IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Volume-3, Issue-2 E-ISSN: 2347-2693

Detailed Survey on Phishing and Anti-Phishing Techniques

Anirban Bhowmick
Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal, India
anirban.bhowmick1993@gmail.com

www.ijcaonline.org

Received: Jan /09/2014 Revised: Feb/08/2014 Accepted: Feb/04/2014 Published: Feb/28/ 2014

***Abstract*—** Internet has been a pathway for cybercrimes. The facility internet has delivered is vast but at the same time data privacy of individuals has been risked.

Phishing is an example of social engineering techniques used to deceive users. In this paper, the authors made an attempt to enlighten readers with the different aspects of phishing. Public awareness is essential to combat such crimes. Different types of phishing techniques, their avoidance and detection has been presented in this paper. Further, a section highlights numerous research works on anti-phishing techniques. The later part of the paper illustrates the phishing scams and statistics for greater understandability of the problem. The readers are also informed about the various anti-phishing groups and where to report any sort of suspicious phishing activity.

***Keywords*—** Phishing, Internet based security, Anti-Phishing

I. INTRODUCTION

Internet plays a vital role in the modern lifestyle offering a wide range of applications including online reservations, sending emails, online shopping, access to social networking sites and education related material is readily available. Internet has provided man immense facilities on the cost of data confidentiality. Security of information over the internet has always been a major concern.

Phishing is an online identity theft in which an attacker, also known as a phisher, tries to deceptively retrieve a user's confidential data. The word phishing appeared around in 1995, when internet scammers by means of email trap to *fish* for passwords and financial information from internet users.

Phishing includes:

- Deceptive attacks, in which users are misled by fraudulent messages or emails into giving out personal information to phishers.
- Malware attacks, in which malicious software installed in the victim's local machine causes data compromises.
- DNS-based attacks, in which the look-up of host names is changed to direct users to a fraudulent website.

Before researching further about phishing it's important to explain what not phishing is. Nigerian 419 scam which involved sending emails to trick receivers into giving money to the scammer and internet auction are not considered phishing since they don't involve gaining user's authorizations.

A phishing attack encompasses three roles of phishers.

1. *Mailers* send out fake emails, which mislead users to fake websites which is the exact copy of a legitimate reputed website.
2. *Collectors* set up these deceitful websites which request users to deliver personal information.
3. *Cashers* use this confidential information to accomplish a pay-out.

United States is the prime host of phishing, accounting for 43% of phishing sites reported in January 2012. Germany 6%, followed by Australia, Spain, Brazil, Canada, the U.K., France, Netherlands, and Russia. A study suggests that women are more vulnerable to phishing than men and users between the ages of 18 and 25 are more susceptible to phishing than other age groups. The effect of phishing on the global economy has been quite significant. RSA estimates that worldwide damages from phishing attacks cost more than $1.5 billion in 2012, and had the potential to reach over $2 billion if the average uptime of phishing attacks had remained the same as 2011

II. TYPES OF PHISHING

Phishing is the method used to steal personal information through deceptive means. There are a number phishing techniques used to obtain confidential data from users. As technology becomes more advanced, more phishing techniques are also being developed. Below are some major types of phishing [1] [2].

A. Deceptive Phishing

In this case, phishers send the same email to millions of users requesting them to fill their personal details such as passwords, usernames, security codes, and credit card numbers on a specific website. The link provided in this email will redirect the user to a fake website which will be a careful replica of a reliable genuine website. The phisher also employs address deceiving so that the email seems to

be from the original source. The email can claim to be a re-send of the original or a restructured version as a trapping strategy.

B. Malware Phishing

Phishing scams involving malware need it to run on the target's local system. The malware is generally attached to the email, in the form of a link, directed to the user by the phishers or may also be attached to downloadable files existing in that email. Once the user clicks on the link or downloads the file, the malware will start functioning. Malware are mounted into victim's computer to gather undisclosed data. In some cases, the malware look to support other techniques. Malware phishing is a rising concern for small and medium businesses as they are not able to keep their software updated the whole time.

C. Keyloggers and Screenloggers

Keyloggers and screenloggers are malwares that read input from the keyboard and forward appropriate data to the hacker over the Internet. The malware can implant themselves into the browsers as small programs known as assistant objects that run spontaneously when the browser is underway. To avert keyloggers from retrieving personal credentials, websites offer choices to use mouse click to make data entries through the virtual keyboard.

D. Session Hijacking

In session hijacking or cookie hijacking, the phisher exploits the web session control mechanism to gain illegitimate access to data or services in a computer system. In this case, the user's activities are observed until they sign in to a target account or try making a transaction which will necessitate filling in their private details like credit card number and password. At that point the software takes over and can start illegal actions, such as transferring funds, without the user's knowledge.

E. Web Trojans

Web trojans pop up when users are trying to log in. They assemble the user's credentials. The user believes to be entering the data on a website but actually it is being entered locally and then transmitted to the phisher for misuse.

F. DNS-Based Phishing

Pharming refers to hosts file alteration or Domain Name System (DNS)-based phishing. When an individual enters a URL of a website it is first decoded into an IP address before it is communicated over the Internet. The bulk of user's PCs running a Microsoft Windows operating system first look up these host names in their hosts file before undertaking a Domain Name System (DNS) lookup. With the pharming system, attackers interfere with a company's host files or domain name system so that requests for URLs return a fraudulent address and subsequent communications are focused to a fake website. The users are ignorant that the website where they are entering personal data is managed by phishers.

G. System Reconfiguration Attacks

This attack causes modifications in the settings in the user's PC for malicious purposes. In this case, phishers send a message whereby the user is requested to reconfigure the settings of the computer. The message may originate from a web address which resembles a trustworthy genuine source. There are cases in which the URLs in a favorites file are altered to mislead users to identical websites. For example - a university website URL may be altered from university.xyz.com to university-xyz.com.

H. Content Injection Phishing

Content injection is the method where the phisher changes a part of the content on the page of a reliable website. This is done to deceive the user to go to a web page which does not belong to the legitimate website. Then, the individual is requested to enter personal information on the illegal website. This is a fake website managed by phishers to gain the confidential information of the individual.

I. Man-in-the-Middle Phishing

This type of phishing is tougher to sense comparatively. In these attacks, hackers place themselves between the user and the genuine website. They study the information being entered by the user but continue to permit the user on to the subsequent steps so that user transactions are not disturbed and the user remains uninformed. Later they can use the information collected when the user is not active on the system.

J. Phone Phishing

This type of phishing refers to messages that claim to be from a trusted legitimate source like a bank asking users to dial a phone number concerning difficulties with their bank accounts. Traditional phone equipment has dedicated lines, so Voice over IP, being easy to manipulate, becomes a decent choice for the phisher. Once the phone number, possessed by the phisher and provided by a VoIP service, is dialed, voice prompts tell the individual to enter their account numbers and PIN. Caller ID deceiving can be used alongside so that the call seems to be from a reliable source.

III. PHISHING DETECTION TECHNIQUES

A few of the phishing detection techniques [3] that can be put into practice are

1) *Content Based Approach*

This method measures the resemblance between two websites by comparing the content elements like text,

images and format of the two websites. Algorithms are used to calculate similarity between two websites to detect the phishing web pages which have higher similarities to phishing targets. It necessitates finding the phishing target prior to the resemblance comparison computation.

2) *Heuristics Based Approach*

This method rates the phishing probabilities of a particular webpage using reputation scores either obtained from the anti-phishing community or computed from the given webpage. However, the consistency of the reputation scoring is a great challenge.

3) *Black Listing Based Approach*

Blacklist is collection of recognized phishing websites or webpages published by reliable entities like Google and Microsoft. These blacklisted websites are non-trusted or banned. Web browsers compare the URL of a website against a blacklist of known fraudulent websites to check for the dependability of the visited webpage. If the user enters the blacklist website, a warning will be displayed. However, this technique is not appropriate to sense the new phishing attacks.

IV. PHISHING AVOIDANCE TECHNIQUES

Phishing can be eradicated by user awareness. Users have to be educated and made alert of such scams. A number of technological solutions have been put forward but if the individual behind the keyboard falls for a phishing attack, the technological solutions won't matter.

A survey on effectiveness of numerous anti-phishing educational materials suggests that educational resources reduced user's inclination to enter data into phishing webpages by 40%; however, this has also discouraged users from clicking on legitimate links. This leads to the acceptance that it is of utmost importance to find a novel and well-organized method of educating a bulky proportion of the population. Once there is awareness among the people, it becomes difficult for any phisher to accomplish the task. Users can combat phishing by the following ways:

1) *Do not click on suspicious hyperlinks in e-mails*

Users should avoid clicking on any distrustful hyperlink in an e-mail, especially from unidentified sources. These hyperlinks can redirect the user to a fraudulent website. The user can inspect the website link by manually typing it into a web browser.

2) *Verify HTTPS*

Modern web browsers have certain built-in security indicators that can safeguard users from phishing scams, including domain name highlighting and HTTPS indicators. While entering personal information like credit card number and password, the individual should make sure the address bar displays *https://* rather than *http://* and that there is a protected lock icon at the bottom right hand corner of the browser.

3) *Secure Host File*

A phisher can compromise the hosts file on the target's system and direct the individual to a deceitful website. Configuring the host file to read-only can solve the problem, but whole security will be determined by having a decent firewall that will provide a guard against interference by external attackers.

4) *Avoid entering confidential data in pop-up windows*

Users should avoid entering vital information like credit card credentials in pop-up windows even if the it seems certified or claims to be protected because there is no way to check the security and legitimacy of these pop-ups.

5) *Don't provide personal credentials over the phone*

The user may be asked to provide financial details over the phone, the caller claiming to be from a reliable and genuine source. The phone call can be from a number which looks legitimate but the area code in the phone call can be altered using VoIP technology. The user should always be aware of phone phishing schemes and should not disclose personal information over the phone unless the user had initiated the phone call.

6) *Keep your softwares enabled and updated*

Anti-spyware and firewall settings should be enabled to avert phishing attacks and users must update their softwares frequently. Further, antivirus software should be updated as most antivirus vendors have signatures that safeguard against some technology exploits. Antivirus softwares scan each and every file which comes through the internet to the system.

V. RELATED WORK

This section highlights the various techniques proposed and implemented by different researchers to protect individuals from getting phished.

Authors in [4] have recommended a technique to avert phishing by using an amalgamation of one time password (OTP) and encrypted token for user machine identification. In the first step, the user receives the password via SMS or by alternative emails. At the same time, the encrypted token is created which has user specific data stored in the user machine. The next step is to access the required website with the password and legitimate token which is necessary for successful authentication.

In [5], authors have suggested a PageSafe model to prevent phishing. This PageSafe model depends on user input to decide the validity of a web page. It uses external information sources on the internet. PageSafe prevents accesses to phishing websites and warns against DNS poisoning attacks.

Authors in [6] describe a new approach to diminish spear phishing attacks through the use of document authorship techniques - Anti-Spear phishing Content-based Authorship Identification (ASCAI). ASCAI notifies the user of likely disparities between the writing styles of a deceptive email and email from reliable authors by detailed study of the email body.

In [7], the authors have proposed a method in which the URLs are examined before visiting the actual website, so as to deliver security against web attacks. This method uses several parsing operations and query processing which uses many techniques to detect the phishing attacks and other web attacks. This approach is entirely based on operation through the browser and hence only affects the speed of browsing.

In [8], the authors have put forward a phishing detection approach, PhishZoo that uses appearances of trusted websites to sense phishing. The method delivers similar precision to blacklisting approaches (96%), with the improvement that it can classify zero-day phishing attacks and targeted attacks against smaller websites.

In [9] the authors suggest approaches to detect replication of website layout and structure through source code (and optionally image) fingerprinting. This Anti phishing method is based on URL and Domain Identity, and Image Based Webpage Matching. It initially recognizes the related authorized URL in which approximate string matching algorithm is used. The image based matching mechanism uses key point's detection and feature extraction methods.

Authors in [10] propose a technique to visually compare a suspected phishing page with the legitimate one. The aim is to determine whether the two pages are similar. Signature based algorithm is used. Further, the proposed algorithm is inspired by two previous open source anti-phishing solutions: the Anti-Phish browser plug-in and its DOM Anti Phish extension.

In [11], the authors put forward a technique which is purely constructed on image comparison using discriminative key point features in web pages. They used an invariant content descriptor, the Contrast Context Histogram (CCH), to calculate the resemblance between suspicious pages and legitimate web pages. This anti phishing tool is highly efficient and error free. It can be used in online banking, online shopping and to maintain the mail accounts.

Authors in [12] have suggested a technique which detects phishing activity without opening a phishing web page. This method makes use of a hybrid technique called Adaptive Neuro Fuzzy Inference System (ANFIS). Neural networks and Fuzzy logic have been used to effectively counter the phishing attacks. The detection rate is 98% .There is no false positives present, which may lift up a false alarm and classify a genuine email as a phishing email.

Authors in [13] have suggested a method which uses four features - HTML crosslink check, false info feeder check, SSL handshake and Certificate Suspicious check. This indicates that use of attribute-based anti-phishing checks can deliver a solid defense against phishing. This technique has been employed in Phish Bouncer tool.

VI. PHISHING SCAMS

Organized forces around the world executed exceptionally sophisticated phishing scams to aim a variety of organizations and leaders. Here are some notorious phishing scams.

- In August 2013, a few days before Iran's national election to choose a successor to President Mahmoud Ahmadinejad, thousands of Gmail account users in Iran were targeted in phishing attack intended to influence the election. The attacks originated inside Iran and had been occurring for about three weeks.
- In late 2013, a man was arrested for his part in a phishing scam targeting UK college students to steal in excess of £1.5m. The scam sent emails inviting students to update their student loan details on a malicious site that took large amounts of money from their accounts.
- In January 2009, a man was tricked in providing his Facebook account details. He was likely a victim of spear phishing. He had responded to an email that had asked him to click on a link to his Facebook account where he provided his username and password. A deceiving message was sent to all his friends informing that he was robbed and was in a need for money. They were asked to send the money to Western Union branch in London.
- In late 2004, a lady received an email from a fake bank website warning that her bank account would be suspended unless she updated her account to meet the company's new anti-fraud techniques. She clicked on the link that came with her email and provided her account particulars. Subsequently, all her money disappeared.

- In April 2013, an AP journalist clicked on a spear phishing email disguised as a Twitter email. The phisher then hacked journalist's Twitter account. Stock markets plunged after a deceiving tweet about an explosion at the White House, erasing $136.5 billion of value from the S&P 500 index.

VII. ANTI PHISHING GROUPS

A. *The Anti-Phishing Working Group (APWG)*

APWG is a global group that brings together businesses affected by phishing attacks, security products and services companies, law enforcement agencies, government agencies, trade association, regional international treaty organizations and communications companies.

Founded in 2003 by David Jevans, the APWG has more than 3200 members from more than 1700 companies and agencies worldwide. Member companies include leading security companies such as BitDefender, Symantec, VeriSign, McAfee, IronKey and Internet Identity. Financial Industry members include the ING Group, VISA, Mastercard and the American Bankers Association.

B. *Phish Tank*

Phish Tank, launched in October 2006, is a collective clearing house for data about phishing on the internet. The company offers a community-based phish verification system where users submit suspected phishes and other users vote if it is a phish or not. It also delivers an open API for developers and researchers to incorporate anti-phishing data into their applications. Phish Tank is supported by OpenDNS, a public DNS resolver; OpenDNS uses Phish Tank data to avert phishing attacks for their users.

VIII. REPORT PHISHING

The best method is to report the fear of being phished to an group that can investigate further and stop such cybercrimes. There are several places on the internet where the reporting can be done [14].

- One is the U.S. government-operated website which delivers information where to direct a copy of the deceiving email or the fraudulent URLs so that they can be examined by authorities. It also contains the particulars on phishing scams and how to identify them and to protect personal information. The link is provided below
  http://www.us-cert.gov/nav/report_phishing.html

- Another website to report such scams is the Anti-Phishing Working Group (APWG). This website features an option where the user can copy and paste the matters of the doubtful email. The link is provided below http://antiphishing.org/report-phishing/

IX. STATICTICS

This section presents the phishing activity statistics published by Anti-phishing Working Group (APWG) for the 1st quarter of 2014 [15]. These months saw the second highest number of phishing attacks ever documented in a first quarter by the APWG in its Phishing Activity Trends Report.

The APWG keeps a track of the number of exclusive phishing websites. This is determined by the distinctive base URLs of the phishing sites. Unique phishing websites detected between January 2014 and March 2014 is shown below, Figure 1.

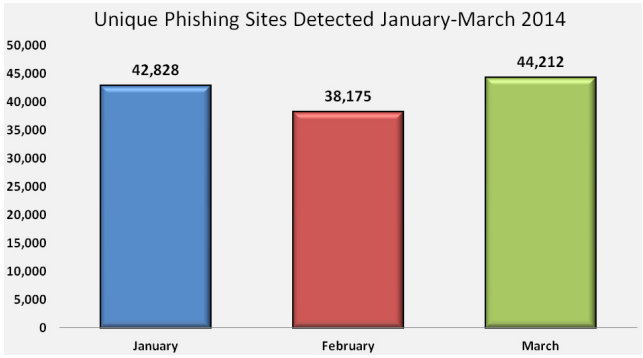

Fig 1: Unique Phishing Websites detected between January and March 2014

The number of exceptional phishing reports given to APWG during quarter 1 of 2014 was 171,792. This was an increase for the 6.8 percent increase from 160,777 received in quarter of 2013.

The number of unique phishing reports provided to APWG rose by nearly 7,000 during the three month period. The chart below (Figure 2) demonstrates the unique phishing reports in the1st quarter of 2014.

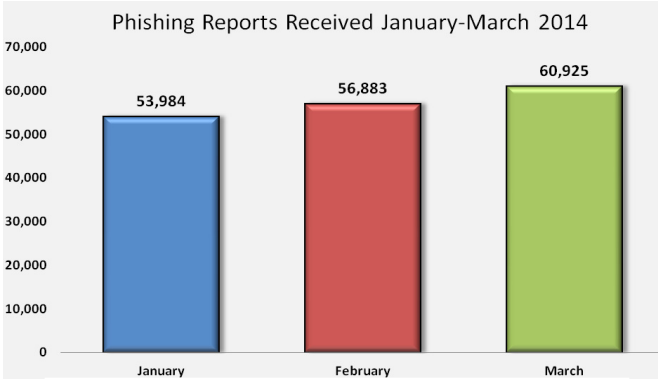

Fig 2: Phishing Reports between January and March 2014

The following figure (Figure 3) chains figures grounded on brands phished, unique domains, unique domain/brand pairs, and unique URLs. Brand/domain pairs count the exclusive occurrences of a domain being used to target a specific brand. If the number of unique URLs is greater than the number of brand/domain pairs, it designates many URLs are being hosted on the same domain to target the same brand. Knowing how many URLs occur with each domain shows the estimated number of attacking domains a brand-holding target needs to trace and counteract. Since a phishing prevention technology needs the full URL so as to prevent over-blocking, it is beneficial to appreciate the overall number of distinctive URLs that happen per domain.

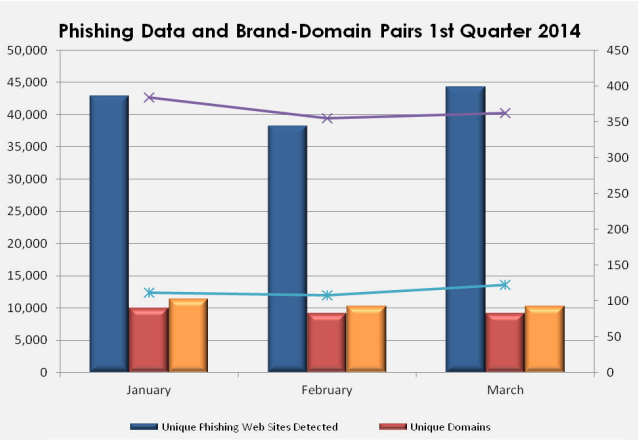

Fig 3: Phishing Data and Brand-Domain Pairs between January and March 2014

The quantity of brands targeted stayed relatively consistent during quarter 2014.

	January	February	March
Number of unique phishing websites detected	42,828	38,175	44,212
Unique Domains	9,918	9,088	9,152
Unique Brand-Domain Pairs	11,351	10,214	10,275
Unique Brands	384	355	362
URLs Per Brand	111.53	107.53	122.13

Table 1: Brands targeted between January and March 2014

A total of 557 brands were targeted by phishers in the first three months of 2014. This was up from the 525 targeted in the fourth quarter of 2013. The number of brands targeted in any given month remained below the all-time high of 441 that was recorded in April 2013. Table 1 illustrates the brands targeted in the 1st quarter of 2014. Further, Figure 4 provides an estimate of the hijacked brands during the months January, February and March.

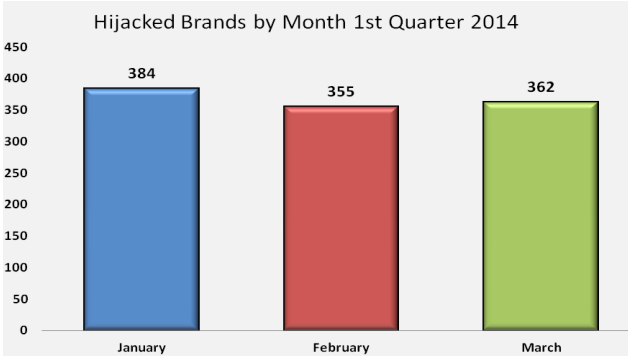

Fig 4: Brands Hijacked between January and March 2014

Payment Services continued to be the most-targeted industry sector at the beginning of 2014, with 46.51 percent of attacks during the three-month period (Figure 5).

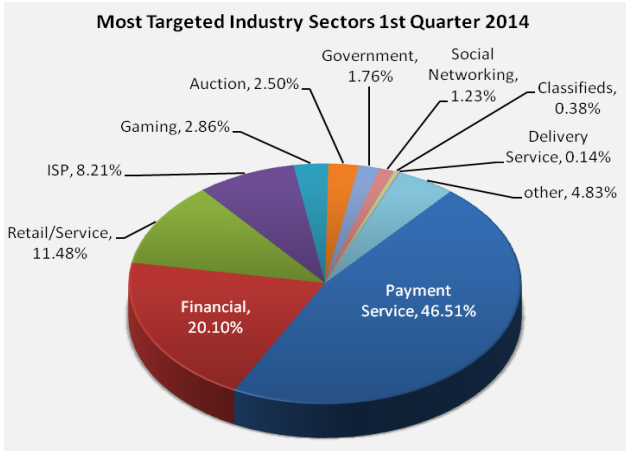

Fig 5: Most Targeted Industry Sectors between January and March 2014

Further, United States continued to be the top country hosting phishing sites during the first quarter of 2014. This is mainly due to the fact that a large percentage of the world's Web sites and domain names are hosted in the United States. Also, the statistical highlights for 1st quarter of 2014 is presented in a tabulated form in Table 2.

	January	February	March
Number of unique phishing websites detected	42,828	38,175	44,212
Number of unique phishing e-mail reports received by APWG from consumers	53,984	56,883	60,925
Number of brands targeted by phishing campaigns	384	355	362
Country hosting the most phishing websites	USA	USA	USA
Contain some form of target name in URL	56.76%	54.31%	64.47%
Percentage of sites not using port 80	0.85%	0.42%	0.56%

Table 2: Statistical Highlights between January and March 2014

X. CONCLUSION

User ignorance is a reason that people are trapped in giving out their personal credentials to phishers. This paper edifies the readers with numerous aspects of phishing so that users can take steps to safeguard themselves from getting phished. The statistics presented in the paper reveal a lot about the losses incurred as a consequence of such cybercrimes in the 1st quarter of 2014. Organizations like Phish Tank and Anti-Phishing Working Group aim to spread awareness among the people and eradicate such crimes. People also need to play their part to discourage phishers from succeeding in such illegal activities.

REFERENCES

[1] Vignesh.M, Gokul Ram.T, Akhil.R, Rajesh Kumar N.S.R, Karthikeyan.R, "Analysis of Phishing in Networks", International Journal of Scientific & Engineering Research, Volume 4, Issue 9, September-2013 196 ISSN 2229-5518

[2] Aanchal Malhotra, Navdeep Kaur, "Browser Prevention Against Phishing Website Security Risk", IJCSC, Vol. 3, No. 1, January-June 2012, pp. 215-219

[3] Kanchan Meena, Tushar Kanti, "A Review of Exposure and Avoidance Techniques for Phishing Attack", International Journal of Computer Applications (0975 – 8887) Volume 107 – No 5, December 2014.

[4] Ahmad Alamgir Khan, "Preventing Phishing Attacks using One Time Password and User Machine Identification", International Journal of Computer Applications (0975 – 8887) Volume 68– No.3, April 2013.

[5] P. K. Sengar, Vijay Kumar, "Client-Side Defense against Phishing with PageSafe", International Journal of Computer Applications (0975 – 8887) Volume 4 – No.4, July 2010.

[6] Mahmoud Khonji, Youssef Iraqi, Andrew Jones, "Mitigation of Spear Phishing Attacks: A Content-Based Authorship Identification Framework", 6th International Conference on Internet Technology and Secured Transactions, 11-14 December 2011, Abu Dhabi, United Arab Emirates.

[7] Gaurav Kumar Tak, Gaurav Ojha, "Multi-Level Parsing Based Approach Against Phishing Attacks with the Help of Knowledge Bases", IJNSA Vol.5, No.6, November 2013.

[8] Sadia Afroz, Rachel Greenstadt, "PhishZoo: Detecting Phishing Websites By Looking at Them"

[9] T.Balamuralikrishna, N.raghavendrasai, M.Satya Sukumar "Mitigating Online Fraud by Ant phishing Model with URL & Image based Webpage Matching", International Journal of Scientific & Engineering Research Volume 3, Issue 3, March - 2012.

[10] A.V.R.Mayuri "Phishing Detection based on Visual-Similarity" International Journal of Scientific & Engineering Research Volume 3, Issue 3, March - 2012.

[11] Mallikka Rajalingam, Saleh Ali Alomari, Putra Sumari, "Prevention of Phishing Attacks Based on Discriminative Key Point Features of WebPages", International Journal of Computer Science and Security (IJCSS), Volume (6): 2012.

[12] Shivender Singh, Anil K. Sarje, Manoj Misra, "Client-Side Counter Phishing Application using Adaptive Neuro-Fuzzy Inference System", 978-0-7695-4850-012 2012 IEEE.

[13] Michael Atighetchi, Partha Pal, "Attribute-based Prevention of Phishing Attacks", 978-1-4673-2104-412 2012 IEEE.

[14] M. Usha, P. Deepika, "Phishing - A Challenge in the Internet", International Journal of Computer Science and Information Technologies, Vol. 5 (1), 2014, 260-26.

[15] An article on "Phishing Activity Trends Report 1st Quarter 2014" available at: docs.apwg.org/reports/apwg_trends_report_q1_2014.pdf.