An Analysis of Phishing Blacklists: Google Safe Browsing, OpenPhish, and PhishTank

Simon Bell
Royal Holloway, University of London
simon.bell.2014@rhul.ac.uk

Peter Komisarczuk
Royal Holloway, University of London
peter.komisarczuk@rhul.ac.uk

ABSTRACT

Blacklists play a vital role in protecting internet users against phishing attacks. The effectiveness of blacklists depends on their size, scope, update speed and frequency, and accuracy - among other characteristics. In this paper we present a measurement study that analyses 3 key phishing blacklists: Google Safe Browsing (GSB), OpenPhish (OP), and PhishTank (PT). We investigate the uptake, dropout, typical lifetimes, and overlap of URLs in these blacklists.

During our 75-day measurement period we observe that GSB contains, on average, 1.6 million URLs, compared to 12,433 in PT and 3,861 in OP. We see that OP removes a significant proportion of its URLs after 5 and 7 days, with none remaining after 21 days - potentially limiting the blacklist's effectiveness. We observe fewer URLs residing in all 3 blacklists as time-since-blacklisted increases – suggesting that phishing URLs are often short-lived. None of the 3 blacklists enforce a one-time-only URL policy - therefore protecting users against reoffending phishing websites. Across all 3 blacklists, we detect a significant number of URLs that reappear within 1 day of removal – perhaps suggesting premature removal or re-emerging threats. Finally, we discover 11,603 unique URLs residing in both PT and OP – a 12% overlap. Despite its smaller average size, OP detected over 90% of these overlapping URLs before PT did.

CCS CONCEPTS

• **Security and privacy** → **Malware and its mitigation**; **Phishing**; • **General and reference** → *Empirical studies*; *Measurement*;

KEYWORDS

Security, Phishing, Blacklists, Measurement Study

ACM Reference format:
Simon Bell and Peter Komisarczuk. 2020. An Analysis of Phishing Blacklists: Google Safe Browsing, OpenPhish, and PhishTank. In *Proceedings of the Australasian Computer Science Week Multiconference, Melbourne, VIC, Australia, February 4–6, 2020 (ACSW 2020),* 11 pages.
https://doi.org/10.1145/3373017.3373020

ACSW 2020, February 4–6, 2020, Melbourne, VIC, Australia

ACM ISBN 978-1-4503-7697-6/20/02…$15.00
https://doi.org/10.1145/3373017.3373020

1 INTRODUCTION

Originating in the 1990s on AOL before expanding to other platforms such as email, SMS, and social media: phishing attacks lure their victims into revealing sensitive information such as passwords and credit card numbers by spoofing legitimate organisations. Phishing campaigns are a dangerous threat in the cyber world and the number of these attacks continues to grow: over 180,000 unique phishing websites were detected by the Anti-Phishing Working Group (APWG) in Q1 2019 [1], up from 138,328 in Q4 2018, and 151,014 in Q3 2018.

Phishing blacklists are a popular defence strategy that aim to protect people from phishing attacks. These blacklists typically contain known phishing URLs, providing an access control list which is used to prevent users from visiting these dangerous websites. 3 popular phishing blacklists widely used today are: Google Safe Browsing (GSB) [10], PhishTank (PT) [28], and OpenPhish (OP) [25]. These 3 blacklists are used by the web browsers Chrome, Safari, Firefox, Opera, email provider Yahoo! Mail, antivirus providers McAfee, Kaspersky, Virus Total and Strong Arm, and online reputation and internet safety service web browser plugin Web Of Trust.

For these phishing blacklists to be effective they need to be updated quickly and regularly to protect users from emerging phishing attacks. URLs should be removed from a blacklist when their website is no longer a threat – so as not to impact website visitors once it is safe – but also added again if that same website becomes a threat in the future. The number of URLs contained in a blacklist can contribute to its effectiveness; a small set of niche blacklisted phishing URLs will not provide a user with full protection compared to a large and comprehensive blacklist with a wide net. It is also important to understand the inner-workings of these blacklists as this can help determine how effective they will be at protecting people against phishing attacks.

In this paper we study 3 blacklists: GSB, PT, and OP, to determine uptake, dropout, typical lifetimes, and any overlap of URLs in these blacklists. Over a 75-day measurement period we regularly retrieve the latest copy of the 3 blacklists and store timestamps for when URLs are added and removed from each blacklist. Using this data we can then calculate various differences between timestamps to carry out our measurements. We discover that, in total, 1,731,452 URLs are added to GSB; 52,234 to OP; and 48,473 to PT. Throughout our measurement study the average number of URLs contained within each blacklist is: 1,581,351 for GSB; 3,861 for OP; and 12,433 for PT. This shows that GSB is by far the largest blacklist in our study. We also see 17 times more URLs added to GSB than PT and OP combined. The sheer volume of URLs in the GSB blacklist compared to PT and OP combined will likely make GSB a more effective weapon to protect users against phishing attacks.

By measuring URL durations in blacklists we discover that the OP blacklist removes a significant volume of URLs from its dataset after a duration of 5 and 7 days; no URLs remain in OP for more than 21 days. This potentially limits OP's effectiveness at protecting users from phishing attacks. We see that, across all 3 blacklists, as time increases, fewer URLs remain in the blacklists. This is because, once blacklisted, phishing URLs are often short-lived.

Through analysing URLs that reappear in blacklists we determine that none of the 3 blacklists enforce a one-time-only URL policy in their dataset; URLs reappear in the blacklists if they continue or re-emerge as a threat. This is good for users because it means they will be protected against reoffending phishing websites. We also show that large numbers of URLs reappear in all 3 blacklists within 1 day of removal – suggesting that these URLs were either removed too soon or that they came back online again.

As a result of comparing the PT and OP blacklists we discover that 11,603 unique URLs reside in both of these blacklists, which is 12% of the total number of URLs added to both blacklists. Despite its smaller average size – seen in the earlier measurement – OP detected over 90% of these overlapping URLs before PT did.

To the best of our knowledge, our paper is the first to analyse uptake, dropout, typical lifetimes, and any overlap of URLs in the blacklists: GSB, PT, and OP.

We organise the remainder of this paper as follows. Section 2 gives a definition of phishing, provides a background to the three main blacklists we use in this study, and also explores previous studies related to our work. Section 3 presents an overview of our infrastructure, explains our data collection process and methodology, and provides an overview of our experiments. Section 4 gives technical details of our infrastructure implementation. Section 5 presents our key measurement results and interpretations. Section 6 discusses our findings in this paper, followed by our conclusion in Section 7.

2 BACKGROUND AND RELATED WORK

2.1 Phishing

The term phishing originates in the 1990s on America Online (or AOL), the number one provider of internet access at the time. A group of hackers, called the warez community, created an algorithm to generate random credit card numbers which were used to create fraudulent AOL accounts. In 1995 AOL stopped the random credit card generators, which caused the warez group move onto alternative methods. Hackers would use the platform's instant messenger system to contact users whilst posing as AOL employees. These messages would lure victims into verifying their accounts or confirming billing information. On January 2, 1996, a Usenet group dedicated to AOL used the term "phishing" to describe what the warez group was doing and to warn the AOL community about the attacks. AOL eventually included the term "phishing" in its emails and messaging software to warn users about these attacks. Since the days of AOL phishing, attackers have moved to other platforms – such as email, SMS, and social media – to lure victims in where it is much harder for them to get caught.

An example of a modern phishing attack might be an email pretending to be from Facebook asking a user to verify their account. A URL contained within this email might direct the user to a spoof Facebook login page, whereby the user inputs their username and password. Phishing attacks are not limited to email – phishing tweets have been reported on the social media platform, Twitter [3]. In this case study, phishing attacks involved malicious users using promoting tweets to lure in their victims by promising verification status on the app. Victims would then hand-over their Twitter password, phone number, and credit card information to these criminals.

Over 180,000 unique phishing websites were detected by the Anti-Phishing Working Group (APWG) during the first quarter of 2019 [1]. This is an increase from the 138,328 seen in Q4 2018, and from the 151,014 seen in Q3 2018. The most-targetted industry sectors of the 2019 phishing emails were SaaS/Webmail (36%), Payment (27%), and Financial Institution (16%). 58% of phishing sites were using SSL certificates (HTTPS) in an attempt to further convince users that they are legitimate websites. These statistics show that the number of phishing attacks and phishing websites continues to grow, therefore research into phishing attacks plays an important role in understanding and reducing the impact of such threats.

2.2 Blacklists

A blacklist is defined as a set of elements to be blocked; an access control list. An example of a blacklist would be an email client that blocks known spam senders (e.g. spam@phishy.org). Any emails received from these senders would be marked as spam and possibly moved to an appropriate spam folder. Our study looks at phishing blacklists that are used to block access to URLs. We focus on 3 key phishing blacklists (GSB, OP, and PT) in this study.

Google Safe Browsing: launched in 2007, GSB is a URL blacklist that contains both malicious and phishing URLs and is used by the web browsers Google Chrome, Safari, Firefox, Opera, and Vivaldi to protect users from dangerous websites. We focus on GSB in our study because of its prominence in popular web browsers: already in 2012 GSB was protecting 600 million users from dangerous websites [34]. In 2015 GSB began using the term "Social Engineering" to categorise phishing websites which also encompass additional types of deceptive content. Google defines a social engineering web attack as occurring when either: "the content pretends to act, or looks and feels, like a trusted entity - like a bank or government" or "the content tries to trick you into doing something you would only do for a trusted entity - like sharing a password or calling tech support" [9]. During the week commencing 3rd September 2017 the total number of sites deemed dangerous by GSB was 573,433 phishing and 500,245 malicious. During that week GSB detected 24,756 new phishing sites and 6,312 new malware sites.

GSB provides two APIs for accessing its blacklist: Lookup and Update. The Lookup API provides a remote service whereby URLs to be checked are sent to Google's servers and a response is returned for each URL stating if the URL is in the blacklist. The Update API provides the user with a local copy of the blacklist; this local copy is stored as a database of SHA-256 URL hash prefixes, the majority of the hash prefixes being 4 bytes. To perform a URL blacklist lookup, the URL hash prefix is checked in the local database and, if there is a prefix match, then the full URL hash is retrieved from Google's servers to determine if there is a match on the full hash.

PhishTank: launched in October 2006, PT provides a community based phishing website reporting and verification system. Users of the website can submit URLs of suspected phishing websites; the PhishTank community then vote as to whether these URLs are phishing or not. PhishTank is used by the web browser Opera, online reputation and internet safety service web browser plugin Web Of Trust, email provider Yahoo! Mail, and antivirus providers McAfee and Kaspersky [27]. The PhishTank blacklist of approved phishing URLs can be downloaded as a JSON file.

OpenPhish: launched in 2014, OP is the result of a 3 year research project on phishing detection that uses autonomous algorithms to detect zero day phishing websites. Our study has access to the academic feed. OpenPhish is used by the antivirus companies Virus Total and Strong Arm. The OpenPhish blacklist can be downloaded as a JSON file.

2.3 Related Work

Existing literature has explored the effectiveness of malware blacklists [14, 15]. Research into phishing attacks has explored why they work [6], the effectiveness of toolbars in protecting users [36, 37], the effectiveness of web browser warnings [7], demographic analysis of phishing susceptibility and effectiveness of interventions [30], phishing website lifecycle [11], and a study to determine a baseline for phishing campaign success [12]. There are also various techniques to prevent phishing attacks including Dynamic Security Skins [5] and Trusted Devices [26] along with educational aspects of phishing training including PhishGuru [16] and the game Anti-Phishing Phil [31]; the effectiveness of these two educational approaches were analysed [17]. Previous studies have also developed techniques to detect phishing websites [2, 13, 29, 35, 38], and explored phishing blacklists, their overlap and effectiveness of takedown efforts by defenders [21–24].

Two key studies, carried out in 2007 [20] and 2009 [32] focused on phishing blacklists and how effective they are at protecting users from phishing email attacks, paying particular attention to the delay from an email containing a phishing URL being received to that URL appearing in a blacklist. A 2019 study [19] measured the characterisation of threat intelligence to understand how effective these methods are as defence mechanisms. Another 2019 study [39] analysed public blacklists to determine characterisation and evolution of reported activities over a 10-year duration.

Existing literature, such as [4, 11, 18], describe various blacklist datasets in terms of size, etc. However, existing studies do not specifically investigate and measure the characterisations of the blacklists themselves - since they are usually part of a broader set of research aims. In our study, we will analyse the phishing blacklists: Google Safe Browsing, OpenPhish, and PhishTank to explore uptake, dropout, typical lifetimes, and any overlap of URLs in these blacklists. Our purpose is to help the research community gain a more detailed understanding of these 3 blacklists. To the best of our knowledge, ours is the first study to analyse these 3 blacklists in such a way.

3 DESIGN

This section provides an overview of the design decisions we made and infrastructure we built to carry out our measurement studies. The design and methodologies described in this section are explained in more technical details in Section 4: Implementation.

3.1 Overview

The infrastructure we built to carry out our experiments consists of 3 systems to retrieve, store, and update 3 phishing blacklists (GSB, OP, and PT). These systems allow us to measure the number of URLs added to and removed from each blacklist and save these in our database. With this data we can analyse the number of times each URL is added to and removed from a blacklist and whether or not this happens multiple times. We can also measure the duration that each URLs remains in a blacklist for.

3.2 Data Collection

We retrieve a copy of each blacklist (GSB, PT, and OP) and store them in our database. The GSB blacklist encrypts each URL within its dataset as a SHA-256 hash prefix, URLs are categorised as either social engineering, malicious, unwanted software, or potentially harmful application. We study just the social engineering URLs in our study. We use GSB's API to to retrieve the latest copy of the blacklist. The blacklists PT and OP are provided as JSON files which we retrieve directly from their websites – the technical details of this can be seen in Section 4: Implementation.

3.3 Overview of Experiments

The key experiments we carry out in this paper are:

(1) Analysis of blacklists: PT, OP, and GSB, to determine number of URLs in each and how their sizes vary over time
(2) Measure how long URLs remain in each blacklist for
(3) Measure and analyse blacklisted URLs that are removed from then re-added to the same blacklist; timings between reappearance
(4) Comparison of URLs between blacklists and detection times of overlapping URLs

3.4 Methodology

Our core methodology involves regularly retrieving the latest copy of the 3 blacklists and storing each URL that is added or removed along with the timestamp of when this occurred. To calculate the total number of URLs in each blacklist we count the total number of entries in each JSON file for PT and OP and the total number of rows in the SQLite Database for GSB. We also remove duplicates to determine how many of these URLs – or hash prefixes in the case of GSB – are unique. To count the total number of domain names in each blacklist we extract the domain from all URLs in the blacklists then group and total these.

We use the aforementioned URL added/removed timestamps to calculate the duration each URL remained in the blacklists for. URLs which have not been removed from a blacklist currently still reside in that blacklist, therefore we use the current timestamp – at time of measurement – to calculate duration in blacklist. Using this, we show the total number of URLs in each blacklist that did not have

a removal timestamp and are therefore still in the blacklist. Since each blacklist only contains a list of URLs – not a list of URLs to be removed – we set the removal timestamp for a given URL to when our system sees that a previously added URL no longer appears in the blacklist.

Our local copy of GSB is stored as a database of SHA-256 URL hash prefixes; the majority of the hash prefixes are 4 bytes (2^{32} bits). Due to these short URL hash prefixes there is likely to be an increase in the number of collisions as the size of the dataset grows. The average number of collisions in k samples, each a random choice among n possible values is: $N(n,k) = k(k-1)/2n$. In our dataset of 1,731,452 SHA-256 URL hash prefixes there will be approximately 349 collisions. Therefore our GSB measurement calculations are accurate to within 0.02%.

To measure URL reappearance in blacklists, we analyse all URLs in each blacklist that have been added more than once. We calculate the duration of time a reappearing URL was included in the blacklist and the duration of time that URL was excluded from the blacklist – we repeat this for the number of times a URL was added to a blacklist. If, on the final inclusion timestamp for a reappearing URL in a blacklist, there is no removal timestamp then we assume the URL is still in the blacklist (as before).

4 IMPLEMENTATION

Our entire system is implemented on a virtual machine running Ubuntu operating system, 8 core CPU, 24 GB RAM. The measurement framework is written in Python programming language.

4.1 Blacklist update systems: GSB, PT, OP

We use 3 blacklists in our system: GSB, OP, and PT. To implement our GSB lookup system, the library *gglsbl* [8] is used. This library allows our system to fetch the latest GSB hash prefixes and also perform lookups against the database. The library uses the SQLite [33] database for storing GSB data. The library contains the method *update_hash_prefix_cache()* which is used to update the URL hash prefix database. This method is called every 5 minutes.

Both the PT and OP blacklist datasets are download as JSON files from their websites. The URL entries from these files are then extracted and saved into our local MySQL database. Metadata stored along with URLs includes discovery timestamps from the blacklists and timestamps for when URLs were added to our database. Both datasets are downloaded every hour and new entries saved in the local database. URL lookups for these two blacklists are completed by importing all URLs from both local blacklist databases and storing them in a Python *dictionary* in order to perform faster lookups, as per our GSB lookup implementation.

5 RESULTS

5.1 Overview of Blacklists

This section analyses the number of URLs added to and removed from each of the 3 blacklist: PT, OP, and GSB, during our 75-day measurement study from March to June 2019. Table 1 provides an overview of the 3 blacklists, showing total number of unique URLs added, removed, not removed, and added and removed once, in each blacklist during our measurement experiments from March to June 2019. The number of URLs not removed shows how many URLs were added to each blacklist but which were not removed during our measurement - therefore we conclude that these URLs remain in the blacklist at time of measurement. These results show that GSB is a considerably larger blacklist; with 33 times as many URLs added to GSB compared to OP and over 17 times as many URLs added to GSB compared to PT and OP combined. The increased size of GSB, compared to PT and OP, suggests that it may detect more URLs and therefore be more effective at detecting phishing websites compared to PT and OP.

	PT		OP		GSB
	URLs	Domains	URLs	Domains	URLs*
Added	48,473	20,458	52,234	14,721	1,731,452
Removed	33,245	15,327	46,866	13,315	633,321
Not removed	15,228	6,729	5,368	1,774	1,098,131
Added and removed once	30,967	14,409	43,103	12,147	113,530

Table 1: Overview of blacklists: PT, OP, and GSB showing total number of unique URLs and domains added, removed, not removed, and remaining in each blacklist. Measured from March to June 2019. *Number of SHA-256 URL hash prefixes for GSB.

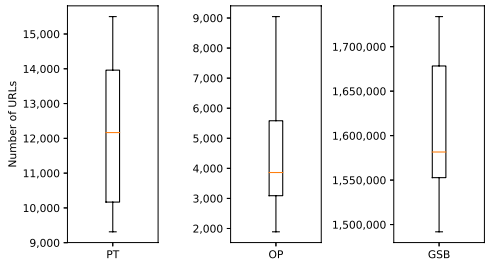


Figure 1: Box plots showing number of URLs in blacklists: PT, OP, GSB, at each update. Measured between March and June 2019.

When comparing the number of domains to the number of URLs for PT and OP, in Table 1, we see that there are at least twice as many URLs compared to domains for each data set. This suggests that each domain has about 2-3 blacklisted URLs. However, the average number of URLs per domain name is 1. The 10 most frequent domain names added to PT consist of 4,566 URLs - 1% of the dataset. The 3 most frequent domain names in PT consist of 1,459; 829; and 388 URLs, respectively. The 10 most frequent domain names added to OP consist of 8,412 URLs - 16% of the dataset. The 3 most frequent domain names in OP consist of 3,487; 1,965; and 6,55 URLS, respectively. This shows that the majority of domain names appear in both the OP and PT blacklists only once but that a small number of domain names contain multiple different blacklsited URLs. The most frequent domain names in both blacklists appear significantly more times than any other domain in the dataset.

The 3 box plots in Figure 1 show the number of URLs in each of the 3 blacklists: PT, OP, and GSB, on each update. PT and OP were updated once per hour and GSB was updated every 5 minutes. All 3 blacklists were updated 24/7 throughout the measurement experiment which ran from March to June 2019. These box plots show that the number of URLs in the PT blacklist ranged from 9,313

Category	URLs
Threat Type	
Threat type unspecified	0
Malware	362,230
Social engineering	8,718,240
Unwanted software	599,546
Potentially harmful application	37,790
Platform Type	
Platform type unspecified	0
Windows	1,609,928
Linux	1,609,928
Android	20,451
OSX	1,609,928
iOS	37,790
Any platform	1,609,927
All platforms	1,609,927
Chrome	1,609,927
Threat Entry Type	
Threat entry type unspecified	0
URL	9,717,803
Executable	0

Table 2: Overview of GSB blacklist showing total number of unique SHA-256 URL hash prefixes in each category.

to 15,500 with a median of 12,433; the lower quartile was 10,174 and the upper quartile was 14,041. The OP blacklist saw a range of 1,889 to 9,047 URLs with a median of 3,861; the lower quartile was 3,096 and the upper quartile was 5,748. Finally, the number of URLs in GSB ranged from 1,491,850 to 1,733,813 with a median of 1,581,351; the lower quartiles was 1,551,780 and the upper quartile was 1,677,262. These figures show that the number of URLs in all 3 blacklists stayed within reasonably consistent ranges throughout the measurement study. The range in number of URLs in these blacklists varied by 6,187 in PT; 7,158 in OP; and 241,963 in GSB. This range difference compared to the median number of URLs for each blacklist is PT: 50%, OP: 185%, and GSB: 15%. This shows that the range of URLs in the OP blacklist, during our measurement study, was considerably greater then the average number of URLs we saw in the blacklist. This is likely due to large numbers of URLs that are removed from the OP blacklist, therefore keeping its average size down, and may suggest that URL durations in OP are relatively short. In GSB, the range in number of URLs was relatively small compared to the average number of URLs - suggesting that URLs may remain in the blacklist for some time.

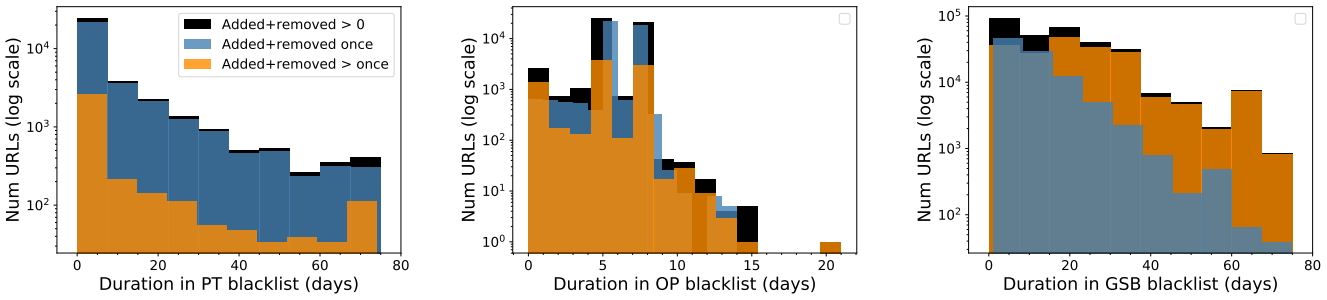
Overall, the findings in Table 1 and Figure 1 show that GSB is by far the largest of the 3 backlists, in terms of number of URLs added, and also sees the greatest number of URLs that remain in the blacklist throughout our measurement study. We see that the median number of URLs contained within GSB, throughout our measurement period, is just over 1.5 million. In comparison, the median number of URLs in PT was 12,433 and in OP was 3,861 - both over 99% less than GSB's median. Also, 97% fewer URLs were added to both PT and OP than GSB - this further illustrates the scale of GSB and how many URLs are added. Interestingly, during our study, the OP blacklist saw 3,761 more URLs added to it than PT. However, the median number of URLs residing in OP was 8,572 less than PT's median - a 69% decrease. This shows that even though 3,761 more URLs were added to OP during our study, the average number of URLs in OP remains low - possibly due to more frequent cleansing of the OP dataset. Due to the higher number of URLs residing in GSB compared to OP and PT, it is likely that GSB would catch a greater number of phishing URLs when deployed to check a random feed of URLs. Therefore GSB may be more effective than PT and OP at protecting users from phishing attacks due to its greater size.

Category	URLs
Malware	362,230
Windows	57,223
Linux	57,223
Android	0
OSX	57,223
iOS	18,895
Any platform	57,222
All platforms	57,222
Chrome	57,222
Social engineering	8,718,240
Windows	1,453,040
Linux	1,453,040
Android	0
OSX	1,453,040
iOS	0
Any platform	1,453,040
All platforms	1,453,040
Chrome	1,453,040
Unwanted software	599,546
Windows	99,665
Linux	99,665
Android	1,556
OSX	99,665
iOS	0
Any platform	99,665
All platforms	99,665
Chrome	99,665
Potentially harmful application	37,790
Windows	0
Linux	0
Android	18,895
OSX	0
iOS	18,895
Any platform	0
All platforms	0
Chrome	0

Table 3: Overview of GSB blacklist showing total number of unique SHA-256 URL hash prefixes in categories: *Threat Type* and *Platform Type*, combined.

5.1.1 GSB categories. This section analyses the total number of URLs in each category of the GSB blacklist. Table 2 provides an overview of the GSB blacklist, showing total number of unique SHA-256 URL hash prefixes in each category. The three main categories within GSB are: ***threat type***, ***platform type***, and ***threat entry type***. The subcategories of ***threat type*** are: *threat type unspecified*, *malware*, *social engineering*, *unwanted software*, and *potentially harmful software*. The subcategories of ***platform type*** are: *Windows*, *Linux*, *Android*, *OSX*, *iOS*, *any platform*, *all platforms*, and *Chrome*. The subcategories of ***threat entry type*** are: *threat entry type unspecified*, *URL*, and *executable*. In Table 3, categories ***threat type*** and ***platform type*** are combined to show the total number of URLs in both.

Tables 2 and 3 show that, of the 4 main threat type categories, *social engineering* contains the greatest number of total URLs at 8,718,240. Of this total, 1,453,040 URLs are unique; this number remains consistent across all platform types within the *social engineering* category. This is because phishing attacks are not software or platform specific; they rely on human presence for the attack be effective. There are 599,546 total URLs categorised as *unwanted software*, of which 99,665 unique; URLs are consistently shared



(a) PT (b) OP (c) GSB

Figure 2: Histograms of URL durations (days) in blacklists: PT, OP, GSB, for URLs that are added to and removed from each blacklist at least once, only once, and greater than once. Logarithmic y-axes. Measured between March and June 2019.

across all platforms except *Android* which sees 1,556 unique URLs. In the *Malware* threat type category: there are 362,230 total URLs, of which 57,223 unique URLs are on *Windows*, *Linux*, and *OSX* platforms, while *iOS* sees 18,895 unique URLs and *Android* sees 0 URLs. Finally, the *potentially harmful application* category has 37,790 total URLs, of which 18,895 unique URLs are categorised to the *Android* and *iOS* platform, while the remaining platforms see 0 URLs. These figures show that GSB contains more social engineering URLs in its blacklist than other threat type. Suggesting that GSB may be more effective at detecting phishing URLs than other types of threats listed within its categories.

Key Findings: The GSB blacklist contained an average of 1,581,351 URLs, compared to 12,433 in PT and 3,861 in OP. We see 17 times more URLs added to GSB than PT and OP combined. Social engineering URLs make up the bulk of all URLs in GSB. This makes GSB the largest phishing blacklist in our study; suggesting that GSB should detect a greater number of URLs – therefore making it a more effective blacklist than PT and OP. The average number of URLs in the OP blacklist is 93% less than than the total number of added URLs to OP – suggesting that OP enforces strict limits on how long URLs remain in its dataset for.

5.2 URL durations in blacklists

In this section we analyse how long URLs remain in each of the 3 blacklists (PT, OP, GSB) for. Figures 2a to 2c are histograms showing URL durations in the 3 blacklists, in days, for URLs which were added to and removed from each blacklist at least once, only once, and greater than once. These results were measured between March and June 2019. The y-axes of these histograms are shown on a logarithmic scale to make the results clearer despite a wide variance in range.

In Figure 2a we see that the most frequent duration for URLs in the PT blacklist, at just over 10,100 URLs, is between 0 and 5 days. Further analysis of this data revealed a frequency of 14,000 URLs with a duration between 0 and 1 day. We see that about 5,000 URLs remain in the blacklist for between 5 and 15 days. The histogram displays a skewed right pattern; we see the frequency of URLs decrease as the duration in the blacklist increases. There is an increase in URL frequency to the right of the graph at 65+ days – this suggests that PT may possibly cleanse its database at this time therefore resulting in a large numbers of URLs being purged. These durations show that a lot of phishing websites that appear in PT are removed within 24 hours. This may be because these websites are taken offline soon after appearing in the blacklist and therefore no longer pose a threat to users. We see that some URLs remained in the blacklist for the entire duration of our experiment which may suggest that phishing websites stay in the PT blacklist while they continue to pose an active threat to users. However, the increase in number of URLs removed at a duration of 65+ days in the blacklist is a concern **if** these websites are still actively serving phishing content at time of removal. An analysis of the contents of these websites may provide a clearer picture of why these URLs are removed at 65+ days. We also see that a greater number of URLs are added to the PT blacklist only once. The median duration for URLs in the PT blacklist is 2 days.

The histogram in Figure 2b shows URL durations in the OP blacklist. The multimodal pattern in this histogram reveals a spike at the 5 and 7 day durations along with a maximum duration of 21 days. For URLs that are added and removed only once, we see 25,578 URLs with a duration of 5 days and 24,790 URLs with a duration of 7 days – all other duration days see less than 650 URLs each. There are significantly fewer URLs in the blacklist after the 7 day duration and just over 10 URLs between the 10 and 14 day durations. This suggests that the OP blacklist may carry out cleansing of its dataset on URLs which have been in the blacklist for 5 days and 7 days along with a maximum duration of 21 days. It may be that websites which have been taken offline are checked and removed after being in the blacklist for 5 days and that this check is carried out again after 7 days. No URLs remain in OP after 21 days – which may reduce the effectiveness of OP if used to protect users from phishing attacks. Since any phishing websites that stays online for over 21 days may no longer appear in the blacklist, a user may believe that these websites are safe - when they are not. This poses serious security issues about using the OP blacklist to protect people from phishing attacks. Compared to PT, OP has significantly more URLs that are added to the blacklist more than once. The median duration in the OP blacklist is 5 days.

Figure 2c shows a histogram of URL durations in the GSB blacklist. We use GSB's Update API to retrieve the latest copy of the blacklist for these experiments. URLs are encrypted as SHA-256 hash prefixes in the local database therefore there are a number of hash collisions in our results. This happens when 2 different URLs share the same hash prefix and would appear in our results as an inaccurate URL duration. For example we saw a number of URLs in our dataset which had negative durations in GSB. This is because the removal timestamp of URL hash prefix has matched a different URL's hash prefix – for a URL that was removed before the original URL was added – therefore showing as a negative duration in the blacklist. We filter our results to only show URLs with positive durations, however, a small number of the results shown are likely to still contain collisions.

Figure 2c shows a skewed right pattern, with the greatest frequency of URLs having the shortest duration in GSB. As duration increases, the frequency of URLs decreases. This is likely because a lot of URLs are taken offline soon after they appear in a blacklist. A possible reason for this is that the URL's hosting provider becomes aware of the blacklisted URL on their server and therefore terminates the related account. Another reason is that blacklisted websites are likely to see a reduction in visitor traffic, due to visitors being unable to access the blacklisted site, therefore attackers may quickly move on and set-up a new website. There is a slight increase in frequency of URLs at the 53 to 60 day duration period, for URLs added once, and 60 to 70 day period for URLs added more than once. This is possibly due to GSB cleansing the dataset at this duration for each URL although we continue to see URLs remain in the blacklist for longer than 60 days - albeit less frequently. As with the PT blacklist, there is no apparent limit on the duration of which URLs remain in GSB other than the potential 53 to 70 day duration cleanse. The median duration in the GSB blacklist is 10 days. A greater quantity of URLs were added to GSB more than once compared to just once. Overall, we see that there is a steady decrease in the frequency of URLs as their durations in the blacklist increase – which is to be expected as blacklisted websites are often taken offline shortly afterwards.

Key Findings: Across all 3 blacklists: as time increases, fewer URLs remain in the blacklists. This is because, once blacklisted, phishing URLs are often short-lived. The OP blacklist limits the majority of URLs in its dataset to a duration of either 5 or 7 days; no URLs remained in OP for more than 21 days – therefore potentially limiting the blacklists effectiveness.

5.3 URL reappearance in blacklists

In this section we investigate URLs that are re-added to a blacklist, at a later date, after having previously been removed. This may happen if an attack website is deemed to be safe and is therefore removed from a blacklist - but then later becomes a threat again so is re-added to the blacklist.

Table 4 shows, for each of the 3 blacklists, the number of times each URL was added. This clearly shows how many URLs were added to a each blacklist only once, and how many URLs were added more than once. We see that, for PT, over 95% of all URLs are added to the blacklist only once, just over 3% are added only twice, and 0.01% are added only 3 times. No URLs were added to PT more than 6 times during our measurement study. Similarly, in the OP blacklist, we see 93% of all added URLs are only added once, and 5% are added only twice. We see an increase in the number of URLs added between 3 and 6 times in OP compared to PT; no URLs were added to OP more than 8 times. For the GSB blacklist, we see that the highest frequency – 30% of all added URLs – were only added once. Interestingly, we see a significant drop in the number URLs that were added to the blacklist twice - just 2% of all URLs added to GSB. The number of URLs added 3 times to GSB increases to 24% of all URLs. The reason for this dip in number of URLs added to GSB between 1 and 3 times may be that if a URL is added to the blacklist twice then there is a significantly higher chance that it will continue to reappear. Hence URLs appearing 3 to 8 times are seen more frequent than appearing just twice.

	Num URLs		
Num times added to blacklist	**PhishTank**	**OpenPhish**	**GSB**
1	47,127	49,128	620,089
2	1,571	2,604	43,524
3	636	726	488,474
4	9	334	256,836
5	1	148	367,686
6	1	24	159,992
7	0	1	64,382
8	0	1	56,166
9	0	0	1,447
10	0	0	69

Table 4: Overview of blacklists: PT, OP, and GSB, showing number of times each URL was added. Measured between March and June 2019.

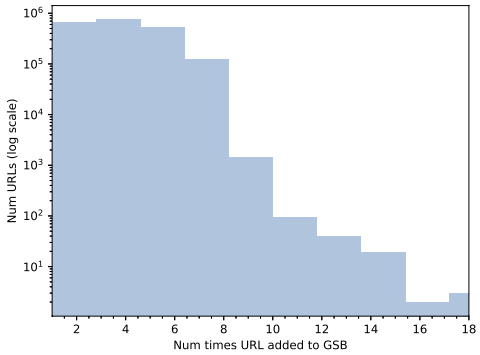
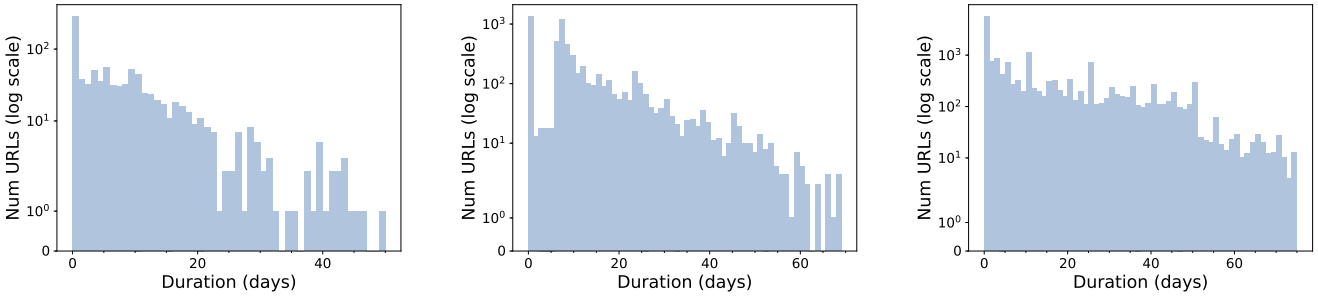


Figure 3: Histogram of number of times each URL was added to GSB blacklist. Logarithmic y-axis. Measured between March and June 2019.

The GSB blacklist is much larger then PT and OP, as a result, we see a significant increase in the number of times URLs were re-added to the to blacklist. To help visualise all of this data, the frequency of URLs that are re-added to the GSB blacklist is represented as a histogram, seen in Figure 3. This histogram shows us that the highest frequency of URLs were added to the blacklist between 1 and 6 times. The maximum number of times URLs were added to the blacklist was 18 and we see over 10 URLs were added to the GSB



(a) PT (b) OP (c) GSB

Figure 4: Histograms of URL durations (days) between URLs being removed and re-added in blacklists: PT, OP, GSB. Logarithmic y-axes. Measured between March and June 2019.

blacklist between 15 and 18 times. Although there will be some URL hash prefix collision within these results, we can still see that GSB allows websites to be re-added to its blacklist multiple times. This may be due to GSB frequently monitoring previously blacklisted websites – that have since been removed from the backlist – to determine if they reoffend. When such websites are found to be hosting malicious content again then they might be re-added to the blacklist

Overall, we see that, for all 3 blacklists, once a URL has been removed it can reappear again at a later date. This shows that none of the blacklists enforce a one-time-only URL policy in their dataset and that all 3 blacklists will re-add URLs if they continue or re-emerge as a threat. This is good for users because they will be protected against reoffending phishing websites.

To further explore URLs that are re-added to blacklists, we calculate the duration between removal and reappearance timestamps for all URLs that are re-added to the 3 blacklists. These durations are shown as histograms in Figures 4a to 4c. We use smaller bin widths in these histograms to produce finer granularity results. Although these reductions in bin widths produce multimodal graphs, we still see a general skewed right multimodal pattern in all 3 of these histograms. This shows that, in all 3 blacklists, fewer URLs reappear as the time since they were removed increases. This is understandable because you would expect the majority of phishing URLs to be taken offline as the duration of time since they were added to a blacklist increases.

We see delays between URL removal and reappearance in the PT blacklist in Figure 4a. In this histogram, 285 URLs reappear in the PT blacklist within 1 day; this is the most frequent reappearance delay representing 12% of all reappearing URLs in the blacklist. In comparison, just 38 URLs reappear in the PT blacklist 1 day after being removed. We still see over 20 URLs re-added to PT after more than 30 days since the URLs were originally removed. No URLs reappeared in PT after 50 days. An interesting example of the make-up of a phishing URL in this dataset is:

http://www.facebook.com.https.s1.[redacted].com

The domain name *[redacted].com* includes the subdomains *www*, *facebook*, *com*, *https* and *s1*. The results of combining these subdomains is that, to a user, this gives the appearance that the URL leads to *facebook.com* - illustrating a very common masquerading technique used by phishers. We have redacted the actual phishing domain from this example URL for privacy and security. This URL was re-added to PT 46 days after removal.

Figure 4b shows delays between URL removal and reappearances in the OP blacklist. 779 URLs are re-added to the blacklist within 1 day of removal and 564 URLs are added 1 day after removal. There is a noticeable drop in the number of URLs re-added to OP between 2 and 5 days after removal. The number of URLs then increases again from 6 days after removal. This relates to the pattern seen in Figure 2b (results section 5.2: *URL durations in blacklists*), where there is a peak in number of URLs that remain in OP for 5 and 7 days. This may suggest that OP does not allow certain URLs to reappear in its dataset, within a certain time period, if they have been previously cleansed.

Delays between URL removal and reappearance in the GSB blacklist are shown in Figure 4c. We see that over 3,200 URLs are re-added to the blacklist within 1 day, and 1,000 URLs reappear 1 day after being removed. Interestingly, there is a peak of over 500 URLs that reappear in the blacklist 26 days after removal. This may be where one specific campaign – which had previously been neutralised – later became a threat again and therefore reappeared in the blacklist.

Key Findings: Overall, we see that none of the 3 blacklists enforce a one-time-only URL policy in their dataset therefore all 3 blacklists re-add URLs if they continue or re-emerge as a threat. This is good for users because they will be protected against reoffending phishing websites. We also see that a large number of URLs reappear in the blacklists within 1 day of removal – suggesting that these URLs were either removed too soon or that they came back online again.

5.4 Blacklist Overlap

In this section we explore how many URLs reside in both the PT and OP blacklists. We do not analyse URLs that also reside in GSB because URLs are encrypted in the GSB blacklist. In total 11,603

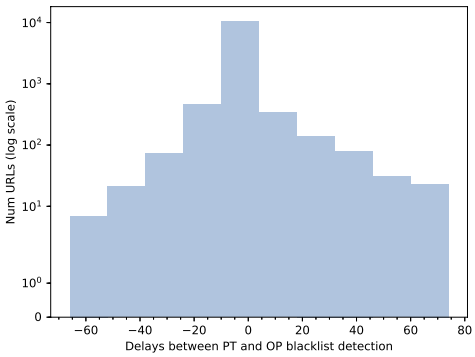


Figure 5: Histogram of delays (in days) between PT and OP detecting URLs. Positive values indicate PT detected URLs before OP, negative values indicate OP detected URLs before PT. Logarithmic y-axis. Measured between March and June 2019.

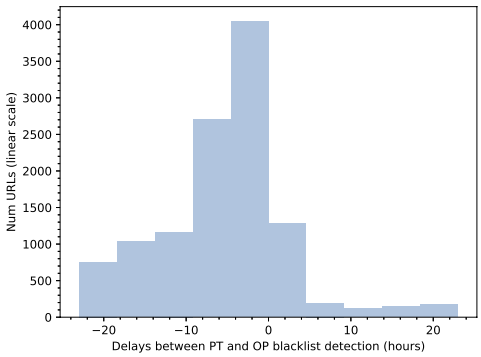


Figure 6: Histogram of delays (in hours) between PT and OP detecting URLs - limited to first 24 hours. Positive values indicate PT detected URLs before OP, negative values indicate OP detected URLs before PT. Linear y-axis. Measured between March and June 2019.

unique URLs – consisting of 6,079 unique domain names – appeared in both the PT and OP blacklists during our measurement study carried out between March and June 2019. The 10 most frequent domain names, appearing in bot PT and OP, consist of 998 URLs; less than 1% of the dataset.

Figure 5 is a histogram showing the difference (in days) between PT and OP first detection times for all URLs residing in both blacklists. Positive values indicate that PT detected a URL before OP, negative values indicate OP detected a URL before PT. For example: if a phishing URL appears in PT on April 1 and then in OP on April 30 then the difference in detection times between PT and OP is 30 days. If a phishing URL appears in OP on April 1 and then in PT on April 30 then the difference in detection times between PT and OP is -30 days (i.e., OP detected the URL before PT). The histogram shows that 807 URLs were first detected by PT and that 9,990 URLs were first detected by OP. Both blacklists detect 814 URLs within 1 day. These results show that OP detected 92% more URLs before PT did – suggesting that OP detects phishing URLs more quickly than OP. However, there are lead and lag times of over 60 days for both blacklists – meaning that both blacklists took at least 2 months to detect some URLs that had already been detected by the other blacklist. Overall, PT saw the greatest number of URLs with delays of over 60 days. Our experiments ran for just over 70 days which defines the upper delay limit for this study.

Figure 6 shows the difference (in hours) between PT and OP first detection times for all URLs residing in both blacklists. This histogram shows the first 24 hours of lead and lag times for both PT and OP and is represented on a linear scale for clarity. We see that, in the first 24 hours, 894 URLs were first detected by PT and 9,697 URLs were first detected by OP. Both blacklists detect 1,020 URLs within 1 hour. We cannot increase out measurement granularity any further than 1 hour because both blacklists are updated once per hour. In these results we see that OP detected 91% more URLs before PT did – again, illustrating OP's faster detection times. We see lead and lag times up to the maximum duration, 24 hours, for both blacklists – i.e., both blacklists see up to a 24-hour delay. However, significantly more URLs were first detected by the OP blacklist.

The reason OP detects over 90% of URLs before PT is likely because the PT blacklist is a community based network that relies on people submitting potential phishing URLs. Members of the community then vote whether these submitted potential phishing URLs are genuinely phishing or not. This process of submitting and then manually verifying phishing URLs takes time. Whereas the OP blacklist uses autonomous algorithms to detect zero day phishing websites. Automated detection is faster and does not require multiple people to vote – therefore detection times are reduced.

As previously mentioned, we only check URLs that reside in both PT and OP. In future work we would like to include GSB so we can compare the detection times between all 3 blacklists.

Key Findings: We see that 11,603 unique URLs reside in both the PT and OP blacklists and that OP was faster at detecting over 90% of these URLs. However, both blacklists have large lead and lag time delays between each other of over 65 days. OP's automated approach to phishing detection likely explains its faster detection rates whereas PT's manual, community-driven verification approach may explain its lag.

6 DISCUSSION

In this paper we carried out 4 key experiments:

(1) Analysis of blacklists: PT, OP, and GSB, to determine number of URLs in each and how their sizes vary over time
(2) Measure how long URLs remain in each blacklist for
(3) Measure and analyse blacklisted URLs that are removed from then re-added to the same blacklist; timings between reappearance
(4) Comparison of URLs between blacklists and detection times of overlapping URLs

Our results show that the average number of URLs contained within each of the 3 blacklist was: 1,581,351 for GSB; 12,433 for

PT; and 3,861 for OP. We see that GSB is by far the largest of the 3 blacklists we analysed. Along with social engineering, GSB also contains URLs categorised as malware, unwanted software, and potentially harmful application. These threat types are organised into different platform types (such as Linux, OSX, Windows etc) and the number of URLs within each threat type and platform type varies. However, the number of URLs under each platform type for social engineering URLs always remains the same – because these attacks rely on human presence and are not platform specific. GSB encrypts all URLs in its blacklist which makes it difficult for us to analyse individual URLs - unless a URL is already known to us.

Across all 3 blacklists, as time increases, fewer URLs remain in the blacklists. This is because, once blacklisted, phishing URLs are often short-lived. We discovered that the OP blacklist removes a significant amount of URLs from its dataset after 5 and 7 days; no URLs remained in OP for more than 21 days. This may potentially limit the effectiveness of OP as a blacklist because users may no longer be prevented from visiting an active phishing website once it has been in the blacklist for over 21 days.

Our results show that none of the 3 blacklists in our study enforce a one-time-only URL policy in their datasets. Therefore all 3 blacklists re-add any URLs to their dataset that become a threat again. This is good for users because they will be protected against reoffending phishing websites. We also see that a large number of URLs reappear in the blacklists within 1 day of removal – suggesting that these URLs were either removed too soon or that they came back online again. Without knowing the response status of these websites we do not know specifically why these websites reappeared in the blacklists. In future work we plan to analyse the status and contents of blacklisted websites.

Our results show that 11,603 unique URLs reside in both the PT and OP blacklists. We also see that OP was faster at detecting over 90% of URLs that eventually resided in both blacklists. However, both blacklists have lead and lag time delays of over 65 days. This may limit the effectiveness if just one of these two blacklists is used. OP deploys an automated approach to phishing detection and this likely explains the faster detection rates seen in our results. Conversely, PT employs a manual, community-driven verification voting system to confirm phishing URLs submitted to its dataset and this may explain its lag behind OP. In future work, we would like to also include GSB in our blacklist overlap analysis to compare GSB's detection times again OP and PT.

6.1 Limitations

All URLs in GSB are encrypted, which means all of our measurements that determine blacklist add and remove timestamps – based on URL matching – will contain a small number of hash collisions. Although our results contain some noise, we still see patterns in the data which are significant beyond the percentage of hash collisions.

In section 5.4: Blacklist Overlap, we do not include the GSB blacklist in our analysis. This is because URLs are encrypted in the GSB dataset. In future work we plan to design an experiment around the GSB API limitations in order to check OP and PT URLs in the GSB blacklist.

Although we show that the OP blacklist detected over 90% of URLs before PT, we did not analyse false positives in either the PT or OP dataset. It may be that the OP blacklist contains false positives due to its automation whereas the PT blacklist contains less false positives because of its human-driven verification process.

In future work we plan to retrieve the status and contents of blacklisted URLs to improve our understanding of the impact of blacklisting on websites. This will allow us to determine if phishing websites are offline when removed from blacklists, among other research questions.

7 CONCLUSION

This measurement study paper analysed 3 key phishing blacklists: Google Safe Browsing (GSB), OpenPhish (OP), and PhishTank (PT). We investigated the uptake, dropout, typical lifetimes, and considered the overlap of URLs in these blacklists. During our 75-day measurement period we observed that GSB contained an average of 1,581,351 URLs, compared to 12,433 in PT and 3,861 in OP. GSB is seen as a ground truth with respect to blacklisting resources and we saw in this study 17 times more URLs added to GSB than PT and OP combined. The sheer volume of URLs in the GSB blacklist could make GSB an effective weapon in the protection of users against phishing attacks through URL blacklisting.

Our measurements revealed that the OP blacklist removed a significant volume of URLs from its dataset after a duration of 5 and 7 days; no URLs remained in OP for more than 21 days. Therefore potentially limiting OP's effectiveness at protecting users from phishing attacks. We saw that, across all 3 blacklists, as time increased, fewer URLs remained blacklisted – as phishing URLs are often short-lived. We determined that none of the 3 blacklists enforced a one-time-only URL policy in their dataset; URLs reappeared in the blacklists if they continued or re-emerged as a threat. This is good for users because they will be protected against reoffending phishing websites. We also showed that a significant number of URLs reappear in all 3 blacklists within 1 day of removal – suggesting that these URLs were either removed too soon or that they came back online. Finally, we compared the PT and OP blacklists and discovered that 11,603 unique URLs resided in both of these blacklists – a 12% overlap. Despite its smaller average size, OP detected over 90% of these overlapping URLs before PT did.

ACKNOWLEDGMENTS

Research supported by the UK EPSRC grant EP/K035584/1 (Centre for Doctoral Training in Cyber Security). We would like to thank the anonymous reviewers for their insightful comments and observations.

REFERENCES

[1] APWG, G. A., and Manning, R. APWG Phishing Activity Trends Report, 1st Quarter 2019, 2019.

[2] Basnet, R. B., Sung, A. H., and Liu, Q. Feature selection for improved phishing detection. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems* (2012), Springer, pp. 252–261.

[3] Bonnington, C. Twitter is promoting a 'get verified' phishing scam. https://www.dailydot.com/debug/twitter-promoted-phishing-site/, 2018.

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

[5] Dhamija, R., and Tygar, J. D. The battle against phishing: Dynamic security skins. In *Proceedings of the 2005 symposium on Usable privacy and security* (2005), ACM, pp. 77–88.

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

[7] Egelman, S., Cranor, L. F., and Hong, J. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2008), ACM, pp. 1065–1074.

[8] Filipovich, A. gglsbl. https://github.com/afilipovich/gglsbl/, 2014.

[9] Google. Safe Browsing protection from even more deceptive attacks. https://security.googleblog.com/2015/11/safe-browsing-protection-from-even-more.html, 2015.

[10] Google. Safe Browsing. https://safebrowsing.google.com/, 2018.

[11] Han, X., Kheir, N., and Balzarotti, D. Phisheye: Live monitoring of sandboxed phishing kits. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (2016), ACM, pp. 1402–1413.

[12] Jagatic, T. N., Johnson, N. A., Jakobsson, M., and Menczer, F. Social phishing. *Communications of the ACM 50*, 10 (2007), 94–100.

[13] Khonji, M., Iraqi, Y., and Jones, A. Phishing detection: a literature survey. *IEEE Communications Surveys & Tutorials 15*, 4 (2013), 2091–2121.

[14] Kührer, M., and Holz, T. An empirical analysis of malware blacklists. *PIK-Praxis der Informationsverarbeitung und Kommunikation 35*, 1 (2012), 11–16.

[15] Kührer, M., Rossow, C., and Holz, T. Paint it black: Evaluating the effectiveness of malware blacklists. In *International Workshop on Recent Advances in Intrusion Detection* (2014), Springer, pp. 1–21.

[16] Kumaraguru, P. *Phishguru: a system for educating users about semantic attacks*. Carnegie Mellon University, 2009.

[17] Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., and Hong, J. Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology (TOIT) 10*, 2 (2010), 7.

[18] Li, F., Ho, G., Kuan, E., Niu, Y., Ballard, L., Thomas, K., Bursztein, E., and Paxson, V. Remedying web hijacking: Notification effectiveness and webmaster comprehension. In *Proceedings of the 25th International Conference on World Wide Web* (2016), International World Wide Web Conferences Steering Committee, pp. 1009–1019.

[19] Li, V. G., Dunn, M., Pearce, P., McCoy, D., Voelker, G. M., and Savage, S. Reading the tea leaves: A comparative analysis of threat intelligence. In *28th USENIX Security Symposium (USENIX Security 19)* (Santa Clara, CA, Aug. 2019), USENIX Association, pp. 851–867.

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

[21] Moore, T., and Clayton, R. Examining the impact of website take-down on phishing. In *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit* (2007), ACM, pp. 1–13.

[22] Moore, T., and Clayton, R. The consequence of non-cooperation in the fight against phishing. In *2008 eCrime Researchers Summit* (2008), IEEE, pp. 1–14.

[23] Moore, T., and Clayton, R. Evaluating the wisdom of crowds in assessing phishing websites. In *International Conference on Financial Cryptography and Data Security* (2008), Springer, pp. 16–30.

[24] Moore, T., and Clayton, R. Evil searching: Compromise and recompromise of internet hosts for phishing. In *International Conference on Financial Cryptography and Data Security* (2009), Springer, pp. 256–272.

[25] OpenPhish. OpenPhish - Phishing Intelligence. https://openphish.com/, 2018.

[26] Parno, B., Kuo, C., and Perrig, A. Phoolproof phishing prevention. In *Financial Cryptography* (2006), vol. 4107, Springer, pp. 1–19.

[27] PhishTank. Friends of PhishTank. https://www.phishtank.com/friends.php, 2018.

[28] PhishTank. PhishTank | Join the fight against phishing. https://www.phishtank.com/, 2018.

[29] Prakash, P., Kumar, M., Kompella, R. R., and Gupta, M. Phishnet: predictive blacklisting to detect phishing attacks. In *2010 Proceedings IEEE INFOCOM* (2010), IEEE, pp. 1–5.

[30] Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., and Downs, J. Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2010), ACM, pp. 373–382.

[31] Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., and Nunge, E. Anti-phishing Phil: the design and evaluation of a game that teaches people not to fall for phish. In *Proceedings of the 3rd symposium on Usable privacy and security* (2007), ACM, pp. 88–99.

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

[33] SQLite. SQLite Home Page. https://www.sqlite.org/, 2018.

[34] WebProNews. Google Discusses Its Safe Browsing Record. https://www.webpronews.com/google-discusses-its-safe-browsing-record-2012-06/, 2012.

[35] Whittaker, C., Ryner, B., and Nazif, M. Large-scale automatic classification of phishing pages.

[36] Wu, M., Miller, R. C., and Garfinkel, S. L. Do security toolbars actually prevent phishing attacks? In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (2006), ACM, pp. 601–610.

[37] Zhang, Y., Egelman, S., Cranor, L., and Hong, J. Phinding phish: Evaluating anti-phishing tools. In *Tech Report: CMU-CyLab-06-018* (2006), ISOC.

[38] Zhang, Y., Hong, J. I., and Cranor, L. F. Cantina: a content-based approach to detecting phishing web sites. In *Proceedings of the 16th international conference on World Wide Web* (2007), ACM, pp. 639–648.

[39] Zhao, B. Z. H., Ikram, M., Asghar, H. J., Kaafar, M. A., Chaabane, A., and Thilakarathna, K. A decade of mal-activity reporting: A retrospective analysis of internet malicious activity blacklists.