ISSN 2229-5518

Phishing Susceptibility and Anti-Phishing Security Strategies-Literature Review

Atul M. Tonge , Surbhi R. Chaudhari

Abstract— Phishing is a con game that scammers use to collect personal information from unsuspecting users. The false e-mails often look surprisingly legitimate and even the Web pages where users are asked to enter their information may look real. Phishing is similar to fishing in a lake, but instead of trying to capture fish, phishers attempt to steal personal information. This paper gives brief information about phishing, its attacks, steps that users can take to safeguard their confidential information. Phishing has seen an alarming trend of increase in both the volume and the sophistication of phishing attacks. This paper also shows a survey conducted by netcraft on phishing. Phishing is a form of online identity theft that aims to steal sensitive information such as online banking passwords and credit card information from users. The effectiveness of this paper is to examined in a large-scale dataset collected from real phishing cases and to study different approaches for handling phishing activities. Anti-Phish, that aims to protect users against spoofed web site-based phishing attacks. To this end, Anti-Phish tracks the sensitive information of a user and generates warnings whenever the user attempts to give away this information to a web site that is considered untrusted. Anti-Phish is based on the premise that for inexperienced, technically unsophisticated users, it is better for an application to attempt to check the trustworthiness of a web site on behalf of the user. Unlike a user, an application will not be fooled by obfuscation tricks such as a similar sounding domain name.

In this paper various methods reviewed to confront those challenges, such as Textual and Visual Content-Based Anti-Phishing: A Bayesian Approach, Image-Based Page Matching And Page Classification with anti-phishing strategies.

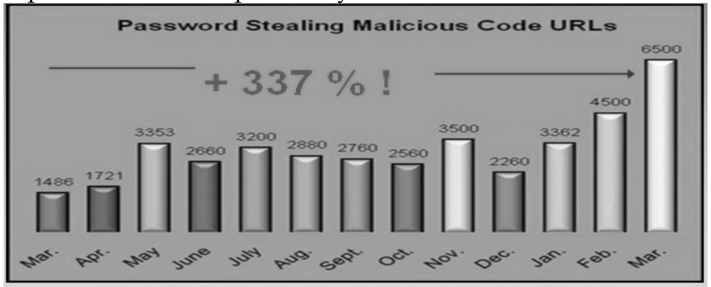
Index Terms—Anti-phishing strategy, Antiphishing Working Group(APWG) ,Awareness and Education, Contrast Context Histogram (CCH), Image Classifier, Netcraft Web Survey, Phishing attacks.

1 INTRODUCTION

PHISHING is a form of online identity theft that aims to steal sensitive information such as online banking passwords and credit card information from users. Phishing scams have been receiving extensive press coverage because such attacks have been escalating in number and sophistication. According to a study by Gartner, 57 million US Internet users have identified the receipt of e-mail linked to phishing scams and about 2 million of them are estimated to have been tricked into giving away sensitive information. The last years have brought a dramatic increase in the number and sophistication of such attacks. Although phishing scams have received extensive press coverage, phishing attacks are still successful because of many inexperienced and unsophisticated Internet users. Attackers are employing a large number of technical spoofing tricks such as URL obfuscation and hidden elements to make a phishing web site look authentic to the victims. 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. However, expecting that all users will understand the phishing threat and surf accordingly is unrealistic. There will always be users that are tricked into visiting a phishing web site. Therefore, it is important for researchers and industry to provide solutions for the phishing threat.

Anti-Phish, that aims to protect users against spoofed web site-based phishing attacks. To this end, Anti-Phish tracks the sensitive information of a user and generates warnings whenever the user attempts to give away this information to a web site that is considered untrusted. Anti-Phish is based on the premise that for inexperienced, technically unsophisticated users, it is better for an application to attempt to check the trustworthiness of a web site on behalf of the user. Unlike a user, an application will not be fooled by obfuscation tricks such as a similar sounding domain name[1,2].

The convenience of online commerce has been embraced by consumers and criminals alike. There has been good progress in identifying the threat, educating businesses and customers, and identifying countermeasures. However, there has also been an increase in attack diversity and technical sophistication by the people conducting phishing and online financial fraud. Phishing has a negative impact on the economy through financial losses experienced by businesses and consumers, along with the adverse effect of decreasing consumer confidence in online commerce. Phishing scams have flourished in recent years due to favorable economic and technological conditions. The technical resources needed to execute phishing attacks can be readily acquired through public and private sources. Some technical resources have been streamlined and automated, allowing use by non-technical criminals. This makes phishing both economically and technically viable for a larger population of less sophisticated criminals .we will identify several of the technical capabilities that are used to conduct phishing scams, review the trends in these capabilities over the past two years.



Source: Anti-Phishing Working Group (Antiphishing.org)

Figure 1: Statistic of password stolen victims until March 2012

2 Literature Review

2.1 Phishing Today

Originally, phishing was identified as the use of electronic mail messages, designed to look like messages from a trusted agent, such as a bank, auction site, or online commerce site. These messages usually implore the user to take some form of action, such as validating their account information. These messages often use a sense of urgency (such as the threat of account suspension) to motivate the user to take action. Recently, there have been several new social engineering approaches to deceive unsuspecting users[3].

These include the offer to fill out a survey for an online banking site with a monetary reward if the user includes account information, and email messages claiming to be from hotel reward clubs, asking users to verify credit card information that a customer may store on the legitimate site for reservation purposes. Included in the message is a URL for the victim to use, which then directs the user to a site to enter their personal information. This site is crafted to closely mimic the look and feel of the legitimate site. The information is then collected and used by the criminals. Over time, these fake emails and web sites have evolved to become more technically deceiving to casual investigation. Recently the definition of phishing has grown to encompass a wider variety of electronic financial crimes. In addition to the widespread use of these fake email messages and web sites to lure users into divulging their personal information, we have also observed an increase in the amount of malicious code that specifically targets user account information. Once installed on a victim's computer, these programs use a variety of techniques to spy on communications with web sites and collect account information.

This method differs from the technical subterfuge generally associated with phishing scams and can be included within the definition of spyware as well. It is important to include them in a discussion on phishing trends for the following reasons:

1] Social component – Criminals often use social engineering along with vulnerabilities in applications such as web browsers or email clients to trick users into installing malicious code on their computer.

2] Common infrastructure – We have observed the use of common tools and techniques for delivering phishing emails and distributing malware. These include the use of botnets, open mail relays, and compromised web sites to host phishing sites and malware.

3] The big picture - As countermeasures are implemented to thwart one method of stealing information, criminals still have additional opportunities available to them. It is important to understand the technical capabilities available to these criminals so that more effective measures for protecting customer information can be developed and law enforcement personnel tasked with tracking down and prosecuting criminals conducting phishing scams can be more effective.

3 Phishing Attack Stages

Phishing attacks involve several stages:

• The attacker obtains E-mail addresses for the intended victims. These could be guessed or obtained from a variety of sources.

• The attacker generates an E-mail that appears legitimate and requests the recipient to perform some action. •
The attacker sends the E-mail to the intended victims in a way that appears legitimate and obscures the true source.

• Depending on the content of the E-mail, the recipient opens a malicious attachment, completes a form, or visits a web site.

• The attacker harvests the victim's sensitive information and may exploit it in the future.

As shown in Figure 2 below, the phishing attack starts with an E-mail to the intended victims. The attacker creates the E-mail with the initial goal of getting the recipient to believe that the E-mail might be legitimate and should be opened. Attackers obtain E-mail addresses from a variety of sources, including semi-random generation, skimming them from Internet sources, and address lists that the user believed to be private [CNET]. Spam filtering can block many of the phishing Emails. If the institution whose customers are being phished regularly uses authenticated E-mail (such as PGP or S/MIME), the recipient may notice that the E-mail does not have a valid signature, thereby stopping the attack. Once the E-mail is opened by the user, the E-mail contents have to be sufficiently realistic to cause the recipient to follow the directions in the Email[4].

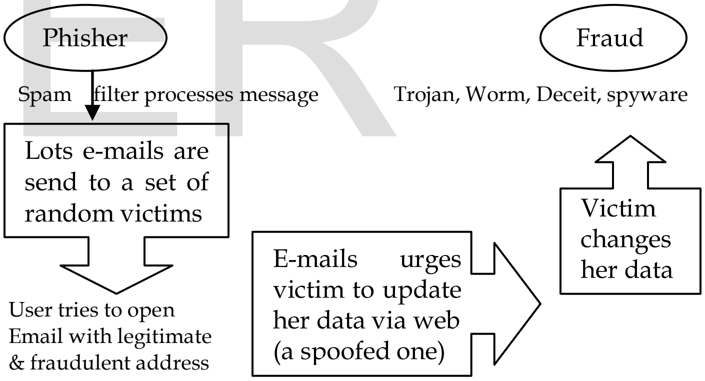

Phisher
Fraud
Spam filter processes message
Trojan, Worm, Deceit, spyware
Lots e-mails are send to a set of random victims
Victim changes her data
E-mails urges victim to update her data via web (a spoofed one)
User tries to open Email with legitimate & fraudulent address

Figure 2: phishing attack stages

4 Anti-phishing Techniques

Anti-phishing technique can be considered as an approach to counter the threats put forth by phishers. This accounts to a number of techniques followed which is categorize as follows.

4.1 List Based Approach

This is probably the most straightforward solution for anti-phishing. A white list contains URL's of known legitimate sites. Many current anti-phishing techniques rely on the combination of white list and blacklist. The representative blacklist/white list based systems include Phish Tank Site Checker, Google Safe Browsing, Fire Phish and CallingID Link Advisor. These anti-phishing solutions are usually deployed as toolbars or extensions of web browsers to remind the users whether they are browsing a safe website. Blacklist suffers from a window of vulnerability between the time a phishing site is launched and the site's addition to the blacklist as it requires frequent updating which is the case for white list also[5].

4.2 Heuristics Based Approach

This technique rates the phishing possibility of a given webpage using reputation scores either obtained from the anti-phishing community or computed from the given webpage. However the reliability of the reputation scoring is a great challenge[5].

4.3 Content Based Approach

This method is used to measure the similarity between two given web pages by calculating the similarity between the content elements (text, image, layout) contained in the web pages. Algorithms are used to compute visual similarity to detect the phishing web pages which have higher similarities to phishing targets. It requires finding the phishing target prior to the similarity comparison computation. It also combines TF-IDF retrieval algorithm to determine the likelihood that a given webpage is a phishing webpage. Words with highest TF-IDF weight on a given webpage can be used to classify the webpage as legitimate or not. This method used CANTINA, a content-based approach to detect phishing websites, which combines a Term Frequency- Inverse Document Frequency (TF-IDF) information retrieval algorithm with heuristics and determines the likelihood that a given webpage is a phishing page. CANTINA uses the five words with the highest TF-IDF weight on a given webpage as the lexical signature of that site and submits them to Google. If CANTINA finds the URL of the site in question within the top results, it classifies that as legitimate webpage or otherwise as phishing webpage.

4.3.1 Image-Based page Matching Approach

Phishers can compose visually similar phishing pages in many different ways with nontext HTML elements, such as images and Flash objects (see the "Current Antiphishing Approaches" sidebar for more about such techniques). To combat this problem, we compute the similarity of phishing pages and authentic pages at their presentation level. Specifically, we treat phishing page detection as an image-matching problem. Figure 3 illustrates the flow of detection scheme, which involves two steps: image-based page matching and page classification. In the scheme, we first take a snapshot of a suspect Web page and treat it as an image throughout the detection process. We use the Contrast Context Histogram (CCH) descriptors proposed in earlier work to capture invariant information around discriminative keypoints on the suspect page. We then match the descriptors with those of authentic pages that are often targeted by phishers, where the pages are stored in a database compiled by users and authoritative organizations, such as the APWG. Matching CCH descriptors yields a similarity degree for a suspect page and an authentic page. Finally, we use the similarity degree between the two pages to determine whether the suspect page is a counterfeit. If the similarity degree between a suspect page and an authentic one is greater than a certain threshold, we consider the suspect page to be a phishing page for the authentic one or genuine if it's not a phishing page for an authentic pages in the database[6].

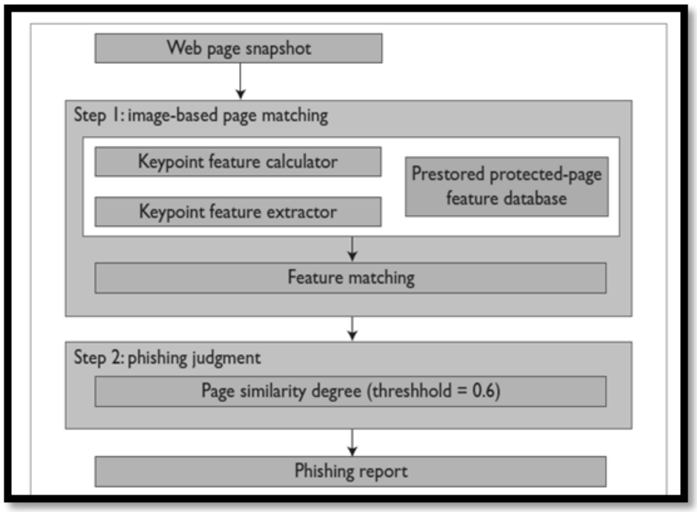

Web page snapshot
Step 1: image-based page matching
Keypoint feature calculator
Keypoint feature extractor
Prestored protected-page feature database
Feature matching
Step 2: phishing judgment
Page similarity degree (threshhold = 0.6)
Phishing report

Figure 3 The flow of the phishing-detection scheme. We first take a snapshot of a suspect page and extract its keypoint feature information. Next, we match the features with the keypoint feature information of protected Web pages. We can then asses the suspect page to determine whether it's a phishing page.

4.3.2 Textual and Visual Content-Based Anti-Phishing: A Bayesian Approach

In this the content of a given web page is transformed into two categories, namely, the textual and the visual, which is addressed by the corresponding classifier. The anti-phishing approach contains the following components.
1) A text classifier using the naive Bayes rules to handle the text content extracted from a given web page.
2) An image classifier using the EMD similarity assessment to handle the pixel level content of a given web page that has been transformed into an image.
3) A Bayesian approach to estimate the threshold used in classifiers through offline training.
4) A data fusion algorithm to combine the results from the text classifier and the image classifier. The algorithm employs the Bayesian approach as well[7].

Figure 3 illustrates an overview of framework. The system includes a training section, which is to estimate the statistics of historical data (i.e., web page training set), and a testing section, which is to examine the incoming testing web pages. The statistics of the web page training set consists of the probabilities that a textual web page belongs to the categories (i.e., phishing and normal), the matching thresholds of classifiers, and the posterior probability of data fusion. Through the preprocessing, content representations, i.e., textual and visual, are rapidly extracted from a given testing web page. The text classifier is used to classify the given web page into the corresponding category based on the textual features. The image classifier is used to classify the given web page into the corresponding category based on the visual content. Then the fusion algorithm is used to combine the detection results delivered by the two classifiers. The detection results are eventually transmitted to the online users or the web browsers[8].

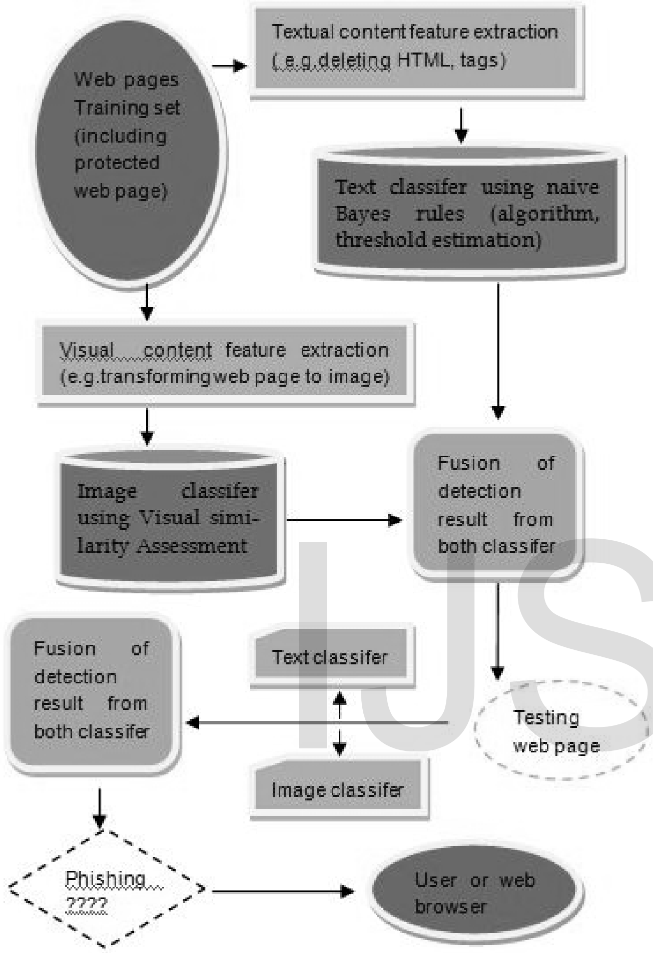

Web pages Training set (including protected web page)
Textual content feature extraction (e.g.deleting HTML, tags)
Text classifer using naive Bayes rules (algorithm, threshold estimation)
Visual content feature extraction (e.g.transforming web page to image)
Image classifer using Visual similarity Assessment
Fusion of detection result from both classifer
Fusion of detection result from both classifer
Text classifer
Testing web page
Image classifer
Phishing ????
User or web browser

Figure 4: Bayesian approach

4. 4 Hybrid Approach

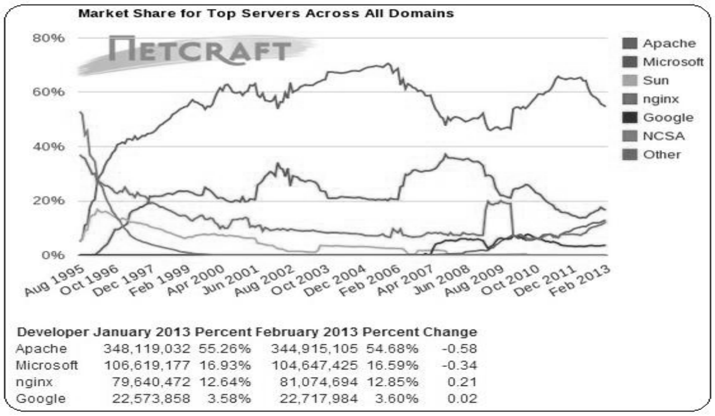
This approach usually combines any of the above mentioned techniques to classify a webpage as legitimate or not.

4.5 Phish Net

Huang *et al*. (2011) states that, a blacklist scheme used to detect phishing attacks is discussed. Blacklist approach based techniques reliance on exact match with blacklisted entries makes it easy for phishers to evade. Phish Net exploits this observation using two components. In the first component, five heuristics to enumerate simple combinations of known phishing sites to discover new phishing URL's. The second component consists of an approximate matching algorithm that dissects a URL into multiple components that are matched individually against entries in the blacklist (Lakshmi and Vijaya, 2012).The first grows blacklists by generating new URL variations from the original ones but after vetting them through DNS and content matching. The second component consists of an approximate matching data structure that assigns a score to each URL based on piece-wise similarity with existing URLs. PhishNet suffers from low false positives and is remarkably effective at flagging new URLs that were not part of the original blacklist. Evaluation with real-time blacklist feeds discovered around 18,000 new phishing URLs from a set of 6,000 new blacklist entries and itleads to very few false positives, 3% and negatives, 5% [5].

5 NETCRAFT'S WEB SERVER SURVEY 2013

Netcraft, an Internet services company that provides web hosting & web server analysis & has launched its February 2013 web server survey after responses from over 630,790,500 web sites 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 in phishing attacks next to Facebook. Netcraft is reporting blocking nearly 6,000 urls targeting taobao users[4].



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

Figure 5: Netcraft's Survey

6 PHISHING COUNTER MEASURES

Although there are multiple recommendations for countering phishing, the following list contains the ones most commonly implemented today to either combat phishing directly or to mitigate phishing-capable threats such as malware[9].

6.1 Awareness and Education

Originally, the primary advantage for criminals conducting phishing-related fraud was the lack of education and awareness of

a) the existence of financial crimes targeting internet users, and

b) the policies and procedures of online sites for contacting their customers regarding account information and maintenance issues.

Both of these issues are being addressed by the online commerce sites and the information security community though various awareness mechanisms. General information on phishing distributed in company email or on a company's web site. Alerts sent to customers about phishing scams directly targeting a specific company. Reminders to customers of corporate policies on contacting customers regarding their account. Papers and talks from the security community targeted to users and businesses.

When companies choose to implement a customer phishing awareness program, it is important that they educate employees as well. In particular, the employees who interact with customers should be knowledgeable about phishing so they can answer customers' questions.Finally, a significant portion of phishing awareness efforts have focused on the threats posed by phishing emails and web sites. However, there is a significant threat from malware that people need to be made aware of as well. When a phishing email or web site is properly identified by a consumer, he or she can easily correlate it with the action of trying to steal account information. However, if malware is detected on a user's computer, the common response is to follow instructions on isolating and removing the threat. The user may not be aware of the functionality of the malware and thus the correlation to the action of trying to steal account information may not be clear.

6.2 Targeting Hosting Sites

One advantage to those who seek to shut down phishing sites is that there is little stealth in the sending of phishing emails. Since phishing emails use the same mass-mailing infrastructure as spam, affected institutions, ISPs, and law enforcement can be made aware of a site hosting a phishing scam and take efforts to get that site shut down. Data from the Anti-Phishing Working Group's trends reports show an approximate decrease of 10% in the average time online for a phishing site from October 2004 to April 2005 (6.4 days to 5.8 days) .Phishers have taken steps to make phishing sites hosting scam pages and the drop sites for compromised data more survivable. This includes the use of dynamic DNS entries and/or port-level redirectors to make networks of phishing sites more resistant to failure of any individual node. If a site that was pointed to by a hostname is taken down, it is possible for a phisher who is using dynamic DNS or a hosting service that provides DNS management to change the hostname to resolve to another compromised machine to serve as a phishing site.

Another form of redirection is the use of a 3rd party to collect information before transmission to its final destination. This could be the use of a 3rd party forms provider used to collect data as part of an e-mail based phishing attack or a script at a drop site for phishing malware that transports the data to a different host. In either case, take down of the publicly-visible site may stop the phishing attack, but not provide access to the cache of stolen information[9].

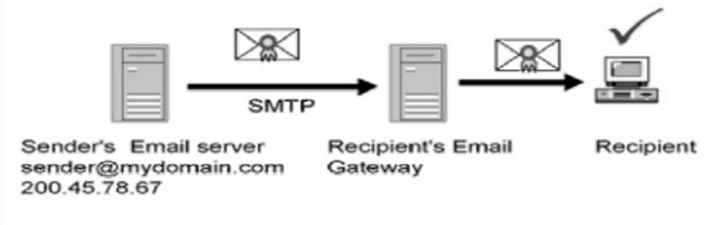


Figure 4: Digitally signed E-mail

6.3 Strong Authentication and Authorization

Two-factor authentication is a mechanism requiring two or more authenticators, usually consisting of something you know (such as a password or PIN) and something you have (such as a credit card or hardware token). For online commerce, two-factor authentication is being implemented by providing the customer with a hardware token for generating a continually changing component for their authentication credentials. The goal is to protect the users if their authentication credentials have been captured by an attacker via electronic surveillance. The timeliness of the ever-changing component limits the attacker's ability to use the credentials in the future. Another countermeasure being implemented by certain banks is the use of transaction numbers (TANs) for authorizing individual transactions. Customers are sent a list of TANs with their monthly statement, and they enter the next unused TAN when authorizing a transaction online. There are also implementations in which users receive a request for their TAN via an out-of-band mechanism, such as an SMS message on their cell phone. As noted previously, at least one Trojan (Bankash) attempts to trick customers of specific banks into divulging their next TAN. The criminal then has a limited but significant time window in which to use that TAN before the customer attempts another transaction.

HTTPS + Secure
Business Website
User's Browser

Figure 11: secure web authentication

5.4 Virus, Spyware, and Spam Prevention

Solutions designed to protect users from viruses , Trojans , spyware, and spam play a role in protecting users from phishing scams. With the marked increase in phishing malware, products that detect and prevent the installation and execution of malicious code are an essential part of an environment for secure home computing. These products must be enabled and in the case of anti-virus and anti-spyware products, must have up-to-date signatures. A large portion of recent malware attempts to disable anti-virus and antispyware software before a detection signature is able to detect and neutralize the malware.

Spam prevention has also contributed to the fight against phishing. Phishing emails use the same distribution mechanism as spam and usually have many of the same characteristics. Email filtering based on content blacklisting, Bayesian filtering, blocking mail from known spamming/phishing relays, anti-forgery solutions such as Sender Policy Framework (SPF) and Sender ID, and other heuristics specific to phishing can help prevent a great many phishing emails from ever reaching potential victims in the first place. However, spammers are continually evolving their tricks for bypassing filters.

7 Acknowledgment

It gives us a great pleasure to submit the paper topic titled—"Phishing Susceptibility and Anti-Phishing Security Strategies ".We wish to take this opportunity to express our heartiest gratitude with pleasure to J.D.I.E.T, Yavatmal, which gave us an opportunity in fulfilling our desire of reaching our goal.We are indebted to our proactive guide Dr. D.N. Chaudhari, H.O.D of cse dept. , J.D.I.E.T because without his valuable guidance this work would not have a success. His constructive, useful, timely suggestions and encouragement in every step immensely helped us to carry out our work.

8 Conclusion

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.

Users have become more aware of phishing crimes and how to identify unsophisticated phishing sites. In response, criminals are using web browser vulnerabilities and obfuscation techniques to create phishing scam pages that are more difficult to differentiate from legitimate sites; thus users can become victims even if they are not aware of phishing scams.

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. However, expecting that all users will understand the phishing threat and surf accordingly is unrealistic. There will always be users that are tricked into visiting a phishing web site. Therefore, it is important for researchers and industry to provide solutions for the phishing threat.

References

[1] The Antiphishing Working Group (2010) Home Page, http://www.anti- phishing.org.

[2] Ollman, G. (2011) The Phishing Guide – Understanding and Preventing, White Paper, Next Generation Security Software Ltd.

[3] The Antiphishing Working Group (2011) Phishing Activity Trends Report, http://www.anti-phishing.org.

[4] Jyoti Chhikara, Ritu Dahiya, Neha Garg, Monika Rani," Phishing & Anti-Phishing Techniques: Case Study", *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 3, Issue 5, ISSN: 2277 128X ,Page 458-465 , May 2013.

[5] C. Emilin Shyni and S. Swamynathan,"Protecting the Online User Information Against Phishing Attacks Using Dynamic Encryption Techniques", *Journal Of Computer Science*,9(4):526-533,ISSN 1549-3636, 2013.

[6] Kuan-Ta Chen, Chun-Rong Huang and Chu-Song Chen,"Fighting Phishing with Discriminative Keypoint Features", Published by *the IEEE Computer Society 1089-7801/09/$25.00* , K . J . Lin Yan Wang ,Institute of Information Science,Academia Sinica Columbia University, 2009 IEEE.

[7] Haijun Zhang, Gang Liu, Tommy W. S. Chow, Senior Member, IEEE, and Wenyin Liu, Senior Member, IEEE, "Textual and Visual Content-Based Anti-Phishing: A Bayesian Approach" ,*Ieee Transactions On Neural Networks* , Vol. 22, NO. 10, pg 1532-1546, October 2011.

[8] Wenyin Liu, Xiaotie Deng, Guanglin Huang and Anthony Y. Fu, "An Antiphishing Strategy Based on Visual Similarity Assessment" ,Published by *the IEEE Computer Society*, 1089-7801/06/$20.00 , City University of Hong Kong, march , april 2006.

[9] Engin Kirda, Christopher Kruegel, "Protecting Users Against Phishing Attacks",2005 Published by *Oxford University Press on behalf of The British Computer Society*. doi:10.1093/comjnl/bxh000.

- *Atul Manohar Tonge is currently pursuing Bachelor degree of four year program in Computer Science & Engineering in SGBAU Amravati University, Maharashtra, India, PH- +917709132024. E-mail: atultonge007e@rediffmail.com*
- *Surbhi Rajendra chaudhari is currently pursuing Bachelor degree of four year program in Computer Science & Engineering in SGBAU Amravati University, Maharashtra, India, PH- +917387308515. E-mail: surbhichaudhari12@gmail.com*