Checkout Checkup:
Misuse of Payment Data from Web Skimming

Phoebe Rouge, Christina Yeung, Daniel Salsburg, Joseph A. Calandrino
Federal Trade Commission
{prouge, cyeung, dsalsburg, jcalandrino}@ftc.gov

***Abstract*—Web skimming poses an increasingly serious threat to online shoppers, with recent cases reportedly occurring on prominent websites like British Airways and Newegg. While prior work largely measures the prevalence of these attacks and explores their technical details, we examine the use of credentials stolen via web skimming. We identified 50 sites apparently compromised to host payment card skimming code. For each site, we attempted to purchase items using a unique payment card. Over an eleven-month study period, we monitored the payment cards for signs of abuse. We observed attempted misuse of 15 of the 50 payment cards. With a single exception, the time from exposure of a payment card until observed misuse was at least 50 days. Thieves tried to use the 15 cards at least 45 times. These attempted payments ranged from $0.10 to $122.44 and totaled $1,342.91. To place our findings in context, we compare these results to a separate study in which we exposed payment data directly via an online paste site. Our observations suggest that the impact of web skimming may not be apparent for an extended period following an incident.**

***Index Terms*—Security, Cybercrime**

I. Introduction

Payment card fraud is as old as the cards themselves. When consumers use credit and debit cards in new circumstances, new opportunities for fraud follow. In the physical world, thieves might obtain payment card data by stealing the physical cards or installing card skimmers on gas pumps. As consumers shop more online, an online analog to physical card skimming, called *web skimming*, *online skimming*, or *e-skimming* (among other names), has emerged. In this paper, we examine how thieves use payment card data from web skimming.

To perform web skimming, a malicious party compromises a website that takes payment information from consumers, installing exploit code on the site. The exploit code siphons user-entered payment details to the malicious party [13]. No sign of an attack may be apparent to a victim consumer, whose purchase may proceed as usual.

The impact of web skimming has been significant and widespread. In 2016, a security company identified more than 5,900 online stores suspected of hosting web skimming code [50]. A web skimming attack on British Airways in 2018 led to the compromise of data for approximately 500,000 consumers [5], [19]. Ticketmaster, Newegg, and Macy's are among the other high-profile targets of web skimming in recent years [1], [52], [58]. In 2019, the FBI and PCI (a payment industry forum) both issued warnings regarding the threat of web skimming to businesses [16], [41].

Existing analyses of web skimming have primarily come from industry and the media. This prior work largely focuses on technical details of the attacks, the prominence of particular campaigns, or high-profile incidents (Section II). We have less insight into how attackers use payment card data that they obtain via these attacks. Information about this use may shed additional light on the attacks and attackers. Such insights are difficult to obtain, particularly without access to large-scale transaction-level data. Outside of financial institutions, few parties beyond the attackers themselves have significant visibility into the resulting misuse of payment card data.

We sought to overcome these challenges. To do so, we developed techniques for locating sites hosting web skimming code, for exposing payment credentials, and for monitoring those credentials (Section III). We identified a large set of sites running the Magento e-commerce platform. Some unpatched versions of this platform are vulnerable to web skimming [21]. Via a series of filtering steps, we narrowed this large initial set to 50 sites hosting suspected payment card skimming code. For each site, we attempted to purchase items using a unique payment card. Over our eleven-month study, we watched the payment cards for signs of abuse.

For 15 of the 50 payment cards, we observed at least one attempted unauthorized use of the card. Section IV details our observations. Besides one case of misuse 16 days after exposure of the card, all observed misuse occurred 50 days or more after exposure. Attackers tried to use the 15 cards at least 45 times, with attempted payments ranging from $0.10 to $122.44 and totaling $1,342.91. Given limitations of the study, these findings are a lower bound: even for sites without resulting observed misuse, we often found additional evidence of compromise in code or web traffic.

To place our observations in context, we compare them to a separate study in which we exposed payment card data directly on an online paste site (Section V). The comparison suggests that web skimming results in far slower misuse of credentials following exposure.

Our results indicate that evidence of abuse might not appear for an extended period following a web skimming incident. Furthermore, the impact of web skimming might be larger than initial signs suggest. Section VI explores these findings and mitigation strategies. We also discuss our disclosure process and its effect.

This report was prepared by staff of the Federal Trade Commission's Office of Technology Research and Investigation and does not necessarily reflect the views of the Commission or any individual Commissioner.

II. Related Work

Our analysis builds on prior work in three primary areas: (1) limited related work exploring web skimming, (2) research examining security threats by intentionally exposing credentials, and (3) work considering the economics of online crime.

Web Skimming. By the year 2000, vulnerabilities in shopping cart software were a known vector for thieves to obtain payment card credentials [59], but modern web skimmers have emerged only in recent years. In 2015, researchers at Sanguine Security reported discovering evidence of then-novel JavaScript-based web skimming code on 3,500 retail websites, tracing the first evidence of the infections to earlier that year [49]. Since that time, numerous others have measured the magnitude of particular web skimming campaigns in terms of the number of sites affected [33], [50]. In some cases, prior work infers the lifetime of infections on websites, using this to estimate the impact of a campaign [32], [45].

Prior work also documents the underlying technical details of particular campaigns [29], [53], [57]. Klijnsma et al. [34] categorize the increasingly diverse variety of web skimming attacks, and they observe that some sophisticated groups monetize stolen credentials on underground forums. Bower [7] proposes browser-based tools for detecting and preventing web skimming, inspired in part by tools to prevent cryptomining. Our work builds on this prior work, providing insights into how thieves misuse payment credentials from web skimming.

Exposing Credentials. To analyze payment credential misuse, we exposed and monitored credentials. Prior work has studied various online threats—particularly in the areas of malware [4], [15] and phishing [9], [23], [24], [39], [43]—by exposing credentials. For example, Bursztein et al. [9], Onaolapo et al. [38], and researchers at Bitglass [6] monitored misuse of online account credentials that they exposed through channels including phishing sites, underground forums, and paste sites. Akiyama et al. [4] exposed credentials to information-harvesting malware, finding that attackers attempted to access 13.2% of the leaked information within 24 hours.

Economics of Online Crime. Researchers have previously considered the economics of online crime, including through analysis of underground forums where credentials are bought and sold. Various work examines the products and prices in these marketplaces [14], [25], [26], [36], [46], [54], [56]. Haslebacher et al. [25] suggest that sellers offer credit card numbers for an average of $10 on some of these forums (as of 2017). Motoyama et al. [36] identify prolific sellers, finding that these sellers are responsible for a disproportionately large volume of sales. Other work evaluates the social structures in these underground marketplaces [3], [20], [60].

Past work also explores the impact of online crime on businesses [2], [11], [47]. Like us, Graves et al. [22] consider the impact of payment credential theft, but they focus on different factors, such as the cost of reissuing cards. In some cases, prior work on the economics of online crime has explored the threat by posing as a victim [31], [35], [40]. For example, Levchenko [35] et al. purchased products advertised through spam, identifying bottlenecks in the attackers' affiliate programs. Kanich et al. [31] used a similar approach to infer the earnings of spammers.

III. Approach

The goal of this study was to explore the misuse of payment data lost in a web skimming attack. To accomplish this, we created payment card data and exposed it through sites hosting suspected web skimming code. We sought websites using the Magento e-commerce platform, since past web skimming campaigns have targeted this platform [21]. In November 2020, Magento was the third most installed e-commerce platform on the web's top million sites, and more than 200,000 live sites used Magento throughout the study period [8].

We analyzed Magento-running websites for evidence of skimming scripts. If we found evidence, we tried to make purchases just like any other shopper, providing actual payment credentials. We attempted purchases that exceeded charge limits on those credentials. As a result, the attempt failed, but we exposed our credentials to attackers. We then monitored the payment credentials for unauthorized purchase attempts.

If we detected web skimming scripts on a website, we engaged in a responsible disclosure process to notify the operators of the site. Section VI-B discusses this process.

A. Creating Credentials and Payment Data

We created 50 synthetic consumer profiles, with each "consumer" having a name, email account, physical address, phone number, and payment card information. Our goal was to create potential victims that look plausibly realistic to an attacker.

We chose names based on common first and last names in recent US Census data. For each synthetic consumer, we created a Gmail, Hotmail, Outlook, or ProtonMail email account, choosing an email address that aligned with the consumer's name (e.g., "janedoe12@example.com"). Using a fake address generation service,[1] we created fake-but-realistic physical addresses that are geographically distributed across the US. Phone numbers were VoIP numbers with area codes roughly matching each profile's assigned physical address.

In creating payment card data, we sought to balance a number of objectives. Beyond our practical financial constraints, we aimed to prevent or severely limit any monetary benefit to malicious parties from our experiment. We wanted charge attempts using the payment data to fail immediately if possible to avoid imposing costs on others. At the same time, we wanted to observe actual misuse of payment data. Generation of realistic payment card data proved to be challenging. Ultimately, we assigned each consumer profile a card number from one of two sources:

- *Unfunded Online Payment Accounts ("Virtual Cards").* We identified a provider of online payment accounts that came with valid Visa credit card numbers. The provider offers an online interface, allowing customers to see

[1] https://www.fakeaddressgenerator.com/usa_address_generator

js/mage
mage.cookies.path
skin/frontend
static/_requirejs
text/magentoinit

Fig. 1. Code search terms for Magento sites.

charges to the card. That interface also shows failed charge attempts. We obtained 18 of these virtual cards but did not fund them. As a result, any charge to the card would fail, but the interface would allow us to see a charge attempt. Note that a failed initial charge may have affected an attacker's future use of the payment data.

- *Gift Cards.* As we generated payment data, the payment provider that we used for the previous option stopped offering new accounts matching our needs. Although existing accounts continued to work, this forced us to consider other options for the remaining 32 profiles. To overcome this issue, we obtained a $20 Visa gift card for each of those profiles. Using the gift card provider's online interface, we could observe attempted transactions up to the remaining balance on these cards.[2] With some exceptions, the interface generally did not list attempted transactions above the remaining balance, and we never observed transactions above the initial $20 balance. As a result of these constraints, we could have missed noteworthy charge attempts, such as a failed initial charge that caused an attacker to give up on a card.

Our assignment of payment source to consumer profile was random. For example, we avoided any relationship between the order of payment data exposure and the payment method. In Section IV, we separate some analysis by payment method, offering some insight into the possible impact of the $20 gift card's constraints on our observations.

The first digits of a payment card number are generally the bank identification number (BIN) [27], [30], also known as issuer identification number (IIN) [30]. A BIN identifies the card's issuer, which can reveal details about the type of card. Public websites allow anyone to learn these details. We checked the BIN values of our cards using several such websites.[3] For both payment methods, the sites identify the cards as "Visa Prepaid Debit" cards. If an attacker similarly checked BINs, this information may have influenced the attacker's behavior (though we note that a reasonable segment of the US population relies on prepaid debit cards in practice [44]).

B. Identifying Sites for Testing

Our first step in identifying compromised sites was finding a set of sites using Magento. Through a variety of steps, we created, filtered, and sorted this set to an ordered list of sites hosting suspected web skimming code. Each step is a heuristic, potentially excluding some compromised sites. While this may have introduced some biases, it allowed us to isolate a subset of real-world sites with evidence of web skimming.

Isolating Magento Installations. To identify Magento installations, we manually compiled a set of search terms based on our analysis of terms common in sites using Magento (see Figure 1). We searched for these terms on PublicWWW.com,[4] which is a search engine for webpage code. This search yielded a large initial set of relevant domains. PublicWWW.com also supplies approximate site rankings derived from Alexa.[5] We stored these site rankings as well.

We processed and filtered this preliminary set to increase our confidence. We resolved each domain to follow redirects or CNAME entries. We then scanned each candidate domain with the node.js version of Wappalyzer.[6] Wappalyzer attempts to identify tools underlying a particular website, returning a confidence score. We removed domains with anything less than a perfect Wappalyzer confidence rating for Magento use.

Filtering to Suspected Compromised Sites. Next, we sought to filter our candidate set to sites showing signs of compromise, even compromise unrelated to web skimming.

We downloaded the home page of each remaining candidate website along with dependencies, including third-party resources. To reduce the retrieval and storage of data likely unnecessary to detect web skimming, we limited the download to files with extensions `htm`, `html`, `ini`, `js`, `jsx`, `php`, `php3`, `php4`, `php5`, `php6`, `php7`, `phtml`, `sh`, and `txt`.

We scanned each site's downloaded files with Magento-Malware-Scanner,[7] an infection scanning tool for operators of Magento sites. The tool identifies signs of potential compromise broadly, not just web skimming. We narrowed our candidate set to sites that the scanner flagged as infected. While this intermediate step did not strictly identify cases of web skimming, it reduced the burden of the remaining steps.

Narrowing by Region and Sorting. We further filtered our candidate set to sites that have a `.com` top-level domain and a WHOIS registrant country code of "`us`." This helped reduce the possibility of issues with our US-based payment data. We then sorted the set using the PublicWWW.com-provided site rank. The remaining identification and testing steps were largely manual, and the resulting ordered candidate list allowed us to prioritize sites based on this rank.

Identifying Web Skimming. Our analysis process was lengthy, and administrators may remedy a web skimming attack before we enter credentials. Therefore, we sought confirmation of a site's compromise immediately prior to entering credentials on the site. In other words, we would test a candidate site for web skimming and, if identified, enter

[2] Note that an attacker aware of the issuer website could also have checked the balance using the card number.

[3] https://binlist.net/, https://quickbinlookup.com/, and https://www.exactbins.com/

[4] https://publicwww.com/

[5] https://www.alexa.com/

[6] https://www.wappalyzer.com/

[7] https://github.com/gwillem/magento-malware-scanner

credentials promptly (see Section III-D) before testing the next candidate site from our rank-ordered list.

To test a potentially compromised site, we used MageReport.com,[8] manually entering the site's URL. We confirmed that MageReport.com detected a "Credit Card Hijack" compromise (i.e., web skimming) of the site. If not, we skipped to the next candidate site in our list. Beyond requiring manual input, scans routinely took more than 30 seconds, which is why we sought to filter sites prior to this step.

C. Collecting and Refreshing Candidate Sites

We began to collect candidate sites in mid-March 2018. We initially identified 279,960 sites that passed our tests for Magento use, filtering from that starting point. We refined our processes over the following six weeks prior to entering credentials on sites.

When beginning the experiment, we were uncertain how many working sites would ultimately host suspected web skimming code. To compensate, we sought to expand our candidate set with a fresh search in late-April 2018, prior to credential entry. This yielded a new batch of 20,897 sites that passed our tests for Magento use.[9] We filtered, sorted, and combined the March and April data. Given the two separate searches, the same site could appear multiple times. We did not analyze a site more than once. The March data contained slightly older rankings than the April data. We assumed that relative ranks of these sites did not change substantially from March to April. In May 2018, we began final manual tests (and data entry) with candidate sites.

Refreshing Sites. The lifetime of a web skimming attack may be limited. Even if the infection goes undiscovered by site administrators, attackers may lose domains that host malicious scripts or receive exfiltrated data. As a consequence, we sought to identify sites in an efficient, automated manner. Unfortunately, final testing and data entry remained a manual process. To increase our chances of catching still-active attacks, we refreshed our list of candidate sites in mid-July 2018.

We repeated our full identification, filtering, and sorting process, beginning by flagging 19,832 sites for suspected Magento use. Following this refresh, we stopped using our old candidate list and switched to the new list. We started at the top of the new list, skipping any previously tested sites. This kept our focus on the top-ranked and freshest compromised sites. Consumers are more likely to encounter top sites in practice, and fresher attacks are more likely to be active.

D. Entering Credentials

We attempted to enter credentials on sites identified by the previous steps. Recall that (1) we tested sites in order based on the Alexa-derived, PublicWWW.com-provided ranking and (2) we tested for the presence of web skimming immediately prior to credential entry (see Section III-B). Credential entry began in May 2018. We successfully entered credentials on our fiftieth and final site by the end of July 2018.

Setup and Data Collection. All testing occurred in Firefox on Linux. We cleared browsing data between tests and modified the User Agent string to the equivalent Firefox version for Windows. We took a screenshot and saved the HTML source of each successfully tested site's home page and final shopping cart page. We configured mitmproxy[10] to capture all web traffic when loading the site. A mitmproxy misconfiguration issue resulted in a failure to collect data for 12 tested sites. This did not impact the behavior of malicious scripts or the transmission of data for any sites.

Testing. We navigated to each candidate site in order. Using our synthetic profile data, we manually attempted a purchase exceeding $20 from the site. Our payment sources were either unfunded virtual cards (18 card numbers) or $20 gift cards (32 card numbers), so none of our transactions succeeded. Nevertheless, we followed the checkout process and entered payment data like a typical shopper, exposing our credentials.[11]

On some tested sites, we were unable to complete payment data entry. Issues ranged from technical navigation errors to sites that sell only to businesses. In these cases, we skipped the site and moved to the next site on our list. To enter each of our 50 payment credentials on unique sites, we navigated to 104 candidate sites, ultimately skipping 54 sites. The PublicWWW.com-provided rankings of the 50 test sites ranged from 48,963 to 2,824,061 with a median of 877,695.

E. Monitoring

Following the entry of credentials, we monitored future payment activity via the online portals for the two payment methods. We attribute any additional transactions to web skimming.[12] Our monitoring began with entry of the first set of payment data in May 2018 and continued until March 2019.

IV. Observations and Results

We entered unique payment information on each of 50 sites hosting suspected web skimming code. For brevity, we refer to charge attempts as simply charges regardless of whether those attempts succeeded. We saw unauthorized charges on 15 payment cards (30%). Overall, 45 observed charges occurred on these 15 cards, totaling $1,342.91.

In this section, we explore the timing and number of observed charges (Section IV-A), the amounts of those charges (Section IV-B), and their recipients (Section IV-C). We split these results in two ways. First, we distinguish between initial charges and overall charges given the possibility that a failed initial charge influenced future behavior. Second, we distinguish between payment methods—virtual card or gift card—given the varying constraints on each (see Section III-A).

[8] https://www.magereport.com

[9] We had access to a temporary paid subscription to PublicWWW.com in March 2018. For the remaining months, we used the public version of PublicWWW.com, which returned fewer search results.

[10] https://mitmproxy.org/

[11] Anecdotally, we did not observe evidence of suspicious scripts that monitor whether a transaction succeeds.

[12] Circumstances strongly support this assumption, but it is an assumption. Other avenues for misuse exist, such as misuse by the retail site itself.

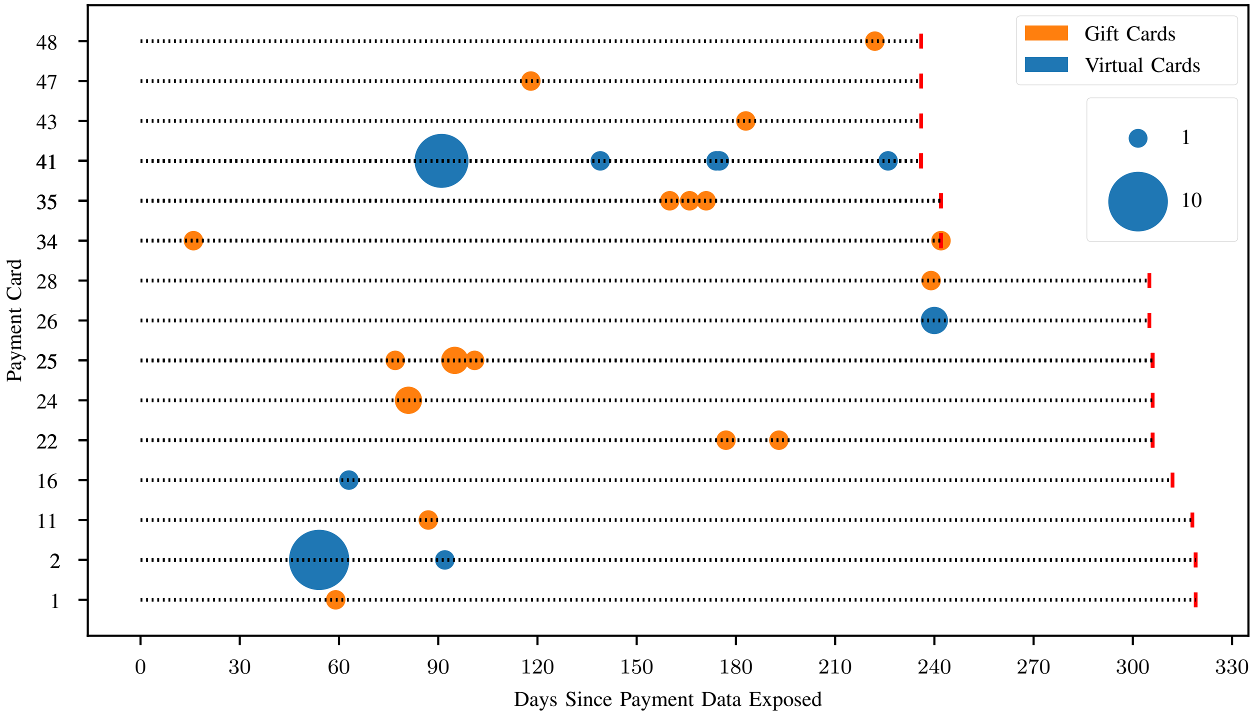

Fig. 2. Timing and number of observed charges. The size of a circle represents the number of transactions occurring on the day. Bars differ in length because we exposed some credentials later, resulting in a shorter observation period.

Beyond our analysis of charges, we also examine network traffic from tested sites (Section IV-D) and details of suspicious scripts (Section IV-E). This reveals further evidence of compromise even for many sites not associated with observed charges, suggesting our findings are a lower bound on misuse. Table VI in the Appendix compiles observations for sites on which we exposed cards.

A. Timing and Number of Charges

We observed charges on 4/18 virtual cards (22%) and 11/32 gift cards (34%). Of the 45 observed charges overall, 26 were to virtual cards, and 19 were to gift cards. This equates to 6.5 charges per misused virtual card and 1.7 charges per misused gift card. Figure 2 shows the timing and number of all charges, distinguishing between virtual cards and gift cards.

Initial Observed Charges. Table I provides details of initially observed charges, including the time from exposure until that charge. For virtual cards, observed charges began between 54 and 240 days after exposure, with a mean of 112 days and a median of 77 days. For gift cards, observed charges began between 16 and 239 days after exposure, with a mean of 129 days and a median of 118 days. Across both payment methods, the mean time to initial misuse was 124 days, and the median time was 91 days.

All Observed Charges. For virtual cards, observed charges occurred from 54 to 240 days after exposure, with a mean of 101 days and a median of 91 days. For gift cards, observed charges occurred from 16 to 242 days after exposure, with a mean of 135 days and a median of 118 days. Figure 3 provides the cumulative distribution of observed charges by date across both payment methods. The overall mean and median were 115 days and 91 days respectively. Table VII in the Appendix provides details of all observed charges.

TABLE I
TIME UNTIL INITIALLY OBSERVED MISUSE OF EACH CARD, AMOUNT OF THE CHARGE, AND PAYMENT DESTINATION CATEGORY.

Type	Days	Amount	Category
Gift Card	16	$1.07	Music
Gift Card	59	$15.93	Hosting
Gift Card	77	$10.00	Telecom
Gift Card	81	$10.00	Hosting
Gift Card	87	$19.99	Gaming
Gift Card	118	$20.00	Retail
Gift Card	160	$20.00	Charity
Gift Card	177	$16.37	Telecom
Gift Card	183	$20.00	Retail
Gift Card	222	$15.08	Transportation
Gift Card	239	$16.11	Telecom
Virtual Card	54	$1.17	Gaming
Virtual Card	63	$122.24	Retail
Virtual Card	91	$100.00	Gaming
Virtual Card	240	$0.10	Food Delivery

The timing of observed charges varied, but several patterns emerged. Only once did an observed charge occur fewer than 50 days after exposure. For three cards, the first observed misuse occurred after more than 200 days. Misuse tended

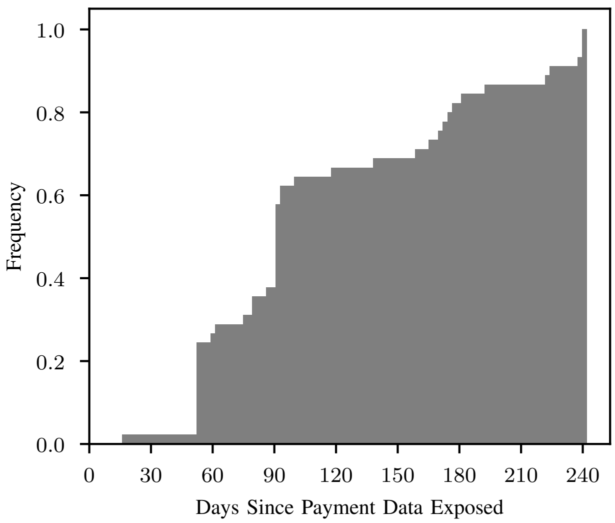

Fig. 3. Cumulative days until charge for all observed charges.

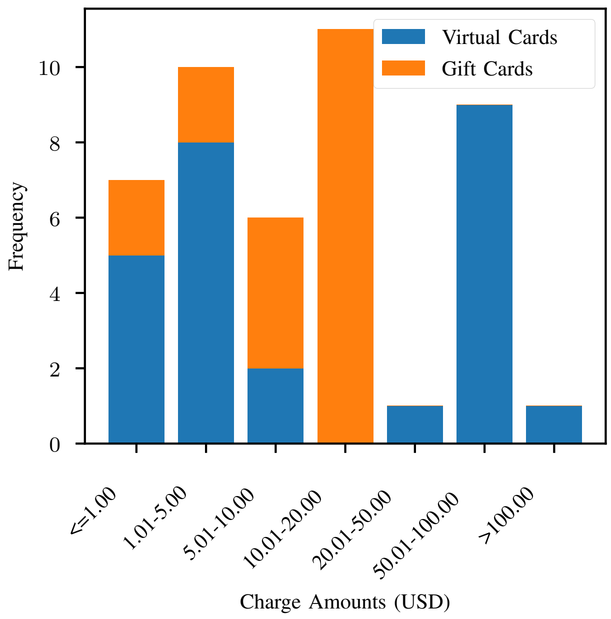

Fig. 4. Observed unauthorized charge amounts. Recall our limitations on observed charges, including a $20 limit for gift cards (see Section III-A).

to be clustered—all observed charges fell within a 40-day span for 13/15 cards—but 226 days separated the first and second observed misuse of one card. These results suggest that thieves do not race to use credentials, particularly initially. Web skimming may not result in fraudulent charges for many months, so its impact might appear minimal at first.

One interesting pattern occurred for virtual cards. In two cases, large bursts of repeated charges followed the initial observed charges to these cards. In one case, ten charges of either $1.17 or $1.00 occurred on the same day for the same virtual card. In the other case, eight consecutive charges of $100.00 occurred for the same virtual card. In both cases, the destination was the same for all charges in the burst of activity. We did not see similar initial bursts with gift cards.

This lack of initial bursts for gift cards may hint at how attackers respond to a failed charge attempt. We could not have seen repeated $100.00 charges to gift cards—this exceeds the $20 initial card balance—but we could have seen initial bursts of charges around $1.00. We did not. A key difference is that an initial charge under $20 would succeed for gift cards and fail for virtual cards, suggesting that the failure drove repeated charge attempts. The cause of this persistence is unclear. Attackers might assume failures stem from transient issues, or they might simply be frustrated.

B. Charge Amounts

Differences between observed charge amounts for virtual cards and gift cards are relatively large. This almost certainly stems from our inability generally to see charges above the remaining balance of gift cards.

Initial Observed Charges. For virtual cards, initial observed charges varied from $0.10 to $122.24, with a mean of $55.88 and a median of $50.59. For gift cards, initial observed charges varied from $1.07 to $20.00, with a mean of $14.96 and a median of $16.11. Given the differing constraints between payment methods, we do not present a combined mean or median here. Table I includes the initial charge amount for each payment card with observed misuse.

All Observed Charges. Figure 4 shows the distribution of observed charge amounts broken down by payment method. These charges varied from $0.10 to $122.24 for virtual cards (mean: $42.21; median: $3.94) and from $1.00 to $20.00 for gift cards (mean: $12.92; median: $15.93). As with initial observed charges, we omit combined mean and median here.

While the constraints of gift cards may have skewed our results, thieves attempted relatively large initial charges on both gift cards and virtual cards. For virtual cards, the median observed charge amount for all charges is much lower than for initial charges. A single large burst of $1.00-$1.17 charges may have influenced this: we did not observe a decreasing pattern more generally for virtual cards (see Table VII).

C. Payment Destination

Payment destinations varied from food delivery services to charities. We may not know what an attacker bought or why, but we can sometimes make educated guesses. For example:

- Three charges were to domain hosting services. These could have been new domains and infrastructure to host skimming scripts or receive siphoned data.
- Three charges were to charities. Criminals purportedly use charities to test payment cards [18], [28]. While disagreement exists over the prevalence of this practice [42], our results suggest that charities are sometimes recipients of attempted charges regardless of the purpose.

We manually categorized payment destinations. Table VII includes both the payment destination and category for all observed charges.

Initial Observed Charges. Table I includes the destination category of each initial observed charge. The Gaming, Telecom, and Retail categories each received three observed charges. Retail received the highest total charge amount ($162.24). For virtual cards, Gaming was the most popular category (2 charges), but a single large Retail charge gave it the highest overall monetary amount ($122.24). For gift cards, the Telecom category was the most popular by both number (3 charges) and monetary amount ($42.48).

All Observed Charges. Figure 5 shows the overall distribution of observed charges, including destination category, number of purchases, and total charge amount. Gaming had the largest number of observed charges (19 charges) and monetary amount ($831.35) by far. Besides a single $19.99 charge to a gift card, all but one of the Gaming charges occurred in two single-day bursts of activity, each on a single virtual card. As a result, Gaming was the most popular category for virtual cards by number of observed charges (18 charges) and monetary amount ($811.36). Retail was the second most popular overall category by both observed charges (7 charges) and monetary amount ($281.41). This category was the most popular for gift cards (5 charges totaling $62.00) and the second most popular for virtual cards (2 charges totaling $219.41).

We previously discussed bursts of charges. Recall that we observed two such bursts for virtual cards: one with ten charges of $1.00-$1.17 and one with eight charges of $100.00. In the former case, all eight attempts were to “blizzard entertainment;” in the latter case, all were to “Steam.” In both cases, these charges fell in the Gaming category, which may have skewed our results to some degree.

D. Analysis of Site Traffic

As we entered credentials on sites, we used mitmproxy to capture web traffic. We analyzed these traffic captures for evidence of payment data theft. In particular, we looked for transmission of payment data to third parties. In spite of a mitmproxy misconfiguration issue (see Section III-D), we have traffic data for 38 of 50 sites where we attempted purchases. Recall that we observed unauthorized charges on 15 cards, which we exposed via 15 sites. For 10 of these 15 sites, we have traffic captures. We separately analyze captures for these 10 sites and for the 28 other sites.

Observed Misuse. After capturing traffic, we filtered those captures to remove communication with the site’s own domain. We then searched request data for our payment credentials, including those credentials encoded or hashed using common methods. For all ten of these sites, we identified transmission of the credentials to third-party domains. Table II lists these third-party domains, which may include legitimate payment processors. Note that a site may transmit credentials to suspicious third parties in an unknown format. For example, one site transmitted payment credentials to `paypal.com`, but we also observed transmission of unknown data to `upartman.com`. We were unable to determine the contents of the unknown data. Superficially, many of the domains in Table II resemble ones that the Magento-Malware-Scanner tool (see Section III-B) associates with Magento malware.

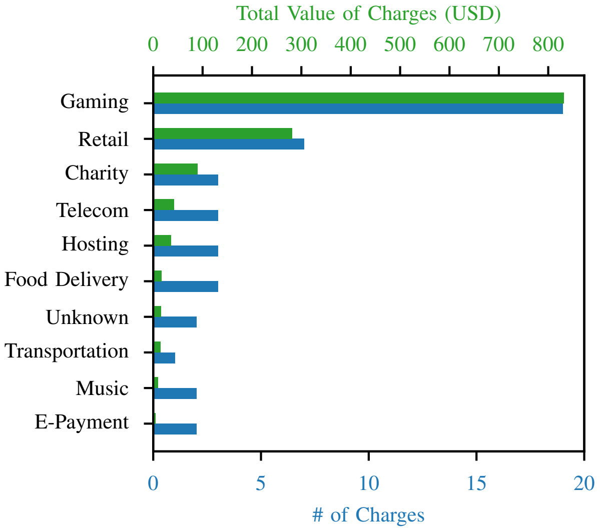

Fig. 5. Destination of observed charges by amount charged (top/green bars) and number of charges (bottom/blue bars).

TABLE II
THIRD-PARTY DOMAINS OBSERVED RECEIVING PAYMENT DATA FOR SITES ASSOCIATED WITH UNAUTHORIZED OBSERVED CHARGES.

# Sites	Domain
2	`cdnapis.com`
2	`magentocore.net`
1	`g-analytics.com`
1	`logistic.tw`
1	`nearart.com`
1	`paypal.com`
1	`simcounter.com`
1	`trafficanalyzer.biz`

In six cases, the transmitted payment data was Base64-encoded in a POST request. In one case, the data was Base64-encoded and AES-encrypted in a POST request. For the final three cases, the data was URL-encoded in a GET request.

For one site, we saw both transmission of an empty data array to the known-suspicious domain `magentocore.net` [12] and transmission of payment data to `cdnapis.com`. We do not know the cause of this particular behavior, but attackers have previously sabotaged each other when targeting the same site [37].

Other Sites. We also collected traffic for 28 sites with no observed unauthorized charges. Like before, we filtered first-party traffic. Of these 28 sites, we identified transmission of payment credentials for 9 sites, each to a single third-party domain (see Table III). In a single additional case,

TABLE III
THIRD-PARTY DOMAINS OBSERVED RECEIVING PAYMENT DATA FOR SITES NOT ASSOCIATED WITH UNAUTHORIZED OBSERVED CHARGES.

# Sites	Domain
2	magentocore.net
1	authorize.net
1	brewtees.com
1	cloud-update.top
1	dnsden.biz
1	heartlandportico.com
1	paypal.com
1	trafficanalyzer.biz

we believe that a suspicious script failed to recognize and send payment credentials simply because the site put spaces between numbers. The transmitted payment credentials were Base64-encoded in one case, custom-encoded in one case, and URL-encoded in the remaining cases.

For each tested site, Table VI lists third parties that we identified receiving payment credentials. Based on our traffic analysis and our analysis of suspicious scripts themselves (see Section IV-E), all identified or suspected transmission of payment data to unexpected third parties occurred via HTTPS. This may reduce the likelihood that network scanners flag the traffic [10].

E. Analysis of Suspicious Scripts

To obtain suspicious scripts, we began by extracting scripts from our mitmproxy traffic captures when available. To complement this data, recall that we saved home pages—including some dependencies—when filtering/testing sites. Particularly for sites where the mitmproxy capture failed, we relied on scripts from these saved files.

From this preliminary set of scripts, we identified suspicious scripts through a combination of manual and automated analysis. We filtered our traffic captures (see Section IV-D) to requests containing data. After compiling the URLs of these requests, we removed first parties, known payment processors, and other known-benign third parties. We flagged scripts containing the remaining URLs. In addition, we used the Magento-Malware-Scanner (see Section III-B) to flag scripts. We manually reviewed flagged scripts and associated traffic for evidence of payment data siphoning. Given such evidence, we treated these flagged scripts as suspicious.

For 49 of 50 sites, we identified suspicious scripts.[13] Our entry of payment data on 15 of these 49 sites resulted in observed misuse. As we discuss below, the suspicious scripts on many of these 49 sites look similar to each other.

We identified 83 URLs in the suspicious scripts. Many load files with names resembling common web services like Google Analytics. For example, we saw 11 occurrences of files named "ga.js" or "ga2.js." We also found five cases in which the site might superficially appear to be loading innocuous jQuery. The domain names in these URLs also sometimes resemble web service or CDN names, such as cdnanalytics.net (9 cases) or g-analytics.com (1 case). Presumably, these choices are to appear innocuous on casual inspection of files and traffic.

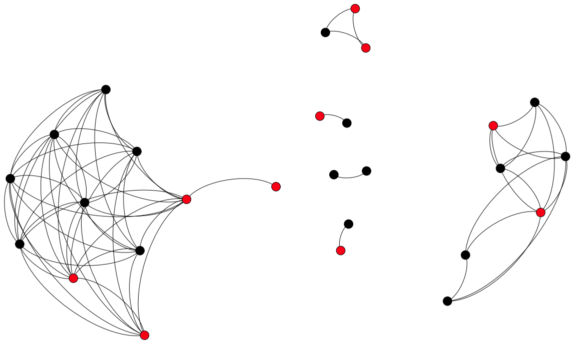

Fig. 6. Clustering of sites by associated suspicious scripts. Red nodes indicate sites associated with observed unauthorized charges.

Whether suspicious scripts were embedded in a web page (17 sites), loaded from the site (3 sites), or hosted elsewhere (29 sites), the scripts on tested sites often superficially appeared similar. To analyze this, we clustered sites using Moss.[14] Moss measures code similarity and is often used for plagiarism detection. It generates a percentage value for similarity. Many suspicious scripts were lightly obfuscated. Before using Moss, we applied basic deobfuscation tools (de4js and JS NICE[15]) and techniques to reduce obfuscation.

For each of the 49 sites, we calculated the average similarity between suspicious scripts on that site and all 48 other sites. Using these values as a weighting factor, we created a force-directed graph of the sites. Figure 6 shows the results, containing the 27 sites that were similar to at least one other site. The ten red nodes are for sites associated with misuse. At minimum, the clusters hint at common code—prior work suggests a market for web skimming code [29]—and they may indicate common attackers.

Scripts on these sites were similar regardless of whether we observed misuse associated with a site. This suggests that attackers could have conducted web skimming on many other identified sites. Although limitations may have prevented us from observing some misuse, some attacks may have no longer been active when we tested sites, or patient thieves might not have misused our credentials during the study period.

As expected, suspicious scripts monitor events to detect a purchase, or they monitor elements of the page that correspond to typical payment information form fields (e.g., a text input with identifier "cc_num"). For one site that yielded observed misuse, a script performs a regular expression match on a form field. If a user enters a payment number matching a particular pattern, the script transmits the number.

[13] The remaining case included a heavily obfuscated script. We were unable to determine its behavior with sufficient confidence to label it suspicious.

[14] https://theory.stanford.edu/~aiken/moss/

[15] https://lelinhtinh.github.io/de4js/ and http://www.jsnice.org/

TABLE IV
DOMAINS IN SUSPICIOUS SCRIPTS ON SITES ASSOCIATED WITH UNAUTHORIZED OBSERVED CHARGES.

# Sites	URL Domain
4	magentocore.net
2	[First Party Domain]
2	cdnanalytics.net
2	cdnapis.com
1	ap76rmx3.accountant
1	brewtees.com
1	cdnstorecontent.com
1	cloud-update.top
1	cloudservice.tw
1	constantincreations.com
1	g-analytics.com
1	go.onclasrv.com
1	logistic.tw
1	mayning.online
1	melissatgmt.us
1	neartart.com
1	simcounter.com
1	trafficanalyzer.biz
1	upartman.com

For suspicious scripts on sites associated with observed charge activity, Table IV contains the list of domains (suspicious or not) that we observed in those scripts.

V. COMPARATIVE STUDY

In 2017, we conducted an unpublished preliminary study, aiming to understand how identity thieves misuse credentials like payment data. That study helped us refine the present study. This preliminary study explored the misuse of credentials that we posted to an online paste site. We sought to mimic a small data breach that resulted in the public disclosure of customers' personal information. To place our web skimming results in context, we describe this prior study and compare the results.

Credentials. As part of our prior study, we created 33 consumer profiles that included all details from Section III-A.[16] We followed the same process for creating those details in both studies. For payment method, 23 of the profiles included virtual card numbers, and 10 contained gift card numbers. These payment methods have the same constraints that we discuss in Section III-A.

Exposing Data. After creating this data, we "leaked" it via Pastebin.com,[17] a public paste site. We posted the credentials and began to watch for misuse in late-April 2017.

Our plan was to monitor exposed credentials for two weeks. Although we observed payment card misuse three days after exposure, evidence of misuse was unexpectedly limited for the first week. Because a goal was to learn more about the nature of misuse, we decided to post the same data a second time. Prior to this second posting, we examined the code behind the "dumpmon" Twitter handle.[18] That code[19] monitors paste sites for credential dumps, which the Twitter account announces. We reformatted our data to increase the likelihood of its identification by credential dump detection code. We exposed the same (but reformatted) data a second time on the eighth day of the study. We continued to track misuse for the remaining week of the study.

TABLE V
COMPARATIVE STUDY STATISTICS FOR TIMING AND AMOUNT OF OBSERVED CHARGES.

			Gift Cards	Virtual Cards	Both
Initial Observed Charges	Days	Min	7	3	3
		Max	8	8	8
		Mean	7.8	6.5	6.8
		Median	8	7	7
	Amount	Min	$1.99	$0.38	–
		Max	$19.74	$97.90	–
		Mean	$14.15	$13.64	–
		Median	$16.45	$2.47	–
All Observed Charges	Days	Min	7	3	3
		Max	13	13	13
		Mean	8.9	8.6	8.6
		Median	8	8	8
	Amount	Min	$0.82	$0.01	–
		Max	$19.74	$2,697.75	–
		Mean	$10.36	$32.68	–
		Median	$11.51	$2.38	–

Details of Charges. Over the two-week monitoring period, we observed 402 charges totaling $12,736.33. Of these charges, 384 occurred on 22/23 virtual cards (96%) and 18 occurred on 8/10 gift cards (80%).[20] This equates to 17.5 charges per misused virtual card and 2.3 per misused gift card. Figure 7 shows the timing and number of all charges.

Table V contains statistics for the timing and amount of observed charges. We suspect that most misuse stemmed from our second credential posting, but we cannot establish this with certainty. Therefore, we conservatively use the date of initial exposure when calculating time until misuse. As with our primary study, we do not present combined charge amount statistics given the different constraints on the payment methods. Figure 8 shows the distribution of observed charge amounts.

Observed payment destinations were diverse, ranging from pizzerias to charities. For initial charges, the E-Payment category received the most observed charges (12 charges) and second-highest total charge amount ($120.12). Seven of these twelve charges were to gift cards, making it the top gift card category by both number and charge amount ($111.22). Six of these charges fell in a single cluster: each was to a different gift card, but all were to PayPal over a single minute, with amounts from $14.68 to $19.74. Over both card types, the Retail category received the second-most observed charges (8

[16] For this preliminary study, profiles included additional details, and we created additional profiles without payment card numbers. We exclude these additional profiles and details from this analysis.

[17] https://pastebin.com/

[18] https://twitter.com/dumpmon

[19] https://github.com/jordan-wright/dumpmon

[20] For the two remaining gift cards, we observed limited pending charges that later disappeared. Because we did not monitor pending charges in our primary study, we do not include those cases in our analysis here.

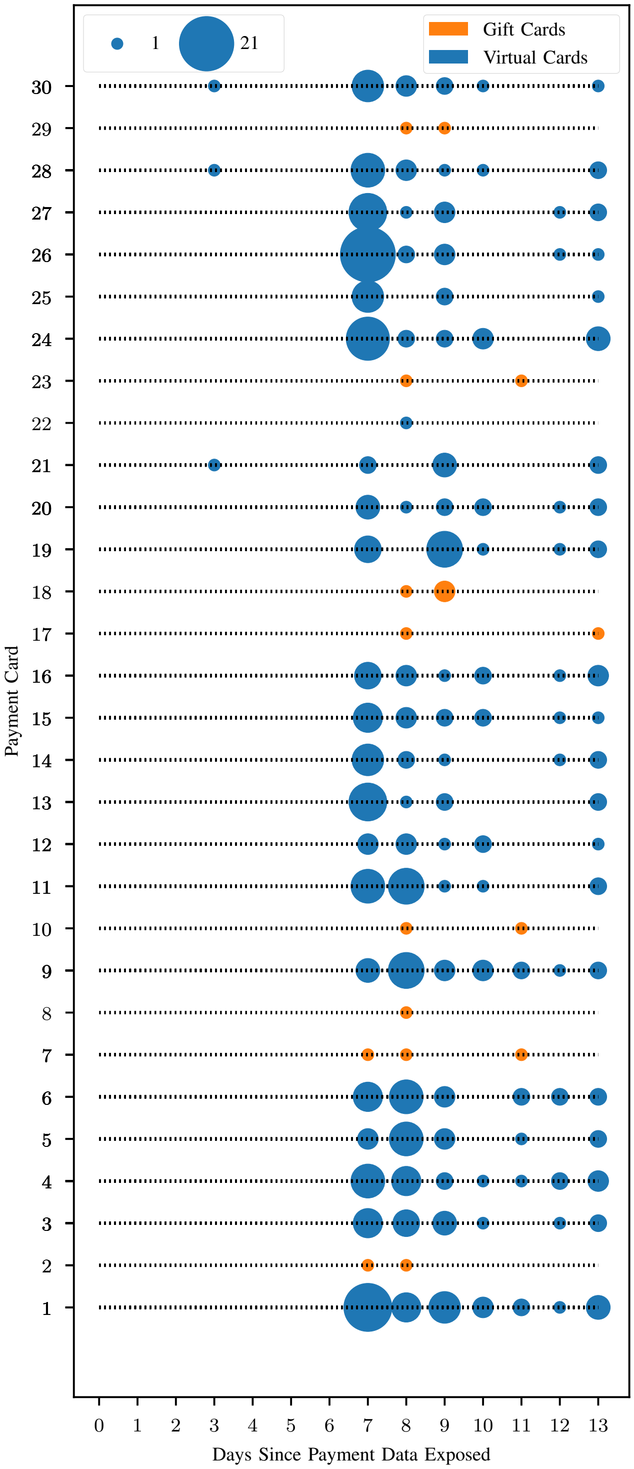

Fig. 7. Timing and number of observed charges on cards exposed in comparative study.

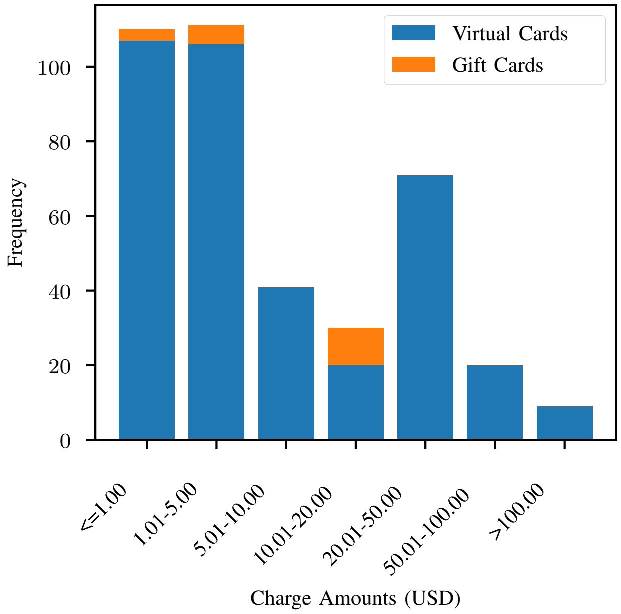

Fig. 8. Observed unauthorized charge amounts for comparative study.

charges) but the highest charge amount ($184.72). All Retail charges were to virtual cards, making this the top category by number and charge amount for these cards.

For all charges and card types, the Retail category had both the most observed charges (142 charges) and the highest total charge amount ($7,484.06). All of the 142 observed Retail charges were on virtual cards, making Retail the most popular category for these cards by all measures. Beyond having the two largest observed charges ($2,697.75 and $1,461.36), we saw several large clusters of Retail charges. For these clusters, we suspect that thieves tried to purchase an item with one card after another. The charges were for the same amount and vendor, and each charge occurred within two minutes of the prior one. For example, we observed 25 charges of $1.08 over 11 minutes, 21 charges of $24.96 over 17 minutes, and 21 charges of $1.09 over 11 minutes. For gift cards, the E-Payment category had the most charges (11 charges) and highest charge amount ($173.71) overall. We observed clustering here too: 8 of 11 charges were to PayPal over the same minute, with charge amounts from $11.01 to $19.74.

Comparison. Even with our conservative estimate of days until use, observed misuse occurred far earlier for a far higher proportion of cards in this comparative study. In the primary study, initial observed misuse typically took months. In this comparative study, such misuse happened within days. Observed charge amounts in the comparative study also tended to be smaller: the majority of attempted charges in the comparative study were under $5; the majority in the primary study were $10 or more.

Incentives may explain differences in the speed of observed misuse. Posting credentials on a paste site may trigger a gold rush among rational thieves. Each misuse of a payment card risks a fraud alert, so even sophisticated attackers might race to extract whatever value they can. This also may help explain the smaller charges in the comparative study, which may have been test charges or charges calibrated to avoid suspicion.

Nevertheless, these results do suggest that web skimming attackers are more willing to delay misuse. For web skimming, patience also might stymie attempts by financial institutions to tie misuse to a compromised site. Once that association occurs, the institution can alert the retailer and shut down other cards that consumers used on the site.

In the comparative experiment, we observed similar early "bursts" of charges for virtual cards (see Figure 7), but these bursts were less dramatic than in the primary study. This difference is likely due to the fact that we leaked multiple cards at the same time, doing so publicly. As a result, an attacker may have tried multiple cards as opposed to the same one repeatedly. The clusters of charges in our comparative experiment support this possibility. In addition, different attackers may have tested the same card on different days.

VI. Discussion

Our findings provide a first look into payment credential misuse resulting from web skimming. The observed behavior suggests that those behind web skimming can be patient and methodical. Stereotypically, we associate these attributes with highly sophisticated adversaries. While the attackers behind web skimming may be sophisticated, the contrasting patterns between our primary and comparative study hint that circumstance may be a determining factor here. If circumstances do not require speed, an attacker may not exploit credentials quickly following the theft.

We do not know precisely what drives this delay, but one benefit is additional time for sites and others to notice web skimming, contact financial institutions, and ensure that exposed payment cards are canceled. Unfortunately, this delay also may make it harder for financial institutions, investigators, and consumers to determine the source of credential misuse. In addition, minimal immediate misuse may lead us to underestimate the impact of web skimming.

Delayed misuse of payment credentials creates other risks for consumers as well. Web skimmers sometimes steal personal information beyond payment card numbers. Theft of these additional credentials may allow the attacker to perform other forms of identity theft, such as applying for credit under the consumer's name. These forms of identity theft may be harder for consumers to discover and address than noticing a fraudulent charge and requesting a new credit card. As a result, later discovery of identity theft may result in greater harm to consumers.

In this section, we discuss web skimming mitigation strategies (Section VI-A), our disclosure process (Section VI-B), and additional limitations of our study (Section VI-C).

A. Mitigation

Website operators, consumers, and financial institutions can take steps to protect against web skimming and its impact.

Website Operators. For online retailers large and small, web skimming poses a serious threat to their customers, their reputation, and their profits [5]. Thankfully, these retailers can take steps to prevent and detect web skimming. For the sites that we analyzed, we suspect that attackers exploited out-of-date software to install web skimming scripts. Installing updates and applying security patches promptly can help prevent or mitigate web skimming and other threats. Website operators also can use tools like those we used to scan their sites periodically for threats or compromise (see Section III-B).

Recall that we saw bursts and clusters of charge attempts. Some of these may have been attempts to verify that the cards were active [42]. To prevent misuse of their sites for these and other malicious purposes, online retailers might wish to review their anti-fraud approaches like rate limiting.

Consumers. While a non-technical consumer can likely do little today to detect and avoid sites compromised by web skimming, this may change. Existing research has explored ways that consumers' browsers can help protect against web skimming [7], potentially complementing existing browser-based URL blocklists. Future work may yield new tools and features that help consumers defend themselves against this threat.

Financial Institutions. Given the impact of fraud on financial institutions, these businesses take considerable steps to identify and prevent misuse of payment credentials [48]. For banks that issue payment cards, the patterns that we observed may give these institutions more factors to consider. For example, suppose that a consumer spends the full balance of a gift card at the retail site `https://example.com/`. Several months later, the bank that issued the gift card notices a burst of failed charge attempts.[21] The bank could immediately investigate `https://example.com/` and possibly freeze other cards that consumers used on the site, potentially reducing losses.

B. Disclosure

Recall that we successfully entered payment data on 50 sites. For each of those sites, we contacted the site's operators in September 2018, alerting them that we had detected a suspected vulnerability on their site. We used both electronic communication and postal mail.

- *Electronic communication.* In all cases, we looked for contact email addresses on the site itself. We also obtained email addresses from the domain's WHOIS information. For sites that used domain registration proxy services, we entered the text of our email notification into the proxy service's contact box in lieu of an email.
- *Postal mail.* If a physical address for a site's operator was available via the site or WHOIS data, we also sent the operator a letter via overnight delivery or certified mail.

In all cases, the emails and letters informed the operators that we had detected a vulnerability on their site. For the

[21] The bank would not have the same constraints that prevented us from observing some charge attempts.

15 sites where our payment attempts resulted in observed payment card misuse, we also alerted site operators that we had evidence of exploitation. With the physical letters, we included a printed copy of data breach guidance from the US Federal Trade Commission [17]. Electronic communication included a link to the same guidance.

The emails and letters provided site operators with the name and address of a researcher they could contact for more information. Twelve responded. Nine recipients indicated that they had fixed the issue or forwarded the notification to the appropriate team for remediation. Two said they were in the process of migrating away from Magento. The remaining recipient left a voicemail, but we were unable to reach them. In an additional case, we received no response, but after our notification, a note that "ordering capabilities were under maintenance and have now been fixed" appeared on the site.

Recheck of Compromised Sites. In February 2020, we used MageReport.com to rescan each of the 50 sites. Recall that we used MageReport.com to detect evidence of web skimming when identifying sites to test (see Section III-B). Unfortunately, our disclosures did not always permanently result in secure practices. The rescan identified evidence of web skimming on only a single site, but MageReport.com found evidence of other vulnerabilities on 22 additional sites. MageReport.com did not flag the 9 remaining Magento sites (16 other sites appeared to have migrated from Magento, and 2 sites were down).

While repeated compromises have been known to happen [51], we do not know whether some site operators failed to act, took inadequate remediation steps, or neglected to maintain good security practices. Prior work has explored the impact of mass vulnerability disclosures (e.g., [55]). Future work will ideally continue to improve the efficacy of mass disclosures.

C. Additional Limitations

We generally raised limitations in context (e.g., limitations of payment credentials), but several others are worth noting. Our insights are mostly limited to what we can infer from observed charges. We do not know if and when attackers checked siphoned data, especially if we saw no charges. In addition, observed charges do not tell us whether thieves made charges themselves or sold card numbers. We also cannot see steps attackers took to check credentials. In spite of that, observed charges provide a useful perspective.

We did not monitor all credentials for the same period of time. For the primary study, the monitoring period concluded on a given day, but we exposed credentials over months. For the comparative study, we monitored credentials for two weeks. Nevertheless, the monitoring periods were sufficient to yield meaningful observations.

Challenges obtaining payment cards and locating compromised sites resulted in a limited number of sites and observed charges for our analysis. While these numbers are limited, the study provides novel insights into an understudied threat.

Finally, we observed misuse of 15 payment cards, but we selected all 50 sites based on evidence of compromise, uncovering additional evidence for many (see Table VI). Where we did not observe misuse, we cannot discern between cases in which a previously successful attack was no longer active, thieves had not yet used credentials, study limitations prevented us from seeing misuse, etc. Future work could build on this initial exploration, further illuminating the life cycle of web skimming attacks and resulting credential misuse.

VII. Conclusions and Future Work

We identified web sites that attackers apparently compromised to host web skimming scripts. By entering real payment credentials on those sites, we exposed those credentials to attackers and could monitor resulting misuse. This yielded novel insights into how and when thieves make use of stolen payment credentials. Among other findings, our results suggest a moderately long delay between exposure and misuse.

Future work could further illuminate web skimming and the surrounding ecosystem. Future studies—possibly in collaboration with financial institutions—could overcome limitations on our ability to obtain and monitor credentials. Studies could also examine when attackers misuse credentials themselves or sell them to others, as has reportedly occurred following past web skimming campaigns [34]. Deeper examination of script similarities, URLs, and other details could yield additional insights into various parties and their relationships. More generally, additional research into the funding, economics, and logistics of web skimming could be valuable. Future work could also explore whether and how attackers use contact information or other details that web skimming can expose.

Remediation is another area that would benefit from additional study. Insights into the practices of compromised site operators could suggest ways of supporting improved operator security practices, such as refinements to the software update process. Research could also expose ways to help consumers avoid or minimize their harm from web skimming.

Our comparative study suggests that circumstances influence attacker behavior, such as the speed with which attackers misuse credentials. Examining attacker behavior for a wider range of factors and threats may yield novel insights and taxonomies for both attackers and threats.

This paper builds on considerable prior work exploring the possible harms from security threats in practice. Future work should continue to offer engineers, consumers, policymakers, and others information that helps them understand and address the true risks of attacks.

Acknowledgments

We thank Aaron Alva, Laura DeMartino, Hugh Huettner, and Min Hee Kim for feedback on drafts of this work. We also thank our anonymous reviewers for their helpful feedback.

References

[1] "Macy's becomes the latest victim of a Magecart skimming attack," *The Inquirer*, Nov. 19 2019 (accessed Jan. 23, 2020), https://www.theinquirer.net/inquirer/news/3084034/macys-magecart-data-breach.

[2] A. Acquisti, A. Friedman, and R. Telang, “Is there a cost to privacy breaches? An event study,” in *ICIS 2006 Proceedings*, 2006, p. 94, https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1215\&context=icis2006.

[3] S. Afroz, V. Garg, D. McCoy, and R. Greenstadt, “Honor among thieves: A common’s analysis of cybercrime economies,” *2013 APWG eCrime Researchers Summit*, pp. 1–11, 2013.

[4] M. Akiyama, T. Yagi, T. Yada, T. Mori, and Y. Kadobayashi, “Analyzing the ecosystem of malicious URL redirection through longitudinal observation from honeypots,” *Computers & Security*, vol. 69, pp. 155–173, Aug. 2017, http://www.sciencedirect.com/science/article/pii/S016740481730007X.

[5] BBC, “British Airways faces record £183m fine for data breach,” *BBC News*, Jul. 2019, https://www.bbc.com/news/business-48905907.

[6] Bitglass, “"Where’s your data?" experiment,” https://www.bitglass.com/blog/the-bitglass-wheres-your-data-experiment, Dec. 2015.

[7] T. Bower, “Identifying JavaScript skimmers on high-value websites,” Master’s thesis, Imperial College London, Jun. 2019.

[8] BuiltWith, “eCommerce usage distribution in the top 1 million sites,” (Accessed Nov. 5, 2020), https://trends.builtwith.com/shop/.

[9] E. Bursztein, B. Benko, D. Margolis, T. Pietraszek, A. Archer, A. Aquino, A. Pitsillidis, and S. Savage, “Handcrafted fraud and extortion: Manual account hijacking in the wild,” in *Proceedings of the 2014 Conference on Internet Measurement Conference*, ser. IMC ’14. Vancouver, BC, Canada: Association for Computing Machinery, Nov. 2014, pp. 347–358, https://doi.org/10.1145/2663716.2663749.

[10] K. Cabaj, L. Caviglione, W. Mazurczyk, S. Wendzel, A. Woodward, and S. Zander, “The new threats of information hiding: The road ahead,” *CoRR*, vol. abs/1801.00694, 2018. [Online]. Available: http://arxiv.org/abs/1801.00694

[11] K. Campbell, L. Gordon, M. Loeb, and L. Zhou, “The economic cost of publicly announced information security breaches: Empirical evidence from the stock market,” *Journal of Computer Security*, vol. 11, pp. 431–448, 07 2003.

[12] C. Cimpanu, “MagentoCore malware found on 7,339 Magento stores,” *BleepingComputer*, Aug. 30 2018 (accessed Feb. 25, 2020), https://www.bleepingcomputer.com/news/security/magentocore-malware-found-on-7-339-magento-stores/.

[13] L. Constantin, “Web payment card skimmers add anti-forensics capabilities,” *CSO*, Nov. 14 2019 (accessed Jan. 23, 2020), https://www.csoonline.com/article/3453940/web-payment-card-skimmers-add-anti-forensics-capabilities.html.

[14] G. Durrett, J. K. Kummerfeld, T. Berg-Kirkpatrick, R. S. Portnoff, S. Afroz, D. McCoy, K. Levchenko, and V. Paxson, “Identifying products in online cybercrime marketplaces: A dataset for fine-grained domain adaptation,” 2017.

[15] B. Farinholt, M. Rezaeirad, P. Pearce, H. Dharmdasani, H. Yin, S. L. Blond, D. McCoy, and K. Levchenko, “To catch a ratter: Monitoring the behavior of amateur DarkComet RAT operators in the wild,” in *2017 IEEE Symposium on Security and Privacy (SP)*, May 2017, pp. 770–787.

[16] Federal Bureau of Investigation, “Oregon FBI tech tuesday: Building a digital defense against e-skimming,” Oct. 22 2019 (accessed Jan. 23, 2020), https://www.fbi.gov/contact-us/field-offices/portland/news/press-releases/oregon-fbi-tech-tuesday-building-a-digital-defense-agaist-e-skimming.

[17] Federal Trade Commission, “Data breach response: A guide for business,” May 2019 (accessed Mar. 3, 2020), https://www.ftc.gov/system/files/documents/plain-language/pdf-0154_data-breach-response-guide-for-business-042519-508.pdf.

[18] Y. Gable, “Scammers make friends with charities,” http://www.symantec.com/connect/blogs/scammers-make-friends-charities, Jul. 2007.

[19] S. Gallagher, “British Airways site had credit card skimming code injected,” *Ars Technica*, Sep. 11 2018 (accessed Jan. 27, 2020), https://arstechnica.com/information-technology/2018/09/british-airways-site-had-credit-card-skimming-code-injected/.

[20] V. Garg, S. Afroz, R. Overdorf, and R. Greenstadt, “Computer-supported cooperative crime,” in *Financial Cryptography and Data Security*, R. Böhme and T. Okamoto, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 32–43.

[21] D. Goodin, “Brace yourselves: Exploit published for serious Magento bug allowing card skimming [updated],” *Ars Technica*, Mar. 29 2019 (accessed Jan. 23, 2020), https://arstechnica.com/information-technology/2019/03/severe-magento-bug-opens-300k-commerce-sites-to-card-skimming-attacks/.

[22] J. T. Graves, A. Acquisti, and N. Christin, “Should payment card issuers reissue cards in response to a data breach,” in *2014 Workshop on the Economics of Information Security*, 2014.

[23] R. Graves, “Honeytokens and honeypots for web ID and IH,” https://www.sans.org/reading-room/whitepapers/attacking/paper/35962, May 2015.

[24] X. Han, N. Kheir, and D. Balzarotti, “PhishEye: Live monitoring of sandboxed phishing kits,” in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS ’16. Vienna, Austria: Association for Computing Machinery, Oct. 2016, pp. 1402–1413, https://doi.org/10.1145/2976749.2978330.

[25] A. Haslebacher, J. Onaolapo, and G. Stringhini, “All your cards are belong to us: Understanding online carding forums,” in *2017 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2017, pp. 41–51.

[26] T. J. Holt and E. Lampke, “Exploring stolen data markets online: Products and market forces,” *Criminal Justice Studies*, vol. 23, no. 1, pp. 33–50, 2010. [Online]. Available: https://doi.org/10.1080/14786011003634415

[27] ISO/IEC, “ISO/IEC 7812-1:2017(en) identification cards — identification of issuers — part 1: Numbering system,” 2017 (accessed Feb. 1, 2020), https://www.iso.org/obp/ui/#iso:std:iso-iec:7812:-1:ed-5:v1:en.

[28] Jeremy Milk, “Why credit thieves love giving to charity,” https://blog.wepay.com/2016/01/08/why-credit-thieves-love-giving-to-charity-2/, Jan. 2016.

[29] R. Joven, “Inter: Skimmer for all,” *Fortinet*, Jun. 27 2019 (accessed Feb. 20, 2020), https://www.fortinet.com/blog/threat-research/inter-skimmer-for-all.html.

[30] J. Kagan, “Bank identification number (BIN),” *Investopedia*, May 21 2019 (accessed Feb. 1, 2020), https://www.investopedia.com/terms/b/bank-identification-number.asp.

[31] C. Kanich, N. Weavery, D. Mccoy, T. Halvorson, C. Kreibichy, K. Levchenko, V. Paxson, G. Voelker, and S. Savage, “Show me the money: Characterizing spam-advertised revenue,” 08 2011, pp. 15–15.

[32] M. Kersten, “Ticket resellers infected with a credit card skimmer,” Jan. 20 2020 (accessed Feb. 24, 2020), https://maxkersten.nl/2020/01/20/ticket-resellers-infected-with-a-credit-card-skimmer/.

[33] Y. Klijnsma, “Spray and pray: Magecart campaign breaches websites en masse via misconfigured Amazon S3 buckets,” *RiskIQ*, Jul. 10 2019 (accessed Feb. 24, 2020), https://www.riskiq.com/blog/labs/magecart-amazon-s3-buckets/.

[34] Y. Klijnsma, V. Kremez, and J. Herman, “Inside Magecart,” https://www.riskiq.com/research/inside-magecart/, Nov. 2018.

[35] K. Levchenko, A. Pitsillidis, N. Chachra, B. Enright, M. Felegyhazi, C. Grier, T. Halvorson, C. Kanich, C. Kreibich, He Liu, D. McCoy, N. Weaver, V. Paxson, G. M. Voelker, and S. Savage, “Click trajectories: End-to-end analysis of the spam value chain,” in *2011 IEEE Symposium on Security and Privacy*. Berkeley, CA: IEEE, May 2011, pp. 431–446, http://ieeexplore.ieee.org/document/5958044/.

[36] M. Motoyama, D. McCoy, K. Levchenko, S. Savage, and G. M. Voelker, “An analysis of underground forums,” in *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, ser. IMC ’11. New York, NY, USA: Association for Computing Machinery, 2011, p. 71–80. [Online]. Available: https://doi.org/10.1145/2068816.2068824

[37] P. Muncaster, “Magecart black hats battle it out on infected site,” *Infosecurity Magazine*, Nov. 21 2018 (accessed Feb. 24, 2020), https://www.infosecurity-magazine.com/news/magecart-black-hats-battle-it-out/.

[38] J. Onaolapo, M. Lazarov, and G. Stringhini, “Master of sheets: A tale of compromised cloud documents,” in *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS PW)*, June 2019, pp. 414–422.

[39] J. Onaolapo, E. Mariconti, and G. Stringhini, “What happens after you are pwnd: Understanding the use of leaked webmail credentials in the wild,” in *Proceedings of the 2016 ACM on Internet Measurement Conference - IMC ’16*. Santa Monica, California, USA: ACM Press, 2016, pp. 65–79, http://dl.acm.org/citation.cfm?doid=2987443.2987475.

[40] Y. Park, D. McCoy, and E. Shi, “Understanding Craigslist rental scams,” in *International Conference on Financial Cryptography and Data Security*. Springer, 2016, pp. 3–21, http://fc16.ifca.ai/preproceedings/01_Park.pdf.

[41] PCI Security Standards Council, “Press release: Two leading cybersecurity organizations issue joint bulletin on threat of online skimming

to payment security," Aug. 1 2019 (accessed Jan. 23, 2020), https://www.pcisecuritystandards.org/about_us/press_releases/pr_08012019.

[42] T. Peacock and A. Friedman, "Automation and disruption in stolen payment card markets," in *Workshop on the Economics of Information Security*, 2014.

[43] 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," 07 2019, pp. 181–192.

[44] Pew Charitable Trusts, "Why Americans use prepaid cards: A survey of cardholders' motivations and views," Feb. 2014 (accessed Feb. 13, 2020), https://www.pewtrusts.org/~/media/legacy/uploadedfiles/pcs_assets/2014/prepaidcardssurveyreportpdf.pdf.

[45] J. Pimental, "Olympic ticket reseller Magecart infection," Jan. 25 2020 (accessed Feb. 24, 2020), https://www.goggleheadedhacker.com/blog/post/14.

[46] R. S. Portnoff, S. Afroz, G. Durrett, J. K. Kummerfeld, T. Berg-Kirkpatrick, D. McCoy, K. Levchenko, and V. Paxson, "Tools for automated analysis of cybercriminal markets," in *Proceedings of the 26th International Conference on World Wide Web*, ser. WWW '17. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2017, p. 657–666. [Online]. Available: https://doi.org/10.1145/3038912.3052600

[47] P. Rosati, M. Cummins, P. Deeney, F. Gogolin, L. Van der Werff, and T. Lynn, "The effect of data breach announcements beyond the stock price: Empirical evidence on market activity," *International Review of Financial Analysis*, vol. 49, 01 2017.

[48] J. Ryoo, "How banks use machine learning to know a crook's using your credit card details," *Gizmodo*, Nov. 26 2015 (accessed Feb. 25, 2020), https://gizmodo.com/how-banks-use-machine-learning-to-know-a-crooks-using-y-1744771152.

[49] Sanguine Security, "Widespread credit card hijacking discovered," Nov. 17 2015 (accessed Feb. 20, 2020), https://sansec.io/labs/2015/11/17/widespread-credit-card-hijacking-discovered/.

[50] ——, "5900 online stores found skimming [analysis]," Oct. 11 2016 (accessed Jan. 23, 2020), https://sansec.io/labs/2016/10/11/5900-online-stores-found-skimming/.

[51] ——, "Merchants struggle with MageCart reinfections," Nov. 12 2018 (accessed Feb. 25, 2020), https://sansec.io/labs/2018/11/12/merchants-struggle-with-magecart-reinfections/.

[52] T. Seals, "Magecart strikes again, siphoning payment info from Newegg," *Threatpost*, Sep. 20 2018 (accessed Jan. 23, 2020), https://threatpost.com/magecart-strikes-again-siphoning-payment-info-from-newegg/137576/.

[53] J. Segura, "New golang brute forcer discovered amid rise in e-commerce attacks," *Malwarebytes Labs*, Feb. 26 2019 (accessed Feb. 20, 2020), https://blog.malwarebytes.com/threat-analysis/2019/02/new-golang-brute-forcer-discovered-amid-rise-e-commerce-attacks/.

[54] A. Shulman, "The underground credentials market," *Computer Fraud & Security*, vol. 2010, pp. 5–8, 2010. [Online]. Available: https://doi.org/10.1080/14786011003634415

[55] B. Stock, G. Pellegrino, C. Rossow, M. Johns, and M. Backes, "Hey, you have a problem: On the feasibility of large-scale web vulnerability notification," in *25th USENIX Security Symposium (USENIX Security 16)*. Austin, TX: USENIX Association, Aug. 2016, pp. 1015–1032. [Online]. Available: https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/stock

[56] K. Thomas, D. Huang, D. Wang, E. Bursztein, C. Grier, T. J. Holt, C. Kruegel, D. McCoy, S. Savage, and G. Vigna, "Framing dependencies introduced by underground commoditization," in *Proceedings (Online) of the Workshop on Economics of Information Security*, 2015, https://cseweb.ecsd.edu/~savage/papers/WEIS15.pdf.

[57] VISA, "New JavaScript skimmer 'Pipka' targeting eCommerce merchants identified," *VISA*, Nov. 2019 (accessed Feb. 20, 2020), https://usa.visa.com/dam/VCOM/global/support-legal/documents/pfd-identifies-new-javascript-skimmer.pdf/.

[58] Z. Whittaker, "Ticketmaster breach was part of a larger credit card skimming effort, analysis shows," *ZDNet*, Jul. 10 2018 (accessed Jan. 23, 2020), https://www.zdnet.com/article/ticketmaster-breach-was-part-of-a-larger-credit-card-skimming-effort-analysis-shows/.

[59] Wired Staff, "Backdoor exposes credit cards," *Wired*, Apr. 27 2000 (accessed Feb. 20, 2020), https://www.wired.com/2000/04/backdoor-exposes-credit-cards/.

[60] M. Yip, C. Webber, and N. Shadbolt, "Trust among cybercriminals? Carding forums, uncertainty and implications for policing," *Policing and Society*, vol. 23, no. 4, pp. 516–539, 2013. [Online]. Available: https://doi.org/10.1080/10439463.2013.780227

Appendix

Table VI compiles observations for sites on which we exposed payment card numbers, and Table VII provides details of all observed charges.

TABLE VI

Observations for tested sites. ID = identifier for both site and exposed payment card; Card Type = type of payment card exposed; Date Exposed = date payment card exposed; Misuse = whether we observed misuse of payment card following exposure; 3rd-Party Trans. = third-party URLs that we observed receiving credentials (see Section IV-D; "–" where mitmproxy error occurred, see Section III-D); Suspicious Script = whether we identified a suspicious script loaded by the site (see Section IV-E); Misuse Cluster = whether scripts on the site were clustered with an observed misuse case (see Section IV-E).

ID	Card Type	Date Exposed	Misuse	3rd Party Trans.	Suspicious Script	Misuse Cluster	Notes
1	Gift Card	2018-05-02	✓	–	✓	✗	*a*
2	Virtual Card	2018-05-02	✓	–	✓	✓	
3	Gift Card	2018-05-02	✗	–	✓	✓	*b*
4	Gift Card	2018-05-02	✗	–	✓	✗	*b*, *c*
5	Gift Card	2018-05-02	✗	–	✓	✗	
6	Virtual Card	2018-05-02	✗	–	✓	✗	*b*
7	Gift Card	2018-05-02	✗	–	✓	✗	
8	Gift Card	2018-05-02	✗	✗	✓	✓	*d*
9	Gift Card	2018-05-02	✗	–	✓	✗	
10	Gift Card	2018-05-02	✗	–	✓	✗	
11	Gift Card	2018-05-03	✓	logistic.tw	✓	✗	
12	Gift Card	2018-05-03	✗	✗	✓	✓	*b*
13	Virtual Card	2018-05-03	✗	✗	✓	✗	
14	Gift Card	2018-05-08	✗	✗	✓	✗	
15	Gift Card	2018-05-09	✗	✗	✓	✗	
16	Virtual Card	2018-05-09	✓	simcounter.com	✓	✓	
17	Gift Card	2018-05-09	✗	✗	✓	✓	
18	Virtual Card	2018-05-10	✗	paypal.com	✓	✓	
19	Gift Card	2018-05-10	✗	✗	✗	✗	*e*
20	Virtual Card	2018-05-10	✗	✗	✓	✗	
21	Gift Card	2018-05-10	✗	✗	✓	✓	
22	Gift Card	2018-05-15	✓	trafficanalyzer.biz	✓	✓	
23	Gift Card	2018-05-15	✗	✗	✓	✗	
24	Gift Card	2018-05-15	✓	paypal.com	✓	✓	*b*
25	Gift Card	2018-05-15	✓	cdnapis.com	✓	✓	
26	Virtual Card	2018-05-16	✓	magentocore.net	✓	✓	
27	Virtual Card	2018-05-16	✗	heartlandportico.com	✓	✓	*a*, *f*
28	Gift Card	2018-05-16	✓	–	✓	✗	
29	Virtual Card	2018-05-16	✗	✗	✓	✗	
30	Gift Card	2018-06-28	✗	✗	✓	✗	
31	Virtual Card	2018-06-28	✗	✗	✓	✗	
32	Virtual Card	2018-06-28	✗	trafficanalyzer.biz	✓	✗	*a*
33	Gift Card	2018-07-18	✓	g-analytics.com	✓	✗	*c*
34	Gift Card	2018-07-18	✗	✗	✓	✓	*b*
35	Gift Card	2018-07-18	✓	nearart.com	✓	✓	
36	Gift Card	2018-07-24	✗	brewtees.com	✓	✓	
37	Virtual Card	2018-07-24	✗	✗	✓	✗	
38	Virtual Card	2018-07-24	✗	dnsden.biz	✓	✓	
39	Virtual Card	2018-07-24	✗	magentocore.net	✓	✓	
40	Gift Card	2018-07-24	✗	authorize.net	✓	✓	*b*
41	Virtual Card	2018-07-24	✓	–	✓	✓	
42	Virtual Card	2018-07-24	✗	✗	✓	✗	
43	Gift Card	2018-07-24	✓	cdnapis.com	✓	✓	*d*
44	Virtual Card	2018-07-24	✗	✗	✓	✗	
45	Gift Card	2018-07-24	✗	magentocore.net	✓	✓	
46	Gift Card	2018-07-24	✗	✗	✓	✓	
47	Gift Card	2018-07-24	✓	–	✓	✓	*b*
48	Gift Card	2018-07-24	✓	magentocore.net	✓	✗	
49	Gift Card	2018-07-24	✗	cloud-update.top	✓	✗	
50	Virtual Card	2018-07-24	✗	✗	✓	✓	

[a] Identified suspicious code in otherwise apparently benign code (e.g., jQuery)

[b] Identified suspicious code that checks for initialization of window.Firebug (and possibly other browser developer tools)

[c] Identified suspicious code that appears to encrypt payment credentials before transmission (excluding HTTPS)

[d] Identified multiple, apparently unrelated (and possibly interfering) suspicious scripts on the same site

[e] Identified obfuscated code but unable to determine behavior with sufficient confidence to label suspicious

[f] Identified suspicious code that apparently fails to extract credentials in unexpected format (e.g., spaces between numbers)

TABLE VII
ALL OBSERVED UNAUTHORIZED CHARGE ATTEMPTS ON EXPOSED PAYMENT CARDS (V = VIRTUAL CARD; G = GIFT CARD).

Card ID	Type	Date Exposed	Time of Activity	Days Elapsed	Amount	Destination	Category
1	G	2018-05-02	2018-06-30 00:01	59	$15.93	Name-cheap.com	Hosting
2	V	2018-05-02	2018-06-25 12:49	54	$1.17	blizzard entertainment	Gaming
2	V	2018-05-02	2018-06-25 12:49	54	$1.17	blizzard entertainment	Gaming
2	V	2018-05-02	2018-06-25 12:49	54	$1.17	blizzard entertainment	Gaming
2	V	2018-05-02	2018-06-25 12:49	54	$1.17	blizzard entertainment	Gaming
2	V	2018-05-02	2018-06-25 12:49	54	$1.17	blizzard entertainment	Gaming
2	V	2018-05-02	2018-06-25 12:49	54	$1.17	blizzard entertainment	Gaming
2	V	2018-05-02	2018-06-25 12:49	54	$1.17	blizzard entertainment	Gaming
2	V	2018-05-02	2018-06-25 12:49	54	$1.17	blizzard entertainment	Gaming
2	V	2018-05-02	2018-06-25 12:50	54	$1.00	blizzard entertainment	Gaming
2	V	2018-05-02	2018-06-25 12:50	54	$1.00	blizzard entertainment	Gaming
2	V	2018-05-02	2018-08-02 13:49	92	$50.00	bb makeawish americ	Charity
11	G	2018-05-03	2018-07-29 17:57	87	$19.99	Playstation network	Gaming
16	V	2018-05-09	2018-07-11 07:16	63	$122.24	Loreal Kiehl's ecomm	Retail
22	G	2018-05-15	2018-11-08 23:57	177	$16.37	DING *33744664	Telecom
22	G	2018-05-15	2018-11-24 00:29	193	$3.60	PayPal *Exclusivida	E-Payment
24	G	2018-05-15	2018-08-04 00:29	81	$10.00	Name-cheap.com	Hosting
24	G	2018-05-15	2018-08-04 00:29	81	$10.00	Name-cheap.com	Hosting
25	G	2018-05-15	2018-07-31 19:57	77	$10.00	COX LAS VEGAS COMM SV	Telecom
25	G	2018-05-15	2018-08-18 04:53	95	$1.00	Aaron's Direct	Retail
25	G	2018-05-15	2018-08-18 11:28	95	$8.68	APL* ITUNES.COM/BILL	Music
25	G	2018-05-15	2018-08-24 12:30	101	$1.00	Aaron's Direct	Retail
26	V	2018-05-16	2019-01-11 19:12	240	$0.10	DoorDash	Food Delivery
26	V	2018-05-16	2019-01-11 19:12	240	$0.10	DoorDash	Food Delivery
28	G	2018-05-16	2019-01-10 17:54	239	$16.11	Verizon Wireless N8322-01	Telecom
34	G	2018-07-18	2018-08-03 00:50	16	$1.07	APL* ITUNES.COM/BILL	Music
34	G	2018-07-18	2019-03-17 05:44	242	$16.66	Chipotle Online	Food Delivery
35	G	2018-07-18	2018-12-25 23:00	160	$20.00	Depositagift.com	Charity
35	G	2018-07-18	2018-12-31 12:53	166	$20.00	Depositagift.com	Charity
35	G	2018-07-18	2019-01-05 01:57	171	$20.00	Staples Direct	Retail
41	V	2018-07-24	2018-10-23 01:35	91	$100.00	Steam	Gaming
41	V	2018-07-24	2018-10-23 01:35	91	$100.00	Steam	Gaming
41	V	2018-07-24	2018-10-23 01:35	91	$100.00	Steam	Gaming
41	V	2018-07-24	2018-10-23 01:35	91	$100.00	Steam	Gaming
41	V	2018-07-24	2018-10-23 01:35	91	$100.00	Steam	Gaming
41	V	2018-07-24	2018-10-23 01:35	91	$100.00	Steam	Gaming
41	V	2018-07-24	2018-10-23 01:36	91	$100.00	Steam	Gaming
41	V	2018-07-24	2018-10-23 01:36	91	$100.00	Steam	Gaming
41	V	2018-07-24	2018-12-10 17:50	139	$8.74	Hkbqzj Fwyfa Chartley, MA	Unknown
41	V	2018-07-24	2019-01-14 11:18	174	$1.00	Paypal	E-Payment
41	V	2018-07-24	2019-01-15 03:03	175	$6.71	Uoeieqocyqyfahy Ptdzmto	Unknown
41	V	2018-07-24	2019-03-07 03:31	226	$97.17	American Eagle	Retail
43	G	2018-07-24	2019-01-23 20:00	183	$20.00	Amazon.com*MB62E48B1	Retail
47	G	2018-07-24	2018-11-19 17:26	118	$20.00	Amazon.com*M82XM1RA2	Retail
48	G	2018-07-24	2019-03-03 05:10	222	$15.08	Uber Trip	Transportation