IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, January 2022
Impact Factor: 6.252

An Practical Investigation of Phishing Blacklists

Vivek Prasad[1] and Rohit Vishwakarma[2]
Assistant Professor, BSC IT, Suman Education Society's LN College, Borivali East, Mumbai, India[1]
Student, BSC IT, Suman Education Society's LN College, Borivali East, Mumbai, India[2]

Abstract: *In this research, we investigate the efficacy of phishing blacklists. We utilised 191 fresh phishes that were less than 30 minutes old to do two tests on eight anti-phishing toolbars. We discovered that 63% of the phishing efforts in our sample lasted less than two hours. Blacklists were originally useless in safeguarding users, since most of them detected fewer than 20% of phish at hour zero. We also discovered that blacklists were updated at different rates and had varying coverage, with 47% - 83% of phish appearing on blacklists 12 hours after the original test. We discovered that two programmes that used heuristics in addition to blacklists captured substantially more phish from the start than those that merely used blacklists. However, phish spotted by heuristics took a long time to emerge on blacklists. Finally, we ran the toolbars against a database of 15,345 authentic websites. URLs for false positives and found no evidence of mislabelling for blacklists or heuristics. We show our findings and explain how anti-phishing solutions might be improved.*

Keywords: Phishing, Blacklists, Investigation, Evidence, research.

I. INTRODUCTION

Phishing is a common issue that affects both businesses and individualities. MessageLabs estimated in November 2007 that0.8 of emails passing through their system were spam. Each day, around3.3 billion phishing emails were transferred. According to Microsoft Research, phishing assaults affected0.4 of donors. The monthly cost of phishing to consumers and enterprises in the United States is estimated to be between$ 350 million and$ 2 billion Stakeholders have enforced their own countermeasures to limit phishing detriment. Internet service providers, correspondence service providers, cybersurfed makers, registers, and law enforcement are each involved. All of them play vital places. Web cybersurfed suppliers play an important part because of the CyberSource's strategic position and the attention of the cybersurfed assiduity. Web cybersurfs are in a crucial position where they may incontinently and efficiently advise consumers. likewise, the cybersurfed business is veritably consolidated, with two cybersurfs counting for 95 of the overall requests (27). The results used by these two cybersurfs cover the vast maturity of druggies from phishing. According to a recent laboratory disquisition, when Firefox 2 displayed phishing warnings, no druggies put sensitive information into phishing websites (10).

To effectively maximise their eventuality to guard druggies, cybersurfs' warnings must be accurate (many false cons) and timely. presently, the maturity of cybersurfs with erected- in phishing protection ornate-phishing cybersurfed toolbars calculate on blacklists of phish and, occasionally, algorithms to descry phishing websites. Blacklists are preferred over heuristics due to their minimum false cons, which may be related to toolbar suppliers' desire to help implicit action from mislabelling websites. In this exploration, we probe the efficacity of phishing blacklists. We ran two tests on eight phishing toolbars using 191 new phishes that were lower than 30 twinkles old. In our sample, 63 of the phishing sweats lasted less than two hours. Blacklists were firstly useless at securing druggies, since utmost of them detected smaller than 20 of phish at hour zero. We also discovered that blacklists were streamlined at different rates and with varying degrees of content, with 47- 83 of phish appearing on blacklists 12 hours after the original test. We discovered that two technologies that used heuristics to supplement blacklists captured vastly more phish from the launch than those that solely used blacklists. still, phish spotted by heuristics took a long time to crop on blacklists. Eventually, we examined the toolbars for false cons on a set of,345 genuine URLs and set up no cases of mislabelling for either blacklists or heuristics. To the stylish of our knowledge, this is the first composition to attempt to quantify the length of phishing juggernauts as well as the update pace and content of phishing blacklists. Grounded on these criteria, we punctuate possibilities for protectors and suggest results to strengthen phishing blacklists. The rest of the document

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, January 2022
Impact Factor: 6.252

is structured as follows Section 2 provides the environment and applicable work, Section 3 outlines the test design, and Section 4 summarises our findings. Section 5 goes over how phishing blacklists and toolbars may be enhanced.

II. BACKGROUND AND RELATED WORK

Detection and filtering efforts can be undertaken at the phishing e-mail and phishing website levels. To prevent potential victims from receiving phishing emails, Traditional spam-filtering approaches including Bayesian filters, blacklists, and rule-based ranking can be used. Recently, various phishing-specific filters have also been created [1, 11]. In addition to these attempts, certain systems for verifying the identities of email senders have been developed [9, 33]. Despite these hopeful attempts, many users remain vulnerable. Because filtering algorithms are poor, many phishing emails continue to get in users' inboxes. As a result, we must also make an effort to recognise phishing websites.

In general, research to detect phish at the website level is divided into two categories: heuristic techniques that employ HTML or content signatures to identify phish, and blacklist-based methods that use human-verified data. To prevent false positives, use phishing URLs. Our research on blacklist measurement helps us determine how successful blacklists are at filtering phish at the website level.

2.1 Heuristics Against Phishing

The majority of these heuristics for detecting phishing websites identify phish using HTML, website content, or URL signatures. To categorise new webpages, develop classification models on top of heuristics. Gareca et al., for example, established a collection of fine-grained heuristics just from phishing URLs [13]. Based on the page structure of phishing webpages, Ludl et al. determined a total of 18 attributes [21]. To identify phish, Zhang et al. suggested a content-based technique based on TF-IDF and six additional heuristics [39]. Pan et al. developed a technique for compiling a list of phishing webpage features by extracting chosen DOM attributes such as the page title, meta description field, and so on [29]. Finally, Xiang and Hong proposed a hybrid phish detection approach that includes an identity-based detection component as well as a keyword-retrieval component of detection [35]. True positive rates range between 85% and 95%, with false positive rates ranging between 0.43% and 12%.

The heuristics technique has advantages and disadvantages. Heuristics can identify attacks immediately after they are launched, eliminating the need to wait for blacklists to be updated. Attackers may, however, be able to tailor their attacks to circumvent heuristic detection. Furthermore, heuristic techniques may create false positives, misclassifying a valid site as phishing. Heuristics are used in phishing filters in several programmes, including Internet Explorer 7 and Symantec's Norton 360. Our study investigates the accuracy of these heuristics in terms of their capacity to identify phishing and reduce false positives. Furthermore, we investigate how anti-phishing technologies apply heuristics to supplement their blacklists.

2.2 Blacklists for Phishing

Another way used by web browsers to detect phish is to compare URLs to a blacklist of known phish. Other fields have long employed blacklist techniques. One of the most common spam filtering approaches has been the use of blacklists of known spammers. There are now over 20 commonly used spam blacklists in operation. These blacklists may include known spammers' IP addresses or domains, IP addresses of open proxies and relays, nation and ISP netblocks that convey spam, RFC violators, and virus and exploit attackers. Although a spam blacklist of known IP addresses or domain names can be used to prevent phishing emails from being delivered, it is often insufficient to prevent phishing websites from being delivered.

One reason for this is because certain phishing websites are hosted on compromised domains. As a result, blocking the entire domain due to a single phish on that site is not conceivable. In the phishing scenario, a blacklist of certain URLs is a preferable approach. It takes several steps to compile and distribute a blacklist. To begin, a blacklist provider engages into contracts with multiple data sources to analyse suspected phishing emails and URLs. Emails obtained through spam traps or discovered by spam filters, user complaints (e.g., Phish tank or APWG), or validated phish collated by third parties such as takedown vendors or financial institutions are examples of data sources. Additional verification processes may be required depending on the quality of these sources.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, January 2022
Impact Factor: 6.252

Human reviewers are frequently used for verification. As in the case of Phish tank, the reviewers might be a dedicated staff of specialists or volunteers. To limit false positives even further, several reviewers may be required to agree on a phish before it is placed to the blacklist. For example, phish tank requires four votes from users in order to qualify a URL as a phish. Once validated, the phish is put to the central blacklist. The blacklist is sometimes downloaded to local computers. In Firefox 3, for example, phish blacklists are downloaded to browsers every 30 minutes [32]. While this has the benefit of decreasing network requests, performance may degrade between blacklist updates. Several of these blacklists are utilised in integrated browser phishing prevention [4, 15, 25] as well as in web browser toolbars [6, 7, 28]. Despite the fact that blacklists have low false positive rates, they often need human interaction and verification, which may be time-consuming and prone to human mistake. Nonetheless, this is the most often utilised type of phishing protection. Our research looks on the speed with which blacklists are updated and their correctness.

III. RELATED WORK

Several academics have investigated the efficacy of phishing toolbars. In November 2006, Ludl et al utilised 10,000 phishing URLs from Phish tank to assess the efficacy of Google's blacklists. They discovered that the Google blacklist had more than 90% of the live phishing URLs, whereas Internet Explorer only had 67%. According to the authors, blacklist-based methods are "very successful in safeguarding users from phishing attempts." One weakness of this study is that the data feed's freshness was not stated. We address this flaw by employing a fresh phishing feed that is less than 30 minutes old, as well as an automated testbed that visits phishing websites nine times in 48 hours to analyse the coverage and update speed of blacklists. In a separate research, Zhang et al. examined the performance of ten prominent anti-phishing solutions using data from Phish tank and APWG in November 2006. Using a total of 100 URLs from each site, as well as 516 real URLs to test for false positives They discovered that just one technology could accurately detect more than 90% of phishing URLs on a continuous basis. but with a 42% false positive rate. Among the remaining instruments, only one accurately recognised more than 60% of the phishing URLs. Both sources are valid. This research suffered from the same flaw as the previous. It also included a tiny sample of false positive URLs in the study. We based our research on this configuration, but made the following changes: First, we utilised a new phish less source. less than 30 minutes. Second, we broaden the technique by examining phish detected by heuristics vs blacklists individually. Third, we ran phish nine times in 48 hours. to investigate the coverage and updating rate of blacklists; Finally, we tested for false positives with a considerably larger sample size.

Other studies have investigated the efficacy of spam blacklists [18, 30, 16]. Ramachandran et al., for example, tested the efficacy of eight spam blacklists in real time by evaluating a 17-month trail of spam messages gathered at a "spam trap" site [30]. Whenever a host spammed their domain, they checked to see if the host IP was listed in a set of DNSBLs in real time. They discovered that around 80% of the spam they received was listed in At least one of the eight blacklists was used, however even the most aggressive blacklist had a 50% false negative rate. In addition to the studies described above, a variety of industry initiatives were employed to assess the effectiveness of phishing toolbars.

IV. METHODOLOGY

This section describes our anti-phishing testbed, how we gathered phishing URLs for testing, and our assessment approach.

4.1 Testbed for Anti-Phishing

Yue et al. [39] created an anti-phishing testbed, which we employed. The testbed is built on a client-server model. It consists of a task manager and a group of workers, each of whom is in charge of assessing a particular tool. The task manager began the test by retrieving a list of probable phishing sites to test against. The task manager then distributed each URL to a group of workers, each of whom was running a different tool. We operated each worker on a virtual computer to decrease the number of machines required. Each worker downloaded the required web page, used a basic image-based comparison technique to determine whether or not the web page had been tagged as phishing, and provided that value to the task manager. The image-based comparison technique works as follows: each tool has many known states (for example, a red icon if it has identified a phishing site and a green symbol if it has not), and each tool may be configured to be in one of these states. In the web browser, navigate to a well-known location. We take

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, January 2022
Impact Factor: 6.252

screenshots of the tools and compare them to screenshots of the tools in each of their known states. The task manager calculated general data and consolidated all of the employees' results, including true positives, true negatives, false positives, false negatives, and sites that no longer exist.

4.2 Phishing Feed

The phishing URLs for this study were collected from the data repository of the University of Alabama (UAB) Phishing Team. As part of the UAB Spam Data Mine, UAB has partnerships with many providers who contribute their spam. One of the most significant sources is a spam-filtering organisation that offers services ranging from small businesses to Fortune 500 corporations. 500 businesses in more than 80 countries. This organisation examines approximately one billion emails every day, employing a combination of keyword searches and unique algorithms to detect possible phishing. They then extract the URLs from these emails and transmit them in batches to UAB every four minutes. UAB evaluated the URLs they received from the spam-filtering business manually to see if they were phishing URLs. If a URL was a phishing attempt that had not been reported to UAB was already placed on a list to be examined by the testbed. Every 20 minutes, UAB submitted this list to the testbed. 1 Within 10 minutes of receiving each batch of URLs, the testbed began testing them. UAB was able to mark each URL with the four-minute time periods in which it was seen since they got phish URLs every four minutes. As a result, they were able to identify the initial time segment in which a URL was viewed and subsequent time segments in which the same URL was reported. This method of storing phishing URLs helps us to identify the duration of each spam campaign — the time period during which phishers send out emails with the identical subject line. URL for a phishing scam. If the spam campaign only lasts one day, the success of anti-phishing techniques on future days is less essential than the effectiveness on day one. While some users will receive phishing emails days after they are sent, the majority of users will read them within a few hours. As a result, the most essential period to safeguard is while the spammer is still actively sending emails.

4.3 Evaluation Procedure

Tools tested: We tested eight anti-phishing toolbars that use various blacklists and heuristics. They are Microsoft Internet Explorer version 7 (7.0.5730.11), version 8 (8.0.6001. 18241), Firefox 2 (2.1.0.16), Mozilla Firefox 3 (3.0.1), Google Chrome (0.2.149.30), Netcraft toolbar (1.8.0), McAfee Sitead.

1URLs were sometimes randomised in an attempt to defeat precise matching. We do not take two into account. URLs are considered unique if the sole change between them is in the attribute component of the URLs. visor (2.8.255 free version), and Symantec Norton 360 (13.3.5). With the exception of Internet Explorer 7 and Symantec, all of these solutions rely only on blacklists. When a phish is spotted by heuristics rather than blacklist, the two toolbars that utilise heuristics to supplement their blacklists issue distinct warnings. We utilised the default settings for all tools except Firefox 2, which we used the "Ask Google" option to contact the central blacklist server every time instead of downloading phishing blacklists every 30 minutes. Configuration of the testbed: We set up four PCs with Intel Core 2 CPU 4300 @ 1.80 GHz processors. Each PC ran two instances of VMware, each with 720MB RAM and 8GB storage. hard disc. To prevent network delay, we operated the task manager and workers on the same system for each toolbar. We left every browser open for six to eight hours before each test to download blacklists, and we left the browser open for 10 minutes between each run of the test since some of the toolbars employ local blacklists. We picked an eight-hour interval since the relevant blacklists will dependably download during this time. As a result, we're looking at the best-case scenario for blacklist efficacy.

We did the test for two to three hours on October 2, 8, and 9, 2008, as well as on December 3, 4, 5, and 15, 2008. During this period, additional batches of distinct phish were created. Every 20 minutes, data is delivered to the testbed. The testbed started testing them 10 minutes after getting the phish, for a total lag period of about 30 minutes. Before taking the screenshot, each worker opened the required browser with toolbars for 30 seconds. We assessed the toolbars' performance at hour 0, 1, 2, 3, 4, 5, 12, 24, and 48 for each URL.

Every hour, we cleaned the browser cache. In October, we gathered and tested 90 URLs, and in December, we collected and tested 101 URLs. After compiling the data, we thoroughly inspected every website that the toolbars identified as credible. This step was required since certain hosting firms did not generate 404 errors when removing a phish. They

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, January 2022
Impact Factor: 6.252

instead replaced it with their top page. The toolbar will mark the webpage as valid in this situation, although it was not. The fraudulent website has been removed.

V. RESULTS

5.1 Duration of the Phishing Campaign

The length of a phishing campaign (LPC) is defined as the period elapsed between the first appearance of a phish in our source report and the last appearance of that phish in our source report. We got reports from our source every 4 minutes, as specified in Section 3.2. The LPC for 127 of the 191 phish we utilised to test phishing blacklists was less than 24 hours, suggesting that their respective phishing campaign lasted less than 24 hours. The LPC for 25 URLs ranged between 24 and 48 hours, while the LPC for the remaining URLs ranged between 3 and 23 days. When we looked more thoroughly at the first day's data, we discovered that 109 URLs were spammed in only two hours, accounting for 63% of the URLs in this dataset. We estimated the LPC for 5491 phish given from the same source and confirmed by UAB from February 17 to April 13, 2009, to validate our findings. Similarly, to our testbed dataset results, we discovered that 66% of these phish had an LPC of less than 24 hours, 14.5% had an LPC of 24 to 48 hours, and the remaining 19% of URLs had an LPC of 3 to 47 days.

We discovered that 44% of the URLs had LPCs of less than two hours. It is crucial to remember that the LPC does not always coincide to the period that a phishing site is operational. In reality, we discovered that the time it takes to take down websites is often far slower when compared to the length of a phishing campaign. By hour 2, 63% of the phishing efforts in our sample had been completed, but just 7.9% of those phishes had been removed. on average, 33% of the websites were taken down after 12 hours, about half were taken down after 24 hours, and 27.7% were still operational after 48 hours. Our LPC results reveal that our data is current and that current takedown attempts lag behind phishing operations. Ludl et latest.'s revealed that 64% of the phish were already down when they did their test [21], but just 2.1% of phish in our sample were already down in our original test.

5.2 Blacklist Security

We show the findings of two experiments conducted in October and December of 2008 (Figures 2 and 3). We discovered that blacklists were useless at first in protecting users, since most of them captured less than 20% of phish at hour zero. We also discovered that blacklists were in use. As of 12 hours after the initial test in October, 47% to 83% of phish appeared on blacklists, which varied in speed and coverage.

We discovered that the coverage rates of various blacklists were substantially connected. The same blacklists appear to be used by Firefox 2, 3, and Google Chrome. Internet Explorer 7 and 8 are also supported. Exchange a blacklist. We merged the data for tools that utilise the same blacklists in our research. In our October test, all of the blacklists originally contained less than 20% of the phish. Every hour, new phish appeared on the blacklists, implying that the blacklists were updated at least an hour. The Symantec blacklist is one significant enhancement. In hour 0, their blacklist captured the same amount of phish as the others, but in hour 1, it caught 73% of the phish, which was two to three times higher. than the others. Until 12 hours after the original test, this difference is likewise statistically significant. 3 One probable explanation is that Symantec employs heuristic findings to speed up blacklist updates [2]. The coverage of the Firefox and Netcraft blacklists is consistently strongly connected, as we discovered. Five hours after our initial test in October, 91% of the URLs on the Netcraft blacklist also appeared on the Firefox blacklist, and 95% of the URLs on the Firefox blacklist also appeared on the Netcraft blacklist. Except for our initial test in December, the two blacklists are constantly strongly connected every hour. This shows that the two blacklists share certain data sources or contain data that is similar. sources with comparable properties. Others were less connected, with phish being on the Internet Explorer blacklist just 45% of the time and 73% of the time on the Firefox blacklist, indicating that they utilise separate feeds with little overlap. Up to the first 5 hours, we discovered that the Firefox blacklist was more thorough than the IE blacklist, and the Symantec blacklists performed much better than the rest of the toolbars from hour 2 to 12. The differences were no longer statistically significant after 12 hours. Figure 2 depicts this conclusion in further detail. We saw similar changes in coverage for various toolbars in our December dataset. However, Firefox and Netcraft did significantly better than in October. The

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, January 2022
Impact Factor: 6.252

The Firefox blacklist originally comprised 40% of phish, and by hour 2, 97% of phish were already on the blacklist. One reason for this disparity might be because the two tools gained additional sources that were comparable to our feed during this time period. Finally, we found no statistically significant improvement in the other toolbars. Finally, we looked at phishes that the IE 8 and Firefox blacklists had missed five hours after our original test in October. At hour 5, the IE 8 blacklist missed 74 phishes, 73% of which targeted overseas financial institutions. The Firefox blacklist failed to detect 28 phishes, 64% of which targeted international financial institutions. However, due to our small sample size, we did not find a statistically significant difference. There was a considerable difference in how quickly phish targeting US institutions and international organisations were added to the blacklist. There were several significant discrepancies between the phish overlooked by the IE8 blacklist and those missed by Firefox. For example, IE8 missed 21 Abbey Bank phishes whereas Firefox missed only four.

5.3 Positive Errors

	Detected by blacklist at hour 0	Detected by heuristics	false posi-tives
IE7 - Oct 08	23%	41%	0.00%
Symantec - Oct 08	21%	73 %	0.00%
IE7 - Dec 08	15%	25%	0.00%
Symantec - Dec 08	14%	80%	0.00%

Table 1: Accuracy and false positives of heuristics

To test for false positives, we prepared a list of 15,345 valid URLs. The URLs were gathered from three different sources, which are shown below. Using Google's inurl tool, a total of 2,464 URLs were collected by picking the login pages of sites. We specifically searched Google for pages with one of the following login-related strings in the URL: login, logon, signing, sign on, login.asp. A script was used to visit each URL to see if it was active and if it included a submission form. These pages were chosen to test if technologies can distinguish between phishing sites and real sites that are frequently spoofing. Ludl et al. also utilised this method to collect their samples. Extraction of 1000 emails reported to APWG on August 20, 2008 yielded a total of 1015 URLs. We collected 1401 URLs from the 1000 emails we examined.

URLs known to be good against a whitelist were deleted. This left us with 1015 URLs containing a variety of phish, malware, and spam. Each of these URLs was personally reviewed and phishing URLs were eliminated, leaving 851 confirmed no phishing URLs. We did the false positive test within 24 hours after retrieval. The list was chosen because it represented a source of phishing feeds used by several blacklist providers, and so we expected more false positives than from other sources. Similarly, we collected 10,000 URLs by removing no phishing URLs from a list of spam/phishing/malware URLs supplied to UAB's spam data mine between December 1 and December 15, 2008. We evaluated these URLs within a week of receiving them. Once again, this represents a source of phishing feeds that blacklist vendors are likely to receive; as a result, we expect this source to have more false positives than other sources. We found no instances of valid login sites being mislabelled as phish. There was one occasion when a malware website was identified as a phish by the Firefox blacklist among the 1,012 URLs from APWG. Finally, we found no false positives among the 10,000 URLs from the UAB spam data mine. Our study evaluated an order of magnitude more authentic URLs for false positives than earlier studies, however our findings on false positives were the same: phishing False positives are almost non-existent on blacklists. Our findings differ from those of a 2007 HP study in which the author received the Google blacklist and evaluated each entry to determine if it was a false positive. According to this analysis, the Google blacklist contains 2.62% false positives. However, the mechanism for checking false positives is not adequately defined, and the report does not offer a list of false positives. In our false positives test, we personally confirmed each URL labelled as phish and double-checked it with one of the Internet's recognised phish archives. It's also conceivable that Google's methodology or sources for phishing URLs have changed since 2007. We would like to validate the Google blacklist in the future using the same way as in the HP research [24]. Google's blacklist, on the other hand, is no longer publicly accessible.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, January 2022
Impact Factor: 6.252

5.4 Heuristic Accuracy

Symantec's Norton 360 toolbar and Internet Explorer 7 both employ heuristics. We report on their performance in this area. We discovered that programmes that apply heuristics detect substantially more phish than those that merely use blacklists. Symantec's heuristics recognised 70% of phish at hour 0, whereas Internet Explorer 7's heuristics detected 41% of phish. This is two to three times the quantity of phishing detected by blacklists during that time period. Furthermore, none of the 15,345 URLs we evaluated produced false positives. We also discovered that IE 7 and Symantec employ heuristics in somewhat different ways. Both programmes provide a brief and mild warning for potential phishing recognised by heuristics. However, Symantec's toolbar added a feedback feature. When a user accesses a potentially malicious website that is recognised by heuristics but is not on the blacklist, the URL is transmitted to Symantec for human review [2]. In our test, 95% of the phishes discovered by Symantec heuristics were on the Symantec blacklist at hour one, but none of the phishes detected by IE7 heuristics were on the IE blacklist at hour one. At the user interface level, this feedback loop is critical. When heuristics identify a phish, toolbars present less severe, passive warnings to avoid potential responsibility. However, after the phish has been validated as a phishing site by a person, toolbars can totally block the web page's content (active warnings). According to a recent laboratory research [10], Users only pay attention to active phishing warnings and disregard passive ones.

VI. DISCUSSION

6.1 Limitations

Our research has a few drawbacks. First, because we acquired all of our URLs from a single anti-spam provider, the URLs we received may not be indicative of all phish. Second, all of the URLs were likely discovered by a spam provider and never reached users who were protected by that vendor. However, because not all users are protected by commercial spam filters, browsers must identify these phishing URLs as well. Second, these URLs were solely taken from emails and did not include additional attack routes like Internet messenger phishing.

6.2 Opportunities for Defenders

Defenders' window of opportunity is defined as the length of the phishing campaign plus the time gap between the time a user gets a phishing email and the time the phishing email is opened. Whenever the user opens the email. Users are protected if they do not get any phish or if the website is banned or taken down by the time, they click on a phish. Section 4.1 shows that 44% of phishing attacks lasted less than 2 hours. According to recent study, the period between when a user gets and opens a phishing email is less than two hours for a minuscule part of the Internet population. They discovered that two hours after the phishing emails were sent, at least half of those who would eventually click on the phishing link had already done so; after eight hours, virtually all (90%) of those who would click had already done so. Their research also shown that persons with technological abilities were just as susceptible as nontechnical folks to fall for phish. AOL recently surveyed 4,000 email users aged 13 and up on their email usage in a countrywide poll. According to the poll, 20% of respondents check their email more than 10 times per day, and 51% check it four or more times each day (up from 45% in 2007) [3]. Assuming that individuals check their emails at a consistent pace, 20% check their emails once every hour and a half, and 51% check their emails once every four hours According to these data, the important window of opportunity for protection is between the beginning of a phishing effort and 2 to 4 hours afterwards. Our findings have various implications for anti-phishing defences. First, anti-phishing efforts should be directed more on upstream safeguards, such as phishing prevention at the email gateway level This work should be directed at the browser level on updating the blacklist more rapidly or making greater use of heuristic detection. Second, additional to address the initial limited blacklist coverage problem, research and industry development initiatives to effectively educate users (e.g. [20, 34]) and build trusted user interfaces (e.g. [8, 36, 31, 37]) are required.

6.3 Enhancing blacklists

The first step in improving blacklists is to detect more phishing URLs early. Potential phishing URLs can be acquired from URLs extracted from spam/phishing filters at e-mail gateways, URLs derived from user reports of phishing emails or websites, and phishing websites recognised by toolbar heuristics, as illustrated in Figure6 (Figure6). Each of these sites provides unique coverage. We begin by discussing how to enhance each source. When it comes to phishing emails,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, January 2022
Impact Factor: 6.252

email gateway filters are the initial point of contact. Given the defenders' narrow window of opportunity, as mentioned in section 4.1, suppliers should concentrate their efforts here. Regular spam filters, on the other hand, are insufficient since they include a large amount of spam that would take significant human work to filter. We advocate utilising spam filters as the initial line of security, followed by algorithms created to detect phishing websites as a second layer, to increase phish detection at this level. Once a suspect URL has been identified by both sources, it should be removed. Human review has been requested. Because residential and corporate email accounts receive various email distributions, providers should gather URLs from a number of sources to ensure maximum coverage. Phishing emails and websites reported by users are likely to include phish that spam filters missed. As a result, user 4Assumes an eight-hour sleep period. Reports should be used to supplement data from email gateway spam filters. However, consumers may be unmotivated to report and check phishing. User incentives (e.g., points, awards) may be beneficial. Overcome this issue. Finally, we propose that browser anti-phishing programmes strengthen their blacklists using algorithms. This technology is similar to disease outbreak early warning systems. When a user hits a probable phish that is spotted by heuristics but is not on the blacklist, the tool may transmit the URL for human review and, once validated, adds the URL to the blacklist. This technique is likely to work since some individuals read their email significantly more frequently than others.

VII. CONCLUSION

The two systems that used heuristics to supplement blacklists captured considerably more phish at first than those that solely used blacklists. Given the situation, Because of the brief duration of phishing attacks, heuristics are quite useful. However, suppliers may be worried about the increased likelihood of false positives and potential responsibility for mislabelling websites when utilising heuristics. Associated Bank-Corp sued Earthlink in 2005 when the Earthlink anti-phishing software ScamBlocker disabled the bank's official page [5]. Earthlink was able to defend itself against the action by claiming that it was utilising a blacklist of phish given by a third party, and so it cannot be held accountable as a publisher when that information is incorrect under a clause in the Communication Decency Act. If, on the other hand, a toolbar utilises heuristics to detect and stop a phish that turns out to be a false positive, the toolbar seller may be held liable. As a "publisher" under the CDA, he is not immune.

In our testing, neither the blacklists nor the heuristics produced any false positives. However, suppliers are concerned about the possibility of false positives. To avoid liability, we propose that providers apply heuristics to detect phish and then have specialists check them. We also welcome additional debate regarding the risks of supplying phishing blacklists and heuristics. There has been no test case on this subject so yet. Lack of understanding on these issues may diminish suppliers' motivation to use heuristics even more. Major suppliers, such as Microsoft and Firefox, who provide security to the vast majority of users, do not suffer direct financial losses as a result of phishing. If, however, they use heuristics and are sued, they might face millions of dollars in damages and legal expenses.

REFERENCES

[1]. S. Abu-Nimeh, D. Nappa, X. Wang, and S. Nair. A comparison of machine learning techniques for phishing detection. In eCrime ’07: Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit, pages 60–69, New York, NY, USA, 2007. ACM.

[2]. Andy Patrizio. Symantec readies phishing protection software, august 7, 2006. visited jan 1, 2009. http://www.smallbusinesscomputing.com/news/ article.php/3624991.

[3]. AOL Press Release. Its 3 a.m. are you checking your email again? july 30, 2008. visited jan 1, 2009. http://corp.aol.com/press-releases/2008/07/ it-s-3-am-are-you-checking-your-email-again.

[4]. Apple Inc. . Visited jan 1, 2009. http: //www.apple.com/safari/features.html#security.

[5]. ASSOCIATED BANK-CORP v. EARTHLINK, INC. Memorandum and order, 05-c-0233-s. http://www.iplawobserver.com/cases/2005-09-14_ Associated_Banc_Corp_CDA_Section_230.pdf.