SCAMMAGNIFIER: Piercing the Veil of Fraudulent Shopping Website Campaigns

Marzieh Bitaab*, Alireza Karimi*, Zhuoer Lyu*, Adam Oest†, Dhruv Kuchhal†, Muhammad Saad‡,
Gail-Joon Ahn*, Ruoyu Wang*, Tiffany Bao*, Yan Shoshitaishvili*, and Adam Doupé*
*Arizona State University, †Amazon, ‡X Corp.
{mbitaab, akarimi6, zlyu15, gahn, fishw, tbao, yans, doupe}@asu.edu
{aoest, dkuchhal}@amazon.com, muhammadsaad@x.com

***Abstract*—In an evolving digital environment under perpetual threat from cybercriminals, phishing remains a predominant concern. However, there is a shift towards fraudulent shopping websites—fraudulent websites offering bogus products or services while mirroring the user experience of legitimate shopping websites. A key open question is how important fraudulent shopping websites in the cybercrime ecosystem are?**

This study introduces a novel approach to detecting and analyzing fraudulent shopping websites through large-scale analysis and collaboration with industry partners. We present SCAMMAGNIFIER, a framework that collected and analyzed 1,155,237 shopping domains from May 2023 to June 2024, identifying 46,746 fraudulent websites. Our automated checkout process completed 41,863 transactions, revealing 5,278 merchant IDs associated with these scams. The collaborative investigations with one of major financial institutions also confirmed our findings and provided additional insights, linking 14,394 domains to these fraudulent merchants. In addition, we introduce a Chromium web extension to alert users of potential fraudulent shopping websites. This study contributes to a better understanding of e-Commerce fraud and provides valuable insights for developing more effective defenses against these evolving threats.

I. INTRODUCTION

E-commerce is constantly under threat by cybercriminals who employ diverse tactics to illicitly siphon funds from unsuspecting Internet users [31]. These threats have evolved over time and can be manifested through various means like phishing, pet scams, fake charities, and fraudulent shopping websites [10,15,51]. Across all these scam methods, fraudsters bait victims to either surrender personal information that can be monetized or make direct payments for products that will never arrive.

Fraudulent shopping websites typically involve fraudsters displaying commonly purchased items (e.g., clothes and purses) at discounted rates to attract customers. Those who purchase these items will receive counterfeit products or no product at all [11, 13]. Fraudulent shopping websites are distinctly different from phishing websites in their attack modality because the latter harvest user credentials for monetization while fraudulent shopping websites directly solicit payments from customers without necessarily stealing their personal information. Moreover, unlike phishing websites that disguise as specific brands and target their users, fraudulent shopping websites indiscriminately target online shoppers, thus potentially impacting more online users. According to Federal Trade Commission (FTC), online shopping scams ranked as the second most frequently reported fraud type in 2023 [24]. In 2023 alone, users reported $392 million loss to these scams. This is a significant increase from 2020 with $246 million of losses [22]. The rise in scams has had a profound impact on both consumers and businesses. This not only affects individual consumers but also undermines the entire society's trust in online shopping.

The risk of fraudulent shopping websites has not gone unnoticed. Researchers have recently proposed novel techniques for detecting fraudulent shopping websites at scale [13, 34]. These techniques collect multiple features (e.g., number of CSS IDs in the HTML content, likes of the associated Facebook page, and the age of the domain) for a website and use machine learning models to identify if the website is a fraudulent shopping website using these features. Unfortunately, fraudsters can alter their websites to bypass most (if not all) of these features. To facilitate robust and more timely detection of fraudulent shopping websites, we focus on finding a *stable*, difficult-to-bypass feature of fraudulent shopping websites.

The main challenge that our study faces is composing a comprehensive dataset of fraudulent shopping websites. Obtaining a comprehensive fraudulent shopping websites dataset and designing an effective detection tool is a chicken-and-egg problem: Designing an effective fraudulent shopping websites relies on a good understanding of fraudulent shopping websites, yet the collection of such a dataset relies on an effective detection tool. In fact, the state-of-the-art detection tools [3,13] only yield detection rates of 59.3% and 23.25%, respectively.

We find that the most critical component of fraudulent shopping websites is the "checkout" process where the fraudulent shopping websites interact with payment processors to process their victims' payments. This interaction uses a unique identifier called *merchant ID*, which uniquely identifies the fraudulent shopping websites with a payment processor. Merchant IDs play a crucial role in the detection and analysis of fraudulent shopping websites. These unique identifiers, assigned by payment processors, serve as a valuable tool for tracking and linking fraudulent activities across multiple domains. Merchant IDs are considered public information and cannot be arbitrarily set or changed by merchants, making

Network and Distributed System Security (NDSS) Symposium 2025
24-28 February 2025, San Diego, CA, USA
ISBN 979-8-9894372-8-3
https://dx.doi.org/10.14722/ndss.2025.24763
www.ndss-symposium.org

them a reliable indicator to identify a domain's merchant. Moreover, the rigorous verification processes that payment processors implement make it challenging for a fraudster to obtain multiple merchant IDs, thus providing a robust foundation for studying fraudulent shopping websites.

To verify our assumption, we build SCAMMAGNIFIER, a novel approach aiming to unveil merchant IDs for fraudulent shopping websites and help understand how these websites operate under the hood. During the course of our study, SCAMMAGNIFIER collected 1,155,237 shopping domains from May 2023 to June 2024, and applied a machine learning model to detect 46,746 fraudulent shopping websites [13]. Our Auto-Checkout performed 41,863 successful checkout and collected 5,278 merchant IDs, with 4,484 merchant IDs linked to a single legitimate payment service provider. While collecting data, we observed a peculiar functionality deployed by fraudulent shopping websites where the domain would be redirected during the checkout phase. This domain redirection hides the initial fraudulent shopping websites domain from the payment processor, thereby obfuscating the nature of the website from the payment processor.

Surprisingly, the merchant IDs that SCAMMAGNIFIER collected enabled us to *look back in history*: We can study past fraudulent shopping websites operations before May 2023 because payment processors keep records of historical merchant IDs and their corresponding websites. To obtain scam confirmation and insights, we shared our findings with financial Org A. Financial Org A affirmed the validity of our approach and disclosed domains associated with the set of merchant IDs (4,484). Intriguingly, 14,394 domains are connected to these merchants. Across these fraudulent shopping websites, we found that 97.73% of them had a short lifespan of less than one year with a high payment volume occurring merely a day after their domain registration. Note that high payment volume for new and unpopular websites is rather unusual. To understand this behavior, we next share our results with technical Org B, which informed us that scam sites derive their traffic through advertisements on social media platforms. Particularly, 28.78% of the fraudulent shopping websites' users were routed through advertisements on Facebook (including Instagram), while 21.10% and 9.38% of these users were from advertisements on Google and Bing, respectively. By consolidating all these insights, we conclude that fraudulent shopping websites are indeed the work of a coordinated and cunning criminal enterprise. These criminals manipulate many merchant IDs on several payment processors and have several procedures in place to evade detection.

To help users defend against fraudulent shopping websites, we created a Chromium web extension, SCAMCHECKER, that alerts users when they visit a potential fraudulent shopping website. SCAMCHECKER uses the Auto-Checkout module in SCAMMAGNIFIER to extract merchant ID about the merchant who runs a website. The extension will warn the user if the associated merchant ID belongs to a previously seen fraudulent shopping website. In our experiment, SCAMCHECKER improved the detection rate of the state-of-the-art fraudulent shopping website detector from 59.30% to 76.74%.

Contributions. This paper makes the following contributions:

- We measure the interconnectedness of fraudulent shopping websites with SCAMMAGNIFIER, which scans newly registered scam domains and automatically performs the checkout process to extract fraudulent merchant IDs.
- We use SCAMMAGNIFIER to analyze 46,746 fraudulent shopping websites and complete 41,863 successful automated checkouts to obtain 5,278 scam merchant IDs. We note that the scammers operate in large groups with a few merchant IDs controlling a large number of scam sites.
- We present a Chromium web extension that can be used in any Chromium-based browser to warn users when they visit a potential fraudulent shopping websites. The extension detects fraudulent shopping websites not only based on the classifier results but also according to commonality of the associated merchant with known fraudulent merchants.
- To validate our findings and obtain additional insights about the lifecycle of fraudulent shopping websites, we collaborate with prominent industry organizations. Through those partnerships, we obtained actionable information about evading mechanisms employed by fraudulent shopping websites as well as their preferred advertising platforms.

To foster open science, we plan to publicly release our source-code and data upon the acceptance of this paper.

Ethics. All data that we gathered was crawled from public websites. In addition, we ensure that the crawler did not send excessive traffic, and we did not have access to personally identifiable information. Financial Org A's analysis was within its routine fraud prevention efforts consistent with the platform's usage policy. Furthermore, all actions adhered to user privacy regulations and were within the context originally intended. Our study does not involve completing actual checkouts or making purchases from any suspected fraudulent shopping websites. We collected all necessary data from the checkout pages without finalizing transactions or confirming purchases. Therefore, no websites or businesses were adversely affected during the course of our research.

We also worked to block and stop as many fraudulent shopping websites as possible. We reported all merchant IDs found to the payment processors when possible. We also reported all fraudulent shopping domains to Google and Microsoft. We attempted to report the fraudulent shopping domains that were hosted on Shopify, the largest e-Commerce platform, to Shopify, however we were unable to successfully reach a corresponding unit, despite several attempts over multiple channels.

II. BACKGROUND

In today's online environment, phishing websites are well-known for imitating trusted entities to steal users' credentials [53]. However, a less explored but equally hidden threat is that of fraudulent shopping websites [13]. Unlike phishing counterparts, fraudulent shopping websites imitate genuine e-Commerce experiences, tricking users into purchasing fictitious or counterfeit products. In this section, we aim to show the distinct characteristics of both types of websites, shedding

light on their unique operational tactics and drawing clear distinctions between them.

A. Phishing Websites

Phishing is a social engineering attack that involves deploying websites to steal users' sensitive information, such as credit card numbers or account credentials [43, 53]. Phishing websites are made to look similar to well-known brands or organizations' websites to gain users' trust and steal their sensitive or personally identifiable information. Fraudsters use several evasion [38, 42] and cloaking techniques [55] to hide from crawlers and phishing detection tools. These techniques include using captcha, URL redirection, and being hosted on trusted domains.

Phishing detection has been studied comprehensively by prior research [12, 26, 30, 36]. Researchers proposed several feature-engineering approaches for machine learning-based methods to detect phishing URLs based on URL features [37, 46] or visual similarity with known brands [7, 16]. A substantial body of research is dedicated to understanding the effectiveness of the ecosystem and the lifecycle of phishing websites [41, 42].

Browser-based phishing detection such as Google Safe Browsing (GSB) [52] and Microsoft Defender SmartScreen [39] identify and alert users of potential threats. This type of phishing detection, due to its scale and always-on nature, serves a particularly important role [47]. Furthermore, organizations have increasingly adopted "takedown" services that work to remove phishing websites, curtailing their malicious activities [4, 5]. To enhance these technical measures, many institutions also invest in awareness training, equipping individuals with the knowledge and skills to recognize and avoid phishing attempts [6]. However, while these measures are effective against phishing, they often fail to detect fraudulent shopping activities. This is largely because fraudulent shopping websites can closely mimic legitimate ones in design and functionality, making them harder to detect. Moreover, these sites may not necessarily exhibit malicious behavior like phishing sites. Rather, they deceive users through counterfeit products, false advertising, or other subtle scams, rendering conventional anti-phishing tools ineffective in this domain.

B. Fraudulent e-Commerce Websites

Fraudulent e-Commerce websites represent a category of deceptive websites meticulously crafted to trick users into transacting despite offering counterfeit merchandise or non-existent products. These websites, in their bid to emulate the operational dynamics of legitimate e-Commerce entities, present extensive product listings, social media logos, and legitimate payment gateways. Contrary to traditional phishing websites that often employ techniques to evade detection, fraudulent e-Commerce websites strategically harness Search Engine Optimization (SEO) methodologies [18] and online advertising to enhance their visibility, enticing users to engage with them. Rather than stealing users' sensitive information, the primary objective of fraudulent e-Commerce websites is to steal the victim's money.

A salient distinction between phishing and fraudulent e-Commerce websites lies in their *modus operandi*: fraudulent e-Commerce websites platforms do not seek to impersonate well-known brands, rendering detection based on brand similarity [7, 16] moot. Fraudulent e-Commerce websites use false advertising to attract victims [35]: They may look similar to any other e-Commerce website at a glance, but they deliver fake or no products. They use various SEO techniques to advertise themselves and lure unsuspecting customers [40]. Some of the techniques include "keyword stuffing" [35] where they use a large number of keywords in content and/or meta tags to be detected by search engine crawlers. Moreover, they leverage social media ads to advertise themselves (as we show in Section V-C).

In the context of fraudulent e-Commerce websites, *online shopping websites*, which we study in this paper, are a significant sub-category [13, 14]. These platforms mimic the behavior of legitimate e-Commerce websites, often offering discounted or scarce items to captivate users, especially deal-seekers [14]. The strategic positioning of such offerings plays a crucial role in luring users, especially those who are in pursuit of deals or popular items. The illusion of scarcity or discounted pricing can create a sense of urgency among potential buyers, nudging them toward making impulsive purchasing decisions. Moreover, fraudulent shopping websites employ sophisticated tactics to enhance their credibility. This includes the use of professional-grade website design and secure payment gateways. However, despite their seemingly authentic appearance, these platforms primarily aim to carry out fraudulent transactions. Therefore, it is an important and crucial task to study how fraudulent shopping websites work and how we can leverage their mechanism to prevent users from interacting with such websites.

As mentioned previously, one important aspect of most fraudulent shopping websites is the role of payment service providers, such as PayPal, Google Pay, Stripe, or Venmo, in these scams. Unlike phishing websites, which only aim to collect personal information, fraudulent shopping websites actually process payments from buyers using legitimate payment gateways. Therefore, the payment service providers offer an avenue for studying and measuring fraudulent shopping websites, as they are where the scam touches the financial system. In addition, payment service providers can be involved in detecting and preventing fraudulent transactions, as well as providing refunds or chargebacks to the victims.

C. Detecting and Measuring the Impact of Fraudulent e-Commerce Websites

While government organizations, including the Federal Trade Commission and the Federal Bureau of Investigation, regularly warn users of the threat of such scams, underscored by escalating reported financial losses [19, 20, 23], the cybersecurity community appears to be lagging in its comprehensive investigation of fraudulent e-Commerce websites.

Historically, prior research has predominantly focused on one specific type of fraudulent e-Commerce websites, be it pet scams, cryptocurrency scams, or counterfeit online retail websites. Two notable works in this domain are PREDATOR [28] and Beyond Phish [13]. Hao et al. [28] introduced

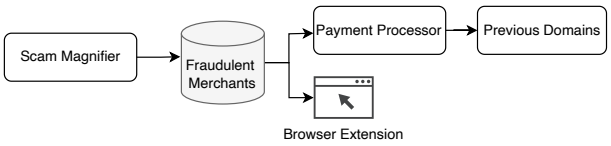


Fig. 1: Usage of SCAMMAGNIFIER in fraudulent shopping websites detection and validation through financial Org A.

PREDATOR, a scalable solution for domain fraud detection, leveraging features to identify fraudulent domains at the time of registration. A notable limitation of PREDATOR is its reliance on datasets predominantly associated with spam for model training. The methodology encompasses data collection, harnessing features derived from DNS, content, and social media paradigms to achieve large-scale fraud detection.

Bitaab et al. [13] address the issue of fraudulent e-Commerce websites, such as counterfeit shopping websites, deceptive charities, and cryptocurrency scam sites. They define fraudulent e-Commerce websites as websites directly defrauding users by mimicking legitimate e-Commerce experiences (and we use this definition in our paper). The authors introduce an automated approach to collect samples through crowdsourcing, identifying the defining characteristics; they develop an open-source classifier, Beyond Phish [13], which has a high detection rate and low false positive rate.

III. OVERVIEW

The methodology of this study is designed to comprehensively detect and analyze fraudulent shopping websites. Initially, we manually analyze several fraudulent shopping websites to identify patterns and indicators of fraudulent behavior. This initial manual analysis is crucial for creating the subsequent automated processes.

Based on our observations from manual analysis, we develop SCAMMAGNIFIER, that gathers 1,155,237 domains between May 2023 and June 2024. Following data collection, the model performs an automated checkout procedure that successfully completes 41,863 checkout processes. To validate our findings, we collaborate with financial Org A, which confirms our results and provides additional insights by linking 14,394 domains to the identified fraudulent merchants. This multi-layered analysis not only ensures the accuracy of our detection mechanisms but also highlights the extensive reach and impact of these fraudulent activities within the cybercrime ecosystem.

The detection phase serves both as an application of our research and a validation of our measurement results, as shown in Figure 1. By introducing a Chromium-based browser extension, we provide a practical tool for fraudulent shopping websites detection and alerting users about potential merchants and fraudulent shopping websites. This extension embodies the practical implications of our study, offering a direct defense mechanism for consumers while also serving as a testament to the effectiveness and accuracy of our detection framework. Our study contributes to a deeper understanding of fraudulent shopping websites and paves the way for developing more robust defenses against these evolving threats.

IV. DATA COLLECTION AND DATASET

As a preliminary step in our study, we manually analyzed several fraudulent shopping websites. This involved interacting with fraudulent shopping websites and observing URLs browsing logs, functionality, and overall operation.

We realized that some fraudulent shopping websites use payment processors, and payment processors have unique IDs that link to the payment gateway users (i.e., the users or organizations accepting payment). We standardize on *merchant IDs* as a term for these unique IDs.

Merchant IDs are unique identifiers assigned to businesses by payment processors when these businesses set up accounts to accept electronic payments. Merchant IDs are not considered private or hidden information. For most payment processors (e.g., PayPal and Google Pay), we can retrieve merchant IDs from the checkout page or the payment gateway page. Each payment processor uses its own merchant ID format, and merchants cannot arbitrarily set or change their assigned IDs. Unlike credit card numbers, merchant IDs are constant, and they are not reassigned to random strings during the payment process. [2, 27, 45]

For example, a website on e-Commerce platform *A* has a dedicated field `merchant-id`, while a website on e-Commerce platform *B* has a different field called `merchantID`. Through in-depth analysis we identified that the merchant IDs can be obtained from either the payment gateway's URL or the HTML of the landing page, depending on the payment gateway that is used by the fraudulent website.

Notably, we found that different fraudulent shopping websites sometimes share the same merchant ID, suggesting they funnel payments to the same account, indicating operation by a single entity or group. Our findings reveal a pattern of shared domain registrars and hosting providers among the samples, hinting at a common control. This pattern extends to website design, suggesting the use of shared templates or scripts.

Some fraudulent shopping websites accept payments through server-side credit card processing (i.e., they do not use any externally observable payment processor). Credit card payments differ significantly from online payment processors such as PayPal: They are processed entirely server-side, with each card-issuing bank handling its own transactions. This architecture makes it infeasible to collect relevant transaction information without involving the issuing banks. Because we cannot observe the merchant IDs involved in server-side credit card processing (without cooperation with credit-card companies), these are out of scope of our research.

Our observations on fraudulent shopping websites operations are summarized in Figure 4. We highlight the reuse of merchant IDs across multiple fraudulent shopping websites, underlining a probable central control. This assertion is supported by our collaborative efforts with financial Org A, offering further insights into the scam operations (in Section IV-D). To measure the interconnectedness of the criminal ecosystem related to fraudulent shopping websites, we must focus our efforts on studying where cybercrime meets the financial ecosystem. For fraudulent shopping websites, this exists when the victim purchases the fraudulent goods. Therefore, we design a system to collect merchant IDs from payment

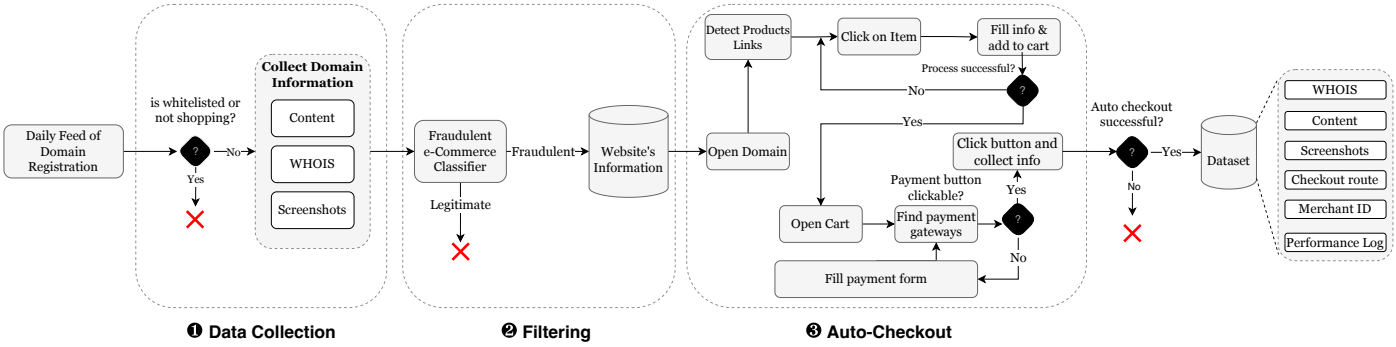


Fig. 2: SCAMMAGNIFIER pipeline design comprises three pivotal stages: 1. Collecting domains daily, 2. Classifying the websites to find potential fraudulent shopping websites using Beyond Phish [13], and 3. Using our Auto-Checkout system to extract fraudulent shopping websites merchant IDs from payment processors. We finally add all of the collected data during this process to a database.

processors on fraudulent shopping websites, and we do so at scale by using a daily feed of domains.

The main reason scammers use a single merchant ID on multiple domains is that payment processors implement rigorous verification processes when a merchant obtains a merchant ID. These processes require merchant ID applicants to supply detailed business information, personal identification, and bank account verification. While a legitimate business may have multiple merchant IDs for different revenue streams or international operations, individual scammers face obstacles because of the extensive verification procedures [44].

Our aim is to create a comprehensive framework that effectively mimics human interactions and strategically skips steps when necessary. During the development of our framework, SCAMMAGNIFIER, we faced several challenges including dealing with diverse fraudulent shopping websites templates, evading bot detection techniques, dynamic content loading, and handling session timeouts. We tackled these issues by designing a system that mimics human interactions.

Figure 2 outlines the SCAMMAGNIFIER framework. This data collection pipeline comprises three pivotal stages: ❶ collecting domains on a daily basis, ❷ classifying the websites on these domains to find potential fraudulent shopping websites, and ❸ using our Auto-Checkout system on identified scam websites to extract the merchant IDs from payment processors.

A. SCAMMAGNIFIER *Overview*

Following our observations from manual analysis, we design a framework to automate the process of data collection and analysis.

In our research methodology, we initiate the process by systematically collecting domain data [1] on a daily basis (❶). Given that a significant proportion of newly registered domains may initially be inactive or not yet hosting any fraudulent shopping websites, we strategically delay the content collection phase to start fourteen days following the registration of these domains. This approach allows for a more accurate assessment of active and functional websites. For each domain we additionally collect WHOIS records, the content of the websites, and visual snapshots of their respective homepages. We acknowledge that timelines of registration and business operations may vary across scam sites, thus leading to situations where we might skip a few. However, the main goal of this work is to measure the interconnectedness of fraudulent shopping websites.

To identify potential fraudulent shopping websites instances within the collected dataset, we use the Beyond Phish classifier [13]. Beyond Phish [13] is an *open-source* machine learning-based classifier that uses various features to detect potential fraudulent shopping websites. This classifier uses manually extracted features from website's content, WHOIS data, URL, and social media to classify them into legitimate or fraudulent. We conducted a manual analysis to validate the classifier's low false positive rate. This step is taken to ensure that our framework will not misclassify legitimate merchants. Our security experts analyzed a subset of $1,000$ websites identified as a scam by Beyond Phish. This method had 23 false positives which is insignificant.

When the classifier identifies the website on a domain as potentially fraudulent (in step ❷), the next step is to run the Auto-Checkout process (❸) and gather data that can provide insights into the modus operandi of fraudulent e-Commerce sites. Upon successful checkout completion, we extract data including the URL of payment gateways, corresponding page source code, and merchant IDs observed.

B. *Auto-Checkout*

A core element of SCAMMAGNIFIER is the Auto-Checkout (AC) component. This component was developed following an extensive manual analysis of the structural design common to shopping websites. Our observations indicate that the typical layout includes a landing page displaying items for sale. Selecting an item redirects the user to a detailed description page, where options to add the item to a shopping cart or proceed directly to purchase are presented. Finally, the user can navigate to the website's checkout page.

In designing the AC component, our aim was to ensure compatibility with a diverse range of shopping website structures. We manually analyzed websites from various e-Commerce platforms to draw the general structure of shopping websites, and the in-the-wild efficacy and performance

metrics of the AC component are discussed in greater detail in Section IV-C.

The AC component uses Selenium to automate a web browser and emulate a human navigating through shopping websites to complete the purchase of items. The component locates and interacts with requisite web elements at each step of the checkout procedure, from product selection to payment confirmation. In general, AC follows this process:

1) *Crawling the website:* In this step, we open the shopping website in the Selenium-enabled browser. Because shopping websites often show deals to first-time visitors, we wait for the appearance of any pop-up modal after opening the website. The AC checks the source code for *modal* CSS classes to detect these pop-ups. After several seconds, it clicks on a random edge of the website to attempt to close the modal. The AC also executes the web page's JavaScript tags to ensure the loading of all the page's assets.
2) *Getting items:* This step looks for items on the webpage by collecting all link elements using XPath queries. Each element can have multiple attributes, such as class names or optional attributes such as "*class=product*". AC looks for keywords (such as *prod* or *product*) in each link element, adding them to a list of potential items. Then, it randomly selects one of the items to click on. To increase the efficiency of this step, we use a shallow filtering process to filter out unrelated links based on the following heuristic:
   - Removing external links
   - Removing links with their path starts with keywords not related to products. These keywords are provided in Table IV.

   The script randomly selects and navigates to product pages, repeating this up to 10 times to increase interaction diversity.
3) *Adding item to cart:* In this step, AC first looks for an "Add to Cart" button to proceed to the next step. However, sometimes the website requires filling in a form before adding an item to the cart, such as choosing the item's size or color. This is a general problem known as *Deep Web Crawling* [29], and we do not seek to solve this problem generally. We take a heuristic-based approach, and if there are form elements such as `select` element, then we randomly choose an item or fill the values. AC automates the selection of product variants (e.g., size, color) by interacting with dropdown menus and clickable elements. It loops through all available selection-based HTML elements and selects or fills them randomly. After cycling through all available options for a product, AC tries to add the product to the cart by clicking on the "Add to Cart" button. AC uses text-based matching to identify the "Add to Cart" button; we provide the full keyword list in Table IV.
4) *Checkout:* In this stage, AC opens the cart and traverses the checkout forms. Similar to the previous stage, AC first looks for a quick checkout button to click on; otherwise, it fills the required form items as described in the previous step. We also attempt to identify the semantic context of the form element using keywords derived from manually analyzing fraudulent shopping websites (Section IV), such as address or other equivalent names with a random-looking address. If all these fail, we provide random text input. In this process, AC finds all form items and fills them according to their tags, showing their required information. Then, it follows the checkout process to reach a payment gateway. After this process, AC clicks the checkout button by looking for related keywords all of which in Table IV and captures the payment gateway redirection URL (and every other URL between the domain and the payment gateway's domain). The process is stopped (the AC does not complete the checkout), and all the information is recorded.

To ensure robustness, we repeat the automated checkout process five times for each URL. This repetition mitigates potential failures due to factors such as unavailable items or incorrect button selections. Examples of websites where the automated checkout process failed to record the checkout path are provided in Figure 5. It is important to note that merchant ID extraction does not require authentication with the payment processor's website. However, if future circumstances necessitate login credentials, an additional authentication step can be incorporated prior to initiating the checkout process. Furthermore, at any stage of the automated checkout procedure, if an express checkout option is available, we utilize this feature to navigate directly to the checkout page for efficient merchant ID retrieval. This methodology ensures comprehensive data collection while maintaining flexibility to adapt to varying e-Commerce website structures and checkout flows.

During the Auto-Checkout process, we collect the following information: (1) the website's main page source, (2) the website's main page screenshot, (3) all visited URLs and network requests made during the process, and (4) the performance log during the process. The performance log in a browser records information such as network requests, JavaScript execution time, and domain redirection. Such information is used to identify redirection chains that may happen during the Auto-Checkout process.

Algorithm 1 provides an overview of AC's approach to collecting the website's merchant IDs. For conciseness, we omit the functions that check for express checkout options or perform logging. AC employs a two-pronged strategy: it first searches for express checkout buttons at each stage of the process, attempting to click them and gather merchant IDs if present. In the absence of express checkout options, AC follows the conventional purchasing route, which involves adding an item to the cart and extracting information from the checkout page. This dual approach ensures maximum efficiency in data collection while simulating realistic user behaviors across various e-Commerce platforms.

C. Dataset

We deployed ScamMagnifier to collect data daily from May 2023 to June 2024. In total, as shown in Table I, we collected 1,155,237 domains, with 46,746 identified as potential fraudulent shopping websites using the ML-based classifier. The AC component successfully navigated and completed

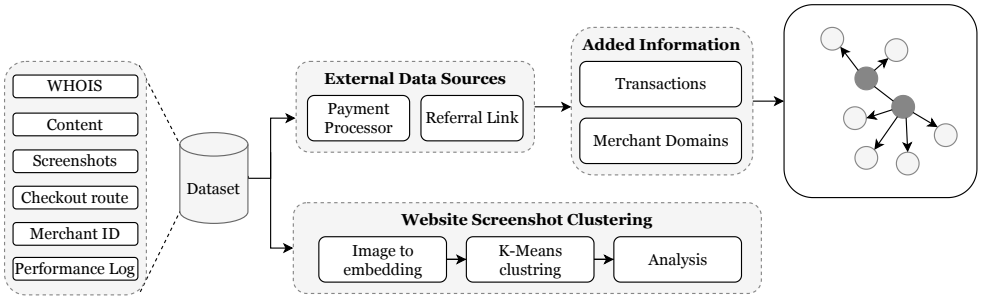


Fig. 3: Expanding the SCAMMAGNIFIER dataset.

Algorithm 1 Auto-Checkout Overall Process

```
Navigate to url
HandlePopups()
product_links ← FindProductLinks()
for i ← 1 to 10 do
    random_product ← Random(product_links)
    NavigateToProduct(random_product)
    HandleSecondaryPopups()
    SelectProductOptions()
    AddToCart()
    if ExpressCheckout() then
        InteractWithExpressCheckout()
    else
        NavigateToCart()
        button ← SearchForPaymentButtons()
        if button then
            FillCheckoutForms()
            AttemptPaymentSelection()
            ClickPaymentButton()
            checkout_success ← True
        end if
    end if
    if checkout_success then
        break
    end if
end for
```

TABLE I: Dataset statistics. Each row shows the number of data points that remain after each filtering step in our data processing pipeline.

Filtering Process	# of Domains
Operative Shopping Domains	1,155,237
Fraudulent shopping websites	46,746
Checkouts Completed	41,863

TABLE II: Statistics of extracted merchant IDs from domains that AC completed the checkout process.

Extracting Step	# of Domains
Total Completed Checkouts	*41,863*
Total Extracted Merchant IDs	5,278
Financial Org A Extracted Merchant IDs	4,484
Unique financial Org A Extracted Merchant IDs	2,790

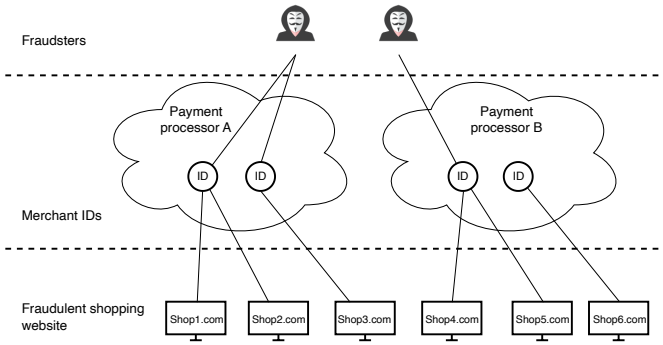
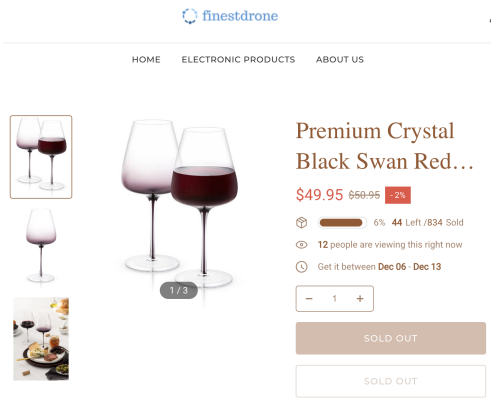


Fig. 4: The operation structure of fraudulent shopping websites based on our observations.

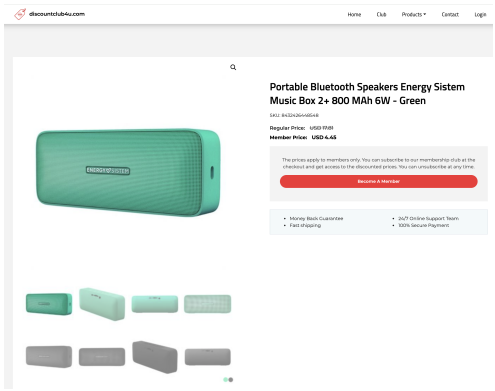
the checkout process for 41,863 domains. Ultimately, out of 41,863 completed checkouts, SCAMMAGNIFIER was able to extract merchant IDs for three[1] different payment processors for 5,278 total. Of these, 4,484 (the vast majority) were for financial Org A, and there were finally 2,790 unique financial Org A merchant IDs. Table II provides a breakdown of the extracted merchant IDs. Section V-A discusses the results of this analysis (in the context of the additional data we receive in our collaborations).

In our study, we manually reviewed 500 websites to find out why our system did not automatically extract merchant information. We found several key reasons. The most common problem, occurring in 46.12% of cases, was that websites only accepted credit card payments for server-side credit-card processing. Next, in 28.64% of cases, our checkout component fails to extract existing merchant information from the websites. Issues with checkout buttons not working were found in 12.15% of the websites. Some sites, about 9.35%, used unusual payment methods such as WhatsApp or direct payments. Finally, 3.74% of the sites required users to create and log into a member account, either free or paid. This analysis helps us understand the various challenges in automating the collection of merchant IDs from online stores. Examples of such websites are shown in Figure 5.

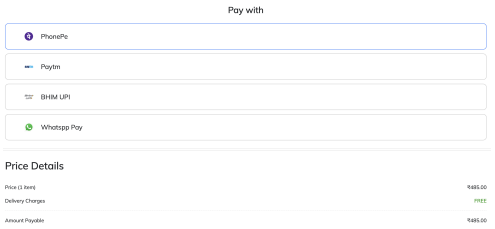
[1]We do not name these other payment processors as this information could easily deanonymize financial Org A.



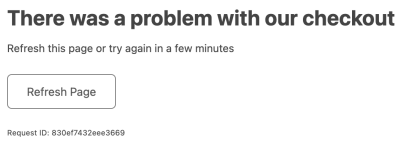
(a) Example of a website AC fails because of all items are sold out.



(b) Example of a website AC fails to proceed because a membership is required.



(c) Example of a website AC fails to proceed because of unusual payment methods.



(d) Example of a website AC fails because the checkout button does not work.

Fig. 5: Examples of various website AC failures.

D. Merchant ID-Oriented Historical Data Collection

During the Auto-Checkout process, we observed a peculiar behavior among scam sites where the checkout button redirected to an intermediary website before eventually navigating to the payment processor's domain. Since there is no prior work on scam site URL redirection, and we did not have end-to-end visibility, we decided to share our findings with the payment service provider financial Org A. Collaboration with financial Org A had multiple benefits for us including (1) scam confirmation to validate our methodology and findings, and (2) additional insights that could be leveraged to further enhance our study. Moreover, it also enabled us to share our perspective and novel findings were useful for financial Org A to improve their scam monitoring controls. In Section V, we briefly share the outcomes of our collaboration with financial Org A.

We then shift our focus to investigate the mechanisms through which users encounter fraudulent shopping websites. This investigation was conducted in collaboration with technical Org B, as detailed in Section V-C. Our findings reveal that advertisements play a pivotal role in directing a substantial volume of traffic towards these fraudulent shopping websites. This collaboration not only facilitated a comprehensive analysis of user pathways to fraudulent shopping websites but also provided valuable insights into the role of advertisements in this process.

With the merchant IDs collected by ScamMagnifier, we attempted to measure different attributes of fraudulent shopping websites with the unique perspective of our collaborators financial Org A and technical Org B, and Figure 3 outlines this additional analysis. This also enabled us to expand our dataset with an additional 14,394 fraudulent shopping websites.

V. Fraudulent Shopping Websites Economics

We study the economics of fraudulent shopping websites by answering the questions following the lifetime of fraudulent shopping websites, from setting up the websites to earning profits:

- How fraudulent shopping websites are structured? (Section V-A)
- How do fraudulent shopping websites evade detection? (Section V-B)
- How do they promote their website to get users' attention? (Section V-C)
- How fast and how much do they profit? (Section V-D)

A. Fraudulent Shopping Websites Structure

The primary objective of this research is to understand the operational tactics of fraudulent shopping websites. During our initial manual analysis, we observe a notable pattern: a singular entity, identified by a unique merchant ID, appears to

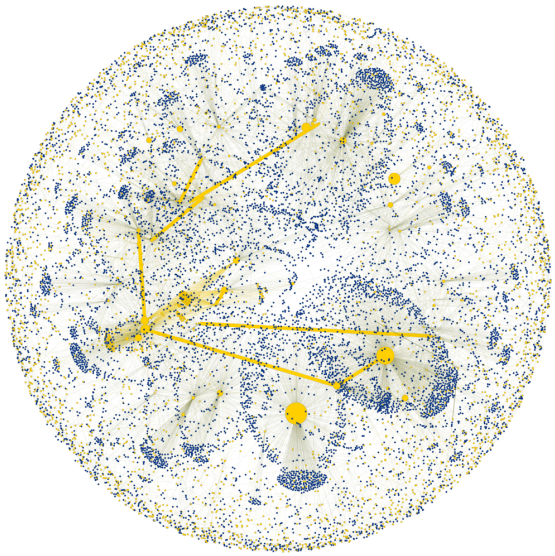


Fig. 6: Merchants (in yellow) and their registered domains (in blue). Highlighted links between merchants indicate they are connected to each other (more details in Section V-A).

administer multiple fraudulent shopping websites. Therefore, we constructed an attributed graph based on the 2,790 initial merchant IDs seen on 4,484 domains.

Our analysis reveal an intriguing landscape where the majority of merchants are linked tomultiple fraudulent shopping websites. Certain merchants exhibit control over a significantly larger number of these domains. Specifically, only 11 merchant IDs, representing 0.55% of the total, are connected to a single domain. In contrast, the most connected merchant ID is associated with 974 domains. This disproportionate domain control by a single merchant suggests the possibility of a more complex and extensive operation, potentially indicative of a large-scale fraudulent scheme. This observation catalyzes further investigation into the nature and extent of these operations, underpinning the necessity for a deeper and more comprehensive analysis.

In collaboration with financial Org A we expanded the domains controlled by each merchant. Also, financial Org A was able to link the vast majority of merchant IDs as controlled by the same entity by leveraging links in domain ownership:

- If two merchant IDs from two unrelated domains are registered by the same entity[2], then these merchant IDs are considered linked (i.e., controlled by the same entity).
- In a handful of cases, financial Org A was able to link merchant IDs because the WHOIS information of the domains matched each other, were unique, and also matched financial Org A's internal information.
- In a few cases, financial Org A was able to link merchant IDs when the merchant IDs had several financial Org A transactions between each other (financial Org A used their own expert knowledge to determine that the same entity controlled both merchant IDs).

Based on this analysis by financial Org A, we analyzed the links between merchant IDs and the collected domains. We then discovered 34 connected merchant IDs through our analysis pipeline, forming 6 connected components.

We create an attributed graph based on the expanded dataset[3] as shown in Figure 6, that illustrates the nodes of merchant IDs (in yellow) and the registered domains (in blue), and an edge between them when they are associated. The merchant ID nodes are larger according to the monetary value of their financial Org A transactions. In addition, the yellow edges indicate that financial Org A considers the two merchant IDs to be controlled by the same entity (using the process described previously).

Figure 6 emphasizes two key points: (1) Fraudulent shopping websites are very interconnected, and (2) Fraudulent shopping websites are created in campaigns and seem to be orchestrated at a large scale.

Figure 9 shows the statistics of the link of 14,394 domains to a merchant ID. The results show that 54.55% of all collected fraudulent websites are managed by only 10 merchant IDs. This finding is surprising and reveals that a small number of merchants are responsible for a large proportion of fraudulent websites. This reflects a similar trend that is observed in large-scale phishing campaigns [43], and suggests that fraudulent shopping websites are conducting similar large-scale campaigns.

These observations show that ScamMagnifier can help financial institutions build a proactive mitigation system by flagging related merchants and blocking their future transactions. Having information extracted through ScamMagnifier, financial organizations can prevent fraudulent activities and protect customers. This has the added benefit of reducing the profitability of fraudulent shopping websites. However, it is important to note that this approach may not be foolproof, as fraudsters may find ways to circumvent the system.

Another interesting observation is that our analysis of fraudulent shopping websites' contents revealed significant usage patterns among popular e-Commerce platforms. Specifically, we found that 58.74% of fraudulent shopping websites were created using Shopify, 19.13% using Wix, and 18.69% using Squarespace, and the rest are not using any well-known e-Commerce platform (3.42%). This data highlights that fraudsters predominantly exploit e-Commerce platforms due to their ease of use and widespread adoption, making them prime targets for creating deceptive online storefronts.

B. Evasion Technique Analysis

Now that we have established that fraudulent shopping websites are interconnected, we turn our attention to the evasion mechanisms employed by fraudulent shopping websites. We use the browser's performance log (described in Section IV-B) for this analysis. During the AC process, we record the performance logs for each URL visited, and we

[2]Financial Org A cannot disclose how this information was determined, however, we and financial Org A are confident in this technique.

[3]The SVG format of the graph is available here: osf.io/5uegv

Listing 1: Example of a request header from the performance log while visiting website **A**

```
'request': {
        'headers': {
            'Referer': 'WEBSITE B',
            ...
        },
        'initialPriority': 'VeryHigh',
        'isSameSite': False,
        'method': 'GET',
        'url': 'PAYMENT GATEWAY'
        ...
```

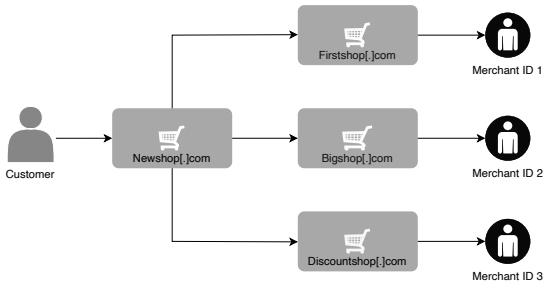


Fig. 7: Example of a website evading detection by the payment processor using redirection to a user-invisible domain.

use them to find redirections that happen during the checkout process of fraudulent shopping websites.

Through manual analysis, we observed a noteworthy pattern where some fraudulent shopping websites redirect to a *different* domain prior to reaching the payment processor. Listing 1 shows an example of such a redirect. To automate the analysis, we perform the following steps: (1) Extraction of log entries pertaining to the navigation towards the payment processor, (2) retrieval of the "referer" header, and (3) checking if the domain in the "referer" header matches the original website's domain.

The results showed that 263 fraudulent shopping websites redirected to an intermediary domain, which we refer to as **B**. This is important because domain **B** initiates the payment request to the payment processor. We hypothesize that fraudulent shopping websites use this indirect approach to reduce the likelihood of being directly detected by the payment processor: The payment processor's visibility is limited to domain **B**, thereby hiding the origin domain **A**. This intricate evasion mechanism is graphically represented in Figure 7.

Throughout our analysis, we did not encounter any instances of a redirection chain exceeding one hop. However, we did observe that a single website may redirect to multiple different domains before ultimately reaching the payment processor. This observation was made possible by conducting the Auto-Checkout process multiple times for each fraudulent shopping websites, thereby revealing the complex and dynamic nature of their operational strategies. These results underscore the critical role of ScamMagnifier, particularly in scenarios where financial institutions are unable to directly observe the actual fraudulent shopping websites events that precipitate the transactions.

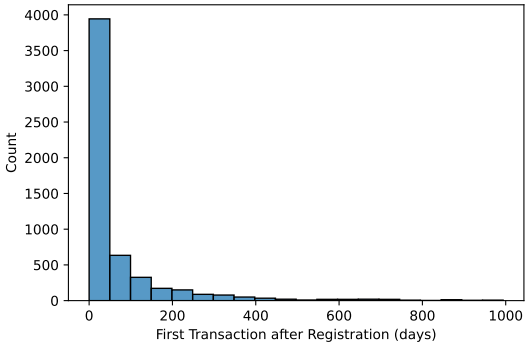


Fig. 8: Time delta between a fraudulent shopping websites domain's creation date and the merchant's first transaction date on financial Org A.

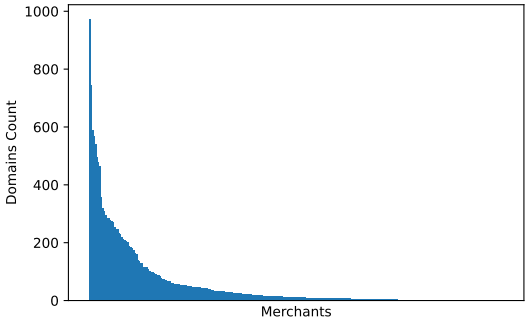


Fig. 9: Number of domains linked to every merchant ID in our dataset.

C. Marketing Strategies

For fraudulent shopping websites to monetize, they need a way to introduce themselves to users. In this section, we try to answer the question of how fraudulent shopping websites monetize. This question cannot be solely answered through the lens of a payment service provider or ScamMagnifier data collected through publicly available resources. Our intuition was that these scam sites must be aggressively advertising themselves to scam customers by offering discounted price offerings. However, to empirically confirm our intuition, we reached out to technical Org B because they have visibility into the online advertisement ecosystem.

We sent technical Org B the 14,394 fraudulent shopping websites domains, and they used their internal information to indicate referral URLs on visits to the domain. They sent us back the aggregated statistics on referral URLs. The results showed that 28.78% of users reached fraudulent shopping websites through advertisements on Facebook (including Instagram), 21.10% from Google (Ads or search results), and 9.38% from Bing. This analysis seems to indicate that fraudsters mostly use advertisements on popular social media websites.

We also find the use of `<meta>` tags as an SEO strategy to promote and enhance their visibility on the most common search engines to reach users. Figure 10 shows an example of an ad for a fraudulent shopping website. Analyzing fraudulent shopping websites ads from the Facebook Ad Library indicates

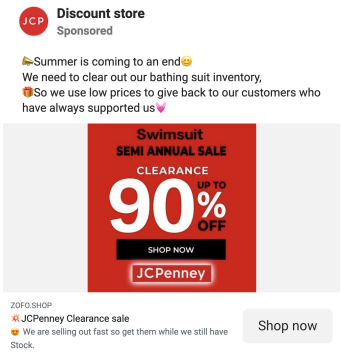


Fig. 10: Example of an advertisement that links to a fraudulent shopping website.

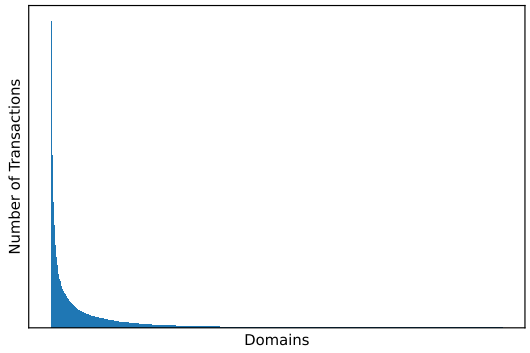


Fig. 11: Total number of transactions for each domain as reported by financial Org A. Note that the axis are deliberately unspecified—financial Org A considers transaction volume to be sensitive, therefore this figure only shows the distribution.

that fraudulent shopping websites create short-term ads several times.

Results from analyzing technical Org B's data are inline with analysis conducted by the Federal Trade Commission in 2023 about scam trends on social media platforms [25]. Their findings reveal that online shopping fraud dominates the landscape of reported scams, accounting for 44% of all social media-related fraud reports. The majority of these cases involve consumers who placed orders in response to advertisements on Facebook or Instagram but never received the promised items. According to this report, since 2021 users report losing $2.7 Billion to social media scams.

D. Time-To-Revenue

One aspect that suggests a complex criminal enterprise is how quickly criminals can monetize a fraudulent shopping domain. Therefore, we measure the *time-to-revenue*—that is, the time between domain registration and the first financial Org A transaction from that domain. This number is calculated by combining the WHOIS information (indicating when the domain was registered) with the financial Org A fraudulent shopping websites transactions.

Figure 8 shows the histogram of the time-to-revenue. Note that these figures are only an estimate, as a shopping website's first transaction may be through a payment processor that we could not collaborate with. Even with this limitation, we find that most fraudulent shopping websites become monetized shortly after being created. Specifically 20.17% of fraudulent shopping websites have transactions within 10 days after their creation date and 97.73% are monetized in less than a year.

VI. Detecting Fraudulent Shopping Websites

In this research, we have developed a browser extension to effectively detect fraudulent shopping websites that are often undetectable by payment processors and ad service providers. These fraudulent websites can evade detection because payment processors cannot always view the actual website users are attempting to purchase from, and ad service providers face challenges due to the sheer volume of ads requiring review. By focusing on client-side detection, our extension leverages economic insights, particularly the repetitive use of merchant IDs by fraudulent shopping websites. We have constructed a blocklist database of fraudulent merchant IDs. The browser extension tries to navigate the website to extract the merchant ID to classify. If the merchant ID does not exist in our blocklist, we rely on Beyond Phish classifier's results. Our evaluation highlights the extension's potential to enhance online transaction security by preemptively identifying and mitigating fraud risks. We provide the detailed information about our browser extension below.

We developed a Chromium-based extension that leverages ScamMagnifier framework to warn users about potential fraudulent shopping websites. The extension is designed to integrate seamlessly with the user's browsing experience, providing real-time assessments of websites as they are visited. Another aspect of our extension is to understand the effectiveness of our approach in blocking merchants.

First, the extension classifies the domain to get the likelihood of it being fraudulent shopping websites. Next the Auto-Checkout component of our data collection pipeline (component ❸ of Figure 3) is activated. This pipeline is designed to gather and analyze detailed information about the website's merchant, scrutinizing various aspects such as the merchant's historical data, associated websites, and any previous fraudulent shopping activities. If the analysis confirms the merchant's involvement in fraudulent operations, it then presents the user with additional information about the potential fraudulent shopping website. Moreover, our extension allows users to provide their feedback. Users can report any issues or discrepancies they observe, contributing to the continuous improvement of the pipeline's accuracy and reliability.

To evaluate the performance of the browser extension, we first distributed the browser extension through our professional channels. Out of 497 user-checked websites, Beyond Phish flagged 51 as potential scams. Of the remaining, Auto-Checkout extracted 54 unique merchant IDs, and 17 of these matched known scam websites.Overall, users checked 497 websites, out of which, the Beyond Phish classifier flagged 51 as potential scams.

We manually verified all 497 websites with three independent security experts per website. We employed a team consisting of two authors and one non-author as reviewers. Each reviewer independently examined the websites' features

TABLE III: Detection results by the browser extension and baselines. The results are calculated based on expert labels. Experts label 86 of the websites as fraudulent shopping websites and 411 as legitimate shopping websites. Each cell shows the rate (in percentage) and the number of matching websites (in parentheses).

Method	Detection Rate	False Positive Rate	True Negative Rate	False Negative Rate
Beyond Phish	59.30%(51)	0.49%(2)	99.51%(409)	40.70%(35)
Beyond Phish + Auto-Checkout	**76.74% (66)**	0.49%(2)	99.51%(409)	23.26%(20)
NetCraft Extension	23.25%(20)	0.24% (1)	99.76% (410)	76.74% (66)

and used search to determine legitimacy. The evaluation process followed clear guidelines: analyzing user reviews, checking DNS records, investigating website history through Archive.org, verifying location information, and examining associated social media accounts.

Our analysis ultimately confirmed 86 fraudulent shopping websites, showcasing the system's ability to catch sophisticated scams that might bypass initial classification. During our manual analysis, of the 497 domains labeled there were 8 domains that were not in unanimous agreement.

Table III shows the results, which demonstrate the effectiveness of Auto-Checkout and Beyond Phish for improved fraudulent shopping websites detection. The integrated approach achieved a much higher detection rate of 76.74% (66 detected fraudulent shopping websites) compared to Beyond Phish's standalone performance of 59.30% (51 detected fraudulent shopping websites). This improvement was achieved while maintaining the same False Positive Rate of 0.49% and True Negative Rate of 99.51%, indicating that the Auto-Checkout model successfully identified 15 additional fraudulent shopping websites that Beyond Phish missed, reducing the False Negative Rate from 40.70% to 23.26%.

When compared to the NetCraft Extension, both Beyond Phish and the combined approach showed substantially better performance, as NetCraft only achieved a 23.25% detection rate with 66 missed fraudulent shopping websites. The consistent False Positive and True Negative rates across Beyond Phish and the combined approach suggest that the Auto-Checkout model effectively complements Beyond Phish by identifying additional fraudulent shopping websites through merchant ID analysis without introducing new false positives.

The enhanced performance of "Beyond Phish + Auto-Checkout" can be attributed to the AC component's ability to detect previously missed websites through shared merchant ID associations with known fraudulent shopping websites detected by Beyond Phish classifier. However, the system still fails to detect 20 fraudulent websites due to their merchant IDs not being previously associated with previously detected fraudulent shopping websites. While the AC component shows potential for improving detection, its effectiveness is contingent upon the false positive rate of the initial detections.

VII. Discussion

Fraudulent shopping websites are not well-understood cyber threats. These websites masquerade as legitimate shopping websites and use various techniques to deceive users into purchasing non-existent or fraudulent items. Online advertising is one of the techniques that allows miscreants to easily publish their deceptive ads on sites that are likely already popular among the targeted victims [54].

The monetary loss [21, 32] and our findings in this paper highlight the importance of detecting fraudulent merchants and taking action against them. Moreover, financial Org A also observed "significant activity" associated with the collected merchants[4].

Further insights on our side reveal that fraudulent sites strategically maintain low refund rates by circumventing formal payment processes, opting instead for direct reimbursements or store credits. Additionally, we identified a pattern of fraudsters creating multiple fraudulent shopping websites, often linked by common merchant IDs, indicating centralized control and a strategy to expand their victim pool. We believe that our findings represent potential directions for understanding fraudulent shopping websites and creating a defense mechanism to protect users.

The proactive extraction and analysis of merchant IDs from fraudulent websites is a critical task. These unique identifiers enable the identification of fraudulent entities and the scope of their operations. Analyzing these IDs can reveal patterns, such as the recurrence of the same merchant ID across multiple fraudulent websites, indicating an interconnected network of related fraudulent activities. Identified fraudulent IDs can be blocklisted or flagging their future transactions, and reported to financial institutions and law enforcement agencies. Integrating this knowledge into existing fraud detection systems enhances their ability to respond to similar threats, improving overall security against financial frauds. This proactive approach also helps prevent financial losses by recognizing and responding to threats early, thereby protecting potential victims' financial assets.

From our analysis, we believe that the following aspects are critical to mitigate fraudulent shopping websites: (1) victimization happens quickly, so early detection is key, (2) payment processors should consider how to verify the provenance of the purchase-originating domain, (3) advertisers need to improve their detection of fraudulent shopping websites, and (4) coordination between payment processors is key to identifying the complex criminal activities.

Data Sharing. Fraudulent merchants use various platforms such as domain registrars, payment gateways, and e-Commerce platforms to create and manage fraudulent shopping websites. This creates an opportunity for data sharing across the ecosystem to better identify fraudulent merchants based on proactive intelligence indicative of fraudulent shopping websites.

[4]While financial Org A cannot publicly share the numbers, the fraud seems to be significant and profitable.

Our experimental results indicate that merchant transactions are crucial in detecting fraudsters. In contrast, relying solely on publicly available information is insufficient to create a detection method that can protect users from fraudulent activities (i.e., we cannot predict a merchant's future domains' legitimacy). Miscreants can create many different types of fraudulent shopping websites using different information, which makes it difficult to detect and prevent such attacks. For example, our results from Section V-A indicates that miscreants are connected to each other and create fraudulent shopping websites at scale. This suggests that proactively detecting fraudulent shopping websites requires creating a data-sharing protocol between vetted entities within the ecosystem to mitigate fraudulent shopping websites accurately at scale.

We have disclosed our findings to financial Org A, Microsoft, Google, and prominent e-Commerce platforms. Our collaboration with these entities aims to bolster the security and trustworthiness of online transactions. Moreover, financial Org A has incorporated SCAMMAGNIFIER to run it internally, so that fraudulent merchant IDs can be identified proactively. This adoption will result in detecting and taking action against fraudulent merchants in the future, ensuring a safer online financial ecosystem for users.

Shared Website Content. Fraudulent shopping websites are typically fabricated by malicious actors who lack ownership of an authentic product. Consequently, they tend to use generic stock images to represent the merchandise they purport to sell [17]. We performed a simple clustering method on fraudulent shopping websites screenshots to cluster them into ten categories. As evidenced in Figure 12, most of these assets are recurrently used across multiple fraudulent shopping websites. Based on an analysis of a selected subset from our compiled dataset, it is often discernible that these resources originate from legitimate shopping websites. Similarly, the textual content of the fraudulent shopping websites can be shared among them with minimum change. However, while such techniques may identify the sites today, it is crucial for the security ecosystem to begin planning mitigations before the sites evolve to a level of sophistication where they are challenging to detect. One particular section of fraudulent shopping websites that caught our attention was the *about us* page. This page usually includes information about the business and legitimate websites provide their story, phone number, address, etc. We observed that fraudulent shopping websites use a *about us* section that includes information from a legitimate shopping website. Figure 13 shows an example of such information on fraudulent shopping websites *about us* section.

Limitations. Our analysis should be considered alongside certain limitations. In our analysis, we collaborated with two organizations and targeted collection of merchant IDs from three major payment processors, which may skew our findings. However, our proposed data collection pipeline is not limited or tied to a specific payment processors, and future analysis can be done on a broader ecosystem. Moreover, in our data collection process, upon checking the number of payment processors, we observed that less than 10% of the websites had payment processors other than financial Org A. This means that the results here should be taken as a lower bound, as our collaboration with financial Org A allows some visibility into transactions for the observed fraudulent shopping websites.

In the course of our study, it is important to note the inability to directly collaborate with credit and debit card providers. This constraint arises from the fact that these providers do not offer an accessible payment gateway for our research purposes (to easily extract a comparable merchant ID). Access to credit/debit card providers could allow us more insight into the fraudulent shopping websites ecosystem, as many of the fraudulent shopping websites did not use any payment gateway, as they instead perform server-side credit card processing.

Our study is also limited by the information that our collaborators financial Org A and technical Org B can share with us and also release publicly. While these restrictions limit the information from our study, without these collaborations we would not understand the scope, scale, and *coordinated* nature of fraudulent shopping websites.

VIII. RELATED WORK

Extensive research has been conducted on phishing attacks, with numerous studies examining their evolution [41, 42], temporal progression [43], and detection mechanisms [8, 33]. However, there is a notable gap in the literature regarding fraudulent shopping websites- a relatively unexplored domain that needs further investigation. fraudulent shopping websites often employ deceptive practices that can mislead users and result in significant financial losses, as discussed in Section I.

This section provides a comprehensive review of the existing literature, first addressing studies related to phishing websites, followed by an examination of the limited research on fraudulent shopping websites. By providing research works on these two areas, we aim to highlight the critical need for more extensive analysis of fraudulent shopping websites and their impact on the internet users.

Phishing Websites: Heijden et al. [50] employed a methodology that correlates URLs detected in phishing emails (as reported to a designated entity) with the timestamps of individual target clicks. The study highlighted cognitive and technical characteristics that distinguish successful phishing emails by amalgamating this click data with an email content analysis. Such findings offer strategic insights that can aid in prioritizing and neutralizing particularly potent phishing URLs. In a parallel effort, Oest et al. [43] proposed a framework tailored to passively monitor victim traffic directed to phishing pages while simultaneously safeguarding tens of thousands of accounts. An important observation from their work is the frequent request by many phishing pages for web resources from third-party entities, including websites they impersonate. Capitalizing on this behavior, they tracked victim traffic to active phishing pages, recording visits from an 4.8 million victims. Their research disclosed that the median phishing campaign lasts 21 hours, with a minute subset of notably successful campaigns accounting for a 89% of all victims.

Fraudulent Shopping Websites: Bitaab et al. study social engineering attacks during the initial phases of the pandemic [14]. By consolidating and analyzing a myriad of data-sources—ranging from DNS records, phishing URLs, phishing website source codes, and phishing emails to web

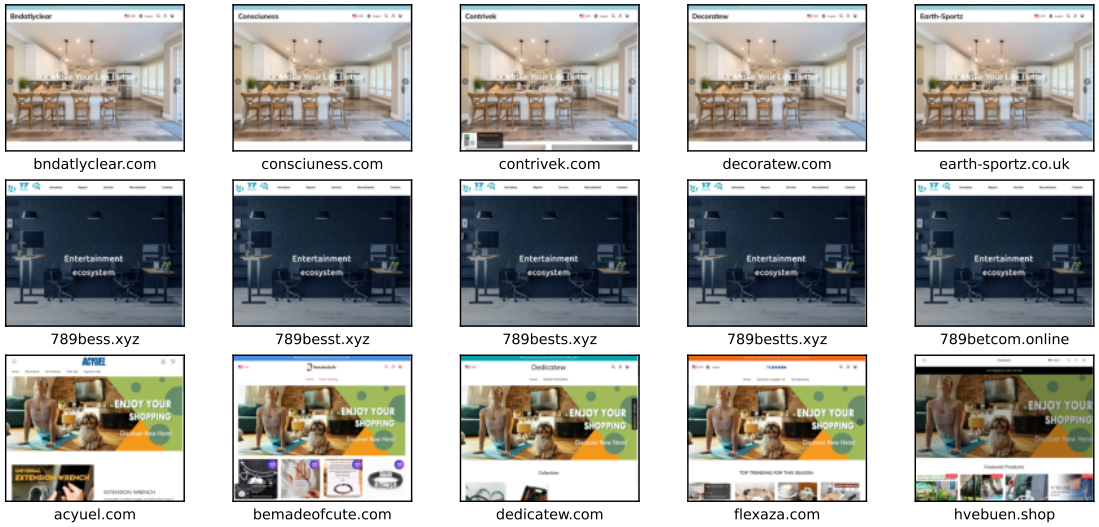


Fig. 12: Random samples from three different clusters of fraudulent shopping websites.

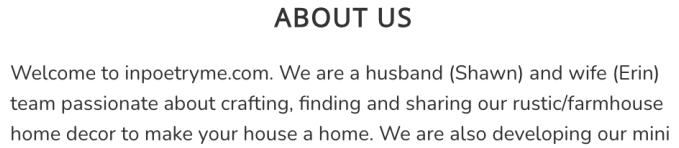


Fig. 13: An example of a commonly used "About Us" section in fraudulent shopping websites.

traffic data for phishing sites, news articles, and official government announcements—they found that phishing attack traffic swelled to 220% of its pre-pandemic rate during March and April 2020. This surge overshadowed typical seasonal upticks. Cybercriminals leveraged the heightened pandemic-related anxieties to entice internet users. The authors' findings not only indicate an escalation in fraudulent shopping websites scams but also spotlight the inadequacies of contemporary defensive mechanisms in detecting such schemes.

Another method by Kotzias et al. introduces a method for detecting scam shopping websites [34]. They extract 111 fetures from each collected domain to create a random forest-based classifier capable of determining the legitimacy of an online shopping website. Upon analyzing this method, we found the data collection to be biased as it relies on four specific scam repositories, filtering out popular domains in the Tranco top 100K, and assumes they are less likely to be scams. Additionally, their real-world application tested 760,000 online shopping domains, but the manual verification was limited to only 100 websites, raising concerns about the accuracy and comprehensiveness of their validation process. Finally, this method is not open-sourced and we do not have access to the trained model to reproduce the results.

In the same domain, Beyond Phish [13], a fraudulent shopping websites detection method by Bitaab et al., has introduced a novel method for detecting fraudulent shopping websites. This approach utilizes manually extracted features based on content, URL, WHOIS information, and social media to assess the legitimacy of fraudulent shopping websites. While such methods show promise, there remains a critical need for detection and mitigation strategies. Current approaches often struggle to keep pace with rapidly evolving fraudulent techniques. Our study aims to address these limitations by examining the underlying mechanisms of fraudulent shopping websites. We seek to develop a deeper understanding of fraudulent shopping websites characteristics and behaviors, identifying key indicators that can be used for early detection, and proposing a mitigation strategies to prevent fraudulent shopping websites from reaching a wide user base.

By focusing on these objectives, we aim to contribute to the development of more robust and adaptive detection systems that can effectively combat the evolving threat of fraudulent shopping websites. Another work by Anderson et al. [9] investigate the expansive infrastructure underpinning internet scams, focusing on web servers promoted through spam. Although spam emails, which accounted for over 80% of all internet emails in 2006, act as the conduit, the more pressing issue centers on the underlying scams designed to exploit unsuspecting users. These scams, which span a spectrum from product sales to the deployment of malicious software, are critically dependent on specific infrastructures. Remarkably, a singular spam campaign may employ thousands of mail relay agents yet depend on just one server to handle responses from its recipients. This makes the infrastructure an indispensable element for the financial success of such spam-driven endeavors. Through the analysis of a daily influx of roughly 150,000 spam emails, the authors discerned more than 2,000 distinct scams hosted on in excess of 7,000 servers.

Phishing and Scam Analysis Techniques: Starove et al. focus on how malicious actors reuse analytics IDs across different websites and platforms [48]. They developed a system that can automatically identify and extract analytics IDs from malicious websites, browser extensions, binaries, and mobile apps. Their findings show that attackers often reuse the same analytics

ID across multiple malicious pages and even across different platforms. This information can be used to discover previously unknown malicious content, cluster malicious content into campaigns, and even deanonymize malicious actors who are hiding behind WHOIS privacy protection services. Another similar study by Subramani et al. investigates patterns employed by modern phishing websites that provide a sense of legitimacy and evade detection mechanisms [49]. They developed an intelligent crawler that can automatically interact with phishing websites, identify their UI elements, and simulate user interactions. Their research found that modern phishing sites often impersonate a brand without closely mimicking the design of legitimate websites, elicit personal information using a multi-step process, embed modern user verification systems (including CAPTCHAs), and sometimes conclude by reassuring users that their private data was not stolen.

Given the inherently distributed nature of fraudulent shopping websites across diverse infrastructures, obtaining empirical data on the interconnections amongst fraudulent shopping websites poses challenges. Nevertheless, such measurements can unveil insights that might remain obscure when scrutinized at a finer granularity. To the best of our knowledge, our study represents the first effort to provide a comprehensive perspective on fraudulent shopping websites at scale, bridging the gap between fraudulent shopping websites and associated merchant information and analyzing the interconnectedness of merchants across various channels.

IX. Conclusion

In the contemporary digital landscape, scams remian a significant threat to Internet users. This is because fraudulent shopping websites are not isolated one-off instances; rather, we find that they are likely part of a sophisticated cybercrime operation, with the criminals potentially making significant profits. The deployment of ScamMagnifier, designed for the automated identification of fraudulent merchants, underscores the pressing need for further proactive countermeasures within the ecosystem. Our framework, which is adaptable to any payment provider, presents a promising solution to strengthen defenses by preemptively blocking fraudulent merchants and accelerating the detection of scams. This acceleration is achieved by detecting one fraudulent merchant and subsequently blocking other websites connected to the same merchant. Finally, to enhance these efforts, we have introduced a Chromium-based extension that utilizes our proposed auto checkout component on top of the ML based classifier to alert users about potential scam websites they visit.

Acknowledgements

We thank the anonymous reviewers for their valuable feedback. Our appreciation also extends to the two industry organizations for their insightful contributions and collaboration. This work was supported in part by the Advanced Research Projects Agency for Health (ARPA-H) under Contract No. SP4701-23-C-0074, National Science Foundation (NSF) Grant No. 2232915, Institute of Information & communications Technology Planning & Evaluation (IITP) grants RS-2024-004398199 (AI-Based Automated Vulnerability Detection and Safe Code Generation) and RS-2024-00442085 (Development of V2X Infra Security Core Technologies for Autonomous Vehicle Services), Department of Navy award N00014-23-1-2563 and N00014-24-1-2193 issued by the Office of Naval Research. We gratefully acknowledge their support.

References

[1] https://www.whoisds.com/newly-registered-domains.

[2] "Find your merchant id," https://community.shopify.com/c/authentication-and-access/store-id/td-p/2099999.

[3] 2023, https://www.netcraft.com/apps-extensions/browser-extension/.

[4] 2023, https://www.phishlabs.com/about/.

[5] 2023, https://www.markmonitor.com/.

[6] 2023, https://www.sans.org/security-awareness-training/.

[7] S. Abdelnabi, K. Krombholz, and M. Fritz, "Visualphishnet: Zero-day phishing website detection by visual similarity," in *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security*, 2020, pp. 1681–1698.

[8] L. M. Abdulrahman, S. H. Ahmed, Z. N. Rashid, Y. S. Jghef, T. M. Ghazi, and U. H. Jader, "Web phishing detection using web crawling, cloud infrastructure and deep learning framework," *Journal of Applied Science and Technology Trends*, vol. 4, no. 01, pp. 54–71, 2023.

[9] D. S. Anderson, C. Fleizach, S. Savage, and G. M. Voelker, "Spamscatter: Characterizing internet scam hosting infrastructure," Ph.D. dissertation.

[10] L. Beltzung, A. Lindley, O. Dinica, N. Hermann, and R. Lindner, "Real-time detection of fake-shops through machine learning," in *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 2020, pp. 2254–2263.

[11] Better Business Bureau, "Online shopping fraud," 2021, https://www.bbb.org/all/scamstudies/fake_online_retailers_study/online_shopping_fraud_study.

[12] 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 *30th USENIX Security Symposium (USENIX Security 21)*, 2021, pp. 3757–3774.

[13] M. Bitaab, H. Cho, A. Oest, Z. Lyu, W. Wang, J. Abraham, R. Wang, T. Bao, Y. Shoshitaishvili, and A. Doupé, "Beyond phish: Toward detecting fraudulent e-commerce websites at scale," in *2023 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, 2023, pp. 2566–2583.

[14] M. Bitaab, H. Cho, A. Oest, P. Zhang, Z. Sun, R. Pourmohamad, D. Kim, T. Bao, R. Wang, Y. Shoshitaishvili *et al.*, "Scam pandemic: How attackers exploit public fear through phishing," in *2020 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2020, pp. 1–10.

[15] C. Carpineto and G. Romano, "Learning to detect and measure fake e-commerce websites in search-engine results," in *Proceedings of the international conference on web intelligence*, 2017, pp. 403–410.

[16] J. Corbetta, L. Invernizzi, C. Kruegel, and G. Vigna, "Eyes of a human, eyes of a program: Leveraging different views of the web for analysis and detection," in *Research in Attacks, Intrusions and Defenses: 17th International Symposium, RAID 2014, Gothenburg, Sweden, September 17-19, 2014. Proceedings 17*. Springer, 2014, pp. 130–149.

[17] DataDome, "Web scraping protection: How to prevent scraping & crawler bots," 2023, https://datadome.co/learning-center/scraper-crawler-bots-how-to-protect-your-website-against-intensive-scraping/.

[18] K. Du, H. Yang, Z. Li, H. Duan, and K. Zhang, "The {Ever-Changing} labyrinth: A {Large-Scale} analysis of wildcard {DNS} powered blackhat {SEO}," in *25th USENIX Security Symposium (USENIX Security 16)*, 2016, pp. 245–262.

[19] Federal Bureau of Investigation, "Holiday Scams," https://www.fbi.gov/how-we-can-help-you/safety-resources/scams-and-safety/common-scams-and-crimes/holiday-scams.

[20] ——, "Scammers using computer-technical support scams to target victims," 2022, https://www.ic3.gov/Media/Y2022/PSA221110.

[21] Federal Trade Commission, "Consumer sentinel network data book 2022," https://www.ftc.gov/system/files/ftc_gov/pdf/CSN-Data-Book-2022.pdf.

[22] ——, "FTC Data Book 2020," https://www.ftc.gov/system/files/documents/reports/consumer-sentinel-network-data-book-2020/csn_annual_data_book_2020.pdf.

[23] ——, "Scammers are hijacking job ads," https://consumer.ftc.gov/consumer-alerts/2023/05/scammers-are-hijacking-job-ads-heres-how-spot-fakes.

[24] ——, "FTC Data Book 2023," 2023, https://www.ftc.gov/system/files/ftc_gov/pdf/CSN-Annual-Data-Book-2023.pdf.

[25] ——, "FTC press releases," 2023, https://www.ftc.gov/news-events/news/press-releases/2023/10/ftc-data-shows-consumers-report-losing-27-billion-social-media-scams-2021.

[26] M. N. Feroz and S. Mengel, "Phishing url detection using url ranking," in *2015 ieee international congress on big data*. IEEE, 2015, pp. 635–638.

[27] Google, "Find your merchant id," https://support.google.com/googleplay/android-developer/answer/7163092?hl=en.

[28] S. Hao, A. Kantchelian, B. Miller, V. Paxson, and N. Feamster, "Predator: proactive recognition and elimination of domain abuse at time-of-registration," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 1568–1579.

[29] I. Hernández, C. R. Rivero, and D. Ruiz, "Deep web crawling: a survey," *World Wide Web*, vol. 22, pp. 1577–1610, 2019.

[30] 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 *28th USENIX Security Symposium (USENIX Security 19)*, 2019, pp. 1273–1290.

[31] Internet Crime Complaint Center, "Internet crime report 2022," 2023, https://www.ic3.gov/Media/PDF/AnnualReport/2022_IC3Report.pdf.

[32] C. Kanich, N. Weaver, D. McCoy, T. Halvorson, C. Kreibich, K. Levchenko, V. Paxson, G. M. Voelker, and S. Savage, "Show me the money: Characterizing spam-advertised revenue," in *20th USENIX Security Symposium (USENIX Security 11)*, 2011.

[33] A. Karim, M. Shahroz, K. Mustofa, S. B. Belhaouari, and S. R. K. Joga, "Phishing detection system through hybrid machine learning based on url," *IEEE Access*, vol. 11, pp. 36 805–36 822, 2023.

[34] P. Kotzias, K. Roundy, M. Pachilakis, I. Sanchez-Rola, and L. Bilge, "Scamdog millionaire: Detecting e-commerce scams in the wild," in *Proceedings of the 39th Annual Computer Security Applications Conference*, 2023, pp. 29–43.

[35] K. Levchenko, A. Pitsillidis, N. Chachra, B. Enright, M. Félegyházi, C. Grier, T. Halvorson, C. Kanich, C. Kreibich, H. Liu *et al.*, "Click trajectories: End-to-end analysis of the spam value chain," in *2011 ieee symposium on security and privacy*. IEEE, 2011, pp. 431–446.

[36] Y. Lin, R. Liu, D. M. Divakaran, J. Y. Ng, Q. Z. Chan, Y. Lu, Y. Si, F. Zhang, and J. S. Dong, "Phishpedia: A hybrid deep learning based approach to visually identify phishing webpages," in *30th USENIX Security Symposium (USENIX Security 21)*, 2021, pp. 3793–3810.

[37] J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Beyond blacklists: learning to detect malicious web sites from suspicious urls," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009, pp. 1245–1254.

[38] S. Maroofi, M. Korczyński, and A. Duda, "Are you human? resilience of phishing detection to evasion techniques based on human verification," in *Proceedings of the ACM Internet Measurement Conference*, 2020, pp. 78–86.

[39] Microsoft, "Microsoft defender smartscreen," https://learn.microsoft.com/en-us/windows/security/operating-system-security/virus-and-threat-protection/microsoft-defender-smartscreen/.

[40] T. Nelms, R. Perdisci, M. Antonakakis, and M. Ahamad, "Towards measuring and mitigating social engineering software download attacks," in *25th USENIX Security Symposium (USENIX Security 16)*, 2016, pp. 773–789.

[41] 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 *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2019, pp. 1344–1361.

[42] 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 *2018 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2018, pp. 1–12.

[43] A. Oest, P. Zhang, B. Wardman, E. Nunes, J. Burgis, A. Zand, K. Thomas, A. Doupé, and G.-J. Ahn, "Sunrise to sunset: Analyzing the end-to-end life cycle and effectiveness of phishing attacks at scale," in *29th {USENIX} Security Symposium ({USENIX} Security 20)*, 2020.

[44] PayPal, "How to get started with your business account," https://www.paypal.com/c2/webapps/mpp/how-to-guides/sign-up-business-account?locale.x=en_C2.

[45] ——, "Onboard sellers before payment," https://developer.paypal.com/docs/multiparty/seller-onboarding/before-payment/.

[46] O. K. Sahingoz, E. Buber, O. Demir, and B. Diri, "Machine learning based phishing detection from urls," *Expert Systems with Applications*, vol. 117, pp. 345–357, 2019.

[47] S. Sheng, B. Wardman, G. Warner, L. Cranor, J. Hong, and C. Zhang, "An empirical analysis of phishing blacklists," 2009.

[48] O. Starov, Y. Zhou, X. Zhang, N. Miramirkhani, and N. Nikiforakis, "Betrayed by your dashboard: Discovering malicious campaigns via web analytics," in *Proceedings of the 2018 World Wide Web Conference*, 2018, pp. 227–236.

[49] K. Subramani, W. Melicher, O. Starov, P. Vadrevu, and R. Perdisci, "Phishinpatterns: measuring elicited user interactions at scale on phishing websites," in *Proceedings of the 22nd ACM Internet Measurement Conference*, 2022, pp. 589–604.

[50] A. Van Der Heijden and L. Allodi, "Cognitive triaging of phishing attacks," in *28th USENIX Security Symposium (USENIX Security 19)*, 2019, pp. 1309–1326.

[51] J. Wadleigh, J. Drew, and T. Moore, "The e-commerce market for" lemons" identification and analysis of websites selling counterfeit goods," in *Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 1188–1197.

[52] 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, San Diego, California, USA*, 2010.

[53] P. Yang, G. Zhao, and P. Zeng, "Phishing website detection based on multidimensional features driven by deep learning," *IEEE access*, vol. 7, pp. 15 196–15 209, 2019.

[54] E. Zeng, T. Kohno, and F. Roesner, "What makes a "bad" ad? user perceptions of problematic online advertising," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–24.

[55] P. Zhang, A. Oest, H. Cho, R. Johnson, B. Wardman, S. Sarker, A. Kpravelos, T. Bao, R. Wang, Y. Shoshitaishvili, A. Doupé, and G.-J. Ahn, "CrawlPhish: Large-scale Analysis of Client-side Cloaking Techniques in Phishing," in *Proceedings of the 42nd IEEE Symposium on Security and Privacy (Oakland)*, San Francisco, CA, May 2021.

APPENDIX

A. *Keywords used in the Auto-Checkout Process*

Table IV shows the set of keywords and URL patterns employed by SCAMMAGNIFIER to identify and extract relevant interface elements during the AC's checkout workflow.

TABLE IV: Keywords used in each step of SCAMMAGNIFIER

Checkout Phase	Keywords
Shadow Filter	`/search`, `/story`, `/live`, `/help`, `/email`, `/account`, `/cookie`, `/about`, `/cart`, `/track`, `/contact`, `/privacy`, `/policy`, `/terms`, `/refund`, `/login`, `/bag`, `/faq`, `/support`, `/customer-service`, `/returns`, `/shipping`, `/signup`, `/register`, `/forgot-password`, `/profile`, `/legal`, `/disclaimer`, `/promo`, `/offers`, `/sale`, `/news`, `/events`, `/blog`
Add to Cart	add to cart, add to basket, add to shopping cart, add to shopping basket, put in cart, put in basket, place in cart, place in basket, buy now, purchase now
Proceed to Checkout	proceed to checkout, checkout, check out, go to checkout, continue to checkout, proceed with purchase, complete purchase, finish order, review order, proceed to payment, confirm order, complete order