IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 September 2023, accepted 14 November 2023, date of publication 16 November 2023,
date of current version 27 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3333883

RESEARCH ARTICLE

Demystifying the Regional Phishing Landscape in South Korea

HYUNJUN PARK[1], KYUNGCHAN LIM[2], (Graduate Student Member, IEEE), DOOWON KIM[2], (Associate Member, IEEE), DONGHYUN YU[1], AND HYUNGJOON KOO[3]

[1]NAVER Corporation, Bundang-gu, Seongnam 13561, South Korea
[2]Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville, TN 37996, USA
[3]Department of Computer Science and Engineering, College of Computing and Informatics, Sungkyunkwan University, Jangan-gu, Suwon 16419, South Korea

Corresponding author: Hyungjoon Koo (kevin.koo@skku.edu)

This work was supported in part by the Basic Science Research Program through NRF funded by the Ministry of Education of the Government of South Korea under Grant 2022R1F1A1074373; in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) funded by the Korean Government [Ministry of Science and Information Communication Technology (MSIT)] under Grant 2022-0-01199; in part by the Graduate School of Convergence Security, Sungkyunkwan University; in part by the National Science Foundation (NSF) under Grant 2210137 and Grant 2335798; and in part by the Science Alliance's StART Program and Gifts from Google exploreCSR and TensorFlow.

ABSTRACT The ever-increasing phishing campaigns around the globe have been one of the main threats to cyber security. In response, the global anti-phishing entity (*e.g.*, APWG) collectively maintains the up-to-date blacklist database (*e.g.*, `eCrimeX`) against phishing campaigns, and so do modern browsers (*e.g.*, Google Safe Browsing). However, our finding reveals that such a mutual assistance system has remained a blind spot when detecting geolocation-based phishing campaigns. In this paper, we focus on phishing campaigns against the web portal service with the largest number of users (42 million) in South Korea. We harvest 1,558 phishing URLs from varying resources in the span of a full year, of which only a small fraction (3.8%) have been detected by `eCrimeX` despite a wide spectrum of active fraudulence cases. We demystify three pervasive types of phishing campaigns in South Korea: i) sophisticated phishing campaigns with varying adversarial tactics such as a proxy configuration, ii) phishing campaigns against a second-hand online market, and iii) phishing campaigns against a non-specific target. Aligned with previous findings, a phishing kit that supports automating the whole phishing campaign is prevalent. Besides, we frequently observe a hit-and-run scam where a phishing campaign is immediately inaccessible right after victimization is complete, each of which is tailored to a single potential victim over a new channel like a messenger. As part of mitigation efforts, we promptly provide regional phishing information to APWG, and immediately lock down a victim's account to prevent further damages.

INDEX TERMS Phishing, regional phishing, South Korea.

I. INTRODUCTION

Phishing is a well-known social engineering attack that lures users by sending a fake message or having them access a spoofed website. This leads to the exfiltration of sensitive information such as credentials (*e.g.*, user IDs, passwords) and financial information (*e.g.*, credit cards). According to a report [17] from the Internet Crime Complaint Center, a division of the Federal Bureau of Investigation, phishing (including smishing [67] and pharming [66]) was the most common type of cybercrime back in 2020 and its frequency doubled from the previous year.

The anatomy and ecosystem of evolving phishing campaigns have been extensively studied, including ① varying evasion techniques [25], [31], [41], [54], [60] used to avoid detection, ② phishing kits [7], [10], [23], [33], [45], [69], [70] that support effective scams, and ③ communication channels are used to exfiltrate user credentials [10], [23], [49]. In response to the ever-growing volume of phishing attacks worldwide, the Anti-Phishing Working Group (APWG) [2] has been established to aggregate known phishing domains and URLs from varying organizations. On the other hand,

The associate editor coordinating the review of this manuscript and approving it for publication was Jiafeng Xie.

web browsers offer a default defense tool to protect users from malicious websites, including those that contain malware, unwanted software, or phishing sites. For example, Google Chrome [18], which dominates the browser market on both desktop and mobile platforms [8], offers Safe Browsing [19], which maintains a blacklist database that is updated in real-time. Similarly, Microsoft Edge [35] provides SmartScreen [36], which performs URL reputation checks.

As proprietary phishing information is inaccessible, most previous works rely on the well-known blacklist database (*e.g.*, eCrimeX [3]) collected by APWG [2]. We investigated 1,558 regional URLs related to a phishing campaign over the last 12 months of 2021. However, unfortunately, only 61 cases (*i.e.*, 3.8%) were detected out of 11.1 million cases in eCrimeX. This indicates that a large number of regional phishing cases (at a country level) have been overlooked or remain a blind spot, despite the significant number of fraud cases [59] (*e.g.*, 121.9K in 2020), raising a question about the effectiveness of the up-to-date blacklist in capturing regional phishing campaigns (Section VI describes our take on why such regional phishing may not have been detected). This study aims to better understand the landscape of phishing campaigns in South Korea by evaluating a phishing dataset in collaboration with the leading Web portal company in South Korea, NAVER [39]. As of writing, the leading Web portal company has 700 million monthly active users [24], making it a highly desirable target for stealing sensitive information.

In this work, we introduce PhishingHunter, a phishing detection framework that can effectively detect even sophisticated scams by utilizing varying resources such as sign-in logging, abuse content reporting, Certstream [9], and open source threat intelligence (*e.g.*, OSINT [42]). Based on the observation and dataset, we identify three pervasive types of phishing campaigns in South Korea, which are characterized by regional characteristics: ① sophisticated phishing campaigns, ② phishing campaigns against second-hand online markets, and ③ phishing campaigns against non-specific targets. In particular, our findings reveal that one reason for the overlook of a phishing campaign (against NAVER) in South Korea is due to its ephemeral and multi-channel nature. For example, it is extremely difficult to detect a phishing campaign when an attacker surreptitiously hands over a link via a private channel, making it invalid when victimization is complete (hit-and-run). As a mitigation, we take immediate action by ① locking down the victim's sign-in with an abuse notice, and by ② reporting the discovered phishing sites to the central blacklist database (*i.e.*, eCrimeX by APWG).

The following summarizes our contributions.

- We reveal that the global blacklist for blocking phishing sites has been overlooking a regional phishing ecosystem.
- We introduce PhishingHunter, which can systematically detect potential phishing sites by utilizing varying resources and analysis techniques.
- We confirm 1,558 phishing campaigns in the span of a year against the largest web portal site (*i.e.*, NAVER) in South Korea, unveiling its landscape by identifying three prevalent types.
- We begin to push regional phishing campaigns to the global anti-phishing association (*i.e.*, APWG), preventing further damages as part of our mitigation efforts.

The rest of this paper is organized as follows. We illustrate the background and motivation of our work in Section II. Then, we take a deep look into the regional phishing landscape in South Korea in Section III. Next, we describe three types of phishing campaigns from Section IV to Section VI. Section VII describes several mitigation techniques with strengths and downsides. Section VIII holds the discussion and limitations of our work, followed by related work in Section IX. Finally, we give the conclusion in Section X.

II. BACKGROUND AND MOTIVATION

This section describes the background of phishing and the motivation for our work.

A. PHISHING CAMPAIGNS AND DEFENSES

1) PHISHING KITS

Phishing attackers often use phishing kits [4], [23], ready-to-deploy software packages that allow for phishing without a strong technical background. The kits typically include a mimicked login form of the target website and its source code (*e.g.*, HTML, JavaScript, CSS) and images, aimed at tricking victims into revealing their credentials. Besides, the kits provide features to automate the phishing campaign, such as an easy installation process that can be quickly relocated in case of discovery and a tool to automatically store victims' information (*e.g.*, IP) and send it back to the attackers.

2) PHISHING DEFENSES

To prevent phishing attacks, security practitioners and Web browser vendors have developed effective defense mechanisms as follows.

- *Anti-Phishing Working Group*: The Anti-Phishing Working Group, APWG [2], is an industry association of anti-phishing entities, which maintains the largest blacklist of phishing URLs (eCrimeX [3]) collected from its member organizations. Many phishing defense mechanisms rely on this blacklist [22], because it efficiently blocks a real-time phishing campaign.
- *Browser Defenses*: Browsers also provide built-in defense tools to protect against malicious online activities like malware, unwanted software, and social engineering. For example, Google Chrome [18] is one of the most popular browsers on both desktop and mobile platforms [57], [58] (*i.e.*, around four billion users). It offers Safe Browsing [20], which maintains a blacklist of web resources containing malicious content. Similarly, the Safari [1],

Firefox [37], and Vivaldi [62] browsers leverage Chrome's blacklist database into blocking potential threats [16]. Meanwhile, the Microsoft Edge browser [35] provides SmartScreen [36] with reputation checks.

- *User Education and Awarenesses:* Another approach to mitigate users from accessing phishing content [48] is through user education or user awareness. However, it cannot be used as a sole method of defense. Khonji et al. [27] observe that equipping with a useful user interface or the enhanced behavior of a system is required for effective user education and awareness.

B. MOTIVATION

Phishing attacks have been well studied and understood. However, prior works have been typically conducted from a global perspective that heavily relies on the largest blacklist of phishing URLs (i.e., eCrimeX [3]). We raise a research question; can a global phishing database capture regional phishing attacks in a timely manner? To answer this question, we confirmed if a phishing campaign against NAVER in Korea, the leading Web portal company [39], has been covered by eCrimeX. We collect 11,097,299 (11.1 million) phishing URLs during our observation period between Jan. 1st, 2015, and Dec. 31st, 2021 (7 years) from the eCrimeX database [3]. Then, we attempt to seek any phishing target that contains both the brand name (*i.e.*, NAVER) and domain squatting [55].

1) TARGET BRAND NAME

The most popular (misused) target brands are Facebook (1,711,698 URLs), Apple (1,635,887 URLs), Paypal (549,819 URLs), Yahoo (556,541 URLs), and Bank of America (284,091 URLs). These top-five brands account for 42.7% (4,738,036) of the whole blacklisted phishing URLs (11.1 M). However, NAVER has been barely captured and reported to the blacklist; merely 61 phishing URLs were reported even with large fraud cases [59].

2) DOMAIN SQUATTING

Domain squatting [55], [64], [68] is one of the most common techniques in phishing attacks by generating the lookalike domains of a target. For example, a homograph [64] is an attack where an adversary can misuse similar-looking characters such as the letter ‘ l’ and the number ‘ 1’: *e.g.*, appIe.com (the letter ‘ l’ is replaced with a number ‘ 1’). With the DNSTwist [61] tool that supports various squatting techniques (*e.g.*, typosquatting [68], combosquatting [68], homograph [64]), we generate more potential phishing domains against NAVER, additionally obtaining 1,400 squatting domains. However, none of them were discovered in the database.

3) MOTIVATION

The above results indicate that a regional phishing target (*e.g.*, NAVER in South Korea) may have remained in a blind spot from a global view, which aligns with the observation from PhishFarm [43]. According to Statista [59], the number of fraud cases reaches up to 122 thousand in 2020 for online second-hand shopping in South Korea alone. This motivates us to look into a regional phishing ecosystem for a better understanding of phishing attacks at a country level.

III. REGIONAL PHISHING LANDSCAPE IN S.KOREA

We aim to better understand the landscape of phishing campaigns in South Korea. To this end, we introduce a comprehensive phishing detection framework, dubbed PhishingHunter. With the framework, we delve into the regional phishing ecosystem in Korea, being able to identify three notable types.

A. PHISHING DETECTION FRAMEWORK

Figure 1 concisely illustrates the workflow of PhishingHunter for identifying a phishing campaign, which leverages varying resources to efficiently disclose even a sophisticated one.

1) OVERVIEW

The PhishingHunter framework comprises three main components: collecting information related to phishing, classifying and analyzing the information, and blocking the phishing attempts to prevent further damage. To this end, PhishingHunter leverages both in-house systems (e.g., user sign-in management, abuse content monitoring, spam detection) and external resources (e.g., open-source intelligence (OSINT), certificate monitoring information) to maximize the data amount to determine a genuine phishing campaign, maintaining the database of an aggregate from every information. Adversaries typically aim to hijack user accounts, centering their attacks around those services. Consequently, phishing attempts against NAVER accounts often exploit various services like e-mails, blogs, and other community features. The in-house systems (*i.e.*, logging and monitoring) incrementally store a variety of records associated with the services. These data trails can be subsequently used to gather valuable information about potential phishing attacks.

2) IN-HOUSE RESOURCE COLLECTION

There are three internal resources: ① the user's sign-in fingerprinting records that aid in profiling phishing adversaries and identifying victims, including IP addresses, geolocation, access device details, and user-agent at the time of user login; ② the Abuse Content Monitoring system that extracts suspicious URLs (i.e., potential phishing attempts) from email contents or user reports; and ③ the Spam Detection system that plays a critical role in recognizing phishing attacks against targeted users. In a nutshell, it collects metadata from a user's phishing report such as a subject, sender, and internal links, and flags a suspicious URL when discovering a known phishing pattern.

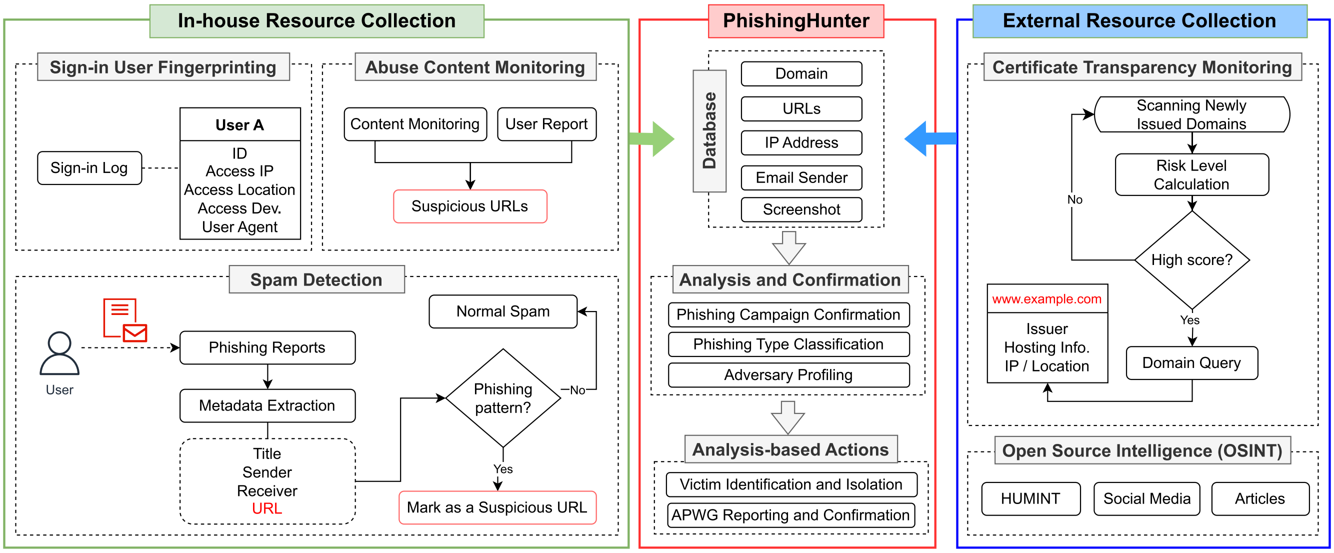

FIGURE 1. Overall PhishingHunter workflow. The PhishingHunter framework consists of three components: central database, monitoring and analysis, and taking actions based on analyses. The database collects information from both in-house resources (e.g., user sign-in management, abuse content monitoring, spam detection) and external resources (e.g., open-source intelligence, certificate monitoring information) to determine a genuine phishing campaign. We identify three types (Section IV, Section V, and Section VI) of regional phishing campaigns that are prevalent in South Korea with PhishingHunter.

3) EXTERNAL RESOURCE COLLECTION

There are two external resources: ① Certificate Transparency Monitoring with the `Certstream` tool [9] to oversee newly registered domains. This is because attackers often register domains mimicking legitimate service certificates to bypass HTTPS browser warnings. Suspected domains are filtered out based on a risk score as an early indicator; and ② open-source intelligence (OSINT) that collects phishing information from various sources such as human intelligence (HUMINT), social media, and reports issued by security firms or organizations. OSINT supplements internal data collection to assist phishing practitioners in conducting further investigations on a potential phishing campaign, thereby reducing false positives.

4) PhishingHunter

The ultimate objective of PhishingHunter is to swiftly detect a phishing campaign and mitigate potential damage. First, PhishingHunter maintains a central database that stores the information from both in-house systems in `NAVER` and external resources. Second, PhishingHunter routinely performs (suspicious) URL surveillance and analysis along with the collected information in the database, confirming a genuine phishing campaign. This aids to identify three pervasive phishing campaign types in South Korea (Section III-B): ① sophisticated phishing campaigns (Section IV), ② the ones against second-hand markets (Section V), and ③ the ones against a non-specific target (Section VI). Third, with a thorough analysis, PhishingHunter makes it possible to take immediate actions by blocking victims' sign-in by informing them of fraudulence once (manually) confirming a phishing URL, and proactively registering the URL into the APWG's database. Note that the three categories are heuristic (rather than an absolute classification) because they explain the regional phishing landscape in South Korea well.

B. PHISHING CAMPAIGN TYPES

In this section, we demystify three pervasive types of regional phishing campaigns against `NAVER` [39].

1) PHISHING CAMPAIGN TARGET

Offering an all-in-one platform service (*e.g.*, search engine, email, blog, map, news, cloud, messaging, interpretation), `NAVER` owns a large number of users.[1] This makes it one of the most preferred targets for a phishing campaign, and it is one of the differentiation factors of the PhishingHunter from other detection frameworks. Such a large number of users naturally form various groups that dive into their interests, hobbies, or passions, leading lots of communities (called a "cafe" service) like Reddit [53]. Of all, the largest one is an online second-hand market community, the `Joonggonara` cafe [38], whose number of users reaches up to 19 million. Such unprecedented popularity of the community becomes a ceaseless phishing campaign target due to the nature of inevitable monetary transactions (*e.g.*, one makes or receives a payment). Indeed, the number of (recognized) online fraud cases in the community has been skyrocketing [59].

2) TYPES OF PHISHING CAMPAIGNS

With PhishingHunter, we identify three pervasive types of phishing campaigns (PCs) in South Korea. First, we characterize sophisticated PCs (Section IV) that involve adversarial tactics to circumvent detection. This type involves in-person

[1] The enrolled Korean users in `NAVER` reach up to 42 million [65] in 2017, which accounts for more than 80% in the whole Korean population. It currently holds 700 million monthly active users [24] around the globe.

TABLE 1. Summary of our collected dataset. We detect 1,588 phishing campaigns (PCs) in South Korea with PhishingHunter, identifying three pervasive types.

Type	Sophisticated PCs	PCs against the Second-hand Market	PCs against Non-specific Targets	Total
Domains	317 (20.3%)	551 (35.4%)	690 (44.3%)	1,558 (100%)

contact from an adversary via spam or junk at all times. Interestingly, every case in our dataset adopts sophisticated (but quite similar) adversarial tactics and techniques, which support our hypothesis that a phishing site may have been generated by an automated tool. Second, we delineate PCs against second-hand online markets (Section V), which adopts the hit-and-run way of a phishing campaign with a phishing toolkit [4], [23]. Notably, unlike sophisticated PCs, this type entails in-person contact with a victim (with a bait item) ahead of accessing a phishing site. Third, we investigate PCs against a non-specific target (Section VI) based on misuse reports received from users, which utilize a suspicious domain name or a clumsy-looking website. The last type of phishing site is not restricted to a victim (i.e., accessible by anyone). Note that there might be other PCs that have not been detected.

C. DATASET COLLECTION

We collected our dataset for 12 months in 2021. Table 1 summarizes the 1,558 unique phishing domains in the wild. Note that we count domain-based PCs because a unique URL has been created based on an individual victim or item.

1) SOPHISTICATED PHISHING CAMPAIGNS

We investigate 145 thousand of TLS certificate updates per hour on average.[2] As the volume of updated information is enormous, we devise a risk score for a domain name that represents potential harmfulness (Section IV-A). We obtain $6,145$ domains in total, whose risk scores are above the pre-defined threshold ($T = 0.5$) configured with our heuristics. Based on potential PCs with the score, we further investigate them using the collected resources like sign-in logging, spam and abuse-contents reporting, and adversary profiling. Finally, we manually confirmed 317 phishing campaigns by leveraging a comprehensive analysis.

2) PHISHING CAMPAIGNS AGAINST A SECOND-HAND MARKET

Unlike other phishing campaigns in general, a fraudulent phishing URL for the second-hand online market (*i.e.*, `Joonggonara` [38]) would not be exposed until interacting with an adversary because the adversary reveals an actual URL to a victim only when expressing the purchase intent. To do so, the adversary posts a bait item, guiding the victim to leave the ID of an instant messenger instead of directly disclosing the phishing URL. Based on our profiling with the central database, we acquire every phishing campaign URL by pretending to buy a certain item and approach the adversary through the messenger. Besides, we leverage a passive DNS (pDNS) [13] to obtain the list of past domains for the phishing campaign, collecting 551 unique domains.

3) PHISHING CAMPAIGNS AGAINST A NON-SPECIFIC TARGET

Revealing a concealed phishing campaign via an individual email is challenging, because monitoring email content is illegitimate under the Personal Information Protection Act (PIPA) [26] in South Korea. Hence, we obtain a phishing campaign URL only in the presence of reporting an abusive case (*i.e.*, Figure 1). We collect 690 for such domains.

IV. SOPHISTICATED PHISHING CAMPAIGNS

This section depicts our strategies for detecting sophisticated phishing campaigns in South Korea, followed by exploring adversarial tactics to avoid detection.

A. STRATEGIES FOR EFFECTIVE DETECTION

As in Figure 1, we aggregate collected information into a single database, including suspicious TLDs, sub-domains, patterns of common words, and misuse of domain squatting for effective detection of a phishing campaign.

1) DOMAIN RISK SCORE

To determine a dubious domain pertaining to a phishing campaign, we define a *risk score* considering the following factors: ① a domain carries a yet-known subdomain (excluding TLD; Top Level Domain), ② each keyword (by splitting a domain with a dot delimiter) has been discovered from the known signatures including a brand name and suspicious words, ③ heuristically a domain contains a hyphen or a dot above our threshold, and ④ a domain certificate has been issued with free of charge like `Let's Encrypt` [15] or `ZeroSSL` [71]. Although free certificates are not problematic in general, our finding shows that the combination of the above factors increases the chances involved with a phishing campaign in practice as the certificates issued by these CAs have been indeed involved in phishing attacks [28]. Table 2 summarizes the factors for computing a risk score with concrete instances. As an example, `members.never.com` has a lookalike domain of the original domain, `naver.com` (domain squatting). Similarly, `navercop.com.co` contains a TLD of `.com` as a subdomain (suspicious TLD). Finally, the aggregate of each risk score per factor evaluates a certain domain, in which the higher score represents a potentially risky domain. In our experiment, we empirically set the threshold ($T = 0.50$) that determines the candidate for a phishing campaign. The risk score of a domain for sophisticated PCs is 0.53 on average.

[2] The number counts all updated notifications including new domain registrations when `Certstream` polls the whole certificate lists with changes in the Merkle tree.

TABLE 2. Primary factors for computing a domain risk score to determine a suspicious domain. The score per each factor represents an empirical weight, which helps to filter out a suspicious domain that may be associated with a phishing campaign. For instance, the brand keyword (*e.g.*, NAVER) in a domain name is the most prevalent factor. Note that the risk score has been carefully adjusted based on a user's phishing report (ground truth), which can be parameterized.

Factors	Risk Score	Concrete Example
Suspicious TLD	0.087	navers.co.{in}
TLD as a subdomain	0.087	navercop.{com}.co
Brand keyword	0.434	{naver.net}.in
Suspicious keyword	0.087	nid.never-{cloud}ing.com
Domain squatting	0.173	members.{never}.com
Number of hyphens	0.044	{nid.naver.com-user06-nidlogin}.me
Number of subdomains	0.044	naver{.}nid{.}coms{.}party
Certificate with free of charge	0.044	(Let's Encrypt or ZeroSSL) certificate

2) ADVERSARY PROFILING

Our monitoring system along with the database (Figure 1) can assist in adversary profiling for further investigation. As the origin of phishing scam emails has been often manipulated (*e.g.*, a popular sender name is no-reply), it is challenging to track a true sender. We discover that one of the notable behaviors is that the sender tends to include oneself as the first recipient to confirm if a phishing email successfully circumvents a spam filter before bombarding the scams. This often allows us to reveal the original attacker who first sends an email that contains a phishing URL, reasoning that the attacker is part of a phishing group. Indeed, we were capable of quick identification with such profiling information (*e.g.*, title, body, sender IP, and sender address) from the years of history.

B. *ANALYSIS OF ADVERSARIAL TACTICS*

Phishing attackers take varying evasion tactics: ① a phishing website appears to be a seemingly benign-looking one, and ② the website avoids (or at least postpones) early detection. This section describes our findings on adversarial tactics for sophisticated phishing campaigns against NAVER.

1) FRAUDULENT WEBSITE WITH HTTPS

A majority of phishing scams employ a secure HTTPS protocol (82% according to PhishLabs [6] and Kim et al. [28]) by adopting TLS certificates mostly free of charge that are issued by ACME CAs (Automatic Certificate Management Environment Certificate Authorities) such as Let's Encrypt [15] and ZeroSSL [71]. It allows an adversary not only to be able to avoid a browser warning of missing a valid certificate but also to prepare another fake site quickly when being blocked without worrying about an additional cost for issuing the certificate.

2) CREDENTIAL REDIRECTION WITH A PROXY CONFIGURATION

Our finding indicates that an elaborate phishing campaign leverages a proxy configuration to actively deceive a victim (❶ in Figure 2). Namely, the victim provides a credential, followed by sending it to a real server after stealing the credential. The victim would receive benign responses (*i.e.*, successful login) back from the server over the end-to-end transaction. A redirection by relaying requests and responses gives another benefit to an adversary because it is possible to exfiltrate a valid credential solely by confirming if a response from the server has been properly responded with the credential. On the contrary, it allows the server to distinguish end users who have been victimized with ease because the source IPs of successful sign-ins would match a proxy IP that is controlled by the adversary (instead of clients' IPs), making a list of victims recognizable.

TABLE 3. Example of email titles to entice users to a phishing website. Most of the titles include attention-grabbing information. Note that the first four categories are from the keywords filtered with an account, sign-in, email, and certificate, respectively.

Category	Title Examples	Count
Account	ID conflict, and Password confirmation	15,065 (39.1%)
Sign-in	New device alarm disabled, Login attempt from abroad, and Login success from a new device	13,369 (34.7%)
Email	Mail backup request, Email update, and Protection disabled	3,506 (9.1%)
Government-issued Certificate	One time password notice, New certificate issued, and Two-factor authentication disabled	3,101 (8.0%)
Other Personal Information	Legal name updated, Cell phone updated, and Private information leaked	2,500 (6.5%)
Impersonation	Contents that disguise a sender as a government, hospital, or banking service	828 (2.1%)
Others	Test, and randomly-looking letters	180 (0.5%)
Total		38,549 (100%)

3) CIRCUMVENTING TECHNIQUES

Attackers utilize bypassing techniques not to be captured by phishing hunters. One of the pervasive techniques is that a phishing site can be accessible solely when a certain condition is met where an empty page or arbitrary website would be returned/redirected otherwise. Such conditions include ① IP blacklist; excluding known spam filtering IPs or disallowing all accesses but a limited number of regional IPs, ② User-agent; avoiding exposure from well-known crawling bots, ③ referrer; checking if access has been made in an intended fashion, and ④ parameter; allowing one to access a phishing campaign only when the combination of certain parameters has been passed.

C. *IN-DEPTH ANALYSIS ON TARGETS AND VICTIMS*

1) EMAIL TITLES

Table 3 shows email titles to entice users to access a phishing website. We classify the titles into six categories including account, sign-in, email, government-issued certificate, personal information, and impersonation. The majority of the titles (73.8%) involve with user's account information like a credential or login to have one pay attention.

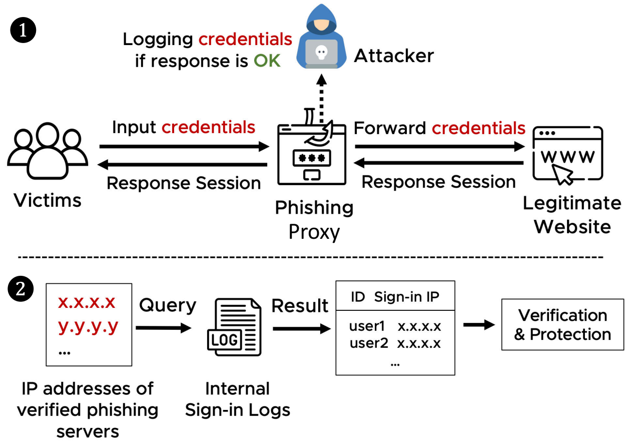

FIGURE 2. Proxy configuration for redirecting a victim's credential. Using the proxy server not only helps an attacker to filter out a true credential but also a defender to recognize a victim as follows. ❶ An attacker obtains a working credential when a victim has successfully signed in to a target website. ❷ Our sign-in logging system returns a list of (proxy) IPs with successful logins, which identifies actual victims.

2) TRUE VICTIMS

Recall that we keep traces of dubious domains with a domain risk score (Section IV-A). Figure 2 demonstrates how we could pinpoint true victims by leveraging a proxy mechanism. It is possible to obtain the list of the victims by looking up a proxy IP (client) because their successful logins have been recorded separately. This aids a defender in recognizing victims, which would be challenging without a proxy configuration otherwise. Note that we took an immediate action to protect those victims by locking their accounts, followed by notifying them.

3) PHISHING DOMAINS AND IPs

We further investigate the geographical distribution of proxy hosting IPs (owned by attackers), source IPs of the senders, TLDs, and domain keywords from sophisticated phishing campaigns. Table 4 shows the number of unique proxy IPs and phishing senders by country. Most scamming emails were sent from servers in the United States whereas four out of five proxy servers are located in South Korea. Figure 3 and Table 5 illustrate the Top 20 domain keywords and Top 10 TLDs that are often used in elaborated phishing scams. nid is the most popular domain keyword presumably because the legitimate login page[3] contains that keyword, followed by navercorp, mail and www. For TLDs, our finding shows that half of Top 10 ranking end with .kr, which evidently differs from other TLDs collected by APWG [21] or Phishlabs [50]. The Top 10 domain keywords from APWG and Phishlabs databases are com, uk, org, net, xyz, live, br, link, info and me, and com, org, ca, io, net, mx, com, uz, monster and ae, respectively. Note that we consider .o-r.kr, .n-e.kr, .r-e.kr, .n-e.kr as TLD, although they are technically not. This is because such domains are offered by a Korean free domain website[4] that attackers can readily obtain such free domains for phishing attacks such as naver-login.o-r.kr.

TABLE 4. Geographical locations (at a country level) by the origin IPs for both proxy servers and phishing senders.

Country	# of Origin IPs	# of Scam Senders
DE	1 (1.64%)	10 (4.35%)
GB	3 (4.92%)	-
GR	-	1 (0.43%)
HK	1 (1.64%)	2 (0.87%)
JP	2 (3.28%)	2 (0.87%)
KR	45 (73.77%)	17 (7.39%)
NL	1 (1.64%)	-
SG	1 (1.64%)	3 (1.30%)
US	7 (11.48%)	195 (84.78%)
Total	61 (100%)	230 (100%)

V. PHISHING CAMPAIGNS AGAINST SECOND-HAND MARKETS

This section elaborates on a pervasive fraud campaign in South Korea against a second-hand online market. We demystify a phishing pattern against the second-hand marketplace in South Korea.

A. SECOND-HAND ONLINE MARKET

As online shopping becomes a norm in our daily lives, so does a second-hand e-commercial market that helps one to buy, sell, or exchange used items in an inexpensive manner. One of the largest second-hand online markets in South Korea, Joonggonara [38], is managed by NAVER [39] as a community service, which is analogous to eBay [14]. Such a virtual market becomes quite attractive to phishing campaigners due to the high volume of community members where the transaction scale of the second-hand market in Korea alone reaches up to approximately 20 billion US dollars in 2021 [12]. Specifically, an adversary posts a *bait* item at a lower cost than its market price, awaiting a victim by captivating those who seek a seemingly great deal.

B. HIT-AND-RUN TEMPORARY PHISHING URLs

Figure 4 briefly illustrates the whole process of a phishing campaign. We hypothesize that an adversary leverages a phishing kit to build a fraudulent scam because its content is equivalent to other campaigns against the second-hand marketplace while a surreptitious domain is *temporarily* available. The phishing kit offers automation of the whole phishing processing including the preparation of a phishing website (❶) and the exhibition of a bait item (❷). When a victim approaches the phisher over an instant messenger (❸), the phishing site (*e.g.*, using a squatting domain) is *temporarily* launched (❹), and then delivered to the potential victim (❺). The lifespan of a scam website is ephemeral because the site is unavailable right after either a credential exfiltration

[3] https://nid.naver.com/nidlogin.login

[4] https://xn–220b31d95hq8o.xn–3e0b707e

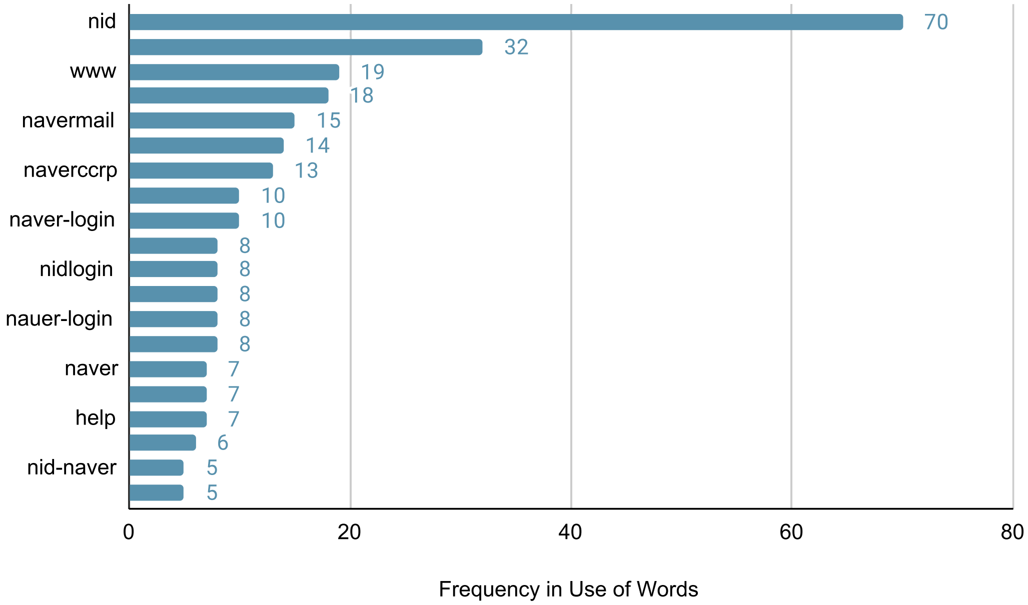

FIGURE 3. Top 20 domain keywords frequently used in sophisticated phishing campaigns.

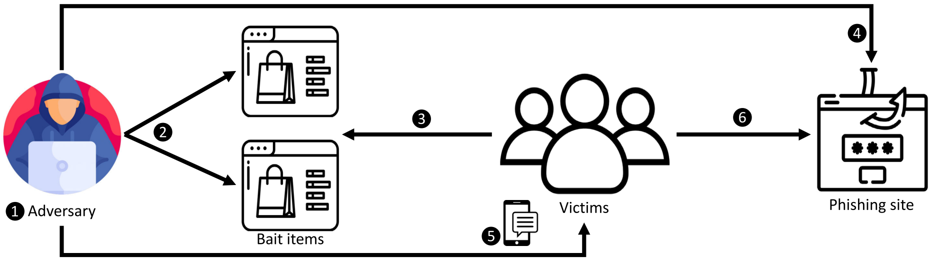

FIGURE 4. Overview of a phishing campaign with a phishing kit against the second-hand online market, the `Joonggonara` community [38]. An adversary sets up a phishing campaign website with the kit (❶), followed by displaying a bait item in the community (❷). The attacker waits for a victim until the victim approaches over a private instant messenger (❸). Then, a fraudulent URL is temporarily launched (❹) and delivered to the victim via the messenger (❺). The phishing URL is immediately closed once a financial transaction (*i.e.*, fraud) is complete (❻).

TABLE 5. Top 10 TLDs that are frequently used in sophisticated phishing attacks. Around 30% of TLDs end with the `xyz.kr` format, differing from global cases [21].

TLD	Count (%)	TLD	Count (%)
.com	110 (43%)	.org	13 (5.1%)
.kro.kr	38 (14.8%)	.info	12 (4.7%)
.ca	21 (8.2%)	.o-r.kr	10 (3.9%)
.ml	17 (6.6%)	.r-e.kr	10 (3.9%)
.net	16 (6.3%)	.p-e.kr	9 (3.5%)
Total			256 (100%)

or a financial transaction is completed (❻). The phishing URLs are seldom revealed due to the nature of such a *hit-and-run* type via a separate channel (*e.g.*, instant messenger). Note that we observed that the sensitive information obtained from the victim can be compromised for a phishing attack in the future. For example, an attacker takes advantage of a stolen credential to post another bait.

VI. PHISHING CAMPAIGNS AGAINST A NON-SPECIFIC TARGET

This section portrays the type of phishing campaigns against a non-specific target (*i.e.*, abuse cases directly filed by users), which differs from the other two types in that a victim has not been destined.

A. ARBITRARY PHISHING DOMAIN

We additionally investigate 690 phishing domains against a non-specific target, being neither sophisticated nor relevant to the second-hand market. A large number of domains (75.07%) do not even include any brand keyword or squatting technique in choosing their domain names. A scam page often resides in a sub-directory (with an arbitrary name) at the irrelevant domain, which we highly speculate that the phishing website has been running on a compromised machine.

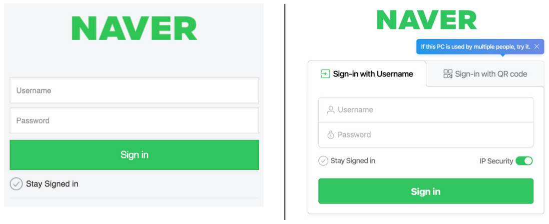

FIGURE 5. Example of a phishing campaign against a non-specific target. Oftentimes the quality of a phishing website (left) looks crude, clearly distinguished from the legitimate site (right) (*e.g.*, lacking a tab of “ sign-in with QR code”). Besides, it does not employ a secure end-to-end communication (*e.g.*, HTTPS connection).

B. POOR QUALITY OF FRAUDULENT SITES

Unlike sophisticated phishing campaigns (Section IV) or the ones against Joonggonara (Section V), the mimicking quality of phishing sites is relatively poor. First, it is trivial to distinguish the appearance of a scam from that of the legitimate NAVER with bare eyes. Figure 5 shows the example of a comparison between a phishing page on the left and the legitimate one on the right. The scam page has replicated an obsolete sign-in page (possibly no updates since then), which lacks multiple sign-in channels. Second, a phishing campaign involves no interaction with an authentic server (recall that a proxy relays incoming and outgoing messages), thus a victim encounters a simple error page or a redirected one once they have been victimized (*i.e.*, offering a credential). Lastly, the scam site with a crude appearance rarely uses HTTPS, which would be easily detectable for those who exercise security awareness training.

C. ADVERSARIAL TACTICS FOR EFFICIENT CAMPAIGNS

A phishing campaign owner often utilizes a compromised server[5] to inject malicious content. A typical adversarial tactic for persistently keeping such content is to separately store them in multiple locations in case a certain (compromised) server is not available (*e.g.*, content elimination, server down). For example, the login page that a victim faces can be stored in a compromised server, whereas its underlying JavaScript file and a PHP source can be requested from other servers. It is noted that the <script> tag is not subject to the same origin policy (SOP) [63], which can execute external content retrieved from foreign origins.

VII. MITIGATIONS

In this section, we describe several mitigation techniques with their strengths and weaknesses, and our efforts to reduce damages from a phishing campaign.

A. BLACKLIST-BASED APPROACH

One of the major issues is that the URLs and domains discovered at the regional phishing campaigns (*e.g.*, in South Korea) have been barely captured in a global blacklist. A simple but quite effective mitigation [43] would be reporting such phishing information to the existing services that maintain the blacklist database such as APWG [2], OpenPhish [47], and PhishTank [51]. As part of our efforts, we have started reporting the collection of phishing campaigns against NAVER to APWG, which has considerably increased up to 1,982 phishing URLs (from 61 cases for the last 7 years) by April in 2022 alone. We believe that it would globally help to reduce potential phishing attempts by blocking them. However, one downside of the (well-maintaining) central blacklist database can come at a (non-negligible) cost. For example, APWG allows only membership organizations to pull and push phishing sites. Besides, multiple blacklist databases across different browsers make it difficult to maintain up-to-date phishing information interchangeably.

B. BROWSER-BASED APPROACH

Similar to Safe Browsing [20] or SmartScreen [36], NAVER provides an additional blacklist database in response to a quick rise and demise of regional phishing campaigns. [34], [44] In particular, NAVER developed the Whale Browser [40] based on Chromium [52], which has been equipped with the (on-the-fly) database by default. As of writing, Whale takes up around nine percent (9.12%) of the browser market share [56] in South Korea. The browser-based approach (under the same jurisdiction) offers an alternative means with quick updates against regional phishing cases, however, it may not be ideal because its effectiveness relies on the usage of a certain browser. We confirmed that 1.1M Whale Browser users were prevented from accessing phishing campaigns due to PhishingHunter in 2022 alone.

C. MITIGATION STRATEGIES

Various mitigation strategies can be implemented to minimize both phishing attacks and the associated victim count. The specific strategies depend on the nature of the phishing attempts. Browser defense represents a prevalent method for mitigating a broad range of phishing attacks. This strategy involves identifying domains associated with phishing activities, sharing this information with the Anti-Phishing Working Group (APWG), and subsequently incorporating these domains into the Safe Browsing blacklist. This process effectively limits further user access to these identified phishing sites. Nonetheless, phishing attacks specific to certain regions are often not successfully blocked by global browsers, necessitating the use of region-specific browsers and corresponding blocking databases.

The utilization of account lockouts serves as an effective measure to prevent further exploitation of an account that has been compromised due to a phishing attack. Sophisticated phishing methodologies can identify potential victims based on the receipt of phishing emails and discrepancies in their login history. Unspecified target phishing often leverages compromised servers, where an attacker's inadvertent error may lead to a directory listing, thus exposing the amassed account list in the form of a TXT file. If an account

[5] An attacker may either directly compromise a server or purchase it from elsewhere, which is beyond the scope of this work.

is flagged as a phishing victim or suspect under such circumstances, immediate account lockout can mitigate further damage, including personal information disclosure. Conversely, when phishing attacks target secondary markets, victim identification proves to be challenging, as the phishing domain is typically disseminated through private messaging platforms.

A final mitigative strategy entails obstructing the delivery path of the phishing domain. Sophisticated phishing campaigns are typically propagated through phishing emails. Hence, information associated with these emails, such as the sender's email address, SMTP IP, email subject, and phishing links within the email content, can be employed to profile the attackers and delay the spread of phishing emails by blocking their transmission. In the context of secondary markets, post monitoring can serve to limit the dissemination of phishing sites by obstructing posts that redirect users to messenger platforms or posts from users who have previously partaken in illicit activities. However, when it comes to phishing campaigns against unspecified targets, the application of these methods becomes challenging due to the indeterminate source of dissemination.

D. OTHER EFFORTS

As part of our mitigation efforts, it is of significance to stop further damage from stolen information as well as the detection of phishing campaigns. We watch the Joonggonara community against bait items, proactively trace a culprit who posts them, and lock down victims' accounts in a timely manner. Meanwhile, we keep informing credential leaks to the victims of a sophisticated phishing campaign, blocking further misuse. Lastly, we look up additional locations that store a certain malicious content (*e.g.*, JavaScript file) by leveraging the crawling result database of the NAVER search engine, followed by blocking them with the Safe Browsing engine in the Whale Browser [40].

VIII. DISCUSSION AND LIMITATION

This section covers in-depth discussions and limitations of our work.

A. SECURITY IMPLICATIONS

Our study reveals that one of the reasons that overlook a phishing campaign (against NAVER) in South Korea is mostly because of its ephemeral and multi-channel properties. Such a region-oriented characteristic is well aligned with previous work [4], [23]. However, as shown in Table 6, there has yet been regionally focused work in phishing analysis. From a global perspective, it is well known that a phishing ecosystem has been evolving by adopting various adversarial tactics such as efficient exfiltration of sensitive information and evasion techniques from detection. However, there is a slight deviation in that non-negligible attempts target the second-hand marketplace community in Korea. Particularly, a hit-and-run approach makes it difficult to hunt down scam URLs and domains because they are both ephemeral (*i.e.*, no longer accessible immediately after a phishing attack is complete) and tailored to a potential victim (*i.e.*, surreptitiously handing phishing URLs in over another channel). It is prevalent to employ a phishing toolkit for handy deployment that enables an adversary to apply the hit-and-run tactic. Besides, a victim has been promoting another phishing campaign without one's perception by providing a victim's credential (*i.e.*, an attacker utilizes the credential to post a bait item). On the other hand, non-sophisticated phishing campaigns are still prevalent against a non-specific target, such as running a fraudulent site on a compromised server. Based on our key findings, regional phishing campaigns have been underestimated, necessitating a global blacklist database via better mutual assistance. Furthermore, we claim that a feature of filtering out the blacklist is *required to be equipped with a messenger and other communication tools* (as well as a browser) toward comprehensive protection.

TABLE 6. Comparison of ours with previous works.

	Region Focus	Target Specific	Ecosystem Focus
PhishingHunter	●	●	●
Oest et al. [46]	○	○	●
Cova et al. [10]	○	●	●
Han et al. [23]	○	●	●
Zawoad et al. [70]	○	●	●

●= Focused, ○= Not Focused

B. KIMSUKY APT GROUP

As a case study, we hypothesize that Kimsuky [29], [30] (also known as Velvet Chollima or Black Banshee), the advanced persistent threat group (APT) in North Korea, may have been involved based on our investigation with intelligence resources and supporting evidence from prior study [29], [30], [32]. According to the report from Cybersecurity and Infrastructure Security Agency [11], 34 out of 111 (around 31%) domains owned by Kimsuky contain NAVER in a domain name. It states that Kimsuky specifically targets individuals identified as experts in various fields, think tanks, and South Korean government entities for the purpose of exfiltrating sensitive information (rather than destroying a computer or disrupting a network). This aligns with our findings from the adversary profiling (Section IV-A) that some of the public individual emails indeed belong to the above targets.

C. DATA REPRESENTATIVENESS

Although we collect varying internal (*e.g.*, sign-in records, filing phishing reports, spam monitoring) and external (*e.g.*, Certstream [9], OSINT [42]) resources, it is not feasible to detect the whole phishing campaigns against NAVER in South Korea. This is because ① an adversary may harness other email platforms for luring a victim, ② a phishing campaign had been terminated considering the nature of its temporariness, and ③ the adversary could utilize a private messenger. Despite the popularity of NAVER in South Korea, it may not be fully representative to describe every aspect of a regional phishing campaign. Evolving a phishing campaign is possible by combining emerging technology with another communication channel in a more elaborated manner. In a similar vein, the volume and type of a regional phishing campaign may be substantially different depending on culture, social issues (*e.g.*, voting, politics), and service as well as a country.

D. AD-HOC RISK SCORE

A domain risk score associated with each factor (Table 2) has been heuristically pre-defined based on a user's phishing report. Note that such a score can be parameterized that can be adjusted depending on the regional characteristics.

E. LIMITATIONS AND FUTURE WORK

Unfortunately, a direct comparison with a browser-based phishing detection technique is infeasible, because, for example, Google Safe Browsing [19] merely offers the hash of a URL. Likewise, the comparison with other browser-oriented filtering is not possible due to their proprietary blacklisting databases. However, based on our experimental results, inherent limitations are evident in the utilization of regional data; thus advocating for enhanced collaboration with additional regional datasets would contribute to augmenting the comprehensiveness and inclusiveness of our findings. Despite Naver's prominence, bolstering collaborative efforts with other phishing-collecting services would lead to more comprehensive coverage of phishing detection in the country. Notably, a worldwide pandemic has engendered a surge in online engagement on a global scale, consequently precipitating a rise in instances of phishing attacks [5]. Thus, analyzing a dataset that contains pandemic-related phishing attacks is part of our future work to see whether regional phishing plays a different role from both regional and global perspectives. As a final note, more exploration is needed in the emerging types of phishing campaigns against improving defenses like passwordless logins and multi-factor authentications because the current phishing detection tools could not be adequately captured.

IX. RELATED WORK

We discuss related work in two key areas: the ecosystem and techniques of phishing campaigns, and phishing kits. Table 6 summarizes the comparison of our work focusing on regional-specific analysis with others mostly on ecosystem analysis.

A. ECOSYSTEM AND TECHNIQUES OF PHISHING CAMPAIGNS

The phishing attack ecosystem has been well understood [25], [31], [41], [46], [54], [60] including ① phishing techniques to circumvent the current phishing detection systems and to lure more victims to phishing campaigns, and ② new phishing detection mechanisms to effectively identify them. Particularly, Oest et al. [46] measured the end-to-end life cycle of a phishing campaign. With their Golden Hour framework that monitors a network, 4.8 million phishing victims were captured. According to their finding, the duration of phishing attacks on average is around 21 hours - from the first victim visit time to the last victim visit time. On the other hand, communication channels for exfiltrating user credentials are widely studied [10], [23], [49]. In previous studies, the effectiveness of blocklists was typically assessed by creating phishing websites and evaluating their detection rates. These studies typically focused on the detection coverage and speed of one or two blocklists. Instead of creating phishing websites, we gather a list of URLs from six different blocklists and compare them to assess their effectiveness. This work fills the gap that has overlooked a regional phishing attack by focusing on phishing cases in South Korea (in collaboration with `NAVER`). This method allows for a more comprehensive evaluation compared to previous studies.

B. PHISHING KITS

Prior studies [7], [10], [23], [33], [45], [69], [70] focus on phishing kits in the wild to help an in-depth understanding of phishing campaigns from the beginning to the end. Analyzing a phishing kit allows one to understand the general design (including source code), evasion techniques, and communication methods (*e.g.*, how victims' credentials and personal information were sent to the attackers). Cova et al. [10] first attempt to better understand the phishing kits written in PHP, measuring varying backdoor exfiltration methods that steal phishing victims' sensitive information. Han et al. [23] take a honeypot approach so that a phishing attacker could deploy phishing kits by exploiting the vulnerabilities of honeypot servers. Collecting 643 phishing kits, they reveal that the life cycle of scam pages with a phishing kit lasted less than 10 days. Moreover, Zawoad et al. [70] discover that 10% of collected scam sites in the wild have been deployed with a phishing kit. The observation in this work well aligns with the pervasiveness of a phishing kit in South Korea.

X. CONCLUSION

In the past, previous phishing studies have been mostly taken from a global perspective, relying on a blacklist database such as `eCrimeX`. However, regional phishing campaigns are often underrepresented in both detection and prevention efforts. This paper aims to fill this gap with 1,558 phishing campaigns targeting `NAVER` (that has 41 million Korean users) during a full year span of 2021 in South Korea. Our finding shows that those have barely been captured in a global database. Our PhishingHunter system allows us to efficiently capture fraudulent URLs by consolidating various information sources and techniques into a single database. We uncover three common types of phishing campaigns in South Korea; sophisticated phishing schemes that involve varying adversarial tactics (*e.g.*, a proxy configuration, circumvention techniques), phishing campaigns against the second-hand online marketplace, and campaigns against a non-specific group. Our findings reveal that a hit-and-run tactic, in which a phishing link is sent to a victim via a private channel, is a widespread means. We mitigate these threats by promptly locking down a victim's account to prevent the exfiltration of sensitive information, and by providing regional phishing information to APWG for preventing potential damages in a timely manner.

ACKNOWLEDGMENT

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the sponsor's views.

REFERENCES

[1] Apple. (2022). *Safari Browser*. [Online]. Available: https://www.apple.com/safari/

[2] APWG. (2022). *Anti-Phishing Working Group*. [Online]. Available: https://apwg.org

[3] APWG. (2022). *eCrime Exchange*. [Online]. Available: https://apwg.org/ecx/

[4] H. Bijmans, T. Booij, A. Schwedersky, A. Nedgabat, and R. van Wegberg, "Catching phishers by their bait: Investigating the Dutch phishing landscape through phishing kit detection," in *Proc. 30th USENIX Secur. Symp.*, 2021, pp. 3757–3774.

[5] M. Bitaab, H. Cho, A. Oest, P. Zhang, Z. Sun, R. Pourmohamad, D. Kim, T. Bao, R. Wang, Y. Shoshitaishvili, A. Doupé, and G.-J. Ahn, "Scam pandemic: How attackers exploit public fear through phishing," in *Proc. APWG Symp. Electron. Crime Res. (eCrime)*, Nov. 2020, pp. 1–10.

[6] P. Blog. (2021). *Top 10 TLDs Abused*. [Online]. Available: https://www.phishlabs.com/blog/top-10-tlds-abused

[7] J. Britt, B. Wardman, A. Sprague, and G. Warner, "Clustering potential phishing websites using DeepMD5," in *Proc. 5th USENIX Workshop Large-Scale Exploits Emergent*, San Jose, CA, USA, 2012, pp. 1–8. [Online]. Available: https://www.usenix.org/conference/leet12/workshop-program/presentation/britt

[8] BrowserStack. (2023). *Understanding Browser Market Share: Which Browsers to Test on in 2023*. [Online]. Available: https://www.browserstack.com/guide/understanding-browser-market-share

[9] CERTSTREAM. (2022). *Real Time Certificate Transparency Log Update Stream*. [Online]. Available: https://certstream.calidog.io/

[10] M. Cova, C. Kruegel, and G. Vigna, "There is no free Phish: An analysis of 'free' and live phishing kits," in *Proc. 2nd Conf. USENIX Workshop Offensive Technol.*, 2008, pp. 1–8.

[11] Cybersecurity UC, Agency IS. (2020). *North Korean Advanced Persistent Threat Focus: Kimsuky*. [Online]. Available: https://us-cert.cisa.gov/ncas/alerts/aa20-301a

[12] K. J. Daily. (2021). *Second-Hand Becomes a Lifestyle for MZ Generation*. [Online]. Available: https://koreajoongangdaily.joins.com/2021/04/15/business/industry/secondhand/20210415200907188.html

[13] DomainTools. (2020). *Passive DNS*. [Online]. Available: https://www.domaintools.com/resources/blog/strengthen-your-investigations-resolve-with-pdns

[14] eBay. (2021). *Multinational E-Commerce Corporation*. [Online]. Available: https://www.ebay.com/

[15] L. Encrypt. (2022). *A Free, Automated, and Open Certificate Authority (CA)*. [Online]. Available: https://letsencrypt.org/

[16] Engadget. (2021). *Apple Puts Additional Walls Between Your Browsing Data and Google on iOS 14.5*. [Online]. Available: https://www.engadget.com/ios-14-5-safari-safe-browsing-173836695.html

[17] FBI. (2020). *Internet Crime Report*. [Online]. Available: https://www.ic3.gov/Media/PDF/AnnualReport/2020_IC3Report.pdf

[18] Google. (2022). *Chrome Browser*. [Online]. Available: https://www.google.com/chrome/

[19] Google. (2022). *Google Safe Browsing*. [Online]. Available: https://safebrowsing.google.com

[20] Google. (2022). *Google Safe Browsing—Google Transparency Report*. [Online]. Available: https://transparencyreport.google.com/safe-browsing/overview?hl=en

[21] APW Group. (2020). *APWG Trends Report Q2 2020*. [Online]. Available: https://docs.apwg.org/reports/apwg_trends_report_q2_2020.pdf

[22] Group-IB. (2020). *Group-IB Enhances Data Exchange Operations by Joining Anti-Phishing Working Group*. [Online]. Available: https://www.group-ib.com/media/gib-apwg/

[23] X. Han, N. Kheir, and D. Balzarotti, "PhishEye: Live monitoring of sandboxed phishing kits," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2016, pp. 1402–1413.

[24] Herald TK. (2022). *Naver's New Goal: 1B Users in 5 Years*. [Online]. Available: http://www.koreaherald.com/view.php?ud=20220413000798

[25] T. Holgers, D. E. Watson, and S. D. Gribble, "Cutting through the confusion: A measurement study of homograph attacks," in *Proc. USENIX Annu. Tech. Conf.*, Berkeley, CA, USA, 2006, p. 24.

[26] Internet Agency. (2021). *Personal Information Protection Act (General Law)*. [Online]. Available: https://www.privacy.go.kr/eng/laws_view.do?nttId=8186&imgNo=3

[27] M. Khonji, Y. Iraqi, and A. Jones, "Phishing detection: A literature survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 2091–2121, 4th Quart., 2013.

[28] D. Kim, H. Cho, Y. Kwon, A. Doupé, S. Son, G.-J. Ahn, and T. Dumitras, "Security analysis on practices of certificate authorities in the HTTPS phishing ecosystem," in *Proc. ACM Asia Conf. Comput. Commun. Secur.*, New York, NY, USA, May 2021, pp. 407–420, doi: 10.1145/3433210.3453100.

[29] J. Kim, K. Kwak, and M. Jang. (2019). *Kimsuky Group: Tracking the King of the Spear Phishing*. [Online]. Available: https://www.virusbulletin.com/virusbulletin/2020/03/vb2019-paper-kimsuky-group-tracking-king-spearphishing/

[30] J. Kim, S. Ryu, and K. Kwak. (2021). *Operation Newton: Hi Kimsuky? Did an Apple (Seed) Really Fall on Newton's Head?* [Online]. Available: https://vblocalhost.com/uploads/VB2021-Kim-etal.pdf

[31] P. Kintis, N. Miramirkhani, C. Lever, Y. Chen, R. Romero-Gómez, N. Pitropakis, N. Nikiforakis, and M. Antonakakis, "Hiding in plain sight: A longitudinal study of combosquatting abuse," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, New York, NY, USA, Oct. 2017, pp. 569–586.

[32] Labs. (2021). *Kimsuky APT Continues to Target South Korean Government Using AppleSeed Backdoor*. [Online]. Available: https://blog.malwarebytes.com/threat-intelligence/2021/06/kimsuky-apt-continues-to-target-south-korean-government-using-appleseed-backdoor

[33] Lazar. (2018). *Our Analysis of 1,019 Phishing Kits | Imperva*. [Online]. Available: https://www.imperva.com/blog/our-analysis-of-1019-phishing-kits/

[34] S. Maroofi, M. Korczynski, and A. Duda, "Are you human? Resilience of phishing detection to evasion techniques based on human verification," in *Proc. ACM Internet Meas. Conf.*, Oct. 2020, pp. 78–86.

[35] Microsoft. (2022). *Microsoft Edge Browser*. [Online]. Available: https://www.microsoft.com/en-us/edge/

[36] Microsoft. (2022). *Microsoft Edge support for Microsoft Defender SmartScreen*. [Online]. Available: https://docs.microsoft.com/en-us/deployedge/microsoft-edge-security-smartscreen

[37] Mizilla. (2022). *Firefox Browser*. [Online]. Available: https://www.mozilla.org/

[38] Naver. (2022). *Joonggonara Cafe*. [Online]. Available: https://cafe.naver.com/joonggonara

[39] Naver. (2022). *Korean Web Portal Service*. [Online]. Available: https://www.naver.com/

[40] Naver. (2022). *Whale—Safe Browsing*. [Online]. Available: https://whale.naver.com/en/details/security/#safebrowsing

[41] N. Nikiforakis, S. Van Acker, W. Meert, L. Desmet, F. Piessens, and W. Joosen, "Bitsquatting: Exploiting bit-flips for fun, or profit?" in *Proc. 22nd Int. Conf. World Wide Web*, New York, NY, USA, 2013, pp. 989–998.

[42] Nordine. (2022). *OSINT Framework*. [Online]. Available: https://osintframework.com/

[43] A. Oest, Y. Safaei, A. Doupé, G.-J. Ahn, B. Wardman, and K. Tyers, "PhishFarm: A scalable framework for measuring the effectiveness of evasion techniques against browser phishing blacklists," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2019, pp. 1344–1361.

[44] A. Oest, Y. Safaei, P. Zhang, B. Wardman, K. Tyers, Y. Shoshitaishvili, and A. Doupe, "PhishTime: Continuous longitudinal measurement of the effectiveness of anti-phishing blacklists," in *Proc. 29th USENIX Secur. Symp.*, 2020, pp. 379–396.

[45] A. Oest, Y. Safei, A. Doupé, G.-J. Ahn, B. Wardman, and G. Warner, "Inside a phisher's mind: Understanding the anti-phishing ecosystem through phishing kit analysis," in *Proc. APWG Symp. Electron. Crime Res. (eCrime)*, May 2018, pp. 1–12, doi: 10.1109/ECRIME.2018.8376206.

[46] A. Oest, P. Zhang, B. Wardman, E. Nunes, J. Burgis, A. Zand, K. Thomas, A. Doupe, and G. J. Ahn, "Sunrise to sunset: Analyzing the end-to-end life cycle and effectiveness of phishing attacks at scale," in *Proc. 29th USENIX Secur. Symp.*, 2020, pp. 1–17.

[47] OpenPhish. (2021). *Timely, Accurate, Relevant Phishing Intelligence*. [Online]. Available: https://www.openphish.com/

[48] G. L. Orgill, G. W. Romney, M. G. Bailey, and P. M. Orgill, "The urgency for effective user privacy-education to counter social engineering attacks on secure computer systems," in *Proc. 5th Conf. Inf. Technol. Educ.*, Oct. 2004, pp. 177–181.

[49] P. Peng, C. Xu, L. Quinn, H. Hu, B. Viswanath, and G. Wang, "What happens after you leak your password: Understanding credential sharing on phishing sites," in *Proc. ACM Asia Conf. Comput. Commun. Secur.*, Jul. 2019, pp. 181–192.

[50] PhishLabs. (2021). *Digital Risk Protection Through Curated Threat Intelligence and Complete Mitigation*. [Online]. Available: https://www.phishlabs.com/

[51] PhishTank. (2021). *Out of the Net, Into the Tank*. [Online]. Available: https://phishtank.org/

[52] Projects TC. (2022). *Chromium Browser*. [Online]. Available: https://www.chromium.org/Home/

[53] Reddit. (2021). *Home to Thousands of Communities, Endless Conversation, and Authentic Human Connection*. [Online]. Available: https://www.reddit.com/

[54] R. Roberts, Y. Goldschlag, R. Walter, T. Chung, A. Mislove, and D. Levin, "You are who you appear to be: A longitudinal study of domain impersonation in TLS certificates," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Nov. 2019, pp. 2489–2504.

[55] J. Spaulding, S. Upadhyaya, and A. Mohaisen, "The landscape of domain name typosquatting: Techniques and countermeasures," in *Proc. 11th Int. Conf. Availability, Rel. Secur. (ARES)*, Los Alamitos, CA, USA, Aug. 2016, pp. 284–289.

[56] Statcounter. (2022). *Browser Market Share Republic Of Korea*. [Online]. Available: https://gs.statcounter.com/browser-market-share/all/south-korea

[57] Statista. (2020). *Global Market Share Held by Leading Desktop Internet Browsers*. [Online]. Available: https://www.statista.com/statistics/544400/market-share-of-internet-browsers-desktop/

[58] Statista. (2020). *Market Share Held by Leading Mobile Internet Browsers Worldwide*. [Online]. Available: https://www.statista.com/statistics/263517/market-share-held-by-mobile-internet-browsers-worldwide/

[59] Statista. (2020). *Number of Fraud Damage Cases When Buying Second-Hand Via Online Shopping in South Korea in 2020, by Platform*. [Online]. Available: https://www.statista.com/statistics/1259630/south-korea-second-hand-online-shopping-fraud-cases/

[60] K. Tian, S. T. K. Jan, H. Hu, D. Yao, and G. Wang, "Needle in a haystack: Tracking down elite phishing domains in the wild," in *Proc. Internet Meas. Conf.*, New York, NY, USA, Oct. 2018, pp. 429–442.

[61] Ulikowski. (2022). *GitHub—Elceef/Dnstwist: Domain Name Permutation Engine for Detecting Homograph Phishing Attacks, Typo Squatting, and Brand Impersonation*. [Online]. Available: https://github.com/elceef/dnstwist

[62] Vivaldi. (2022). *Vivaldi Browser*. [Online]. Available: https://www.vivaldi.org/

[63] W3C. (2022). *Same Origin Policy*. [Online]. Available: https://www.w3.org/Security/wiki/Same_Origin_Policy

[64] Wikipedia. (2022). *Homograph*. [Online]. Available: https://en.wikipedia.org/wiki/Homograph

[65] Wikipedia. (2022). *Naver*. [Online]. Available: https://en.wikipedia.org/wiki/Naver

[66] Wikipedia. (2022). *Pharming*. [Online]. Available: https://en.wikipedia.org/wiki/Pharming

[67] Wikipedia. (2022). *SMS Phishing*. [Online]. Available: https://en.wikipedia.org/wiki/Phishing#SMS_phishing

[68] Wikipedia. (2022). *Typosquatting*. [Online]. Available: https://en.wikipedia.org/wiki/Typosquatting

[69] Wright. (2017). *Phish in a Barrel: Hunting and Analyzing Phishing Kits at Scale | Duo Security*. [Online]. Available: https://duo.com/blog/phish-in-a-barrel-hunting-and-analyzing-phishing-kits-at-scale

[70] S. Zawoad, A. K. Dutta, A. Sprague, R. Hasan, J. Britt, and G. Warner, "Phish-Net: Investigating phish clusters using drop email addresses," in *Proc. APWG eCrime Researchers Summit*, Sep. 2013, pp. 1–13, doi: 10.1109/eCRS.2013.6805777.

[71] ZeroSSL. (2022). *Trusted Certificate Authority for Anyone*. [Online]. Available: https://zerossl.com/

HYUNJUN PARK received the master's degree in cybersecurity from Korea University, in 2020. He is currently a Security Researcher with NAVER Corporation. With more than ten years of industrial experience, he was involved in penetration testing projects for smartphones, smart TVs, and mobile devices with Samsung. At the Defcon Conference, he presented a potential cloud security vulnerability in mobile apps. His research interests include phishing, application security, and usable security.

KYUNGCHAN LIM (Graduate Student Member, IEEE) received the master's degree from the University at Albany. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering and Computer Science, The University of Tennessee (UTK), Knoxville (Advised by Prof. Doowon Kim). Before joining UTK, he had five years of industrial experience in data management and as a Principal Investigator (PI) and a Project Manager (PM) in multiple projects. His research interests include data-driven measurement in security and web security.

DOOWON KIM (Associate Member, IEEE) received the Ph.D. degree in computer science from the University of Maryland, College Park, in May 2020, advised by Prof. Tudor Dumitras. He is currently an Assistant Professor with the Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville. His research interests include computer security (data-driven, usable security), computer networks (internet measurement), and identifying the root causes of security threats by understanding actors.

DONGHYUN YU received the master's degree from Soonchunhyang University. He is currently a Security Researcher with NAVER Corporation, South Korea. His research interests include threat intelligence to collect malicious samples and to trace attacker's behaviors. His research interests include web application security to analyst attack campaigns, such as malware distribution via drive-by download and phishing.

HYUNGJOON KOO received the M.Sc. degree in information security from Korea University, in 2010, and the Ph.D. degree in computer science from Stony Brook University, in 2019. He is currently an Assistant Professor with the Department of Computer Science and Engineering, College of Computing, Sungkyunkwan University (SKKU). Before joining SKKU, he was a Postdoctoral Researcher with the SS Laboratory, Georgia Institute of Technology. He is leading the SecAI Laboratory. His research interests include software security, binary analysis, and security with artificial intelligence.