Using URL Shorteners to Compare Phishing and Malware Attacks

Sophie Le Page
University of Ottawa
slepage2@uottawa.ca

Guy-Vincent Jourdan
University of Ottawa
gjourdan@uottawa.ca

Gregor v. Bochmann
University of Ottawa
bochmann@eecs.uottawa.ca

Jason Flood
CTO Security Data Matrices
floodjas@ie.ibm.com

Iosif-Viorel Onut
IBM Centre for Advanced Studies
vioonut@ca.ibm.co

***Abstract*—In this work we use URL shortener click analytics to compare the life cycle of phishing and malware attacks. We have collected over 7,000 malicious short URLs categorized as phishing or malware for the 2 year period covering 2016 and 2017, along with their reported date. We analyze the short URLs and find that phishing attacks are most active 4 hours before the reported date, while malware attacks are most active 4 days before the reported date. We find that phishing attacks have higher click through rate with shorter timespan. Conversely malware attacks have lower click through rate with longer timespan. We also show the comparisons of referrers and countries from which short URLs are accessed, showing in particular an increased use of social media to spread both kinds of attacks. We also find phishing clicks mainly come from USA and Brazil, while malware clicks mainly come from USA and Russia. Overall based on the observation that 50% of malware attacks are active for several years, while less than 50% of phishing attacks are active past 3 months, we conclude that the efforts against phishing attacks are stronger than the efforts against malware attacks.**

I. Introduction

For several years now, research has looked at phishing website removal times and the number of visitors that the website attracts. Surprisingly very little research has looked at answering the same questions for malware websites. One reason for this may be because the data needed to perform such an analysis is generally hard to come by. Research that looks at phishing website removal times and number of visitors mostly makes use of resources such as log files, third party resources, and honeypots to gather data. However the last two resources are not always easy to get a hold of or to set up. As for the first resource option, log files are becoming scarce since people are instead using tools such as Google analytics to track their website click traffic. These analytics are not public information for understandable reasons.

URL shortening services, which provide users with a smaller equivalent of any provided long URL, are an example of a service which sometimes provide analytics as public information on their shortened URLs. Some URL shortening providers such as bit.ly[1] or goo.gl[2], allow viewing in real time the click traffic of a given short URL, including referrers and countries referring it.

Unfortunately attackers abuse URL shortening services, perhaps to mask the final destination where the victim will land after clicking on the malicious link. In our research we make use of the Bitly URL shortening service, reportedly the most popular service [1], [2], to analyze clicks for phishing and malware attacks. By phishing attacks we mean websites that trick a user into providing personal or financial information by falsely claiming to be a legitimate website. By malware attacks we mean websites that install data transmitting programs without the user's knowledge. We rely on two independent services to gather and classify these attacks: we use the popular community driven portal PhishTank[3] as a main source of phishing attack reports, as well as IBM's threat intelligence sharing platform X-Force Exchange[4] for both phishing and malware attacks.

We first gathered over 300,000 malicious URLs categorized as phishing or malware along with their reported date, and of these we identified over 7,000 Bitly short URLs. We then fetched click analytics from Bitly for each of these short URLs. From our analysis, we find that phishing attacks are most active 4 hours before the reported date, while malware attacks are most active 4 days before the reported date. We also find that phishing attacks have higher click through rate than malware, but that malware attacks have longer timespan than phishing. We also show the comparisons of referrers and countries from which short URLs are accessed, showing in particular an increased use of social media to spread both kinds of attacks. We also find phishing clicks mainly come from USA and Brazil, while malware clicks mainly come from USA and Russia. Overall we find that the efforts against phishing attacks are stronger than the efforts against malware attacks, based on the observation that 50% of malware URLs last for

[1]https://bitly.com/
[2]https://goo.gl/
[3]https://www.phishtank.com/
[4]https://exchange.xforce.ibmcloud.com/

978-1-5386-4922-0/18/$31.00

several years, while less than half of phishing URLs are active past 3 months.

This work provides a new and comparative insight about the life cycle of phishing and malware attacks. Our work adds to the analysis on the life cycle of phishing attacks from the perspective of short URLs, an ever increasing strategy that attackers use to obfuscate malicious URLs. As far as we are aware, our work also constitutes possibly the first analysis on the life cycle of malware URLs with respect to timespan and number of visitors, that is not limited to a single Enterprise or University. In addition this work complements previous short URL analyses, with new findings about click activity for flagged short URLs, and new analyses looking at the timespan of malicious short URLs. From our analysis of short URLs, we also find that Twitter spam click activity has remained consistent since 2010.

Additionally, all of the data used in this work is being made publicly available and can be found at http://ssrg.site.uottawa.ca/ecrime18/.

II. Background

The concept behind URL shortening services is to assist in sharing URLs by providing a short equivalent. URL shortening services provide users with a smaller equivalent of any provided long URL, and redirects to the corresponding long URL by the service provider through an "HTTP 301 Moved Permanently" response. Shortened URLs first appeared in 2001 [3] and initially, the concept was used to prevent breaking of complex URLs while copying text, and to prevent email clients from inserting line breaks in the links which rendered them unclickable. However, its adoption was slow until they became popular in online social networks. Now URL shorteners are almost a requirement due to character limitations in some social media, and due to mobile devices, where space is always at a premium.

Over the years URL shorteners such as Bitly have evolved to provide increased utility to users, such as providing analytics to track clicks. Bitly is one of the most popular URL shortening services, as a few studies have found [1], [2]. Each short URL Bitly issues is unique and will not be re-used, so the Bitly short URL will always direct to the same long URL. When a registered user shortens the same long URL, each instance gets a unique short URL. This way users can keep track of their own click analytics. A long URL may have many short URLs, shortened by different registered users, but an aggregate shortened URL keeps cumulative count of statistics for every click on the long URL through Bitly (see Figure 1).

A short URL is uniquely identified by what Bitly calls a hash. For example, if the following URL is submitted to Bitly "https://www.theweathernetwork.com/ca/hourly-weather-forecast/ontario/ottawa", the corresponding short URL is "http://bit.ly/2lFK4BS" which consists of Bitly's default domain name "bit.ly", and the hash "2lFK4BS" as the backhalf. A hash only contains the characters "a-z,A-Z,0-9", and is a simple iteration across every permutation of the available characters as the URLs come in [4]. For example, given a limit of three character representations, each new URL gets a unique three character combination until no more unique combinations are left, at which point the limit is increased to four character representations.

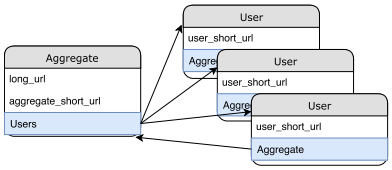


Fig. 1. Bitly mapping of long URL, user short URL and aggregate short URL. A long URL may have many short URLs, shortened by different registered users, but the aggregate shortened URL keeps cumulative count of statistics for every click on the long URL.

Users can also choose to create a customized backhalf which is easier to remember than random letters and numbers, but the backhalf is only accepted if it has not already been used. A customized backhalf can further contain characters "-" and "_" [5]. For example, the short URL "bit.ly/SuperBowl" is a customized backhalf. Further customization of short URLs include Branded Short Domain (BSD), such as "nyti.ms/SuperBowl", where "nyti.ms" is the BSD for New York Times, which allows large organizations to maintain their brand identity while using URL shorteners.

Unfortunately URL shortening services provide attackers with a convenient and free tool to obfuscate their URLs. Through the use of customized backhalfs, attackers can even craft a short URL so as to fit the target's profile (e.g., bit.ly/freemoviesfast). In the case of blacklisting malicious short URLs, for blacklists based only on domains rather than full URLs, false positives pose a threat of blacklisting entire sites such as URL shorteners. This resulted in blacklists having to use crawling in order to resolve shortened URLs, and blacklist the long URL. Nikiforakis *et al.* [6] also give a good overview of other dangers of URL shortening, such as linkrot and hijacking.

Pressure was also put on URL shortening services to put in place countermeasures for spam. For example as shown in Figure 2, Bitly flags malicious short URLs and displays a warning page upon clicking the malicious short URL. Bitly also provides a "preview" of a short URL by allowing users to append a '+' to the end of a short URL when entering it into the browser, as shown in Figure 3. This preview shows the long URL as well as detailed analytics such as number of clicks and percentage of referrer and country clicks. Note that the preview shows a brief overview of the analytics, and in order to see more details such as hourly number of clicks, one needs to use the Bitly API. It is through these public analytics that we propose to compare the life cycle of phishing and malware URLs.

III. Related Work

In this section we look at research related to measurements in phishing and malware attacks, such as the timespan of attacks, and the number of visitors an attack attracts. We then

Fig. 2. Bitly warning page.

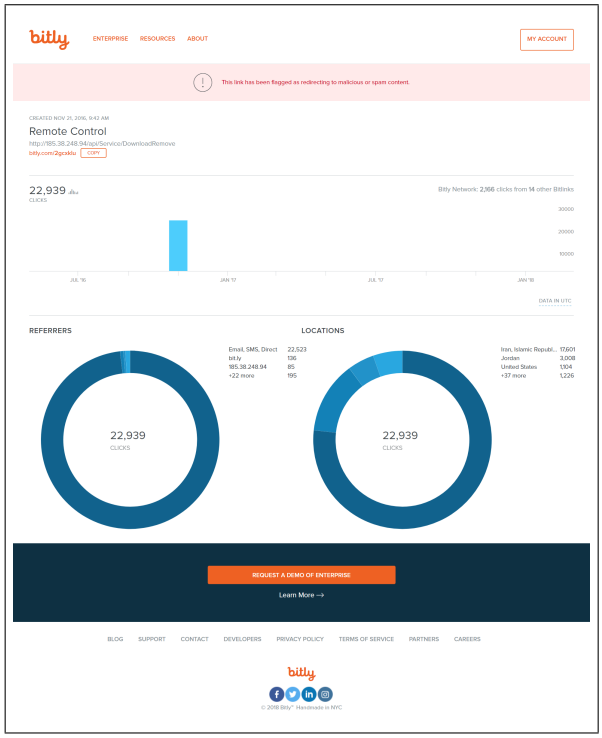


Fig. 3. Bitly short URL preview page.

look at the research around short URL analysis, in which initial research analyzed the acceptance and use of short URLs over long URLs. This research then quickly evolved into analyzing the amount of spam usage for short URLs. In turn this later evolved into the detection of and countermeasures for malicious short URLs.

A. Measurements in Phishing and Malware Attacks

For phishing, in 2007 Ludl *et al.* [7] found for phishing websites not yet blacklisted that the shortest addition to a blacklist was 9 minutes while the longest was 11 days. This was at a time where some blacklists could be debunked if you simply added or removed a '/' to the end of a malicious URL, as shown in their results. In 2007 Moore *et al.* [8] found the mean lifetime of phishing attacks to be 58 hours. They found surprisingly, that the user responses continued at a fairly high level after the reported date, until the site is removed. They could not tell whether this was caused by ongoing spamming activity, or by users catching up with email backlogs in their inboxes. In 2008 McGrath *et al.* [9] showed that some phishing domains last for at least 3 days without being discovered by anti-phishing tools. In 2009 Sheng *et al.* [10] did an empirical analysis of blacklists and found that 63% of the phishing campaigns lasted less than two hours, but that blacklists were ineffective when protecting users initially, as most of them caught less than 20% of phishing attacks within the first hour. In 2014 Gupta *et al.* [11] analyzed anti-phishing landing pages and found a forty six percent decrease in click through from trained users. In 2016 Han *et al.* [12] estimated the success rate of phishing kits by monitoring the activity of real visitors to infected honeypots, of which 9% submitted some data to the phishing page. In 2017 Cui *et al.* [13] analyzed replicas of phishing attacks as clusters, and found that around 80% of the clusters were active for less than one month, while the long-lasting attacks within a cluster stayed active anywhere between 2 months to the entire 10 months of their observations.

For malware, in 2007 Provos *et al.* [14] presented the state of malware on the web and emphasized the importance of this rising threat of injecting malicious content on popular web sites. In part of their analysis they analyzed only the URLs whose presence on the Internet lasted longer than one week to determine how often the binary of the malware changed. From their graph, some of these malware lasted over 20 months. However they did not comment on the portion of their data that is short-lived, and did not give a distribution of the timespan of the malware URLs. In 2014 Yen *et al.* [15] looked at malware in enterprises. Within that web based malware, they looked at encounter rate based on job type and type of website. They found that websites deemed business appropriate accounted for 31% of malware encounters.

Our work adds to the analysis of measuring the timespan and number of visitors for web based malware attacks. Other research not based on URL shortener information has shown that the timespan of phishing attacks has been reduced to hours. We want to find out here whether shortened URL analysis concurs with this observation.

B. Short URL Analysis

In 2008 McGrath *et al.* [9] found evidence that phishers were exploiting URL shortening services as far back as their dataset from 2006. Though they found the numbers not to be large, they warned of phishing continuing to abuse URL shortening services.

In 2010 Kandylas *et al.* [16] performed a comparative study of long URLs and short Bitly URLs on Twitter and found that Bitly short URLs received more clicks than an equal random set of long URLs. To further comprehend short URL distinctive characteristics, in 2011 Antoniades *et al.* [17] studied the lifetime of short URLs which revealed that the timespan of 50% short URLs exceeded 100 days. They also found that short URLs were mostly used on social networks

and propagated through word of mouth, often pointing to news and informative content.

In 2010 Grier *et al.* [18] found that 8% of 25 million URLs posted to Twitter pointed to phishing, malware, and scams listed on popular blacklists. They found that Twitter is a highly successful platform for coercing users to visit spam pages, with a click through rate of 0.13%. They found that blacklists are too slow at identifying new threats, allowing more than 90% of visitors to view a page before it became blacklisted. Around this time other studies started looking at malicious short URLs in emails and highlighted their privacy and security implications [1], [6]. In 2011 Chhabra *et al.* [19] grabbed phishing attacks from PhishTank, and did a target and referrer analysis, focusing on referrer ties to Twitter. In 2012 Klien *et al.* [20] did a geographical analysis and presented the global usage pattern of short URLs by studying the usage logs of their own URL shortening service, and found 80% of short URL content to be spam related. This is possibly because their shortening service did not have spam countermeasures. Conversely in 2013 Maggi *et al.* [21] performed a large scale study on 25 million short URLs over a 2 year period, and found very few short URLs used for spam. This is possibly because they looked at it from a user perspective *vs* service perspective. Their results also highlight that the countermeasures adopted by these services to detect spam are not very effective and can be easily by-passed. Experimental results from their data shows that Bitly allows users to shorten malicious links.

Several research papers exist on the detection of malicious short URLs [2], [22], [23]. In 2013 Wang *et al.* [2], reported results showing that the majority of the clicks were from direct sources and that the attackers utilized popular websites to attract more attention by cross-posting the links. In 2013 Yoon *et al.* [24] proposed an alternative to short URLs by using relative words of target URLs, thus hinting about the target URL, making it then comparatively safe from phishing attacks. In 2014 Gupta *et al.* [22] performed an exploratory study on Bitly's spam URL and account detection mechanism to expose the gaps in security mechanisms. Lastly, what first got us started on the topic of URL shortening was a conference workshop [25] given in 2017. The workshop proposed to use shortened URLs as a representative sample that can be extended to the overall phishing population to measure the impact of attacks.

Our work differs from previous work since we are using URL shortening analytics to compare types of attacks. The only other work from the list above that has done this is Grier *et al.* in 2010 [18], who looked at the blacklist evasion of scam, phishing and malware on Twitter. However, their use of Google Safebrowsing to identify phishing and malware, and Twitter's use of Google Safebrowsing API to filter links, biases their analysis. We also look at a more in-depth analysis of the comparison between types of attacks, including analyzing the click distribution, as well as analyzing the referrers and countries referring click traffic. In addition our dataset includes short URLs whose referrers do not just include Twitter, but also includes several other referrers such as Facebook.

IV. Statement of the Problem

Overall, the problem is: What can URL shortening analytics tell us about the life cycle of phishing and malware attacks?

Sub-problems are then:

- How active are phishing and malware attacks?
- How long do phishing and malware attacks last for?
- What is the distribution of clicks for phishing and malware attacks, before and after the reported date?
- Are there indications of attacks resurfacing after the reported date?
- Which top referrers and countries are referring click through traffic to phishing and malware attacks?

V. Methods

In this section we describe the methods we used to identify Bitly URLs during data collection, and the methods we used to determine unique URLs during analysis.

A. Identifying Bitly URLs

Given a list of malicious URLs, we describe the methods we used to identify those that are Bitly short URLs and those that are long URLs which have been shortened using Bitly's services. From a Bitly long URL, one can recover the corresponding short URL.

Identify Bitly Short URLs

1) Bitly Domain Name: Bitly's default domain name is "bit.ly". Other Bitly domain names are "bitly.com" and "j.mp". The domain name "j.mp" is a domain that Bitly offers to users who prefer a shorter domain name. This method simply checks if a given domain name is equal to either of the three domains mentioned above. Although this method does not identify branded short URLs, this is the easiest and fastest method to identify the majority of Bitly short URLs from a large list of URLs, since the method only requires string matching operations.

2) Bitly Branded Short Domain: When shortening a URL, users can create a Branded Short Domain (BSD) that takes the place of "bit.ly". For example "nyti.ms" is a BSD for New York Times. Since 2011, adding a BSD to a Bitly account is free, although a domain from a third party domain registrar must be purchased and linked to Bitly [26]. This method makes use of Bitly API endpoint "Pro domain" to query whether a given domain is a valid BSD. This method is relatively scalable for a large number of requests, and provides more coverage when used along with identifying by *Bitly Domain Name*. Note that a BSD must be less than 15 characters, including the dot, so we check the length of the domain before checking whether it is a BSD [27].

3) Bitly ASN: An Autonomous System Number (ASN) uniquely identifies the organization that is responsible for a block of IP addresses. This method identifies Bitly short URLs by checking whether a given domain is part of Bitly's ASN, which is "395224". This method will identify all Bitly domain names, including the branded ones. However checking the

ASN of a large list of URLs is usually only feasible with third party resources.

Identify Bitly Long URLs

4) Bitly URL Lookup: Since it is problematic to blacklist the URL shortening service itself, some sources do not blacklist Bitly short URLs but only the redirected URL (long URL). For this reason it is necessary to do a URL lookup to check whether a malicious URL has been shortened, and to record the corresponding short URL. This method identifies a Bitly long URL by checking the Bitly API "LookUp" endpoint, which returns the aggregate shortened URL (http://bit.ly/[aggregate_hash]) for a given long URL if there is any. A long URL may have many short URLs shortened by different registered users, but the aggregate shortened URL keeps cumulative count of statistics for every click on the long URL through Bitly.

B. Determining Unique URLs

We define a unique URL (e.g. http://example.com/path/) by removing duplicates that only add '/' to the end of the URL. In addition, URLs that only add '/' are considered the same URL when shortened using Bitly. There are other cases that Bitly considers as the same URL when shortened, however we do not take into consideration these cases. For example, a URL that begins with "http://www." and one that begins with just "www." are considered the same by Bitly when shortened.

VI. Data Collection

To collect malicious short URLs we use three steps. The first step is to collect malicious URLs from each source, the second step is to identify short URLs from the malicious URLs, and the third step is to fetch Bitly analytics on the short URLs. To collect malicious URLs, we looked for three things from our sources: a large database of verified malicious URLs, malicious URLs categorized as phishing or malware, and a report date. With these requirements in mind, we use the following sources to collect malicious URLs:

- *PhishTank*: A community-driven portal. PhishTank deals with phishing reports verified by users, so all malicious URLs from this source are for category phishing. For PhishTank, we consider the submission date of a malicious URL as the report date.
- *X-Force*: An IBM enterprise security analysis platform. X-Force maintains reports for verified malicious URLs from several categories, and allows querying for specific date ranges for URLs. For X-Force, we consider the created date of a malicious URL as the report date.

A. Collecting Malicious URLs

1) PhishTank: We have been collecting phishing URLs from PhishTank daily since January 1st, 2016. PhishTank is a community-based phish verification system where users submit suspected phishes and other users vote if it is a phish or not. PhishTank uses an adaptive cut-off for number of votes required for a submitted URL to be declared verified. When collecting phishing URLs from PhishTank, we filter those URLs that are verified, and manually verify some of them ourselves, since PhishTank's voting system can sometimes lead to false positives. From the verified URLs, we remove those URLs that are not reachable or are blocked, such as URLs that return a 404 Not Found status code. This is because in our previous research, we were particularly interested in retrieving the DOM of the landing page (final URL). More about our PhishTank dataset and methodology can be found in [13]. From January 1st, 2016 to December 31st, 2017, our PhishTank dataset consists of 51,516 unique phishing URLs.

2) X-Force: We queried using the X-Force endpoint "Get URLs by Category" for categories phishing and malware from January 1st 2016 to December 31st 2017. This resulted in 247,105 unique phishing URLs and 72,681 unique malware URLs in our X-Force dataset.

B. Identifying Short URLs

1) PhishTank: From our PhishTank dataset, we first tried to identify short URLs by *Bitly Domain Name*, but only found 61 URLs this way. This is likely because we removed submissions whose final URL was blocked, and since Bitly short URLs were often blocked by the time we crawled the URL, they were not included in our PhishTank dataset. As a result we searched by *Bitly ASN* directly through PhishTank and identified 1,048 short URLs, 8 of which used a branded domain. Returning to our PhishTank dataset, we used *Bitly URL Lookup* and identified 1,275 long URLs, recording the corresponding aggregrate short URL for each. At the end of this process, we identified 2,207 malicious short URLs from PhishTank for category phishing.

2) X-Force: From our X-Force dataset, we first tried to identify short URLs by *Bitly Domain Name*, but only found 4 short URLs for phishing and 1 short URL for malware. We concluded that X-Force avoids blacklisting Bitly short URLs and only blacklists the redirected URL (long URL). Next we used *Bitly URL Lookup* to identify Bitly long URLs and recorded the corresponding aggregate short URL. Note that X-Force returns malicious URLs without HTTP, so we tried to attach both "http://" and "https://" before looking up the URL. At the end of this process, we identified 5,855 malicious short URLs from X-Force, 2,532 for category phishing and 3,363 for category malware.

C. Fetching Short URL Analytics

As mentioned in Section II, Bitly maintains metrics for each created short URL. After we identified Bitly short URLs from our collected list of malicious URLs, we fetched the following Bitly metrics for each of these short URLs:

- *Clicks*: Returns the number of clicks on a single short URL. Data can be as detailed as the number of clicks for every hour.
- *Countries*: Returns the number of clicks for each country referring click traffic to a single short URL.

TABLE I
NUMBER OF MALICIOUS URLS COLLECTED.

Source	Phishing	Malware
PhishTank	51,516	-
X-Force	247,105	72,681

Category	URLs	Domains	Ratio
Phishing	299,418	98,009	3.1
Malware	72,681	41,772	1.7

TABLE II
NUMBER OF SHORT URLS COLLECTED.

Source	Phishing	Malware
PhishTank	2,207	-
X-Force	2,532	3,363

Category	Short URLs	Long URLs	Domains	Ratio
Phishing	4,324	4,324	3,394	1.27
Malware	3,363	3,363	2,916	1.15

- *Referrers*: Returns the number of clicks for each HTTP referrer referring click traffic to a single short URL. Even the click count for each full referrer URL is available.
- *Information*: Returns the date the short URL was created, as well as a reference to the aggregate short URL.
- *Expand*: Returns the long URL, as well as a reference to the aggregate short URL.
- *Encoders*: Returns the number of users who have shortened a single long URL.

For the sake of completeness, in our analysis we use metrics for the aggregate short URL, which aggregates the metrics from each registered user who shortened the same long URL. Analyzing the aggregate short URLs, we found that over 90% of short URLs were shortened by only one user, *vs* several registered users. This shows that analyzing the aggregate short URLs is effectively the same as analyzing the user short URLs.

Note that for this work, the last day we fetched Bitly metrics was on February 10th 2018.

VII. DATASET

The previously discussed collection process is summarized in Table I and Table II. As shown in Table I, in 2016 and 2017 the number of detected URLs was 299,418 and 72,681 for phishing and malware respectively. We find that for phishing, the URL to domain ratio is 3.1, while for malware the ratio is 1.7. This indicates that more URLs are being hosted on the same domain for phishing than malware. Knowing how many URLs occur with each domain indicates the approximate number of attacks from the domain.

As shown in Table II, from the malicious URLs collected, we identified 7,647 short URLs, 4,324 categorized as phishing and 3,363 as malware. Note there is some overlap between categories phishing and malware, which is expected. In the bottom table we included the number of long URLs, which is the same number as short URLs. This is because we use the aggregate short URL which is a one to one mapping to a long

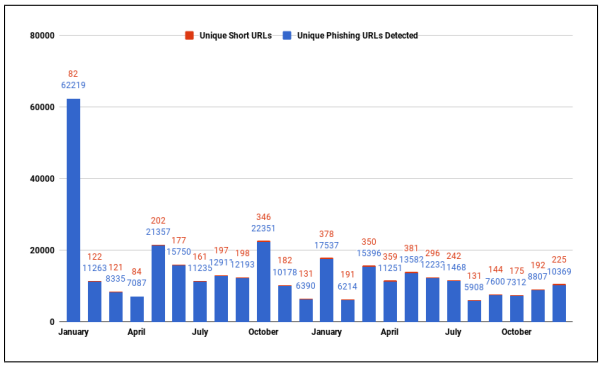


Fig. 4. Unique phishing URLs detected by month 2016-2017 with corresponding number of short URLs.

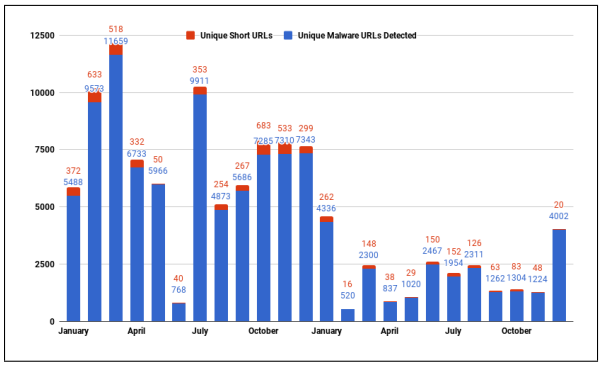


Fig. 5. Unique malware URLs detected by month 2016-2017 with corresponding number of short URLs.

URL. The number of unique domains are of the long URL, for which the URL to domain ratio is 1.27 for phishing and 1.15 for malware. Again this indicates more short URLs are pointing to the same domain for phishing than for malware, but a lot less when compared against the URL to domain ratio of malicious URLs.

Figure 4 and Figure 5 give consolidated phishing URL and malware URL statistics respectively by month for 2016 and 2017, with corresponding number of short URLs. The highest number of detected phishing URLs was in January 2016, with an unusually high number compared to other months. This high number is due to repeated attacks on the same domain, where the number of unique domains in January 2016 is 4,700, so the URL to domain ratio is 13.2 [5]. In contrast, the second highest number of phishing URLs is in October 2016, with 8,823 unique domains and a 2.5 URL to domain ratio. In general there are more unique phishing URLs by month in the first half of 2016 than in the first half of 2017. This follows the same findings when comparing the number of unique phishing attacks in AWPG's 2016 and 2017 reports [28] and follows the same findings as in the Q4 2017 report from McAfee Labs [29].

The highest number of unique malware attacks was in March 2016, followed by July 2016 and February 2016. As

[5]We do not know whether this spike is due to an increase in phishing attacks or because X-Force or one of its sources changed its reporting method.

Figure 5 shows, there are more unique malware attacks in 2016 than in 2017. This does not follow the findings as the Q4 2017 report from McAfee Labs [29], which reported malware increase by 10% in Q3 of 2017, a record high. This may be because the type of malware in our short URL dataset is only the type found on websites, while McAfee looks at a much wider variety of malware such as PowerShell malware and Mac malware. We also find in the 2017 report from IBM X-Force [30] that more malware is being distributed through spam emails, and we suspect these cases are being categorized as spam instead of malware within X-Force.

Comparing the number of short URLs identified in our list of malicious URLs, we find that 1.45% of phishing URLs and 4.62% of malware URLs correspond to short URLs. This follows results in [19] which found in 2011 that 3.13% of phishing URLs from PhishTank corresponded to Bitly short URLs. From this we see that the malware URLs in our dataset are about four times more likely to use URL shortening services. This may be because 91% of malware URLs are reported as base URLs, which may increase the likelihood of a URL having been shortened. This also suggests that malware URLs are more often hosted on maliciously registered domains instead of compromised domains. Conversely only 38% of phishing URLs are reported as base URLs.

VIII. Analysis and Results

In this section we perform an analysis on our previously discussed dataset. We start our analysis by looking at whether clicks are recorded for short URLs flagged by Bitly. With this understanding, we then analyze phishing and malware click through, click timespan, click distribution against reported date, and click sources from HTTP referrers and countries.

A. Flagged Short URL Clicks

Bitly short URL clicks are counted in near real time. One can simply test this by navigating to a short URL, and then previewing the URL by appending '+' to the end of it, to see the click count increased by one. However we found that clicks are not counted for short URLs which Bitly has flagged as malicious, which means click analysis can not be continued once a URL has been flagged. For example research has shown that after a URL has been blacklisted, visits continue to be logged up until the URL is taken down [8]. However based on our findings, this type of analysis would not be observed using short URLs, since clicks are no longer recorded when a short URL is flagged and a warning page is shown, even if a victim continues through the warning page to the malicious site. Therefore this should be taken into consideration when doing click analysis for malicious short URLs.

To check whether a click is counted for flagged short URLs, we gathered flagged short URLs by searching "This link has been flagged as redirecting to malicious or spam content" in Google. This search results in several flagged Bitly short URLs. We then entered these flagged URLs into the web browser, and after continuing through the warning page to the malicious site, we found that our clicks were not counted. It is unclear why clicks are not counted for flagged short URLs. We emailed Bitly to ask, but have not received a reply yet.

To verify our findings, we grabbed from our dataset the short URLs that were submitted on PhishTank as "warning" links along with their submission date. For example some of the short URLs were submitted as follows:

`https://bitly.com/a/warning?hash=1WT7pjU...`

Knowing when a short URL is submitted as a warning link gives us an indication as to when a short URL was flagged. Next we checked the short URL's latest click date against the submission date. If the clicks of a flagged short URL are no longer recorded, then we should observe that there are no clicks after the submission date.

We only have 16 of these "warning" short URLs in our dataset, which were almost all submitted by Veriform [31] to PhishTank. Of the 16 short URLs, 14 are still flagged by Bitly. Of these 14 URLs, 7 have their latest click before the submission date, which follows our understanding that flagged short URLs no longer record clicks. However the other 7 URLs have their latest click after the submission date. Taking a closer look at these 7 short URLs, we find a clear indication of the clicks stopping before the submission date, but then reappearing momentarily after a few more days, before stopping completely.

An example of this case is given in Figure 6, of a short URL "bit.ly/1WT6ZtL" which was created on June 18th 2016 at 9:20 AM, and was submitted on PhishTank as a warning link on June 20th 2016 at 7:57 PM. The graph shows the number of clicks every hour, and we see the clicks stop on June 20th 2016 at 4:00 PM, before the PhishTank submission date. This indicates that the short URL was flagged by Bitly on June 20th at 4:00 PM, at which point clicks are no longer counted. However, we see the clicks continue about 3 days later until stopping on June 25th 2016. After this point no new clicks have been recorded. This indicates that the short URL was flagged, then somehow unflagged by Bitly, before being flagged a final time.

We further manually checked these 7 cases exhibiting this behavior and found 5 of the cases were on compromised domains, hosted on a 3D printing shop, a robotics equipment shop, a kite surfing site, a Christian community radio site, and a travel site. Attacks hosted on a compromised domain would explain why the short URL might have been unflagged momentarily.

As far as we know, none of the previous research on malicious short URL analysis has mentioned clicks no longer being counted for flagged URLs. This may be because Bitly might have brought changes regarding clicks being counted. Based on our analysis we find that within the past two years, Bitly does not record new clicks for flagged short URLs.

B. Click Through

From our 7,647 collected short URLs, only 51.3% generated click traffic within 1 year of reported date, accumulating to over 11.8 million clicks. Of the short URLs generating click

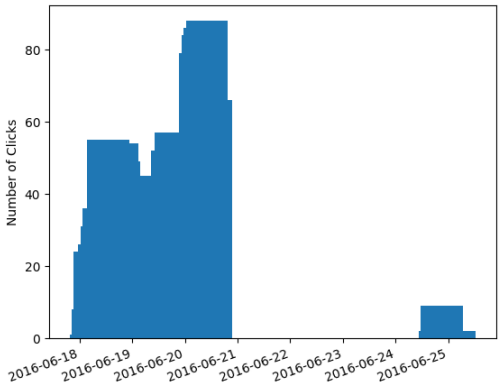


Fig. 6. Hourly clicks of a short URL flagged by Bitly.

TABLE III
NUMBER OF SHORT URLS WITH CLICK THROUGH.

	Phishing		Malware	
	Size	Percent	Size	Percent
Number of unique short URLs	4,324	100.00%	3,363	100.00%
... with clicks	3,535	81.75%	1,042	30.98%
... ... within 1 year of report	3,425	79.21%	520	15.46%
... ... within 3 months of report	3,259	75.37%	313	9.31%
... ... within 50 days of report	3,151	72.87%	264	7.85%
... ... within 48 hours of report	2,346	54.25%	79	2.35%

traffic, we find that 10% of URLs make up 90% of the accumulated clicks. This shows that malicious URLs generate heavy traffic using URL shortening services and that only a few URLs make up the majority of clicks.

Of the short URLs not generating any clicks, most of these are malware URLs, as can be seen in Table III, whose numbers drop by two thirds when considering click traffic. This may be because some of these short URLs may not have been used in actual attacks. Therefore these short URLs do not necessarily need clicks to be reported since it suffices that the long URL be reported. To confirm this, if we look at only the short URLs reported on PhishTank, we find 99.98% of these URLs have click traffic. This indicates that when short URLs are reported they are more likely to have been clicked on, *vs* when the long URL is reported.

Table III shows the number of short URLs as we restrict click through traffic closer to the reported date. When analyzing URL clicks, we restrict our analysis to short URLs with clicks within at least 1 year of reported date since some of the URLs were created as far back as 2009, and do not have any recent click traffic. We consider clicks only occurring years before the reported date are not relevant in our analysis. Looking at Table III, we notice that when we restrict URLs with clicks within 48 hours, 54% of phishing URLs are active while 46% are inactive, and 2% of malware URLs are active while 98% are inactive. The large percentage of inactive URLs suggests that our sources are late at reporting the attacks.

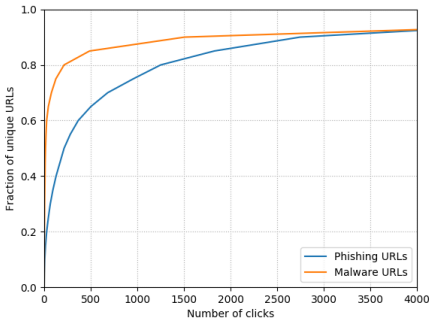


Fig. 7. Click through for phishing and malware URLs. Only the 79% of phishing URLs and 15% of malware URLs generating any traffic within 1 year of report are shown.

Looking at Figure 7, of the URLs that generate any traffic within 1 year of report, 20% of the phishing URLs and 60% of malware URLs receive less than 27 clicks. This shows that malware URLs have less click activity than phishing URLs. This may be because it has been found that phishing attacks work well at getting victims to click on their links [32] [33].

Looking at previous research, Grier *et al.* [18] reported that 2.3% of malicious short URLs have click through traffic, which is much lower than our 51.3%. This may be because their dataset consists of malicious short URLs from Twitter over a period of 3 months, in which 95% of their dataset were scam attacks, and only 5% were phishing and malware attacks. To take this into consideration, of our short URLs generating any traffic within 3 months of reported date, 5.04% of these URLs have clicks from Twitter. This indicates that there is a similar number of malicious short URLs being clicked on from Twitter over the years. Grier *et al.* [18] also showed that 50% of URLs received less than 10 clicks. If we consider only short URLs generating any traffic within 3 months of reported date, looking only at clicks from Twitter, 50% of our URLs receive less than 13 clicks. This indicates that malicious short URLs from Twitter have been receiving about the same number of clicks over the years. More results about Twitter and other referrers for phishing and malware URLs is analyzed in Section VIII-E.

Overall from our click through analysis, we find that phishing receives more click through than malware.

C. Click Timespan

Here we consider the timespan of clicks for each short URL, where timespan is measured from first to last click. We restrict our analysis to short URLs which have clicks within at least 1 year of reported date. Following our findings in Figure 8, we see that 50% of phishing URLs last less than 80 days, while 50% of malware URLs last less than 340 days, just under 1 year. This shows that malware URLs have a longer timespan than phishing URLs.

We also notice that as the graph continues, there becomes more of a difference between the click timespan of phishing and malware URLs. At the extremes, 90% of phishing URLs

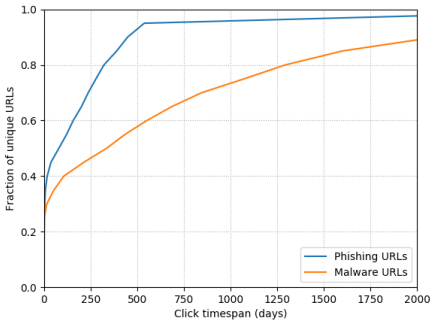


Fig. 8. Click timespan in days from first to last click for phishing and malware URLs. Only the 79% of phishing URLs and 15% of malware URLs generating any traffic within 1 year of report are shown.

last up to 1 year while 90% of malware URLs last less than 5 years. This shows that the upper 50% of malware URLs are lasting for several years, while the upper 50% of phishing URLs are lasting between 3 months and a year. As we look at the portion of the graph beyond 95% of phishing URLs, we notice a horizontal line in click timespan. This line is due to our dataset not having any phishing URLs with a timespan between 500 and 2,000 days, which again highlights the difference in timespan between malware and phishing URLs.

As far as we know, no previous research has looked at the timespan of malicious short URLs. In 2011 Antoniades *et al.* [17] studied the timespan of benign short URLs which revealed that 50% of short URLs exceeded 100 days. This shows that benign short URLs are active longer than phishing URLs, but that malware URLs are actually active three times longer than benign short URLs. Overall from our click timespan analysis, we find that malware lasts for longer timespan than phishing.

D. Click Distribution by Reported Date

We want to look at the click distribution of phishing and malware URLs before and after the reported date. To do this we aggregated the click data for every short URL. The most detailed we can aggregate our clicks is hourly, using Bitly's API. For URLs reported multiple times, we consider each report as a unique, independent event. For example, if we want to aggregate our clicks by hour, we create bins relative to each hour before and after a reported date. Then for each short URL, we normalize the number of clicks so that the hour with the maximum clicks has a weight of 1. This is to avoid short URLs with exceptionally large number of clicks to skew the results. Next we add these normalized values to their respective hour bins. Once we have aggregated all short URLs into bins, we normalize the bins so that their sum is equal to 1. Figure 9 shows the result of our analysis, looking at the distribution of clicks encountered 48 hours, 50 days and 1 year before and after the reported date. The number of URLs for each time frame can be found in Table III.

Looking at Figures 9 (a) and (b), we find a peak in the distribution of clicks for phishing about 4 hours before the reported date. However for malware we find that within 48 hours there is no clear peak in the distribution.

Looking at Figures 9 (c) and (d), as expected from the previous figure, we find a peak in the distribution of clicks for phishing on the day of the reported date. For malware we now see a more clear trend where the clicks have a lower distribution before the reported date, then peak about 4 days before the reported date.

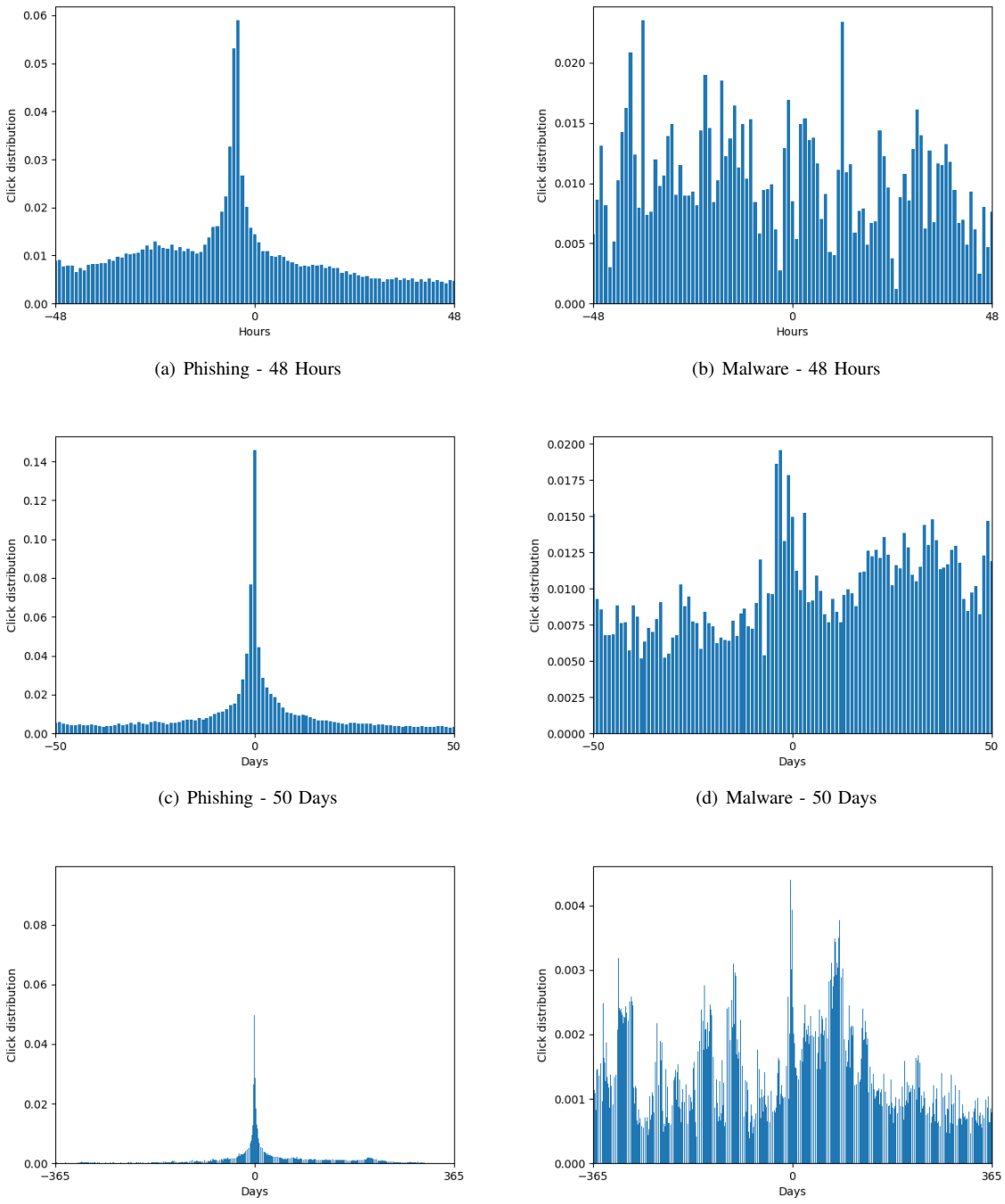
Looking at Figures 9 (e) and (f), here we see a bigger picture of the click distribution over a full year on either side of the reported date. Again as expected from the previous figures, for phishing we see one clear peak, with a shorter tail before the peak and a longer tail after the peak. The shorter tail before the peak indicates that when phishing URL click activity starts, the clicks grow quickly. The longer tail after the reported date indicates that phishing attacks are slow to be completely removed, and we even see a slight bump in clicks at around the half year mark. This bump may indicate an attempt for phishing attacks to resurface. For malware we still see the highest peak near the reported date, but we also see several peaks before the reported date, and surprisingly the second highest peak is after the reported date, after which the clicks slowly start to reduce. The peaks before the reported date show that our sources for reporting malicious URLs are lagging and are not catching active malware URLs. The peak after the reported date indicates that malware attacks are successfully resurfacing and continuing to receive click activity, even after they have been reported. With the click activity remaining constant over a timespan of several years, this view also confirms our findings in Section VIII-C, which found that malware attacks last longer than phishing attacks.

From previous research, Grier *et al.* [18] looked at the performance of several blacklists from when a malicious Twitter URL was posted. They found there was a 20 day lag for malware URLs to be blacklisted and an 8 day lag for phishing URLs to be blacklisted. This shows that blacklists have a larger lag when detecting malware on Twitter. Overall our click distribution analysis also shows that based on the point at which click activity peaks, our reports have a larger lag when detecting malware URLs.

E. HTTP Referrer Clicks

In this section we look at the HTTP referrers referring the most number of clicks. We restrict our analysis to short URLs with clicks within at least 1 year of reported date. We find that phishing URLs were accessed from 5,480 referrer domains, while malware URLs were accessed from 1,468 referrer domains. This makes sense since we have more phishing URLs in our dataset, and since phishing URLs have more click activity, as discussed in Section VIII-B.

As shown in Table V, we find that short URLs are most commonly accessed "directly", that is, from sources including email clients, instant messages, and applications. This also matches other findings [17], [2]. We notice that although "direct" is the most common source for both attacks, it is much more common for phishing URLs. For phishing, the next



(a) Phishing - 48 Hours

(b) Malware - 48 Hours

(c) Phishing - 50 Days

(d) Malware - 50 Days

(e) Phishing - 365 Days (1 Year)

(f) Malware - 365 Days (1 Year)

Fig. 9. Click distribution of phishing and malware URLs encountered 48 hours, 50 days and 1 year before and after reported date.

most common referrer is Facebook. This includes referrers such as m.facebook.com, l.facebook.com, and facebook.com. Interestingly for phishing, Twitter is the source of only 0.25% of clicks *vs* 4.6% for malware. This indicates that Twitter's defense against malware URLs is not as strong as its defense against phishing URLs, as was already noted in [18]. We also notice that a third of clicks for malware URLs are from Blogspot. For phishing we find smikta.net as a referrer with high click accesses; Smikta is an anonymisation service which advertises as a service allowing students to access any websites from school. We identified 11 other anonymisation services, which make up one of the top referrer categories for phishing, as shown in Table IV.

To get a bigger picture, as shown in Table IV, we grouped

TABLE IV
TOP REFERRER CATEGORIES WITH THE LARGEST NUMBER OF CLICKS.

	Phishing		Malware	
Rank	**Category**	**% Clicks**	**Category**	**% Clicks**
1	direct	61.93%	direct	33.07%
2	Social Media	12.14%	Blogs	29.85%
3	Social Networks	3.67%	Social Media	8.33%
4	Anonymisation	3.46%	Social Networks	8.32%
5	Webmail	2.31%	Search Engines	7.07%

TABLE V
TOP REFERRERS WITH THE LARGEST NUMBER OF CLICKS.

	Phishing		Malware	
Rank	**Referrer**	**% Clicks**	**Referrer**	**% Clicks**
1	direct	75.65%	direct	37.76%
2	facebook	10.94%	blogspot	33.30%
3	smikta	2.24%	google	7.43%
4	live	2.15%	twitter	4.60%
5	google	1.12%	facebook	4.08%

TABLE VI
TOP COUNTRIES WITH THE LARGEST NUMBER OF CLICKS.

	Phishing		Malware	
Rank	**Country**	**% Clicks**	**Country**	**% Clicks**
1	US	25.18%	US	24.57%
2	BR	17.40%	RU	14.50%
3	IN	7.13%	RO	7.37%
4	FR	3.25%	ES	5.84%
5	GB	2.71%	TH	3.80%
6	EG	2.54%	BR	3.75%
7	DE	2.27%	LT	3.57%
8	DZ	2.25%	KZ	3.31%
9	CA	2.06%	DE	3.30%
10	A1	1.72%	GB	3.22%

each referrer into categories using X-Force API endpoint "report". We were able to categorize 66% of phishing referrer domains and 72% of malware. For both phishing and malware URLs, we find that the majority of the clicks are from direct sources, social networks and social media, which all together account for between 50% and 80% of clicks. This is likely because these sources allow short URLs to reach a large audience. Similar observations were made in 2013 by Wang *et al.* [2] who investigated Bitly spam short URLs from Twitter. Overall we find an increased use of social media to spread both phishing and malware.

F. Country Referrer Clicks

In this section we look at the countries referring the most number of clicks. We restrict our analysis to short URLs with clicks within at least 1 year of reported date. We find that there are 236 countries referrering click through for phishing and 234 countries for malware. Unlike HTTP referrers, we find phishing and malware have roughly the same number of countries referrering click through, perhaps because the number of possible countries is much less than the number of possible referrers. In total there are 254 possible country codes [34], which shows that nearly all countries are referrering click through traffic for both phishing and malware URLs.

As shown in Table VI, we find that our malicious short URLs are most commonly accessed from US (United States), which matches several other findings [17], [20], [2]. This is likely because of a bias in our data: our data sources tend to be North-American centric, and Bitly is an organization from the United States which presumably has more users from a small pool of countries.

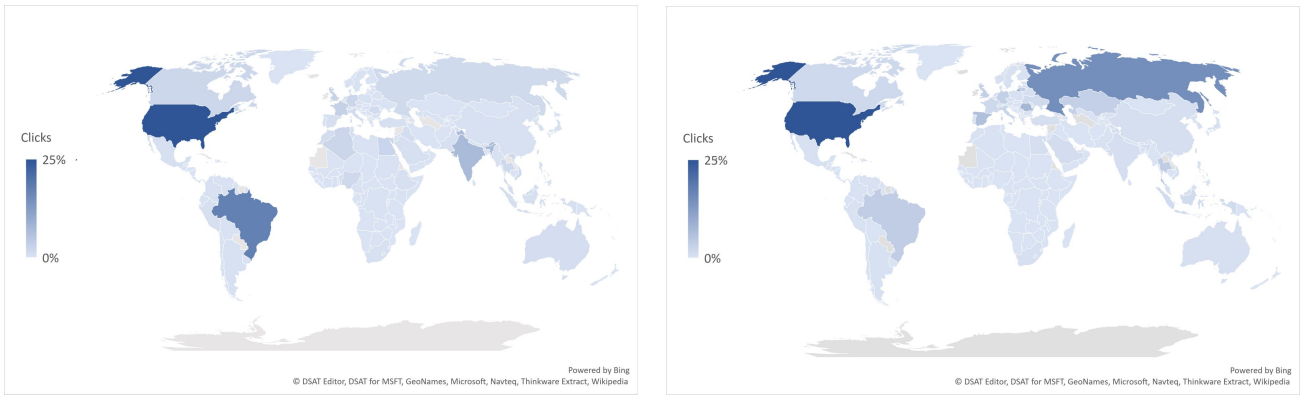
Comparing phishing and malware URLs, we find the top countries for both include US (United States), BR (Brazil), GB (United Kingdom) and DE (Germany). We find the differences to be that phishing includes IN (India), FR (France), EG (Egypt), DZ (Algeria), CA (Canada) and A1 (Anonymous Proxy), while malware includes RU (Russia), RO (Romania), ES (Spain), TH (Thailand), LT (Lithuania) and KZ (Kazakhstan). Some of these similarities and differences can be seen more clearly in Figure 10 where we include all the country click percentages on a world map. Clicks from Brazil account for an especially high number of accesses for phishing. This follows the 1H 2017 APWG report [28], which shows Brazil as the second highest country reporting phishing incidents. Clicks from Russia account for an especially high number of accesses for malware. As the Q4 2016 APWG report shows [28], Russia is ranked #6 in the list of countries with the highest infection rate, right behind China, Turkey, Guatemala, and Ecuador. AWPG also report Scandinavian countries have the lowest infection rates. We find for malware, Scandinavian countries account for 1.03% of clicks, whereas for phishing they only account for 0.43% of clicks.

In 2013, Wang *et al.* [2] investigated Bitly spam short URLs from Twitter collected over a 4 month period, and found Thailand at the top of their list because Thailand was a referrer for one their short URLs which had exceptionally high number of clicks. We found a similar case in our dataset of a short URL with malicious domain "www.pizzahut1150.com", generating over 100,000 clicks from Thailand. This shows that malicious URLs may generate heavy traffic using URL shortening services and confirms our findings in Section VIII-B in which only a few URLs make up the majority of clicks.

Overall we find the United States as the top country referrer for both phishing and malware, where phishing clicks mainly come from USA and Brazil, while malware clicks mainly come from USA and Russia. We also find that Scandinavian countries have higher click percentage for malware than for phishing.

IX. CONCLUSION

In our analysis we found several differences between phishing and malware attacks relative to click through, click timespan, report lag, and click sources. We found that phishing URLs receive more click through traffic, where 60% of malware URLs receive less than 27 clicks while the same is true for only 20% of phishing URLs. We also found that malware



(a) Phishing (b) Malware

Fig. 10. World maps showing countries referring click through for phishing and malware URLs by percentage of clicks. There are 236 countries referring click through for phishing URLs and 234 countries referring click through for malware URLs.

URLs have longer timespan, where 50% of malware URLs are active for several years while less than 50% of phishing URLs are active past 3 months. Although phishing URLs have smaller timespan, those URLs which are active past 3 months have longer timespan than what has been reported in past research, which find that most phishing attack removal times were within days and even hours. Our findings more closely follow what was found in [13], in which several phishing attacks remained active for up to 10 months.

Looking at the click distribution, for phishing we found the highest click activity to be 4 hours before the reported date, with clicks growing quickly before, and then reducing less quickly after, where the clicks remained for up to half a year, with signs of the attacks attempting to resurface. Conversely, for malware we found the highest click activity to be 4 days before the reported date, with several other peaks before and after, indicating that our sources are not catching active malware URLs, and also indicating that malware URLs are successfully resurfacing even after being reported. Overall our click distribution analysis shows that based on the point at which click activity peaks, our reports have a larger lag when detecting malware than phishing.

The results in our analysis of URL timespan and click distribution especially highlight that the efforts against malware attacks are not as strong when compared to phishing attacks. However, for phishing, the domain names might be more deceptive compared to Bitly short URLs, and it can be the attackers' strategy to only use short URLs for a small window of time. Conversely, for malware, the timespan seems to be much longer than expected and it is surprising that the malicious domain is not taken down. It could be that Bitly counts the click even if the domain does not resolve anymore. We would be interested in exploring further to determine why malware URL efforts are not as strong.

During our analysis we also found evidence suggesting that Twitter spam click activity has remained consistent compared to research from 2010 [18]: We found that Twitter receives similar volume of click through traffic, where 5% of short URLs receive clicks from Twitter, as well as similar number of clicks per short URL, where 50% of short URLs receive less than 13 clicks. We also found that Twitter's defense against malware URLs is not as effective as its defense against phishing URLs since Twitter is the source of only 0.25% of clicks for phishing *vs* 4.6% for malware.

In regards to other HTTP referrers, we found the majority of the clicks are from direct sources, social networks and social media, which all together account for 50% and 80% of clicks for phishing and malware URLs respectively. For country referrers, we found United States was the top referrer for both phishing and malware URLs, both with similar click percentages. However, the rest of the top referrers were mostly different, where Brazil was notably the second top country for phishing, and Russia was the second top country for malware. We also found that Scandinavian countries have higher click percentage for malware than for phishing.

Lastly, we also observed the limitations of short URL click analytics, finding that flagged Bitly short URLs no longer record new clicks. This means that click analysis can not be continued once a URL has been flagged. In analyzing country referrers we also acknowledge the limitation of Bitly short URLs since clicks from China are not prevalent in our dataset, in which case it would be interesting to pursue working on a larger and more complete dataset, such as shortening services from China.

All of the data used in this research is being made publicly available and can be found at http://ssrg.site.uottawa.ca/ecrime18/.

Acknowledgements

This work is supported by the IBM® Center for Advanced Studies and the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] A. Neumann, J. Barnickel, and U. Meyer, "Security and privacy implications of url shortening services," in *Proceedings of the Workshop on Web 2.0 Security and Privacy*, 2010.

[2] D. Wang, S. B. Navathe, L. Liu, D. Irani, A. Tamersoy, and C. Pu, "Click traffic analysis of short url spam on twitter," in *Collaborative Computing: Networking, Applications and Worksharing (Collaboratecom), 2013 9th International Conference Conference on*. IEEE, 2013, pp. 250–259.

[3] MetaFilter, "We want 'em shorter." https://www.metafilter.com/8916/We-want-em-shorter, 2001.

[4] Coding Horror, "URL Shortening: Hashes In Practice," https://blog.codinghorror.com/url-shortening-hashes-in-practice/.

[5] Bitly Blog, "How to Create And Customize Your Bitlinks," https://bitly.com/blog/bitly-basics-bitlinks/.

[6] N. Nikiforakis, F. Maggi, G. Stringhini, M. Z. Rafique, W. Joosen, C. Kruegel, F. Piessens, G. Vigna, and S. Zanero, "Stranger danger: exploring the ecosystem of ad-based url shortening services," in *Proceedings of the 23rd international conference on World wide web*. ACM, 2014, pp. 51–62.

[7] C. Ludl, S. McAllister, E. Kirda, and C. Kruegel, "On the effectiveness of techniques to detect phishing sites," in *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*. Springer, 2007, pp. 20–39.

[8] T. Moore and R. Clayton, "An empirical analysis of the current state of phishing attack and defence." in *WEIS*, 2007.

[9] D. K. McGrath and M. Gupta, "Behind phishing: An examination of phisher modi operandi." *LEET*, vol. 8, p. 4, 2008.

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

[11] S. Gupta and P. Kumaraguru, "Emerging phishing trends and effectiveness of the anti-phishing landing page," in *Electronic Crime Research (eCrime), 2014 APWG Symposium on*. IEEE, 2014, pp. 36–47.

[12] 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*. ACM, 2016, pp. 1402–1413.

[13] Q. Cui, G.-V. Jourdan, G. V. Bochmann, R. Couturier, and I.-V. Onut, "Tracking phishing attacks over time," in *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2017, pp. 667–676.

[14] N. Provos, D. McNamee, P. Mavrommatis, K. Wang, N. Modadugu *et al.*, "The ghost in the browser: Analysis of web-based malware." *HotBots*, vol. 7, pp. 4–4, 2007.

[15] T.-F. Yen, V. Heorhiadi, A. Oprea, M. K. Reiter, and A. Juels, "An epidemiological study of malware encounters in a large enterprise," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2014, pp. 1117–1130.

[16] V. Kandylas and A. Dasdan, "The utility of tweeted urls for web search," in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 1127–1128.

[17] D. Antoniades, I. Polakis, G. Kontaxis, E. Athanasopoulos, S. Ioannidis, E. P. Markatos, and T. Karagiannis, "we. b: The web of short urls," in *Proceedings of the 20th international conference on World Wide Web*. ACM, 2011, pp. 715–724.

[18] C. Grier, K. Thomas, V. Paxson, and M. Zhang, "@ spam: the underground on 140 characters or less," in *Proceedings of the 17th ACM conference on Computer and communications security*. ACM, 2010, pp. 27–37.

[19] S. Chhabra, A. Aggarwal, F. Benevenuto, and P. Kumaraguru, "Phish social: the phishing landscape through short urls," in *Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference*. ACM, 2011, pp. 92–101.

[20] F. Klien and M. Strohmaier, "Short links under attack: geographical analysis of spam in a url shortener network," in *Proceedings of the 23rd ACM conference on Hypertext and social media*. ACM, 2012, pp. 83–88.

[21] F. Maggi, A. Frossi, S. Zanero, G. Stringhini, B. Stone-Gross, C. Kruegel, and G. Vigna, "Two years of short urls internet measurement: security threats and countermeasures," in *proceedings of the 22nd international conference on World Wide Web*. ACM, 2013, pp. 861–872.

[22] N. Gupta, A. Aggarwal, and P. Kumaraguru, "bit. ly/malicious: Deep dive into short url based e-crime detection," in *Electronic Crime Research (eCrime), 2014 APWG Symposium on*. IEEE, 2014, pp. 14–24.

[23] R. K. Nepali and Y. Wang, "You look suspicious!!: Leveraging visible attributes to classify malicious short urls on twitter," in *System Sciences (HICSS), 2016 49th Hawaii International Conference on*. IEEE, 2016, pp. 2648–2655.

[24] S. Yoon, J. Park, C. Choi, and S. Kim, "Shrt: New method of url shortening including relative word of target url," *The Journal of Korean Institute of Communications and Information Sciences*, vol. 38, no. 6, pp. 473–484, 2013.

[25] APWG, "Workshop agenda 2017," https://www.apwg.eu/apwg-events/ecrime2017eu/agenda.

[26] L. Dugan, "Bit.ly Pro Is Now Free!" http://www.adweek.com/digital/bit-ly-pro-is-now-free-get-your-custom-short-domain-today/, Jun 2011.

[27] Bitly Blog, "How do I set up a Branded Short Domain (BSD)." https://support.bitly.com/hc/en-us/articles/230898888-How-do-I-set-up-a-Branded-Short-Domain-BSD.

[28] APWG, "APWG Phishing Attack Trends Reports," https://www.antiphishing.org/resources/apwg-reports/.

[29] McAfee, "McAfee Labs Threat Report," https://www.mcafee.com/ca/mcafee-labs.aspx.

[30] IBM, "IBM X-Force Threat Intelligence Index Report," https://www.ibm.com/security/data-breach/threat-intelligence.

[31] Veriform, "VERIFROM - Keep Your Brand Safe," https://www.verifrom.com/.

[32] R. Dhamija, J. D. Tygar, and M. Hearst, "Why phishing works," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 2006, pp. 581–590.

[33] M. Alsharnouby, F. Alaca, and S. Chiasson, "Why phishing still works: User strategies for combating phishing attacks," *International Journal of Human-Computer Studies*, vol. 82, pp. 69–82, 2015.

[34] Laendercode, "List: The two-letter country code / country abbreviation," https://laendercode.net/en/2-letter-list.html.