International Journal of Grid and Distributed Computing
Vol. 9, No. 8, (2016), pp.371-384
http://dx.doi.org/10.14257/ijgdc.2016.9.8.32

# A Comparative Analysis of Phishing Detection and Prevention Techniques

*Shivangi Sharma [1] and Sheetal Kalra[2]

*Department of Computer Science*
*Guru Nanak Dev University, Regional Campus*
*Jalandhar, India.*
*[1]shivangi04@ymail.com, [2]sheetal.kalra@gmail.com*

## *Abstract*

*Phishing assaults use websites and email messages similar to a familiar and genuine organization, so as to trick users into giving their financial or personal information online at the illegal websites. Phishing assaults look out at the various weak points that are present in systems and frameworks because of the human factor. There exist numerous cyber-attacks which are spread through different methods or schemes. They utilize these shortcomings found in end users, thus making the users as the most vulnerable component in terms of security. The phishing problem is extensive, expansive and till now there is no single answer to alleviate all the existing weak points effectively. This leads to implementation of multiple techniques to reduce specific attacks. This paper surveys the literature on the recently proposed anti-phishing techniques designed to detect and prevent phishing.*

**Keywords**: *phishing, security, social engineering, phishing prevention, phishing detection.*

## 1. Introduction

As individuals progressively depend upon the web for individual money, business and venture different web frauds comes into play and poses great threat to users. These internet frauds take many forms. One such internet assault is Phishing.

Phishing assaults use websites and email messages similar to a familiar and genuine organization, so as to trick users into giving their financial or personal information online at the illegal websites. The attacker uses this data for various illegal purposes, for example fraud or identity theft. Clients are deceived to unveil his/her data by downloading and installing hostile software or by giving it through a web form.

Phishing has been constantly on rise from last few years, described in Figure 1. The Anti-Phishing Work Group (APWG) is a non-profit organization that provides anti-phishing education to the people to reinforce the proper understanding of security. Till today there is no such finish arrangement which can catch every last phishing assault. Phishing attacks are becoming more sophisticated and are growing very fast. According to the APWG, there were 289,371 total number of unique phishing websites observed in Q1. The number watched every month raised consistently from the 48,114 detected in October 2015 to the 123,555 detected in March 2016 – identified in March 2016 – a 25 percent expansion over six months. There is usually an outbreak of spamming and online fraud during the holiday shopping season thus, the increase in December 2015 was expected. The proceeding with expansion into 2016 is reason for concern. In the fourth quarter of 2015, the retail/service sector turned into the most-targeted sector with 24.03 percent of attacks, took after nearly by financial services. ISPs had been the most-focused on industry section in the first three quarters of 2015 [1].

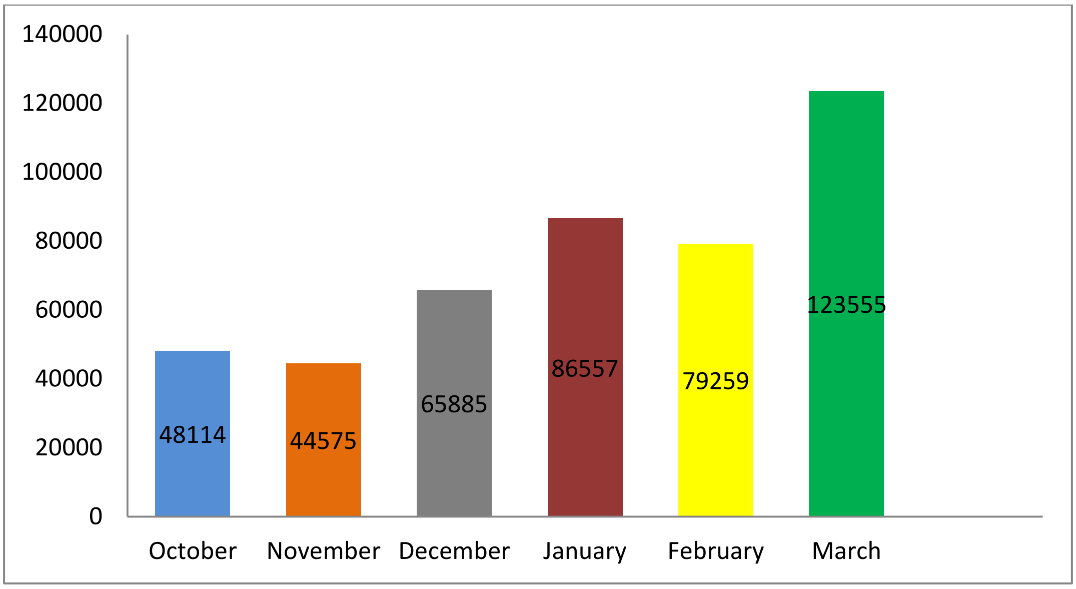

**Figure 1. Number of Unique Phishing Sites Detected October 2015-March 2016**

Phishing is said to be troublesome as it causes enormous budgetary misfortunes for associations that provides online money related services which cause a big negative impact on confidence of consumer about ecommerce. A remarkable rise of nearly 130,000 from January to March was seen in the phishing reports submitted to APWG.

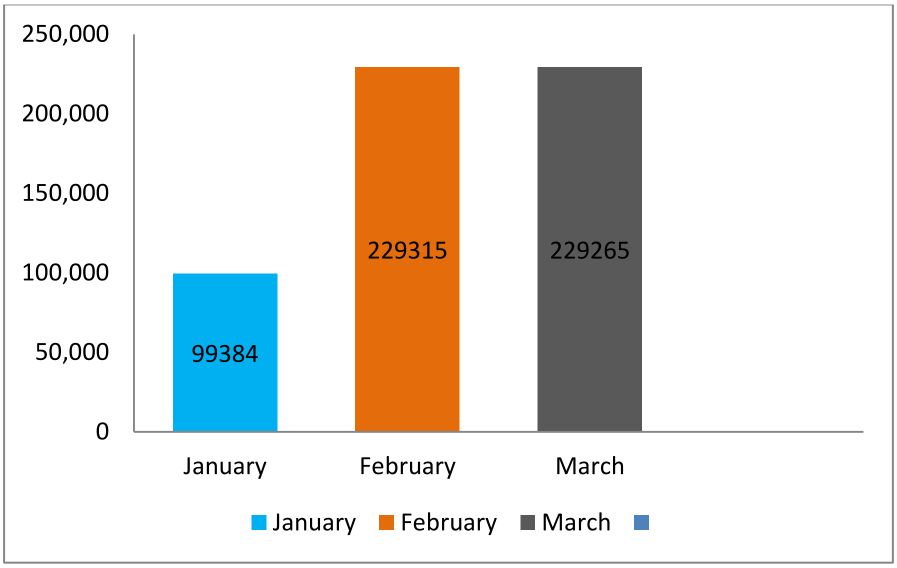

**Figure 2. Phishing Reports Received from January-March, 2016**

Reducing the effect of phishing is significantly critical and increases the value of the general security of an organization.

## 2. Background

### A. History

The word *phishing* originates from *fishing* similar to as fishers (attackers) uses bait (fake email messages) to fish (steal personal information of victims).

According to APWG, the word *phishing* came in 1996 because of social engineering attacks caused against America On-line (AOL) accounts by online scammers [2]. Phishing assaults were generally begun by stealing AOL accounts, and later it moved into assaulting more productive targets/companies, for example, e-trade commerce and online banking.

As of now, phishing assaults don't just pick out end users, additionally specialized representatives that are present at service providers, and bring complex procedures into action, for example MITM assaults.

### B. Phishing Motives

Taking a glance at the way that phishing scammers are getting tremendous monetary gains, it can without much of a stretch be presumed that the inspiration driving phishing is quite often money related. Monetary gain is the major factor that motivates phishers to carry out the attack, other factors such as theft of identity, distributing malware on user's computer, *etc.* also inspires phishers.

i. *Financial gain:* Various researches showed that major targets of phishing attacks are the financial sectors. A phishing assault against any money related association involves destroying the brand of the institution. This is done by creating a fake website page where the phishers asks permission to access the account details of a victim. This is therefore can be said as easy money.

ii. *Identity theft:* Identity theft is identification based crime. It is done to get monetary gains, to carry out fraud and other illegal exercises. It causes immense monetary misfortunes to the ones whose identity is stolen and also to financial institutions and business organizations. Generally the individual whose identity is stolen does not understand until the harm is caused by the phisher. For financial gains - to obtain services, to commit fraud, for criminal activities *etc.* stolen identities can be used by the phisher.

iii. *Malware distribution:* The attack takes place by circulating malware. Phishing messages are normally sent in mass and thus, zombie networks are the most appropriate to dispatch large phishing attacks. These messages contain malware connections which when clicked by a clueless client results into distribution of malware over the casualty's machine.

iv. *Harvesting passwords*: This is carried on by phishers using different techniques like key loggers and other malware. The collected client data is utilized again for monetary benefit, identity theft, fraud or is sold to interested parties for monetary benefit.

v. *Fame and notoriety:* At times phishing assaults are done by individuals for the most part to pick up acknowledgment and reputation among their associates. This is an extremely psychological aspect of phishing wherein data is phished not for monetary benefit but rather just with the end goal of picking up consideration and greatness in the online group.

### C. Challenges

Reducing the effect of phishing is significantly critical and increases the value of the general security of an organization. Since the phishing issue exploits human lack of awareness or naivety concerning their connection with electronic communication channels (e.g. Email, HTTP, and so forth), it is not a simple issue to for all time settle.

The greater part of the proposed frameworks endeavors to reduce the effect of phishing assaults.

Two proposed solutions to reduce phishing attacks:

i. User education: the human is taught so as to upgrade his accuracy to recognize phishing messages, and after that apply appropriate actions on those correctly chosen phishing messages.

ii. Software enhancement: the product is enhanced to better characterize phishing messages. This is done on behalf of the human, or gives data in a more evident manner so that the human would have less opportunity to disregard it.

The disadvantages of these solutions are:

i. Non-specialized individuals oppose learning, and on the off chance that they learn they don't hold their insight for all time. The training should be continuous.

ii. There are some software solutions, like authentication and security warnings, that are still dependent on user behavior. In the event that users disregard security notices, the solution employed can be said as useless.

## 3. Anti-phishing Techniques in Literature

An online communication can take place as:

*i.* *Message claim:* User receives a message in form of an email or web page.

*ii.* *Message introduction:* The message is introduced and shown in the user interface; a mental model is made by the user after seeing the message.

*iii.* *User activity:* From the mental model created by user, he/she performs an activity in the user interface, for example, filling in a form or by clicking on a link.

*iv.* *System operation*: The user's activity is converted into system operations, for example connecting to a web server and then submitting information.

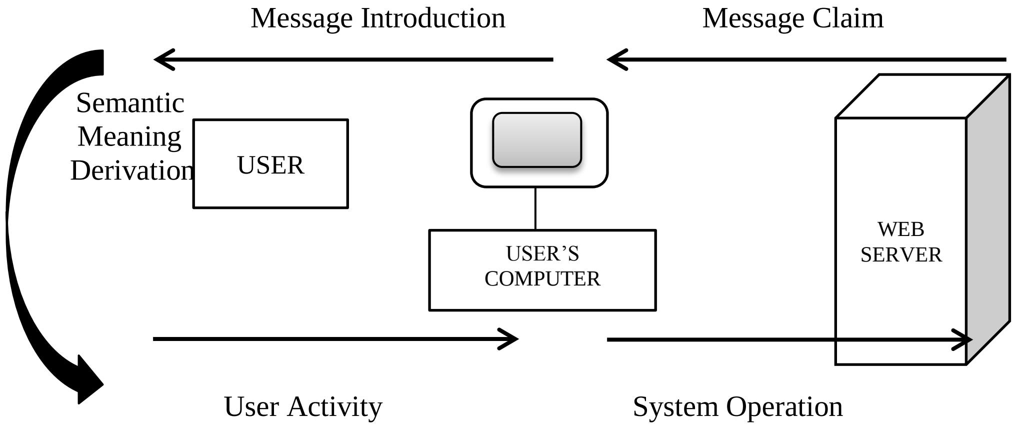

**Figure 3. User's Online Interaction Steps**

### 3.1. Message Claim

Obstructing all phishing messages at this step is the best safeguard against phishing. The vital need is that the computer should have ability to precisely separate phishing messages from honest legitimate ones.

**i.** **Identity of the sender:** It is one of the properties in message claim step.

a) Blacklisting: It is generally used to block harmful messages, for example, spam. A blacklist is a list of URLs that are thought to be malicious [3]. It helps in sorting out the messages that approaches client if the sender's IP location is found in a blacklist. It can be categorized as spam or even just rejected without informing the client. Numerous current security toolbars utilize blacklist to caution clients about phishing sites that have been identified and reported. Some show a notice where as some thwart the whole page.

One of the disadvantages of blacklist is the need to stay up with the latest. Since phishing websites are available for a brief timeframe, a blacklist must be kept updated within hours or minutes so as to block the attack.

There are two approaches to update a blacklist, either checking every new URL that a user is about to browse or periodically synchronizing with a blacklist source. If the procedure of updating the blacklist is slow, this will give phishers the chance to do assaults without being added to blacklist.

Another problem of blacklisting is that it just cautions clients about the deceitful site and encourages them not to continue. Clients may go out on a limb of proceeding on the off chance that they imagine that the present (fake) undertaking is imperative to finish.

Some blacklists for example, Google's Blacklist needs by and large seven hours to be updated [4]. Many solutions have been brought into action depending on the blacklist, one of which is Google Safe Browsing [5]. It utilizes anti-phishing approach, blacklist, to detect phishing. The limitation is that phishing websites which are not present in blacklist are not identified. This approach can result to high false negative rate.

Netcraft [6] depends upon a blacklist which comprises of deceitful sites perceived by Netcraft and those URLs presented by the clients and checked by Netcraft. It is a small software package that gets activated whenever a user browses internet. It shows the location of the server where a website page is facilitated. A warning message is shown if it encounters a website page which is available in the blacklist.

Phish-Tank [7] provides a blacklist, through an API, for the utilization by other tools. Its blacklist is fully filled by crowdsourcing volunteers who submit websites that are likely to be phishing sites and vote on the truthfulness of websites.

b) Whitelisting: It is opposite to blacklisting, permitting users to see messages just from a rundown of worthy sources. The newly created sources are initially marked as unacceptable thus it avoids the new-identity problem. Defining a whitelist is a significant issue. Since it is difficult to anticipate where a user may need to browse, a predefined whitelist blocks users from accessing true websites. Then again, a dynamic whitelist that needs the user's inclusion puts a weight on users because of the fact that, for each site they need to visit, they should first choose whether to place it in the whitelist. This offers vulnerability: if a phishing site can persuade users to submit delicate information to it, it might likewise have the capacity to persuade them to place it into a whitelist.

Kang and Lee [8] proposed worldwide white-list-based strategy that reduces access to phishing sites with a URL similarity check. When a user request a site, the site's URL and IP pair is gone to the Access Enforcement Facility (AEF) to survey if the site is a phishing site. If the site’s URL matches with an entry in the trusted site list, then the program surveys the IP address similarity. If the IP also matches, then the security framework allows the user to proceed; or the framework cautions the user of phishing attack.

A latest work [9] proposes an Automated-Individual-Whitelist (AIWL). It is an anti-phishing tool that is based on user’s whitelist made up of known trusted websites. AIWL follows each login endeavor by individual users through the usage of a Naïve Bayesian classifier. If a repeated successful login for a particular website is accomplished, AIWL instigate the user to add the site to the whitelist. Its whitelist comprises of a Login User Interface (LUI) for each trusted site, a hash of its security certificate, the site URL, a hash of the credential to that website, the HTML DOM path to the username and password input fields and the valid IP address for the URL. Users are cautioned once they present their data to a site that does not exist inside the whitelist. However, this methodology expects that users just present their data to true legal sites, while all others are considered harmful.

**Table 1. Summary of Whitelisting and Blacklisting Approaches**

| Method | Advantage | Disadvantage | Used in |
|---|---|---|---|
| White-listing | Identify legal websites by conserving a whitelist of non-harmful URLs or domains.<br>Experiences no false negative. | High false positive. | IE, Mozilla Firefox browsers. |
| Black-listing | A blacklist of phishing URLs or domains is used to obstruct phishing websites.<br>Experiences no false positive. | High false negatives. | IE, Mozilla Firefox browsers. |

Another technique that fundamentally relies upon the whitelist method was displayed in PhishZoo [10]. PhishZoo has been assessed utilizing 636 phishing sites and 20 true legal site profiles downloaded from Phish-Tank. This procedure fabricated profiles of trusted sites based on fuzzy hash methods. A profile of a site is a combination of several metrics that identify a particular site. The advantage of such whitelisting strategy permits identifying recently launched phishing sites with the capacity of blacklisting and acquiring a heuristic way to warn users.

c) Email Authentication: It avoids phishing by checking if the email is truly sent by sender. Once the email authentication, o is brought into action, it can keep phishing emails from utilizing deceptive addresses, since phishing emails claim to originate from the trusted sources. Attackers can use other channels, for example, blogs to lure victims besides emails.

**Table 2. Comparison of Whitelisting Approaches and Blacklisting Approaches**

| Work | Method | Zero day phishing attacks | Language independent | Image based phishing attack | DNS attack | Shortcomings |
|---|---|---|---|---|---|---|
| J.Kang, D.Lee (2007) | Global whitelist based approach | Yes | Yes | No | Yes | i. If attacker compromise and make a phishing webpage, it cannot detect it. |
| M. Sharifi, S.H. Siadati (2008) | Blacklist generator | No | Yes | No | No | i. Datasets must be selected accurately for greater efficiency.<br>ii. There is a tradeoff between false positive rate and the chance of spamdexing techniques. |
| P.Prakash, M.Kumar, R.R.Kompella, M.Gupta | PhishNet | No | Yes | Yes | Yes | i. Do no detect websites which are not present in the blacklist. |

| | | | | | | |
|---|---|---|---|---|---|---|
| (2010) | | | | | | |
| C.Whittaker, B.Ryner, M.Nazif (2010) | Google safe browsing API | No | Yes | No | Yes | i. Phishing websites which are not listed in blacklist are not detected.<br>ii. This technique leads to high false negative rate. |
| S.Afroz, R.Greenstadt (2011) | PhishZoo | No | Yes | Yes | No | i. It fails on images whose content do not match up well with compared logo name.<br>ii. Human intervention is present. |
| W.Han, Ye Cao, E.Bertino, J.Yong (2012) | Automated Individual Whitelist (AIWL) | Yes | Yes | No | Yes | i. User intervention and knowledge is required.<br>ii. Need a more private device to store whitelist in a more secure environment. |
| R.S.Rao, S.T.Ali (2015) | PhishShield | Yes | Yes | No | Yes | i. Its effectiveness depends upon the input given.<br>ii. It fails to detect phishing when all of the filters are bypassed by phishers.<br>iii. it fails to detect phishing sites when JSoup parsing failure occurs. |
| R.M.Mohammad, F.Thabtah, L.McCluskey (2015) | Community based approach | No | Yes | No | Yes | i. It is dependent upon the user's knowledge.<br>ii. It does not rely on automatically extracting features from webpage. |

**ii. Content of the message:** The next property of message claim step is content of the message

a) Textual content analysis: It is significantly used in antispam and antivirus techniques.

Emigh [11] proposed to heuristically break down the content of email messages and website pages for possibly deceptive links and provide them in a way that detect them as suspicious. Harmful messages are identified via scanning for well-known patterns, such as virus code signatures and spam keywords. Keeping in mind the end goal to beat content analysis, an attacker can change the content to sidestep surely understood filtering rules. For instance, encryption and compression are added to existing viruses in order to avoid antivirus scans. One defense that is applied at the message retrieval step is Spam filtering. Random characters are embedded into spam messages to enable them to sidestep spam filters. Existing complex phishing attacks used images to show text messages so they could overcome content analysis. Getting spam under control may diminish the danger of phishing attacks, since about all phishing attacks at present are launched by spam. Unfortunately, the techniques are insufficient for grouping phishing attacks, as messages are outlined to mimic legitimate mail that come from organizations with which the user as of now has relationship.

Cantina [12] technique relies upon the textual content of the site. Term Frequency–Inverse Document Frequency (TF-IDF) [13] algorithm is applied on the textual content which is combined with additional heuristics used to identify phishing attacks. The main five tokens with most astounding TF-IDF are submitted to search engine which is then followed by comparing suspicious link with search engine results. This methodology comes up short when the text of site is substituted with pictures/images or addition of undetectable content which matches background colour of site.

Spoofguard [14] is a Browser Helper Object that computes an aggregate spoof score for incoming website pages. The figuring depends on basic attributes of known phishing attacks, including:

- Misleading patterns in URLs, for example use of @,
- Presence of similarity in the domain name with respect to popular or previously visited domains,
- Embedded images similar to images from constantly spoofed domains,
- Misleading URLs contained in the page.

b) Visual similarity analysis: Visual similarity assessment has been proposed [15] to detect phishing websites instead of using the textual content analysis. In this method, a suspicious website page is broken down into a set of blocks and then compared with the blocks on legitimate web pages that are stored in a database to see whether it matches. The user is warned if the matching score is above a threshold. This method has some potential issues. The first issue is scalability since there are great many web pages that can be accessed by users. Second, sites constantly keep on changing their visual appearance. Third, the attacker may use visually different but semantically similar phishing messages to sidestep this protection.

Hara *et al*. [16] built up a method which groups the suspicious sites taking into account image similarity. In this ImgSeek application is used for comparing suspicious and legitimate image. This strategy can auto update the whitelist by addition of fake sites that are neither phishing nor legitimate. The restriction of this methodology is high false negative rate. There exists delay in browser’s experience due to image comparison at client's side.

BaitAlarm [17] compares visual features to group legitimate and phishing. Phishers must utilize same styles to mimic the illustrations of legitimate site thus authors considered Cascading Style Sheets (CSS) for identifying phishing sites. Authors picked a legitimate website and then they compared it with a large number of phishing websites. They also indicate the need of whitelist. The disadvantage of BaitAlarm is that in comparison with whitelist database, computation cost of CSS style is too high.

Liu *et al*. [18] proposed a methodology that recognizes phishing in view of the visual features of suspicious sites. The visual elements and equivalent features in legitimate site are compared with each other. These features are for example; block level (content and images/pictures), layout similarity (DOM) and general style (CSS). Every feature is allotted weights as per the priority given when designing a legitimate site. If the visual similarity is above a threshold value then a suspicious website is classified as phishing website. The restriction in this procedure is high response time *i.e.* this plan needs expansive legitimate image database. Also the visual correlation of suspicious fake site with image database is too expensive.

**Table 3. Comparison of Textual Content and Visual Similarity Techniques**

| Work | Techniques | Language Independent | Zero day phishing attacks | Image based phishing attacks | Shortcomings |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| N.Chou, R.Ledesma, Y.Teraguchi, J.C.Mitchell (2004) | SpoofGaurd | Yes | Yes | No | i. The attacker can fool by breaking the password input field into multiple adjacent fields.<br>ii. By slicing an image into adjacent vertical slices and presenting them in order, attacker can carry on phishing attack. |
| L.Weyin, G.Huang, L.Xiaoyue, Z.Min, X.Deng (2005) | L.Weyin | Yes | Yes | Yes | i. There is high response time.<br>ii. There is high false negative rate.<br>iii. Image comparison at client side leads to delay in browser's experience. |
| Y.Zhang, J.I.Hong, L.F.Cranor (2007) | CANTINA | No | Yes | No | i. It does not include a dictionary for languages other than english.<br>ii. TF-IDF does not work well with East Asian languages.<br>iii. It suffers from performance problems due to the time lag involved in querying google. |
| J.Mao, P.Li, K.Li, T.Wei, Z.Liang (2013) | BaitAlarm | Yes | Yes | No | i. The computation cost of CSS style is too high as compared to whitelist database. |

### 3.2. Message Introduction

The client is shown a message, in an email or a web browser, at this time the client interface can give visual prompts to help the client choose whether the message is authentic. Existing browsers shows data related to the trustworthiness and the source of webpage through a group of visual indicators. Proof recommends that users may at present fall for phishing sites who consider that the image database is down for maintenance. Some also do not consider this feature because absence of cue may not trigger their attention. For example, a lock icon that is present in the status bar demonstrates whether the webpage was recovered by an authenticated, encrypted connection. The URL of the webpage is showed in the address bar. These signs are right now the most broadly used defenses against phishing. Numerous browsers toolbars have likewise been proposed to secure against phishing, each with constrained achievement. Security advisories caution clients about phishing to give careful consideration to them every time. Numerous propositions for ceasing phishing assaults depend on a security toolbar that presents notices or security-related data in the web browser's interface.

a) Weider D. Yu, Shruti Nargundkar, Nagapriya Tiruthani [19]: Taking account the broad investigation of the inspiration, causes, methodologies that were utilized to conduct phishing, they proposed an algorithm - PhishCatch - Identify, Defend and Prevent. This algorithm is based on heuristics which will recognize phishing messages and caution the clients about the phishing messages. The language used for execution is Python. This algorithm was executed in Windows XP. Broad testing was done to test the effectiveness of the algorithm in recognizing phishing messages. Due to this reason, honeypots were set up and seeded over the web to draw in phishing messages. PhishCatch was tried against these email ID's. After testing, it was determined that PhishCatch algorithm has a catch rate of 80% and an accuracy of 99%.

b) Martin Husak and Jakub Cegan [20] give a framework for automatic phishing incident processing and work in advancement concerning automatic phishing detection and reporting. The arrangement is part into two sections, phishing incident processing and phishing detection. The center of system is the phishing incident processing part. It is based upon the automatic phishing incident processing tool called PhiGARo which finds clients reacting to phishing assault endeavors and keeps access to phishing websites from the ensured system. Despite the fact that PhiGARo forms the phishing incidents consequently, it relies on upon reports of phishing occurrences from clients. PhiGARo was developed to automate the process. Phishing incidents are accounted by clients who recognize a phishing message in their mailboxes. The report is sent by means of e-mail or web form and then it is accepted by the Request Tracker. A command line API is present for users who do not use the Request Tracker.

c) Taiwo Ayodele, Charles A. Shoniregun, Galyna Akmayeva [21] presents machine learning anti-phishing technique (MLAPT). It is a machine learning framework that applies knowledge intensive approach to identify and prevent phishing attacks on email systems. It is intended to handle identification of phishing like practices with our email systems. The phishing practices are transmitted to the MLAPT such as colour miss-match, change of interface, domain verification, data request verification, re-direction verification. These are identified by means of our machine learning phishing behavioural techniques. Each of the patterns, behaviors, illegitimate similar websites are broken down that disguise as social engineering contents, original, suspicious host names that tends to demand users' personal information and then compare these findings with phishing model judged by humans. This model comprise of a huge number of genuine phishing tests samples, situations, designs, false hostnames and IP addresses many more.

d) Herzberg and Gbara [22] build up an extension for Mozilla called Trustbar. It demonstrates some data about the site, for example, name of the site, its logo, proprietor, CA, or a notice message for sites that are unprotected. This extension is simple to use. The fundamental negative part of Trustbar is that without checking its trustworthiness, it demonstrates the site's Certification Authority (CA) and the client need to do such confirmation. Tragically, most Internet clients have no clue about how to do such confirmation.

e) SpoofStick [23] is a toolbar that demonstrates site's genuine domain name. It can be introduced on both Mozilla Firefox and Internet Explorer. An assault utilizes a legitimate look-a-like name in the sub-space part of the URL to trick the customers. Example, if client enters a URL like "www.ebay.com.spofon.ca" then "spofon.ca" would be consider as the domain of that URL by SpoofStick. Additionally, SpoofStick performs reverse DNS question to show the genuine IP of the site.

f) Netcraft Toolbar [6] shows data about the site, including hosting country, the domain's registration date, and popularity among other toolbar users. This data is thought to be useful in recognizing phishing sites because of the fact that the most phishing sites are short-lived as compared to the legal sites which they imitate.

g) SpoofGuard [14] figures a score called spoof score for the present website page utilizing a gathering of heuristics which are gotten from past phishing assaults. Then it makes an interpretation of this score similar to traffic light as:

i. if scores are over a limit then red. It indicates the page is most likely unfriendly;
ii. if scores are not above or below the limit but in the middle then yellow;
iii. if scores are less then green, showing that the page is safe.

### 3.3. User Activity

Phishing relies upon client being tricked as well as acting under influence. Subsequently, security advisories attempt to debilitate clients from doing unsafe activities.

Security tips recommend that the user ought to disregard links embedded in email, and rather than open up the browser and then write the URL of the real website manually. This guidance is generally not taken after by clients. Taking in account the less recurrence of phishing assaults when contrasted with genuine websites, this proposal gives up the effectiveness of hyperlinks in real messages so as to keep clients from clicking misleading connections in not very many phishing messages.

Hardware token-based technique [3] is viewed as expensive because every client should own his/her device. Also, administration and the training expenses are high, as clients must be taught to utilize these tokens.

Two-factor authentication is a server-side solution. It guarantees that the client knows a thing that nobody knows and in addition there is a security token which is used when he is signing in. This averts phishing in light of the fact that the secret or the password that may be gotten by the assailant is by all account not the only verification component. The phishing attack can take place at any site that does not contain such type of authentication feature.

Mizuno, Yamada [24] gives us the multiple communication channels that are used to validate user identity. Mobile phones can be used as trusted communication channels. This methodology provides ISPs to utilize trusted channels for communication so as to check client’s identity on the modes of communication which are not trusted.

### 3.4. System Operation

In the last stride of a phishing attack, the user's activity is converted into system operation. This is the last opportunity to avoid the assault. The systems operations required in a phishing assault are flawlessly legitimate in light of the fact that phishing does not use system bugs. Notices construct exclusively in light of system operations will definitely produce a high rate of false positive errors which means, cautioning clients about actions that are not inappropriate. These false-positives in the end cause users to stop the warnings or just ignore these warnings.

An all the more intriguing methodology includes adjusting the system operation as indicated by its destination. This thought is applied by Web password hashing, to protect against phishing assaults that take site passwords. The password typed by the user is automatically hashed by the browser with the domain name to which it is being sent, so that a unique password is generated for each website and henceforth sending pointless trash to a phishing website. It is assumed in this method passwords will be typed by users into a password HTML element only. Unfortunately, this element can still be spoofed, and an advanced attack can be carried out which may be able to trap users into uncovering their passwords through other different channels.

Another novel methodology is to check the client's submission with a specific end goal to recognize when a client is submitting private data on a phishing website. SpoofGuard stores the client's passwords’ hashes in the machine that is located near to the client and checks the client's submission to recognize if any password that is being stored is transferred or not. On the off chance that there is recognition, the location where data is submitted will be analyzed to guarantee that the password submission is not taking place at a fake location. The disadvantage of this method is - the checking is opposed to HTTP post information. This checking can be sidestepped by the attacker by utilizing a script to encrypt the information before transferring it that is being input. Submission checking is incapable unless it happens at the keystroke level.

**Table 4. Comparison of Different Anti-phishing Frameworks**

| Authors | Tool/Method | Strength | Weakness |
|---|---|---|---|
| Erwin P. Rathgeb, Dirk Hoffstadt 2008 | E-Mail Honeypot System (EHS) | It is designed specifically to automatically receive, categorize, analyze and archive e-mail spam.<br>The EHS is a combination of widely used freeware components controlled by customized scripts.<br>This allows the EHS to be easily adapted or extended according to specific user needs. | The limited point of view (one domain only).<br>It would be useful to extend the system to more than one sensor. |
| Weider D. Yu 2009 | PhishCatch | Identifies phishing links, alerting the user about them and building an intensive knowledge database containing the information concerning phishing. | Do not interact with the mail servers like Microsoft Outlook. |
| Samuel Marchal 2014 | PhishStorm | It is a system which depends upon search engine query data for phishing URL identification. | Not applicable to all or any sorts of obfuscated URLs.<br>Delay is there caused by multiple HTTP requests for each term that compose a URL.<br>Data that is available publicly through Google Trends and Yahoo Clues is limited. |
| Martin Husak 2014 | PhiGARo | It is the framework for automatic phishing detection. It uses honeypots, a phishing incident and propagation of honeytokens to draw in phishers to a honeypot. | Both methods do not solve initial honeytoken propagation.<br>Less support from honeypots thus decrease in trustworthiness of honeytokens.<br>Problem concerning outsourcing anti-phishing honeypots. |

## 4. Conclusion

Phishing attack deceives user to get private data like passwords and financial information. This has caused lack of faith on people and websites. To get the trust back from users, the problem of phishing should be effectively treated. Phishing attacks have become more technically refined; therefore in detection and prevention of these attacks the currently present anti-phishing toolbars are becoming inadequate. It is necessary to take steps to forestall phishing as it is foretold that phishing attacks are going to be on the rise within the coming years. This paper shows the strengths and weaknesses of current approaches against phishing. The main motivation of this work is to summarize the recent approaches in this field of research, identify major issues and challenges and to encourage further research in this field.

## Acknowledgment

I would like to thank Dr. Sheetal Kalra for her continuous support and all suggestions that were needed to complete my research work. I am very thankful to my parents for their motivation and support.

## References

[1] Anti-Phishing Working Group. Phishing activity trends report, q42015. https://docs.apwg.org/reports/apwg_trends_report_q4_2015.pdf

[2] Mahmoud Khonji, Youssef Iraqi, Andrew Jones, "Phishing Detection: A Literature Survey", in proceedings IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 15, NO. 4, FOURTH QUARTER 2013.

[3] Rami M. Mohammad, Fadi Thabtah, Lee McCluskey, "Tutorial and critical analysis of phishing websites methods", Computer Science Review (2015), http://dx.doi.org/10/1016/j.cosrev.2015.04.001 .

[4] 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).

[5] Google code. Google Safe Browsing.2010. http://code.google.com/p/google-safe-browsing/(accessed 11.12.11).

[6] Netcraft Toolbar. Netcraft. 1995. http://toolbar.netcraft.com/ (accessed 19.12.11).

[7] PhishTank, [Online]. Available at https://www.phishtank.com/

[8] JungMin Kang, Dohoon Lee, "Advanced white list approach for preventing access to phishing sites", in International Conference on Convergence Information Technology, 2007, IEEE, Gyeongju, 2007, pp. 491–496.

[9] Weili Han, Ye Cao, Elisa Bertino, Jianming Yong, "Using automated individual white-list to protect web digital identities", Expert Syst. Appl. 39 (15) (2012) 11861–11869.

[10] Sadia Afroz, Rachel Greenstadt, "PhishZoo: Detecting phishing websites by looking at them", in Fifth International Conference on Semantic Computing, IEEE, Palo Alto, California, USA, 2011.

[11] A. Emigh. Online Identity Theft: Phishing Technology, Chokepoints and Countermeasures. In ITTC Report on Online Identity Theft Technology and Countermeasures., October 2005.

[12] Yue Zhang, Jason Hong, Lorrie Cranor, CANTINA: A content based approach to detect phishing web sites, in: The 16th WorldWideWeb Conference, ACM, Banff, AB, Canada, 2007, pp. 639–648.

[13] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008.

[14] Chou Neil, Robert Ledesma, Yuka Teraguchi, Dan Bon, Client side defense against web based identity theft, in: The 11th Annual Network and Distributed System Security Symposium, (NDSS'04), SpoofGuard, San Diego, 2004, pp. 143–159.

[15] W. Liu, X. Deng, G. Huang, and A.Y. Fu. An Antiphishing Strategy Based on Visual Similarity Assessment. In IEEE Internet Computing, Vol. 10, No. 2, "March/April" 2006.

[16] L. Wenyin, G. Huang, L. Xiaoyue, Z. Min and X. Deng, Detection of Phishing Webpages based on Visual Similarity, In Special Interest Tracks and Posters of the 14th International Conference on World Wide Web, ACM, pp. 1060–1061, May (2005).

[17] J. Mao, P. Li, K. Li, T. Wei and Z. Liang, Baitalarm: Detecting Phishing Sites using Similarity in Fundamental Visual Features, In 5th International Conference on Intelligent Networking and Collaborative Systems, INCoS 2013, IEEE, pp. 790–795, September (2013).

[18] L. Wenyin, G. Huang, L. Xiaoyue, Z. Min and X. Deng, Detection of Phishing Webpages based on Visual Similarity, In 14th International Conference on World Wide Web (WWW): Special Interest Tracks and Posters, (2005).

[19] Weider D. Yu, Shruti Nargundkar, NagapriyaTiruthani, “PhishCatch – A Phishing Detection Tool” in proceedings of the 33rd Annual IEEE International Computer Software and Applications Conference, 2009.

[20] Martin Husak and Jakub Cegan, “PhiGARo: Automatic phishing detection and incident response framework”, in proceedings of the 9Th International Conference on Availability, Reliability and Security, 2014, IEEE.

[21] Taiwo Ayodele, Charles A. Shoniregun, Galyna Akmayeva, “Anti-Phishing Prevention Measure for Email Systems”, In World Congress on Internet Security (WorldCIS-2012).

[22] Amir Herzberg, Ahmad Gbara, Protecting (even) Naive Web Users, or: preventing spoofing and establishing credentials of web sites, DIMACS (2004).

[23] spoofstick. 2005. http://www.spoofstick.com/ (accessed 19.03.12).

[24] Shintaro Mizuno, Kohji Yamada, Kenji Takahashi, Authentication using multiple communication channels, in: The 2005 Workshop on Digital Identity Management, ACM, Fairfax, VA, USA, 2005, pp. 54–62.

[25] Erwin P. Rathgeb, Dirk Hoffstadt, “The E-Mail Honeypot System – Concept, Implementation and Field Test Results” In Second International Conference on the Digital Society.

[26] Samuel Marchal, Jerome Francois, Radu State, Thomas Engel, “PhishStorm: Detecting Phishing with Streaming Analytics”, in proceedings of the IEEE Transactions on Network and Service Management, 2014.

## Authors

**Shivangi Sharma**, She was born on 4h Oct, 1991 in Hoshiarpur, State Punjab, India. She has done B.Tech from Rayat – Bahra College of Engineering and Nano-Technology, Hoshiarpur and M.Tech from GNDU Regional Campus, Jalandhar during 2014-2016. Her research interests are in the field of Network Security.