PhishTank. https://phishtank.com.

[56] G. D. Ruxton. The unequal variance t-test is an underused alternative to student's t-test and the mann–whitney u test. *Behavioral Ecology*, 17(4):688–690, 2006.

[57] S. Sheng, B. Wardman, G. Warner, L. F. Cranor, J. Hong, and C. Zhang. An empirical analysis of phishing blacklists. In *Proceedings of the Sixth Conference on Email and Anti-Spam (CEAS)*, 2009.

[58] H. Siadati, S. Palka, A. Siegel, and D. McCoy. Measuring the effectiveness of embedded phishing exercises. In *10th USENIX Workshop on Cyber Security Experimentation and Test (CSET 17)*, 2017.

[59] A. K. Sood and R. J. Enbody. Crimeware-as-a-service: a survey of commoditized crimeware in the underground market. *International Journal of Critical Infrastructure Protection*, 6(1):28–38, 2013.

[60] StatCounter Global Stats. Desktop vs mobile vs tablet market share worldwide. http://gs.statcounter.com/platform-market-share/desktop-mobile-tablet, 2019.

[61] B. Stone-Gross, T. Holz, G. Stringhini, and G. Vigna. The underground economy of spam: A botmaster's perspective of coordinating large-scale spam campaigns. *LEET*, 11:4–4, 2011.

[62] K. Thomas, F. Li, A. Zand, J. Barrett, J. Ranieri, L. Invernizzi, Y. Markov, O. Comanescu, V. Eranti, A. Moscicki, et al. Data breaches, phishing, or malware?: Understanding the risks of stolen credentials. In *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security*, pages 1421–1434. ACM, 2017.

[63] K. Thomas, D. McCoy, C. Grier, A. Kolcz, and V. Paxson. Trafficking fraudulent accounts: The role of the underground market in Twitter spam and abuse. In *Proceedings of the 22nd USENIX Security Symposium*, pages 195–210, 2013.

[64] K. Tian, S. T. Jan, H. Hu, D. Yao, and G. Wang. Needle in a haystack: tracking down elite phishing domains in the wild. In *Proceedings of the Internet Measurement Conference 2018*, pages 429–442. ACM, 2018.

[65] H. Tu, A. Doupé, Z. Zhao, and G.-J. Ahn. Users really do answer telephone scams. In *Proceedings of the 28th USENIX Security Symposium*, pages 1327–1340, 2019.

[66] E. Ulqinaku, D. Lain, and S. Capkun. 2fa-pp: 2nd factor phishing prevention. In *Proceedings of the 12th Conference on Security and Privacy in Wireless and Mobile Networks*, pages 60–70. ACM, 2019.

[67] US-CERT. Google docs phishing campaign, May 2017. https://www.us-cert.gov/ncas/current-activity/2017/05/04/Google-Docs-Phishing-Campaign.

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

[69] Verizon Enterprise Solutions. Data breach investigations report (dbir). https://enterprise.verizon.com/resources/reports/dbir/, 2019.

[70] L. Von Ahn, B. Maurer, C. McMillen, D. Abraham, and M. Blum. recaptcha: Human-based character recognition via web security measures. *Science*, 321(5895):1465–1468, 2008.

[71] B. Wardman. Assessing the gap: Measure the impact of phishing on an organization. *Annual Conference on Digital Forensics, Security and Law*, 2016.

[72] B. Wardman, M. Weideman, J. Burgis, N. Harris, B. Butler, and N. Pratt. A practical analysis of the rise in mobile phishing. In *Cyber Threat Intelligence*, pages 155–168. Springer, 2018.

[73] C. Whittaker, B. Ryner, and M. Nazif. Large-scale automatic classification of phishing pages. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2010.

[74] M. Wu, R. C. Miller, and S. L. Garfinkel. Do security toolbars actually prevent phishing attacks? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 601–610, New York, NY, USA, 2006. ACM.

[75] G. Xiang, J. Hong, C. P. Rose, and L. Cranor. Cantina+: A feature-rich machine learning framework for detecting phishing web sites. *ACM Trans. Inf. Syst. Secur.*, Sept. 2011.

[76] Y. Zhang, J. I. Hong, and L. F. Cranor. Cantina: A content-based approach to detecting phishing web sites. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 639–648, New York, NY, USA, 2007. ACM.