International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Analysis of Web Phishing Methods

Rishav Shaw[1], Utkarsh Opalkar[2], Nayan Mathur[3], R. Manjula[4]

[1, 2, 3]School of Computing Science and Engineering, Vellore Institute of Technology, Vellore

[2]Professor, School of Computing Science and Engineering, Vellore Institute of Technology, Vellore

Abstract: ***Phishing is considered as a form of web threat and is the act of impersonating a website of an honest enterprise aiming to obtain user's confidential credentials such as usernames, passwords and other private or sensitive information. The Internet today is a medium of our everyday social and financial activities. Internet is also a platform for large and small organizations to expand business via e-commerce. As a result, the number of customers who rely on the Internet to perform procurements is increasing dramatically. Unimaginably large amounts of money are transferred through the Internet every day. This amount of money is very tempting to fraudsters and hackers. Hence, Internet users may be vulnerable to different types of web threats, which may cause financial damages, identity theft, loss of private information, brand reputation damage and loss of customers' confidence in e-commerce and online banking. In this survey paper we shall discuss existing phishing techniques in detail. We shall also take a look at 2 important algorithms that deal with detecting and preventing phishing attacks along with their effectiveness and shortcomings in order to carry out a Comprehensive study. The motivation behind this paper is to contribute in enhancing Anti-Phishing methods and preventing a dominant type of crime in today's day and age – Cybercrime.***

Keywords: Internet, Web Phishing, CyberCrime, E-commerce, Identity theft.

1. Introduction

In 2016, a world without internet is unimaginable. Conducting almost all business and economic transactions online has become an integral part of daily life. Which opened up a new pathway for Cybercriminals. On a global scale, the annual cost lost in cybercrime is easily over $110 Billion. Web phishing is one of the most notorious cybercrimes and is responsible for more than 28% of the cyber-attacks causing e-commerce frauds every day. Sometimes the data obtained by phishing may be employed to cause social defamation, blackmailing and cyber bullying – all of which are illegal offences.

1.1 Phishing Techniques

- The most common phishing technique uses a URL which is masqueraded as if it originated from an authentic source that handles sensitive information such as usernames, passwords, bank account details or similar private data. When users submit this data through these fake URLs, the data is actually sent to the cybercriminals or hackers. With advancements in web technologies and high motivation to practice this form of cybercrime, more and more sophisticated phishing attacks are being carried out all over the world.
- Phishers usually send out emails or links of fake webpages to a huge number of people, of which some are aware of phishing and dodge the bullet but others are likely to fall victim to the trickery. The URLs redirect victims to fake websites or webpages which are designed to look almost exactly like the originals. Next, the victims are asked to submit or update their sensitive and private information which obviously reaches the phishers instead of the supposed authorities.
- Analyzing the URL via the browser or anti-virus and security software is the most direct way to tackle it. We discuss various methodologies used today and judge them critically and mention the shortcomings of major algorithms developed to handle web phishing.
- Till date, it has not been possible to devise a technique to accurately detect and get rid of these phishing attacks. Hence, there is a strong need to analyze and understand current phishing techniques in order to develop and implement newer and better Anti-Phishing techniques.

2. Related Works: Although

Phishing is a relatively new web-threat, it has a massive impact on the commercial and online transaction sectors. Presumably, phishing websites have high visual similarities to the legitimate ones in an attempt to defraud the honest people. Social engineering and technical tricks are commonly combined together in order to start a phishing attack. Typically, a phishing attack starts by sending an e-mail that seems authentic to potential victims urging them to update or validate their information by following a URL link within the e-mail. Predicting and stopping phishing at-tack is a critical step towards protecting online transactions. Several approaches were proposed to mitigate these attacks. Nonetheless, phishing websites are expected to be more sophisticated in the future. Therefore, a promising solution that must be improved constantly is needed to keep pace with this continuous evolution. Anti-phishing measures may take several forms including: legal, education and technical solutions. To date, there is no complete solution able to capture every phishing attack. The Internet community has put in a considerable amount of effort into defensive techniques against phishing. However, the problem is continuously evolving and ever more complicated deceptive methods to obtain sensitive information and perform e-crimes on the Internet are appearing. Anti-phishing tools, or sometimes so called fighting phishing tools, are employed to protect users from posting their information through a forged website. Recognizing phishing websites accurately and within a passable timescale as well as providing a good warning technique reflect how good an anti-phishing tool is. Designing a phishing websites has become much easier and much more sophisticated, and that was the motivation behind looking for an effective anti-phishing technique.

Mixed research methodology has been adopted in our study. Since some previous studies suggest applying protection mechanisms without offering clear experimental results. Hence, qualitative methodology is best fits such researches. On the other hand, some researches taking into consideration experimental analysis, data gathering techniques, testing measures and comparing results, thus it is worthy applying quantitative methodology in such cases.

This article is structured as follows: discusses what phishing is and how it started. , introduces different phishing techniques, describes the phishing websites life cycle, discusses how and why people fall prey to phishing, shows some phishing statistics, describes phishing countermeasures, introduces a detailed discussion of the up to date anti-phishing techniques, compares between human and automatic based protection. Finally, we summarize at the end.

3. Analysis of Phising Techniques

3.1 The Story of Phishing

Deceiving users into giving their passwords or other private information has a long tradition in the cybercrime community. In the early 1990s, with the growing popularity of the Internet, we have witnessed the birth of a new type of cybercrime; that is phishing. In 1987 a detailed description of phishing was introduced, and the first recorded attack was in 1995 [1].

In the early age of phishing; phishers mainly designed their attacks to deceive English-speaking users. Today, phishers broaden their attack to cover users and businesses all over the globe [2]. At the beginning, phishers acted individually or in small and simple groups.

Usually, phishing is accomplished through the practice of social engineering. An attacker may introduce himself as a humble and respectable person claiming to be new at the job, a helpdesk person or a researcher. An example of using social engineering is urgency; by asking the user to submit his information as soon as possible. Risk of terrible results if the user denies complying is another tactic used to start social phishing, for example warn the user that his account will be closed or the service will be terminated if he does not respond. However, some social engineering tactics promise big prizes by showing a message claiming that the user has won a big prize and to receive it he needs to submit his information. Nowadays, as monetary organizations have improved their online investments, the economic benefit of obtaining online account information has become much larger. Thus, phishing attacks became more proficient, planned and efficient.

Phishing is an alternate of the word “fishing” [3] and it refers to bait used by phishers who are waiting for the victims to be bitten [1]. Surveys commonly depict early phishers as mischief-makers aiming to collect information to make long-distance phone calls [4]; such attack was called “Phone Phreaking”. This name was behind the origin of the “ph” re-placement of the character ‘f’ in the word “fishing” [3].

Phishing websites are designed to give an impression that they came from a legitimate party with the aim to deceive users into divulging their personal information. The phishers may use this information for dishonest intentions, for instance money laundering or illegal online transactions. While those phishers focus on individual customers, the organizations that phishers are mimicking are also victims because their brand and reputation is compromised.

There are several definitions of the term “***Phishing*”.** To have a good understanding of phishing and their attacking strategies, several definitions will be discussed. Some definitions believe that phishing demands sociological skills in combination with technical skills. As in the definition from the “*Anti-Phishing Working Group*” [5]:**“*A criminal mechanism employing both social engineering and technical subterfuge to steal consumers’ personal identity data and financial ac-count credentials*”.**

Another definition comes from Ming and Chaebol [6]: **“*A phishing website is a style of offence that network fishermen tempt victim with pseudo website to surrender important information voluntarily*”.** A detailed description stated by

Kirda and Kruegel [7] defines phishing as: **“*creating a fake on-line company to impersonatea legitimate organization; and asking for personal information from unwary consumers depending on social skills and website deceiving methods to trick victims into dis-closure of their personal information which is usually used in an illegal transaction*”.** Some definitions assumed that the success of phishing websites depends on their ability to mimic a legitimate website, because most Internet-users, even those having a good expertise in Internet and information security, have a propensity to decide on a website’s validity based on its look-and-feel which might be orchestrated proficiently by phishers. An example of such definitions comes from James [1], who defines phishing as:

“***Attempts to masquerade as a trustworthy entityin an electronic communication to trick recipients’ into divulging sensitive information such as bank account numbers, passwords, and credit card details***”. Since phishing web-sites request victims to submit their credentials through a webpage, it is necessary to convince these victims that they are dealing with an honest entity, so that a good definition was introduced by Zhang et al. [8] where they define phishing as: Phishing website satisfies the following criteria:

- Showing a high visual similarity.
- Containing at least one login form.

We may outline all previous definitions in one sentence: **“*Phishing website is the practice of creating a copy of a legitimate website and use social skills to fool a victim into submitting his personal information*”.**

3.2 Phishing Techniques

Until recently, phishers relied heavily on spoofed emails to start phishing attacks by persuading the victims to reply with the desired information. These day’s social networking web-

sites are used to spread doubtful links to lure victims to visit phishing websites. A report published by Message Labs [9] estimated that one phishing email occurs every 325.2 emails sent through their system every day. Microsoft Research [10] revealed that 0.4% of email receivers' were persecuted by phishing emails in 2007. A report published by Symantec Corporation [11] substantiate that the amount of phishing web-sites that mimic social networking websites rose by 12% in 2012. If phishers were able to acquire users' social media lo-gin information, they can send out phishing emails to all their friends using the breached account. An email that appears to be originated from a well-known person seems much more trustworthy. Moreover, phishers may send out fake emails to your friends using your account telling them that you face an emergent situation. For example, **"*Help! I'm stuck overseas and mywallet has been stolen. Please send $200 as soon as possible*".** Nowadays, phishing websites have evolved rapidly, maybe at a faster pace than the counter measures. Compromised identifications and phishing toolkit are widely offered for sale on Internet black-markets at low prices [12]. These days, innovative phishing techniques are becoming more frequent, such as malware and ***Man-In-The-Middle attacks (MITM)*** [13].

Phishers use different tactics and strategies in designing phishing websites. These strategies can be categorized into three basic groups those are:

1. ***Mimicking attack:-*** In this attack phishers typically send an email to victims asking them to confirm, update or validate their credentials by clicking on a URL link within the email which will redirect them to a phony webpage. Phishers pay careful attention to designing emails that will be sent to the victims using the same logos of the original website, or sometimes using a fake HTTPS protocol. This type of attack undermines the customer confidence in electronic trading.
2. ***Forward attack:*** This attack starts once a victim clicks on the link shown within an email. He then redirected to a website asking him to submit his personal information. This information sent to a hostile server, and the victim is then forwarded to the real website using MITM technique.
3. ***Pop-up attack:*** Another method used by MITM technique is urging victims to submit their information by means of well-designed pop-up window. The phishers persuade the victims that submitting their information through a pop-up window is considered more secure.

In order to accomplish their job, phishers use a set of intelligent tricks to give the impression to the victims that they are dealing with a legitimate website. Some of these tricks include using IP address in URL, adding a prefix and suffix to a domain name, hiding the true URL shown in the browser address bar, using a fake padlock-icon on the URL address bar and pretending that the SSL is enabled. These tricks make it difficult for the naïve user to distinguish a phishing website from a legitimate one.

Overall, one principle if committed by organizations and customers will guarantee the security of their information; that is: **"*Organizations and consumers should be awareof phishing and anti-phishing methods and take safety measure*".** Theoretically, this principle iseasy, but in practice, it is very difficult to implement since there are new phishing techniques appearing constantly.

3.3 Phishing attack life cycle

To combat phishing, we need to thoroughly investigate the nuts and bolts of the phishing attack. Following, we will describe the phishing attack life cycle.

- ***Planning:*** Typically, phishers start planning fortheir attack by identifying their victims, the information to be achieved and which technique to use in the attack. The main aspect considered by the phishers to pick their targets is how to achieve the maximum profit at the lowest cost and least possible risk. A phisher might need to breach the employee list in an organization, the organization news from a social networking website or the organization calendar. Common social networking such as; email, Voice over IP (VoIP) and Instant Messaging (IM) are used to establish communication between the phisher and the potential victims.

A classic phishing attack consists of two components: a trustworthy-looking email and a fraudulent webpage. The phishing emails contents are commonly designed to con-fuse, upset or excite the recipient. A fraudulent webpage has the look-and-feel of a legitimate webpage that it impersonates, often having a similar logo to the legitimate company, layout, and other critical features.

A survey published in ACM magazine [14] showed that Internet users were 4.5 times more likely to be victims of phishing if they received an invitation to visit a fake URL link from a person they knew. That explains why criminals target social networking websites. Efforts made by webmail providers in filtering phishing emails will de-crease the extent of the problem and reduce the time needed to stop phishing attacks since they are the first point dealing with phishing emails. However, most web-mail providers focus on filtering spam emails and they would be very happy if their spam filter catches phishing emails but without adding any phishing filters that may consume their resources. The main difference between spam emails and phishing emails is that, spam emails are annoying emails sent to advertise goods and services that have not been requested by the user. On the other hand, phishing emails are sent to get your personal information, which will be used later in fraud activities. The authors in [15] recommend stopping phishing attacks at this stage. The authors suggested dividing the email into several parts such as; subject line; email attachments and the salutation line in the email body. Then extracting some structural features from these parts and making some calculation to produce the final decision on the email legitimacy.

- ***Collection:*** As soon as the victim takes anaction making him susceptible to an information theft, he is then urged to submit his credentials through a trustworthy-looking webpage. Normally, the fake website is hosted on a com-promised server, which has been exploited by the phisher for this purpose. A recent survey [16] revealed that 78% of the servers holding phishing websites are either hacked file transfer protocol (FTP) or comprised of software

application susceptibilities. Sometimes, the phishers may use the free cloud applications such as Google spreadsheets in order to host their fake websites [17]. Nobody is going to block "google.com" or even "spreadsheets.google.com", thus, not only naïve users will be deceived, but also ex-pert users are less likely to block this website. In general, to reduce the possibility of being caught, phishers will exploit servers that have weak security or process loopholes operating from countries which have insufficient law enforcement resources [18].

- Fraud: Finally, and once the phisher has achieved his goal, he then becomes involved in fraud by impersonating the victim. Sometimes, the information is sold on the Internet black-market.

The amounts of activities that take place within the first few hours of a phishing life cycle are the most important aspect of any attack.

Once the phishing website has been created and the phishing email has been sent to consumers, the anti-phishing tool should detect and stop the phishing website before the consumer submits his information as shown in Fig. 1. Seconds are important in this situation. Taking down the phishing website is the second line of defense. If we cannot stop the phishing attempt then the spoofed email could reach the victim's mailbox.

An example of such an attack strategy happened when eBay costumers received an email claiming to be from the real eBay company asking customers to update their credentials so as not to freeze their accounts [19]. The email contained a link that seemed to point to the real eBay's website. As soon as the user clicked on that link, he was then transferred to a webpage that asked for his credentials, including credit card number, expiry date and full name. The phishing website had been designed carefully in an attempt to convince the user that he was dealing with a legitimate website.

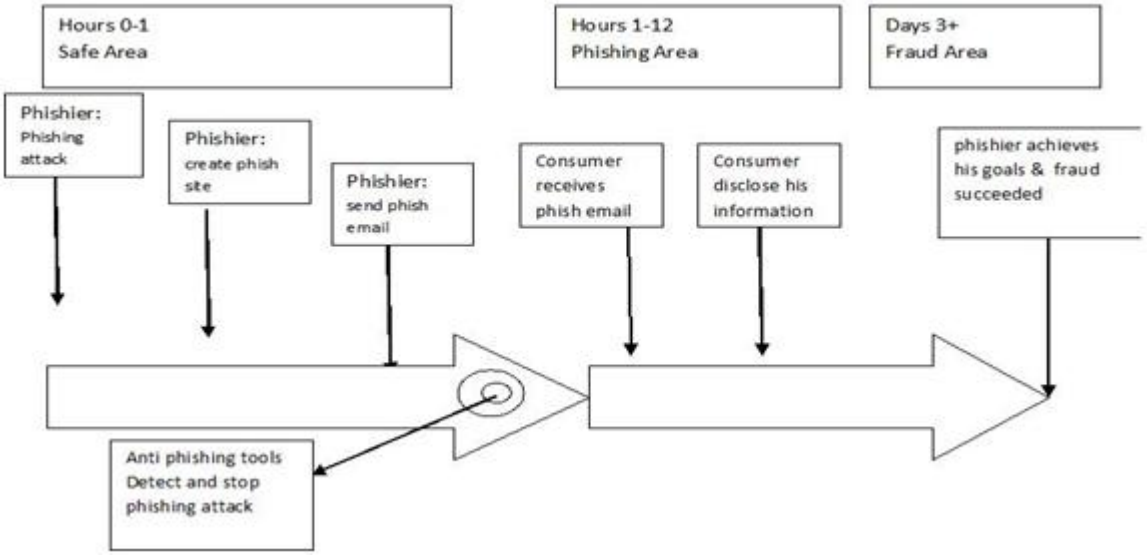
Some phishers stay up to date with the news and design their attacks in conjunction with specific approaching events or disasters. This is what happened during "*Hurricane Katrina*" [20]. Once the announcement of the hurricane was made, phishers initiated their attacks by registering domain-names which masqueraded as donation and victims-aid websites. Phishers sent fake emails masked as Katrina news updates with links directing users to fake websites hosted in the USA and Mexico.

3.4 Why fall prey to phishing?

Phishing is an example of a bigger category of web threats called semantic attacks. Instead of focusing on the technical vulnerabilities, semantic-attacks focus on how humans interact with computers or how they assign meanings to the message contents [21]. A white paper published by Trend Micro [20] which is a worldwide leader in cloud security shows that the victims need on average of about 600 h to resolve the issue of identity theft. Commonly, users have a tendency to trust email messages and websites based on phony clues that in fact offer superficial trust information.

Several researches [22–25,14 ,26–28] have revealed that users are susceptible to phishing for quite a lot of reasons among them:

1) Some people may lack essential knowledge of existing online threats.
2) Although some users may have a good understanding of what does computer viruses, hackers and fraud means, and how to protect themselves from these threats; they may not be familiar of what does phishing means. Therefore, they cannot generalize what they knew to unfamiliar threats.
3) Although some users are wary of falling prey to phishing, they have not developed good strategies for recognizing phishing attacks.
4) Users may focus on their main tasks, while paying attention to security clues is considered a secondary task.
5) Some users may ignore some essential security clues in the URL address bar such as the existence of HTTPS protocol, and as an alternative they used the website contents to decide whether the website is a phishing website or not.
6) Some users are unaware of what does SSL protocol and other security indicators mean.
7) Some users may not notice warning messages, while some other users may notice these messages but they expected that the warnings were invalid.
8) Internet users may lack how the organizations that offer online services are formally contacting their consumers in case of maintenance and information update issues.



3.5 Phishing statistics

A report disseminated by Anti-Phishing Working Group [5], which is a non-profit corporation established in 2003 focuses on reducing the frauds resulting from phishing, crime-ware and email deceiving, shows that 128,387 phishing websites were observed in the second quarter of 2014. This is the second highest number of phishing sites detected in a quarter, eclipsed only by the 164,032 seen in the first quarter of 2012.

The total number of URLs used to host phishing attack increased to more than 175,000 hosts in the second quarter of 2014, while that number was less than 165,000 in the first quarter of the same year. The most affected countries were China with 51% followed by Peru and Turkey with 44%. USA continued the top hosting country of phishing websites. The number of targeted brands decreased to 531 in the second quarter of 2014 after reaching 557 brands in the first quarter of the same year. The average number of phishing URLs per brand decreased to 134 URLs after reaching 147 in the first quarter of the same year. The ratio of IP address based phishing URLs increased in the second quarter to 2.4%. Over 20,000 unique phishing emails were sent monthly during that period. The most industrial sector targeted by phishers was the payment services with 40%, followed by the financial sector with 20%, while the attacks against auctions websites increased from 2.3% in the first quarter to 2.6% in the second quarter. Expert phishers have moved from traditional phishing to a new style of malware attacks in this quarter.

However, Ihab Shraim, the president and Chief Technology Officer (CTO) at MarkMonitor [29] said**, "*It is unlikely that traditional phishing willstop since the cost of producing a phishing attack is almost insignificant*"**. A survey disseminatedby Gartner Inc. [30] which is an advisory and research company, reveals that phishing websites continue to escalate and costs US financial sector an estimated $3.2 billion annually. The same survey estimated that 3.6 million victims fall in such attack. A poll of 2000 US adults carried out by Harris Poll [31] showed that 30% of those surveyed have limited their online transactions, and 24% have limited their e-banking transactions because of phishing.

4. Phishing countermeasures

4.1 Legal Solutions

Followed by many countries, the USA was the first to enact laws against phishing activities, and many phishers have been arrested and sued. Phishing has been added to the computer crime list for the first time on January 2004 by Federal Trade

Commission (FTC), which is a US government agency that aims to promote consumer protection. On May 10, 2006, the US president George W. Bush gave his orders to establish the President's Identity Theft Task Force [32] which aims to ensure that the efforts of federal authorities become more efficient and more effective in the area of identifying and pre-venting cybercrime attempts. On January 2005, the General Assembly of Virginia added phishing to its computer crimes act [33].

On March 2005, the Anti-Phishing Act was introduced in the US congress by Senator Patrick Leahy [34].

In 2006, the UK government strengthened its legal arsenal against fraud by prohibiting the development of phishing websites and enacted penalties of up to 10 years.

In 2005, the Australian government signed a partnership with Microsoft to teach the law enforcement officials how to combat different cybercrimes. Several prosecutions have been made as in 2006; a Florida man has been indicted with development of a phishing website that aimed to take advantage of the victims of Hurricane Katrina [35].

In 2004, Zachary Keith Hill pleads guilty in Texas Federal Court to law breaking related to phishing activity and was penalized to 46 months [36]. Although law enforcement officials successfully arrested, prosecuted and convicted phishers for the past few years [19,20], criminal act does a poor job of preventing phishing since it is hard to trace phishers. More-over, phishing attacks can be performed

quickly and later the phisher may disappear into cyberspace. Therefore, law enforcement authorities must behave quickly because on aver-age the phishing website lives for 54 h only [22].

4.2 Education

The key principle in combating phishing and information security threats is consumer's education. If Internet users could be convinced to inspect the security indicators within the website, then the problem would just go away. However, the most biggest advantage for phishers to successfully scam Internet users is that most Internet users lack basic knowledge of current online threats that may target them and how the online sites are formally contacting their consumers in case of maintenance and information update issues.

In general, although education is an effective counter-measure technique, eliminating phishing via education is a difficult and long-winded process and users have to dedicate a substantial amount of their time to studying the phenomenon. Moreover, phishers are becoming more skilled in mimicking legitimate websites, even to the extent of security experts being deceived.

4.3 Technical solution

Weaknesses that appeared when relying on previously mentioned solutions led to the emergence to innovative solutions. Several academic studies, commercial and non-commercial solutions are offered these days to handle phishing. Moreover, some non-profit organizations such as APWG, PhishTank and MillerSmiles provide forums of opinions as well as distribution of the best practices that can be organized against phishing. Furthermore, some security enterprises, for example, McAfee and Symantec offered several commercial anti-phishing solutions.

The success of anti-phishing techniques mainly depends on recognizing phishing websites accurately and within an acceptable timeframe. Although a wide variety of anti-phishing solutions are offered, most of these solutions were unable to make decisions perfectly on whether the website is phishing or not, causing the rise of false positive decisions. Even worse, a recent study [37] demonstrates that some security providers have fallen victims for phishing attacks.

Hereunder, we preview the most popular approaches in designing technical anti-phishing solutions:

- **Blacklist approach:** Where the requested URLis compared with a predefined phishing URLs. The downside of this approach is that the blacklists usually cannot cover all phishing websites since a newly created fraudulent website takes considerable time before it is added to the list. This gap in time between launching and adding the suspicious website to the list may be enough for the phishers to achieve their goals. Hence, the detection process should be extremely quick, usually once the phishing website is uploaded and before the user starts submitting his credentials.
- **Heuristic approach:** The second technique isknown as heuristic-based approaches, where several features are collected from the website to classify it as either phishing or legitimate. In contrast to the blacklist method, a heuristic based solution can recognize freshly created phishing websites in real time [38]. The effectiveness of the heuristic based methods, sometimes called features-based methods, depends on picking a set of discriminative features that could help in distinguishing the type of website [39]

5. Anti-phishing methodologies in literature

Detecting and preventing phishing websites is an essential step towards shielding users from posting their sensitive information online. Several approaches and comprehensive strategies have been suggested to tackle phishing. Anti-phishing methodologies can be grouped into five categories: blacklist and whitelist based approach, instantaneous based approach, decision supporting tools, community rating based approach, and intelligent heuristic based approaches.

Below, we shed the light on common anti-phishing techniques by evaluating a list of related works and substantiating the need for an automated technique, as oppose to human involvement when fighting against phishing.

5.1 Blacklist and whitelist based approach

A blacklist is a list of URLs thought to be malicious. Blacklist is collected through several methods, for instance heuristics from web crawlers, manual voting, and honeypots. Whenever a website is visited, the browser refers it to the blacklist to examine if the current visited URL is present within the list. If so, this indicates that it is a malicious website and as a consequence, the browser warns users not to submit any sensitive information. Blacklists can be saved either locally on the user's machine or on a server that is queried by the browser for every requested URL.

The main aspects of blacklists are quantity, quality and timing. Quantity refers to the amount of available phishing URLs within the list. On the other hand, quality can be measured in terms of erroneous listing and is commonly known as the false positive rate, which means classifying legitimate websites incorrectly as phishing. This has a negative influence on users as they lose confidence and trust in the blacklist for each false positive reading, thus potentially ignoring the correct warning signals. The third and most significant aspect is timing, which plays a key role to ensure the effectiveness of the blacklist since most phishing websites have a short life span.

If the process of updating the blacklist is slow, this will give website phishers the opportunity to carry out attacks without being added to blacklist. Blacklists are updated at various speeds, in a recent study [40] scholars estimated that approximately 47%–83% of phishing URLs are displayed on blacklists almost 12 h after they launched. The same study ascertained that zero hours defence delivered from most well-known blacklists-based toolbars claimed a TP rate ranges from 15%–40%. Therefore, it is necessary for an efficient blacklist to be updated instantly in order to keep users safe from being phished.

A survey published by APWG [41] found that 78% of phishing domains were hacked domains, and at the same time they were already serving legitimate websites. As a consequence, blacklisting those domains will, in-turn, add legitimate websites to the blacklist. Even if phishing websites are removed from the blacklisted domain, legitimate websites hosted in the same domain may be left on the blacklist for a long period of time, thus causing significant harm to the reputation of the legitimate website or organization.

Some blacklists such as Google's Blacklist needs on average seven hours to be updated [42]. A range of solutions has been deployed depending on the blacklist approach, one of which is Google Safe Browsing [43]. Another solution is Microsoft Ié anti-phishing protection [44]. Site Advisor [45] is a database-backed measure which is designed essentially to defend against malware-based threats such as Trojan horses and Spyware. Site Advisor comprises automated crawlers which browse websites then carry out tests and build threat assessments for every visited website. Regrettably, like other blacklists, Site Advisor cannot recognize newly created threats.

VeriSign [46] has a commercial phishing detection solution. VeriSign has a web-crawler which collects millions of websites to recognize "clones" in order to discover phishing webpages.

One potential drawback with crawling and black-list approaches might be that anti-phishing parties will al-ways race against attackers.

Netcraft [47] is a small software package that is activated every time a user browses the Internet as shown in Fig. 3. Netcraft relies on a blacklist which consists of fraudulent websites recognized by Netcraft and those URLs submitted by the users and verified by Netcraft. Netcraft displays the Anti-phishing methodologies in literature

Detecting and preventing phishing websites is an essential step towards shielding users from posting their sensitive information online. Several approaches and comprehensive strategies have been suggested to tackle phishing. Anti-phishing methodologies can be grouped into five categories: blacklist and whitelist based approach, instantaneous based approach, decision supporting tools, community rating based approach, and intelligent heuristic based approaches.

Below, we shed the light on common anti-phishing techniques by evaluating a list of related works and substantiating the need for an automated technique, as oppose to human involvement when fighting against phishing.

VeriSign [46] has a commercial phishing detection solution. VeriSign has a web-crawler which collects millions of websites to recognize "clones" in order to discover phishing webpages. One potential drawback with crawling and black-list approaches might be that anti-phishing parties will al-ways race against attackers.

Anti-Phishing Best Practice Checklist

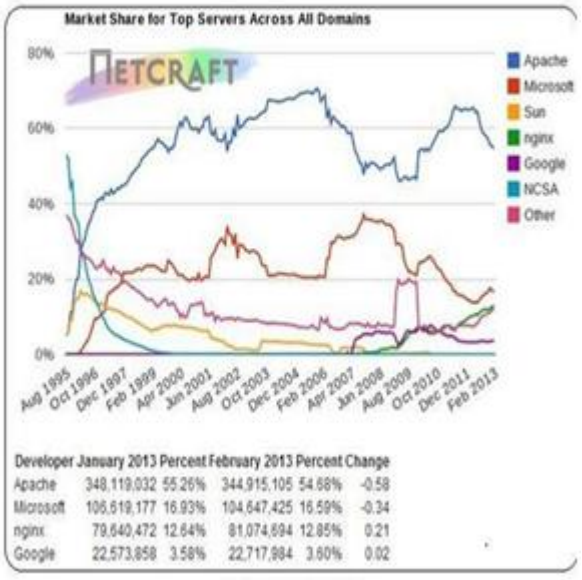
Table 1. Anti-Phishing Best Practice Checklist

BEST PRACTICE	BUSINESS	CONSUMER
Always install, update, and maintain firewall and intrusion detection software including those that provide malware security.	✓	✓
Use latest web browser version and install security patches when available.	✓	✓
Practice awareness when receiving emails asking for account details	✓	✓
Never email financial/personal information.	✓	✓
Only open email attachments from trusted parties.	✓	✓
Never click on links in suspicious emails.	✓	✓
Report suspicious emails to appropriate authority.	✓	✓
Monitor logs from firewalls, intrusion detection systems, DNS servers, proxy servers on a daily basis for a signs of infections.	✓	
Monitor outbounds SMTP connection attempts that do not originate from normal SMTP mail gateways.	✓	
Establish rigorous password policies for clients, servers, routers and enforce them.	✓	
Ensure that approved devices can connect to the organization's network.	✓	
Regularly read the latest news and info regarding phishing.	✓	✓

In terms of specific technologies, businesses and consumers alike should look for layered solutions that protect against both sending—that is, becoming an unwitting accomplice to propagating spam—and receiving phishing emails. From a business perspective especially, layered solutions should also offer content protection at the client side, or end points, and at the network gateway—as well as monitor network behavior. This ensures against "rogue" devices such as laptops and notebooks—which are not always under administrators' control and may not have adequate or updated threat protection installed—infecting the entire network. The following checklist can serve as a guideline in making technology-related decisions to combat phishing:

IV. NETCRAFT's WEB SERVER SURVEY 2013

Netcraft, an Internet services company that provides web hosting & web server analysis & has launched its 2013 web server survey after responses from over 630,790,500 web sites.



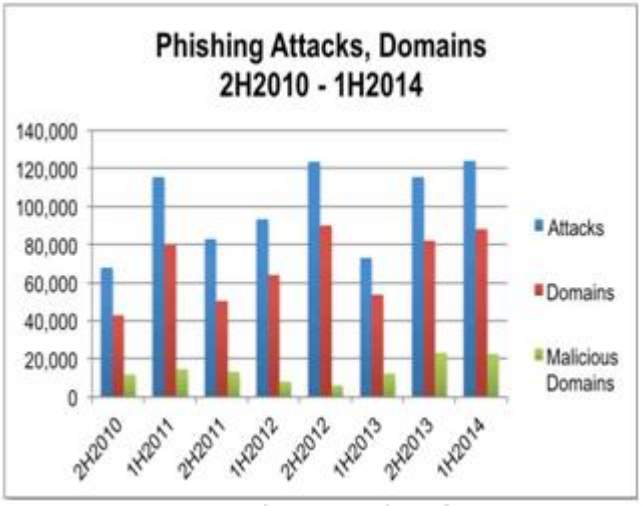
Developer	January 2013 Percent	February 2013 Percent	Change
Apache	348,119,032 55.26%	344,915,105 54.68%	-0.58
Microsoft	106,619,177 16.93%	104,647,425 16.59%	-0.34
nginx	79,640,472 12.64%	81,074,694 12.85%	0.21
Google	22,573,858 3.58%	22,717,984 3.60%	0.02

Fig5. Netcraft's survey

There was a major decline this month of sites that use Microsoft IIS & Apache, with both servers seeing a combination or more than 5 million hostnames. On the other side, nginx saw a 12.85% increase in business in last month January with 1.4 million more hostnames than December. The largest gains in hostnames positions nginx as one of the most well-known webservers, placing it less than 500 individual sites as Microsoft's IIS, which also has under 13 % of business. Tengine, an nginx derivative

managed by China e-tailer Taobao, and now is used for just about 4 million hostnames. For the meantime, Alibaba, which is affiliated with Taobao, has the second largest number of hostnames in China, with more than 11% of the hostnames in China. Although China makes up 19% of world population, only 5.8% of the world's websites are actually hosted in China. Still leading China, Microsoft has 38% of Chinese hosted websites using IIS, followed by 26% using Apache, & lastly 1% that uses nginx, which is considerably above average. In a report, Netcraft stated that taobao draws the second highest number if phishing attacks next to Facebook. Netcraft is reporting blocking nearly 6,000 urls targeting taobao users.

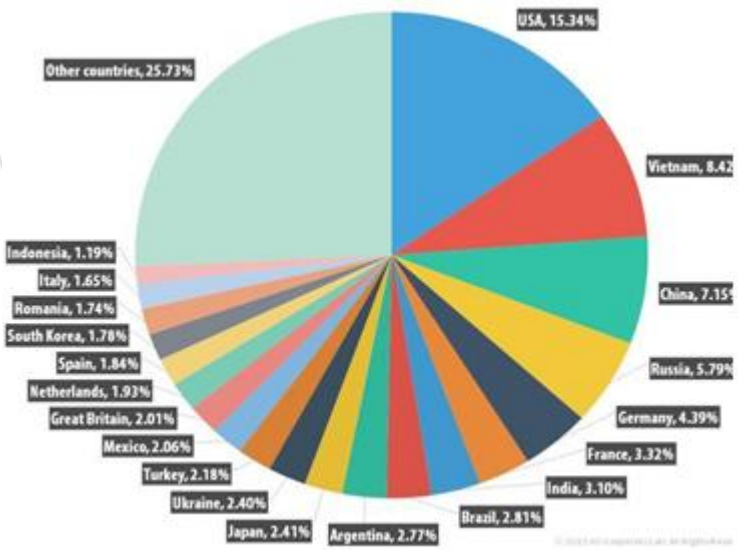
Basic Statistics



	1H2014	2H2013	1H2013	2H2012	1H2012	2H2011
Phishing domain names	87,901	82,163	53,685	89,748	64,204	50,298
Attacks	123,741	115,565	72,758	123,476	93,462	83,083
TLDs used	227	210	194	207	202	200
IP-based phish (unique IPs)	2,317	837	1,626	1,981	1,864	1,681
Maliciously registered domains	22,679	22,679	12,173	5,833	7,712	12,895
IDN domains	112	82	78	147	58	36
Number of targets	756	681	720	611	486	487

The above table contain numbers that explain the statics in the bar graph above it, where 1H means First Half and 2H means Second Half of the respective years.

It was found that almost any business having its footprint on the internet could be a phishing target. Especially if a site handles personal data of users or non-users. There was a noticeable variety among targets in the year 2014. It was observed that the major targets were large and small banks in Latin America, India, and the Middle East.



Sources of spam by country (2015)

The US (15.34%) remained the biggest source of spam in Q3. Vietnam was second with 8.42% of global spam, compared to 3.38% in the previous quarter. China rounded off the Top 3 (7.15%) – its share remained unchanged from the previous quarter. Russia's share (5.79%) dropped by 2.03 p.p., pushing it from second to fourth position. It was followed by Germany (4.39%) and France (3.32%) – their shares changed only slightly compared to Q2

6. Conclusion

Phishing differs from traditional scams primarily in the scale of the fraud that can be committed. In order to combat phishing, business and consumers need to adopt best practices and practice awareness, educate themselves about phishing and anti-phishing techniques, use current security protection and protocols, and report suspicious activities. By doing so, they can reduce their exposure to fraud and identity theft, safeguard their confidential information, and

help fight one of today's most serious and ongoing threats of phishing. The most effective solution to phishing is training users not to blindly follow links to web sites where they have to enter sensitive information such as passwords. The final technical solution to phishing involves significant infrastructure changes in the Internet that are beyond the ability of any one institution to deploy. However, there are steps that can be taken now to reduce the consumer's vulnerability to phishing attacks. Some of those steps are:

For Corporations:

- Establish corporate policies and communicate them to consumers.
- Provide a way for the consumer to validate that the E-mail is legitimate.
- Stronger authentication at web sites.
- Monitor the Internet for potential phishing web sites.
- Implement good quality anti-virus, content filtering and anti-spam solutions at the Internet gateway.

For Consumers:

- Automatically block malicious/fraudulent E-mail.
- Automatically detect and delete malicious software.
- Automatically block outgoing delivery of sensitive information to malicious parties.
- Be suspicious. All of these technologies are available now and can be deployed by both consumers and institutions interested in protecting their customers.

References

[1] Lance James, Phishing Exposed, Syngress Publishing, 2005.
[2] Lauren L. Sullins, Phishing for a solution: Domestic and international approaches to decreasing online identity theft, Emory Int. Law Rev. 20 (2006) 397–433.
[3] Oxford Dictionaries. 1990. http://www.oxforddictionaries.com/definition/english/phishing (accessed 13.10.12).
[4] David Watson, Thorsten Holz, Sven Mueller, Behind the scenes of phishing attacks. Know your Enemy: Phishing. 2005. http://www.honeynet.org/book/export/html/8 7 (accessed 17.01.12).
[5] APWG, Greg Aaron, Ronnie Manning, APWG Phishing Reports. APWG. 2014. http://docs.apwg.org/reports/apwg_ trends_report_q2_2014.pdf (accessed08.02.13).
[6] Q. Ming, Y. Chaobo, Research and design of phishing alarm system at client terminal, in: IEEE—Asian-Pasific Conference on Services Computing, APSCC'06. Asian, 2006, pp. 597–600.
[7] Engin Kirda, Christopher Kruegel, Protecting users against phishing attacks with AntiPhish, in: The 29th Annual In-ternational Computer Software and Applications Confer-ence, IEEE Computer Society, Washington, DC, USA, 2005, 517–524.
[8] Yue Zhang, Jason Hong, Lorrie Cranor, CANTINA: A content-based approach to detect phishing web sites, in: The 16th World Wide Web Conference, ACM, Banff, AB, Canada., 2007, 639–648.
[9] MessageLabs. The MessageLabs Intelligence Annual Secu-rity Report: 2009 Security Year in Review. 2009. http://www.symantec.com/connect/blogs/mes sagelabs-intelligence-annual-security-report-2009-security-year-review (accessed08.05.13).
[10] Dinei Florencio, Cormac Herley, Evaluating a trial de-ployment of password re-use for phishing prevention, in: The Anti-Phishing Working Groups 2nd Annual eCrime Researchers Summit, eCrime'07, ACM, New York, 2007, pp. 26–36.
[11] Symantec Corporation. Internet Security Threat Report 2013. Symantec Corporation, 2013.
[12] Jason Franklin, Vern Paxson, An inquiry into the nature and causes of the wealth of Internet miscreants, in: The 14th ACM Conference on Computer and Communications Security, CCS'07, ACM, New York, 2007, pp. 375–388.
[13] Gregg Keizer, Phishers Beat Bank's Two Factor Authentica-tion, InformationWeek, Manhasset, NY, 2007.
[14] Tom N. Jagatic, Nathaniel A. Johnson, Markus Jakobsson, Filippo Menczer, Social phishing, Commun. ACM (2007) 94–100.
[15] M. Chandrasekaran, K. Narayanan, S. Upadhyaya, Phishing email detection based on structural properties, in: NYS Cyber Security Conference, 2006.
[16] G. Aaron, R. Rasmussen, Global Phishing Survey 2H/2009. Sao Paulo, Brazil.: Counter eCrime Operations Summit IV, 2010.
[17] Larry Seltzer, betanews. 2011. http://betanews.com/2011/06/ 30/phishers-have-found-a-new-use-for-google-docs-stealing-your-identity/ (accessed 20.10.12).
[18] APWG, 2003. http://www.antiphishing.org/ (accessed20.12.11).
[19] BBC News. Jail for eBay phishing fraudster. 2005. http://news.bbc.co.uk/2/hi/uk_news/england/lancashire/4396914. stm (accessed 20.10.11).
[20] TREND MICRO. Threat Reports. 2013. http://www.trendmicro.com/us/secu rity-intelligence/research-and-analysis/index.html (accessed20.05.13).
[21] Bruce Schneier, Inside risks: semantic network attacks, Mag. Commun. ACM 143 (12) (2000) 168.
[22] Rachna Dhamija, J.D. Tygar, Marti Hearst, Why phishing works, in: The SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA, 2006, 581–590.
[23] M. Wu, R.C. Miller, S.L. Gar, Do security toolbars actually prevent phishing attacks? in: The SIGCHI Conference on Human Factors in Computing Systems, ACM, NY, USA, 2006, 601–610.
[24] Markus Jakobsson, The human factor in phishing, in: Privacy & Security of Consumer Information'07, 2007.
[25] Julie S. Downs, Mandy Holbrook, Lorrie Faith Cranor, Behavioral response to phishing risk, in: The Anti-Phishing Working Groups, 2nd Annual eCrime Researchers Summite, Crime'07, ACM, New York, NY, USA, 2007, 37–44.
[26] Ponnurangam Kumaraguru, et al., Getting users to pay at-tention to anti-phishing education: evaluation of retention and transfer, in: The Anti-Phishing Working Groups 2nd An-nual eCrime Researchers Summit, eCrime'07, ACM, Pitts-burgh, PA, USA, 2007, pp. 70–

81.
[27] Huajun Huang, Junshan Tan, Lingxi Liu, Countermeasure techniques for deceptive phishing attack, in: International Conference on New Trends in Information and Service Science, 2009, NISS'09, IEEE, Beijing, 2009, 636–641. Coll. of Comput. Sci., Central South Univ. of Fore.
[28] Steve Sheng, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Faith Cranor, Julie Downs, Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions, in: The 28th International Conference on Human Factors in Computing Systems, CHI'10, ACM, New York, NY, USA, 2010, pp. 373–382.
[29] MarkMonitor. MarkMonitor. 1999. https://www.markmonitor.com/ (accessed 14.01.13).
[30] Gartner Inc. Gartne. 2011. http://www.gartner.com/it-glossary/data-mining (accessed 30.05.11).
[31] Harris Poll. Taking Steps Against Identity Fraud. Harris Pol, 2006.
[32] "Executive Order 13402." Presidential Documents. 2006. http://www.gpo.gov/fdsys/pkg/FR-2006-05-15/pdf/06-4552.pdf (accessed 12.05.13).
[33] General Assembly of Virginia. CHAPTER 2005. http:// leg1.state.va.us/cgi-bin/legp504.exe?051+ful+CHAP0827 (ac-cessed 21.05.13).
[34] Grant Gross, Senator introduces 'phishing' penalties bill. 2004. http://www.informationweek.com/phishers-would-face-5-years-under-new-bill/d/d-id/1030773?(accessed 18.03.11).
[35] John Leyden, Florida man indicted over Katrina phish-ing scam. 2006. http://www.theregister.co.uk/2006/08/18/ hurricane_k_phishing_scam/ (accessed21.05.13).
[36] Lea Goldman, Cybercon. 2004. http://www.forbes.com/forbes/2004/10 04/088.html (accessed 21.05.13).
[37] KrebsonSecurity. HBGary Federal Hacked by Anonymous. 2011. http://krebsonsecurity.com/2011/02/hbgary-federal-hacked-by-anonymous/ (accessed14.05.13).
[38] Daisuke Miyamoto, Hiroaki Hazeyama, Youki Kadobayashi, An evaluation of machine learning-based methods for detection of phishing sites, Aust. J. Intell. Inf. Process. Syst. (2008) 54–63.
[39] [39] Xiang Guang, Jason Hong, Carolyn P.Rose, Lorrie Cranor,CANTINA+: A feature-rich machine learning framework for detecting phishing web sites, ACM Trans. Inf. Syst. Secur. (TISSEC) (2011) 1–28. 09.
[40] [40] Steve Sheng, Brad Wardman, Gary Warner, Lorrie Faith Cranor, Jason Hong, Chengshan Zhang, An empirical analysis of phishing blacklists, in: The 6th Conference on Email and Anti-Spam, CEAS'09. CA, USA, 2009.
[41] Rod Rasmussen, Greg Aaron, Global Phishing Survey: Trends and Domain Name Use 2H2009. Lexington, MA, 2010.
[42] David Dede, Ask Sucuri. 2011. http://blog.sucuri.net/2011/ 12/ask-sucuri-how-long-it-takes-for-a-site-to-be-removed - from-googles-blacklist-updated.html (accessed 17.02.12).
[43] Google code. Google Safe Browsing. 2010. http://code.google.com/p/google-safe-browsing/ (accessed 11.12.11).
[44] Microsoft, Support-. Microsoft IE 9 anti-phishing. 2012. http://support.microsoft.com/kb/930168 (accessed 19.12.12).
[45] McAfee. SiteAdvisor. 1987. http://www.siteadvisor.com/ (ac-cessed19.12.11).
[46] Symantec. Verisign Authentication Services. 1982. http://www.verisign.com/ (accessed19.12.11).
[47] Netcraft Toolbar. Netcraft. 1995. http://toolbar.netcraft.com/ (accessed19.12.11).