International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 09, September 2015
Available at http://internationaljournalofresearch.org

Anti Phishing Framework On Visual Cryptography

[1]Meenuga Jayanna & [2]C V Madhusudan Reddy

ABSTRACT :

Phishing is an attempt by an individual or a group to thieve personal confidential information such as passwords, credit card information etc from unsuspecting victims for identity theft, financial gain and other fraudulent activities. In this paper we have proposed a new approach named as "A Novel Antiphishing framework based on visual cryptography" to solve the problem of phishing. Here an image based authentication using Visual Cryptography (vc) is used. The use of visual cryptography is explored to preserve the privacy of image captcha by decomposing the original image captcha into two shares that are stored in separate database servers such that the original image captcha can be revealed only when both are simultaneously available; the individual sheet images do not reveal the identity of the original image captcha. Once the original image captcha is revealed to the user it can be used as the password.

KEYWORDS: Phishing; Visual Cryptography; Image Captcha; Shares; Security

1. INTRODUCTION

Online transactions are nowadays become very common and there are various attacks present behind this. In these types of various attacks, phishing is identified as a major security threat and new innovative ideas are arising with this in each second so preventive mechanisms should also be so effective .Thus the security in these cases be very high and should not be easily tractable with implementation easiness. Today, most applications are only as secure as their underlying system. Since the design and technology of middleware has improved steadily, their detection is a difficult problem. As a result, it is nearly impossible to be sure whether a computer that is connected to the internet can be considered trustworthy and secure or not. Phishing scams are also becoming a problem for online banking and e-commerce users. The question is how to handle applications that require a high level of security. 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. One definition of phishing is given as “it is a criminal activity using social engineering techniques. Phishers attempt to fraudulently acquire sensitive information, such as passwords and credit card details, by masquerading as a trustworthy person or business in an electronic communication”. Another comprehensive definition of phishing states that it is “the act of sending an email to a user falsely claiming to be an established legitimate enterprise into an attempt to scam the user into surrendering privateinformation that will be used for identity theft”. The conduct of identity theft with this acquired sensitive information has also become easier with the use of technology and identity theft can be described as “a crime in which the impostor obtains key pieces of information such as Social Security and driver's license numbers and uses them for his or her own gain”. So here introduces a new method which

can be used as a safe way against phishing which is named as "A novel approach against Anti-phishing using visual cryptography". As the name describes, in this approach website cross verifies its own identity and proves that it is a genuine website (to use bank transaction, E-commerce and online booking system etc.) before the end users and make the both the sides of the system secure as well as an authenticated one. The concept of image processing and an improved visual cryptography is used. Image processing is a technique of processing an input image and to get the output as either improved form of the same image and/or characteristics of the input image. In Visual Cryptography (VC) an image is decomposed into shares and in order to reveal the original image appropriate number of shares should be combined. This paper is organized as follows: Section 2 deals with the related work on Phishing and Section 3 describes Visual Cryptography. Section 4 & 5 deals with Current and proposed Methodologies. Section 6 presents the implementation and analysis. Section 7 contains the conclusions.

2. RELATED WORK Phishing web pages are forged web pages that are created by malicious people to mimic Web pages of real web sites. Most of these kinds of web pages have high visual similarities to scam their victims. Some of these kinds of web pages look exactly like the real ones. Victims of phishing web pages may expose their bank account, password, credit card number, or other important information to the phishing web page owners. It includes techniques such as tricking customers through email and spam messages, man in the middle attacks, installation of key loggers and screen captures. Emails are one of the most common techniques for phishing, due to its simplicity, ease of use and wide reach. Phishers can deliver specially crafted emails to millions of legitimate email addresses very quickly and can fool the recipients utilising well known flaws in the SMTP. Some of the most common techniques used by phishers include official looking and sounding emails, copying legitimate corporate emails with minor URL changes, obfuscation of target URL information etc. Methods like virus/worm attachments to emails, crafting of 'personalised' or unique email messages are also common. Researchers propose user-based mechanisms to authenticate the server. Automated Challenge Response Method[1] is one such authentication mechanisms ,includes challenge generation module from server which in turn interacts with Challenge-Response interface in client and request for response from user. Challenge-Response module in turn will call the get response application which is installed in the client machine. Once the challenge-response is validated user credentials are demanded from client and it is validated by server to proceed the transaction. Automated Challenge-Response Method ensures two way authentication and simplicity. The proposed method also prevents man-in-the middle attacks since the response is obtained from the executable which is called by the browser and third man interruption is impossible. Here instead of getting response from get-response executable it is better to update the get-response executable automatically from bank server when the responses are about to nullify. Now there are DNS-based anti-phishing approach[2] technique which mainly includes blacklists, heuristic detection, the page similarity assessment. But they do have some shortcomings.

Blacklist is a DNS based anti-phishing approach technique now most commonly used by the browser. Anti Phishing Work Group, Google and

other organizations have provided an open blacklist query interface. Internet Explorer7, Netscape Browser8.1, Google Safe Browsing (a feature of the Google Toolbar for Firefox) are important browsers which use blacklists to protect users when they are navigating through phishing sites. Because every URL in the blacklist has been verified by the administrator, the false alarm probability is very low. However, there are a lot of technical disadvantages. Firstly, the phishing websites we found is a very small proportion, so the failed alarm probability is very high. Secondly, generally to say, the life cycle of a phishing website is only a few days. A website might be shut down before we found and verified it is a phishing website. Heuristic-based anti-phishing technique is to estimate whether a page has some phishing heuristics characteristics. For example, some heuristics characteristics used by the SpoofGuard [3] toolbar include checking the host name, checking the URL for common spoofing techniques, and checking against previously seen images. If you only use the Heuristic-based technique, the accuracy is not enough. Besides, phishers can use some strategies to avoid such detection rules. The user may be deceived by the phishing website because the phishing website imitates a legitimate website. Its pages are often similar with the legitimate sites. Therefore, some researchers proposed a similarity assessment method to detect phishing sites. For example, CANTINA [4] is a content similarity based approach to detect phishing websites. First, it calculates the suspicious page's lexical signature using TF-IDF and then feed this lexical signature to a search engine . According to the suspicious page's sort order in the search results we can determine whether it is a phishing site. Liu Wenyin and Anthony Y. Fu etc. [5][6] proposed a page visual similarity assessment method to detect phishing websites, if a web page is similar to a financial organization's page, but it is not the organization's web page itself, it is considered a phishing site's page. JungMin Kang and DoHoon Lee [7] proposed the URL similarity assessment method, if an URL is similar to a bank's URL, but it is not the bank's URL, it is considered a phishing website's URL. There is low assess accuracy rate for the URL and content similarity assessment techniques. The speed of calculating the visual similarity between pages is too slow, so it is only used for phishing-spam detection generally. A three factor authentication scheme[8] named Phish-Secure focuses to counter attack phishing. Here as a first factor of authentication, an image similarity detection is done which helps in finding out which page the user tends to visit, then it is checked for Phishing .For this purpose a system captures the image of a webpage in a particular resolution in the required format. This image is termed as Visual image. If the attacker is going to create a Phishing site he is going to use the replica of the original webpage in order to fool the users. Now Phish-Secure gets the Visual image of the visited page and collects the mean RGB value of the image. This is termed as V_RGB. The database with Phish-Secure uses consists of details about the page which has to be authenticated. The actual mean RGB of various WebPages is stored in the database which is denoted as A_RGB. Phish-Secure will utilize this information and make a comparison to find out the similarity between the visited page and the page in the database. The similarity is obtained in means of percentage, if the percentage of similarity (PS) is greater than 99 % then Phish-Secure concludes which website the user is tending to visit. This is carried out by taking the corresponding URL in the database and checking is done in order to find whether the site is

Phishing or not. As a second factor of authentication Phish-Secure grabs the destination IP in Layer 3 which gives information about to which IP address the user is getting connected, this is referred as V_IP. If an attacker's web server IP address has already been found guilty the particular IP is blacklisted. Phish-Secure check this Blacklist with the V_IP and will warn the user. On the other hand if the V_IP is not found in Blacklist, further verification is done in the following step.Here in this step Phish-Secure grabs the actual list of IP address of the provider which he tends to connect. This is because any provider may have multiple servers for the purpose of load balancing and the user may be connected to his location accordingly. In order to avoid any confusion Phish-Secure gets the list of IP address which is referred to as actual IP and is checked with the V_IP (i.e.) the IP address to which the user is getting connected. If these two IP address are same Phish-Secure identifies the particular site as genuine and returns a message as authenticated. On the other hand if there is a mismatch in the above verification Phish-Secure identifies the site as Phishing and warns the user. In addition to this the V_IP is added to the black list so that in future if the attacker uses the same web server and tries to attack, PhishSecure detects the site as Phishing in the second step. These popular technologies have several drawbacks: 1. Blacklist-based technique with low false alarm probability, but it cannot detect the websites that are not in the blacklist database. Because the life cycle of phishing websites is too short and the establishment of blacklist has a long lag time, the accuracy of blacklist is not too high. 2. Heuristic-based anti-phishing technique, with a high probability of false and failed alarm, and it is easy for the attacker to use technical means to avoid the heuristic characteristics detection. 3. Similarity assessment based technique is time-consuming. It needs too long time to calculate a pair of pages, so using the method to detect phishing websites on the client terminal is not suitable. And there is low accuracy rate for this method depends on many factors, such as the text, images, and similarity measurement technique. However, this technique (in particular, image similarity identification technique) is not perfect enough yet. An offline phishing detection system named LARX, acronym for Large-scale Anti-phishing by Retrospective data-eXploration [9] to counter phishing attacks has been proposed. First, it uses traffic archiving in a vantage point to collect network trace data. Secondly, LARX leverage cloud computing technology to analyze the experimental data in a way similar to the "divide and conquer" scheme. It used two existing cloud platforms, Amazon Web Services and Eucalyptus. A physical server is also used for comparison. All of LARX's phishing filtering operations are based on a cloud computing platform and work in parallel. Finally, as an offline solution, LARX can be effectively scaled up to analyze a large volume of network trace data for phishing attack detection. To meet the need that user effectively manage more and more accounts and passwords, OpenID was born. OpenID is a convenient, simple, user-centric ID management system. OpenID provides single sign-on (SSO) service, that is, we login only once and can enjoy the service of multiple sites. But OpenID is vulnerable to phishing attacks. To avoid phishing attacks, many methods have been proposed, but there is no satisfactory method. "New Anti-phishing Method with Two Types of Passwords in OpenID System" [10], proposes a model of two types of passwords for anti-phishing which is convenient and safe for OpenID users. An OpenID account has a fixed password and

several temporary passwords. The fixed password can only be used in bound PCs, that is, we must bind the fixed password on several known PC. Users can login on any PC with a temporary password. However, we need to access the mailbox or mobile phone for getting the temporary password, and we only use it in a period of time. Through analysis, this method can effectively avoid phishing. Detecting and identifying any phishing website in real-time, particularly for e-banking, is really a complex and dynamic problem involving many factors and criteria. Because of the subjective considerations and the ambiguities involved in the detection, Fuzzy Data Mining (DM) Techniques can be an effective tool in assessing and identifying phishing websites for e-banking

since it offers a more natural way of dealing with quality factors rather than exact values. "Modelling Intelligent Phishing Detection System for e-Banking using Fuzzy Data Mining"[11], a novel approach to overcome the 'fuzziness' in the e-banking phishing website assessment propose an intelligent resilient and effective model for detecting e-banking phishing websites. The proposed model is based on Fuzzy logic (FL) combined with Data Mining algorithms to characterize the e-banking phishing website factors and to investigate its techniques by classifying there phishing types and defining six e-banking phishing website attack criteria's with a layer structure. The proposed e-banking phishing website model showed the significant importance of the phishing website two criteria's (URL & Domain Identity) and (Security & Encryption) in the final phishing detection rate result, taking into consideration its characteristic association and relationship with each others as showed from the fuzzy data mining classification and association rule algorithms. Our phishing model also showed the insignificant trivial influence of the (Page Style & Content) criteria along with (Social Human Factor) criteria in the phishing detection final rate result. Haijun Zhang, Gang Liu, Tommy W. S. Chow [12] proposed a textual and visual content based antiphishing mechanism using Bayesian approach. This framework synthesizes multiple cues, i.e., textual content and visual content, from the given web page and automatically reports a phishing web page by using a text classifier, an image classifier, and a data fusion process of the classifiers. A Bayesian model is proposed to estimate the threshold, which is required in classifiers to determine the class of web page. It also develop a Bayesian approach to integrate the classification results from the textual and visual contents. The main contributions of this paper are threefold. First, it propose a text classifier using the naive Bayes rule for phishing detection. Second, it propose a Bayesian approach to estimate the threshold for either the text classifier or the image classifier such that classifiers enable to label a given web page as "phishing" or "normal." Third, a novel Bayesian approach to fuse the classification results from the text classifier and the image classifier is proposed . There are various mutual authentication methods using cell phones such as browsing using phones, password generation etc. There are various problems regarding these methods such as hijacking account setup, theft of the trusted device and attacks on the network. Thus there are various methods present in online manipulations for making the systems safe from these types of attacks. But we can see that they have its own problems which make it again unsafe. So a system based on visual cryptography which can perform as a new method can overcome these problems effectively

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 09, September 2015
Available at http://internationaljournalofresearch.org

3. VISUAL CRYPTOGRAPHY One of the best known techniques to protect data is cryptography. It is the art of sending and receiving encrypted messages that can be decrypted only by the sender or the receiver. Encryption and decryption are accomplished by using mathematical algorithms in such a way that no one but the intended recipient can decrypt and read the message. Naor and Shamir [13] introduced the visual cryptography scheme (VCS) as a simple and secure way to allow the secret sharing of images without any cryptographic computations. Visual cryptography schemes were independently introduced by Shamir and Blakley, and their original motivation was to safeguard cryptographic keys from loss. These schemes also have been widely employed in the construction of several types of cryptographic protocols and consequently, they have many applications in different areas such as access control, opening a bank vault, opening a safety deposit box, or even launching of missiles. A segment-based visual cryptography suggested by Borchert [14] can be used only to encrypt the messages containing symbols, especially numbers like bank account number, amount etc. The VCS proposed by Wei-Qi Yan et al., [15] can be applied only for printed text or image

A recursive VC method proposed by Monoth et al., [16] is computationally complex as the encoded shares are further encoded into number of sub-shares recursively. Most of the previous research work on VC focused on improving two parameters: pixel expansion and contrast. In these cases all participants who hold shares are assumed to be honest, that is, they will not present false or fake shares during the phase of recovering the secret image. Thus, the image shown on the stacking of shares is considered as the real secrete image. But, this may not be true always. So cheating prevention methodologies are introduced by Horng et al., [17] and Hu et al., [18]. But, it is observed in all these methodologies, there is no facility of authentication testing. VCS is a cryptographic technique that allows for the encryption of visual information such that decryption can be performed using the human visual system. We can achieve this by one of the following access structure schemes. 1.(2, 2)-Threshold VCS scheme- This is a simplest threshold scheme that takes a secret message and encrypts it in two different shares that reveal the secret image when they are overlaid. 2. (n, n) -Threshold VCS scheme-This scheme encrypts the secret image to n shares such that when all n of the shares are combined will the secret image be revealed. 3.(k, n) Threshold VCS scheme- This scheme encrypts the secret image to n shares such that when any group of at least k shares are overlaid the secret image will be revealed. In the case of (2, 2) VCS, each pixel P in the original image is encrypted into two sub pixels called shares. Figure.1 denotes the shares of a white pixel and a black pixel. Note that the choice of shares for a white and black pixel is randomly determined (there are two choices available for each pixel). Neither share provides any clue about the original pixel since different pixels in the secret image will be encrypted using independent random choices. When the two shares are superimposed, the value of the original pixel P can be determined. If P is a black pixel, we get two black sub pixels; if it is a white pixel, we get one black sub pixel and one white sub pixel.

4. CURRENT METHODOLOGY In the current scenario, when the end user wants to access his confidential information online (in the form of money transfer or payment gateway) by logging

into his bank account or secure mail account, the person enters information like username, password, credit card no. etc. on the login page. But quite often, this information can be captured by attackers using phishing techniques (for instance, a phishing website can collect the login information the user enters and redirect him to the original site). There is no such information that cannot be directly obtained from the user at the time of his login input. International Journal of Distributed and Parallel

5. PROPOSED METHODOLOGY For phishing detection and prevention, we are proposing a new methodology to detect the phishing website. Our methodology is based on the Anti-Phishing Image Captcha validation scheme using visual cryptography. It prevents password and other confidential information from the phishing websites. The proposed approach can be divided into two phases: 5.1.REGISTRATION PHASE In the registration phase, a key string (password) is asked from the user at the time of registration for the secure website. The key string can be a combination of alphabets and numbers to provide more secure environment. This string is concatenated with randomly generated string in the server and an image captcha[19] is generated. The image captcha is divided into two shares such that one of the shares is kept with the user and the other share is kept in the server. The user's share and the original image captchais sent to the user for later verification during login phase. The image captcha is also stored in the actual database of any confidential website as confidential data.

5.2.LOGIN PHASE In the Login phase first the user is prompted for the username (user id).Then the user is asked to enter his share which is kept with him. This share is sent to the server where the user's share and share which is stored in the database of the website, for each user, is stacked together to produce the image captcha. The image captcha is displayed to the user .Here the end user can check whether the displayed image captcha matches with the captcha created at the time of registration. The end user is required to enter the text displayed in the image captcha and this can serve the purpose of password and using this, the user can log in into the website. Using the username and image captcha generated by stacking two shares

6. IMPLEMENTATION & ANALYSIS The proposed methodology is implemented using Matlab. Figure 5.shows the result of creation and stacking of shares. In the registration phase the most important part is the creation of shares from the image captcha where one share is kept with the user and other share can be kept with the server. For login, the user needs to enter a valid username in the given field. Then he has to browse his share and process. At the server side the user's share is combined with the share in the server and an image captcha is generated .The user has to enter the text from the image captcha in the required field in order to login into the website. The entire process is depicted in Figure.5 as different cases.Case1 and Case 2 illustrates the creation and stacking of shares of two image captcha's resulting in original captcha. In Case3 share1 of first image captcha is combined with share2 of second captcha resulting in unrecognizable form of captcha.

7. CONCLUSION Currently phishing attacks are so common because it can attack globally and capture and store the users' confidential information. This information is used by the attackers which are indirectly involved in the phishing process. Phishing websites as well as human users can be easily identified using our

proposed "Anti-phishing framework based on Visual Cryptography". The proposed methodology preserves confidential information of users using 3 layers of security. 1st layer verifies whether the website is a genuine/secure website or a phishing website. If the website is a phishing website (website that is a fake one just similar to secure website but not the secure website), then in that situation, the phishing website can't display the image captcha for that specific user (who wants to log in into the website) due to the fact that the image captcha is generated by the stacking of two shares, one with the user and the other with the actual database of the website. Second layer cross validates image Captcha corresponding to the user. The image Captcha is readable by human users alone and not by machine users. Only human users accessing the website can read the image Captcha and ensure that the site as well as the user is permitted one or not. So, using image Captcha technique, no machine based user can crack the password or other confidential information of the users. And as a third layer of security it prevents intruders' attacks on the user's account. This method provides additional security in terms of not letting the intruder log in into the account even when the user knows the username of a particular user. The proposed methodology is also useful to prevent the attacks of phishing websites on financial web portal, banking portal, online shopping market.

8. REFERENCES

[1] Thiyagarajan, P.; Venkatesan, V.P.; Aghila, G.; "Anti-Phishing Technique using Automated Challenge Response Method'", in Proceedings of IEEE- International Conference on Communications and Computational Intelligience, 2010.

[2] Sun Bin.; Wen Qiaoyan.; Liang Xiaoying.; "A DNS based Anti-Phishing Approach," in Proceedings of IEEE- Second International Conference on Networks Security, Wireless Communications and Trusted Computing, 2010

[3] Nourian, A.; Ishtiaq, S.; Maheswaran, M.;" CASTLE: A social framework for collaborative antiphishing databases", in Proceedings of IEEE- 5th International Conference on Collaborative Computing:Networking, Applications and Worksharing, 2009.

[4] Sid Stamm, ZulfikarRamzan, "Drive-By Pharming", v4861 LNCS,p495-506, 2007, Information and Communications Security - 9th International Conference, ICICS 2007, Proceedings.

[5] Anthony Y. Fu, Liu Wenyin, "Detecting Phishing Web Pages with Visual Similarity Assessment Based on Earth Mover's Distance (EMD)",IEEE Transactions on Dependable and Secure Computing, v 3, n 4, p301-311, October/December 2006

[6] Wenyin Liu, Xiaotie Deng, Guanglin Huang, and Anthony Y. Fu, "An Antiphishing Strategy Based on Visual Similarity Assessment", IEEE Internet Computing, v 10, n 2, p 58-65, March/April 2006.

[7] JungMin Kang, DoHoon Lee, "Advanced White List Approach for Preventing Access to Phishing Sites", 2007 International Conference onConvergence Information Technology, ICCIT 2007, p 491-496, 2007

[8] Nirmal, K.; Ewards, S.E.V.; Geetha, K.; "Maximizing online security by providing a 3 factor authentication system to counter-attack 'Phishing'", in Proceedings of IEEE- International Conference on Emerging Trends in Robotics and Communication Technologies, 2010.