Phishing Free Sequel Applying Image Encryption and Visual Cryptography

K. Hemapriya[1], S. Aiswarya[2], S. Oviya[3]
[1]*Assistant Professor, Department of CSE, Panimalar Institute of Technology, Chennai, India*
[2,3]*Student, Department of CSE, Panimalar Institute of Technology, Chennai, India*

***Abstract*: Phishing is an endeavor by an individual or an organization to steal classified data like passwords, credit card data from innocent users to enforce wholesale fraud, monetary benefit and other deceitful exercises. A basic username and password based validation is not adequate security for sites providing financial transactions. This paper proposes a methodology for protection against phishing sites. Phishing sites are usually identified by certain URL features and web traffic indicators. This proposal combines the usage of image captcha and one-time password to provide improved security. Visual cryptography is employed to protect the image captcha by breaking down the first picture captcha into two shares that are put away in discrete database servers. This way the image captcha can be uncovered just when both shares are available simultaneously and the individual shares cannot be combined to find the original captcha. The original captcha is used as the password by the user.**

***Keywords*: authentication, image processing, phishing, shares, visual cryptography.**

1. Introduction

Phishing is a strategy for obtaining individual data utilizing tricky messages and false websites. The aggressors take on the appearance of a confidant, frequently a genuine or conceivably genuine individual, or an organization the unfortunate user may work with. It's one of the most established sorts of cyber-attacks. It is as yet a standout amongst the most far reaching and obnoxious attacks. That stated, there is an assortment of strategies that fall under the idea of phishing. There are two or three diverse approaches to classify them. One is by the reason for the phishing attack. By and large, a phishing effort endeavors to inspire the injured individual to complete one of two things: hand over sensitive information and download malware. But some phishing assaults mean to get login data from explicit individuals. Assailants devote considerably more vitality to deceiving those unfortunate casualties, who have been chosen in light of the fact of high potential prizes. Common phishing techniques are Spear-Phishing, Whaling, Pharming, Phishing Injection and Domain Hijacking.

Spear phishing includes writing email messages that have all the earmarks of being from a companion or other confidant in sender, for example, a business, a friend, or a nearby organization. Spear phishers exploit the level of proximity between the objective and the alleged sender. Whaling is a sort of phishing that moves one stage up the step of hazard and reward. It alludes to phishing efforts pursued against senior administrators, for example, CEOs or CFOs. The unfortunate casualty is attracted with an exceptionally proficient, corporate-enhanced report, for example, a phony FBI legitimate subpoena, an IRS review declaration, or advertising crisis announcement. Pharming is a phishing device that includes diverting system traffic or adjusting the Domain Name System (DNS) for the objective site. The DNS is in charge of changing over space names, to numerical Internet Protocol (IP) delivers that PCs use to speak with one another.

Phishing injection happens when pernicious substance, (for example, a site not associated with the primary page) is embedded into a generally genuine site through a security hole in the site. Deceitful pages can be made by replicating and transferring a basic HTML page to a bargained site server while introducing explicit back-end abilities to process client entered information, consequently making the information accessible to the programmer. Domain hijacking includes accessing an authentic domain and diverting it from the domain proprietor's web server to another by reconfiguring the domain name to divert clients to the fake website. Domain hijacking does not require the phisher to acquire access to the objective web server itself.

2. Related work

Yicong Zhou, Zhongyun Hua, Chi-Man Pun and C. L. Philip Chen [2] is motivated by the course structure in electronic circuits, this paper presents a general confused system called the course confused framework (CCS). Utilizing two 1-D disordered maps as seed maps, CCS can produce an enormous number of new disorganized maps. Models and assessments demonstrate the CCS's vigor. Contrasted and comparing seed maps, recently created turbulent maps are increasingly eccentric and have better disorganized execution, more parameters, and complex disorganized properties. To research utilizations of CCS, the authors presented a pseudo-arbitrary number generator (PRNG) and an information encryption framework utilizing a disorganized guide produced by CCS.

In paper [3], six features like primary domain, subdomain, path domain, Alexa reputation, Google Index, Page rank has been to classify phishing sites. Particularly, Google Index is used as opposed to utilizing Google indexed lists like customary strategies to enhance the exactness. A five layer neuro-fuzzy network is implemented to classify phishing and legitimate websites. The paper was inspired by an earlier report that utilized the traditional neural organize show. Consolidating the neural system with the fuzzy model, there is a decent outcome regarding ID precision.

In the work published in 2010 [4], the creators have utilized 27 highlights to assemble a display dependent on fluffy rationale. Despite the fact that this is a promising arrangement, it neglects to illuminate how the highlights were separated from the site, decisively includes identified with human-factors. In addition, the standards were built up based on human experience, which is one of the issues considered in this paper. Moreover, the site was arranged into five unique classes that is, (extremely real, authentic, suspicious, phishy and exceptionally phishy), yet the creators did not illuminate what is the scarcely discernible difference that separates between these classes.

This principle motive of paper [5] is to explore the potential utilization of an acclaimed quantum calculation show, i.e., quantum strolls (QW) in picture encryption. It is discovered that QW can fill in as a phenomenal key generator on account of it's in here nt-nonlinear chaotic dynamic conduct. Moreover, a novel QW-based picture encryption calculation is implemented. Reproductions and execution examinations demonstrate that the proposition is sufficiently secure for picture encryption and beats earlier works. It additionally opens the entryway towards bringing quantum calculation into picture encryption and advances the intermingling between quantum calculation and picture handling.

Next paper [7] proposes the (n, k, p)- Gray code, which incorporates a few usually utilized codes, for example, the parallel reflected, ternary, and (n, k)- Gray codes. This code can be extensively applied in electronics and communication systems. This paper centers around three illustrative utilizations of the (n, k, p)-Gray code, specifically, picture bit plane disintegration, picture de-noising, and encryption. Imaging systems respond better to the new code than the conventional systems.

3. Existing System

Phishing website pages are fashioned site pages that are made by obnoxious individuals to copy web pages of genuine sites. The vast majority of these sorts of pages have high visual similarities to trick their exploited people. A portion of these sorts of site pages look precisely like the genuine ones. Casualties of phishing pages may uncover their financial balance, other critical data to the phishing site page proprietors. It incorporates methods, for example, deceiving clients through email and spam messages, middlemen attacks, establishment of pins and screen captures among other attacks. The following graph (Fig. 1) depicts the commonly affected phishing domains.

The three main techniques used worldwide are,

- Blacklist based technique: The websites which are suspicious are matched with a list of websites that are classified as phishing websites. If there is a match, then such a website is classified as a phishing site and prevented from the user's access. The negative part of this plan is that it for the most part does not cover all phishing sites since a newly propelled misrepresentation site sets aside some opportunity to add to the boycott record. Sheng et al. Portrayed that boycotts are normally include the record at differing frequencies, rough 50–80% of phishing spaces included boycott in the wake of playing out some monetary misfortune.

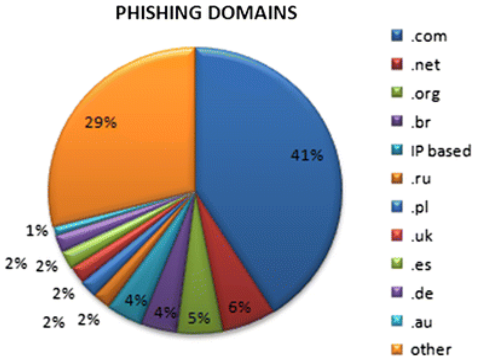

Fig. 1. Commonly affected phishing domains

- *Heuristic-based technique:* In this kind of methodologies, the heuristic plan of suspicious sites matches with the list of capabilities, which are commonly found in phishing sites. Zero-day assault (i.e., assaults that were not seen previously) can be distinguished utilizing heuristic methodology. Zhang et al. proposed a content-based phishing location system called CANTINA, which take a rich arrangement of list of capabilities from different field of a website page.
- *Visual similarity-based approach:* Visual similarity-based approaches compare the visual appearance of a suspicious website and its corresponding legitimate site. Visual similarity-based techniques use features set like text content, HTML Tags, Cascading Style Sheet (CSS), image processing, etc., to make decision. Chen et al. [8] proposed an anti-phishing approach based on discriminative key-point features in a web page.

A. *Disadvantages of Existing System*

- Blacklist-based strategy has low false alarm likelihood, yet it can't distinguish the sites that are not in the marked database. Since the existence cycle of phishing sites is excessively short and the foundation of blacklist has a long slack time, the exactness of the method isn't excessively high.
- Heuristic-based phishing detection method has a high likelihood of false alert, and it is simple for the assailant to utilize specialized techniques to evade such prevention method.
- Similarity based method needs lengthy timespan to ascertain a couple of pages, so utilizing the strategy to distinguish

phishing sites on the customer terminal isn't reasonable. Also, there is low precision rate for this technique relies upon numerous elements, for example, the content, pictures, and similitude estimation.

4. Proposed system

The idea of image captcha and an enhanced visual cryptography is utilized. Image processing is a strategy of handling an original picture and to get the yield as either enhanced type of a similar picture as well as attributes of the information picture. In Visual Cryptography (VC), a picture is disintegrated into shares and so as to uncover the original, number of shares ought to be consolidated. VCS is a cryptographic strategy that takes into consideration the encryption of visual data with the end goal that decoding can be performed utilizing the human visual framework. As the name portrays, in this methodology, site cross confirms its very own personality and demonstrates that it is a certifiable site (to utilize bank exchange, E-business and internet booking framework and so on.) before the end clients and make the both the sides of the framework secure and additionally a validated one. This can be accomplished by one of the accompanying access structure plans.

- (2, 2) - Threshold VCS conspire This is a least difficult limit plot that takes a mystery message and scrambles it in two unique shares that uncover the mystery picture when they are overlaid.
- (n, n) - Threshold VCS plot: This plan scrambles the mystery picture to n offers with the end goal that when all n of the shares is joined will the mystery picture be uncovered.
- (k, n) Threshold VCS conspire: This plan scrambles the mystery picture to n shares with the end goal that when any gathering of in any event k shares are overlaid the mystery picture will be uncovered.

On account of (2, 2) VCS, every pixel P in the first picture is scrambled into two sub pixels called shares. Note that the selection of shares for a white and dark pixel is haphazardly decided (there are two decisions accessible for every pixel). Neither one of the shares gives any insight about the first pixel since various pixels in the mystery picture will be scrambled utilizing free arbitrary decisions. At the point when the two offers are superimposed, the estimation of the first pixel can be resolved. On the off chance that P is a dark pixel, two dark sub pixels are obtained; on the off chance that it is a white pixel, one dark sub pixel and one white sub pixel are obtained.

A. *Architecture Diagram*

The user initially enters the registration details. The user also enters a secret key of specific digits. The captcha is generated from the user's secret key and the secret key from the bank's side. The generated captcha is split into two shares. One share is sent to the user and the other share is stored in the bank's database. When the user logs in, the user has to provide the share which will be overlapped with the share from the bank's side. If the captcha matches the original captcha, the user can proceed to transactions through OTP. If not, the share probably originated from a phishing website and the user should refrain from further communication with the site.

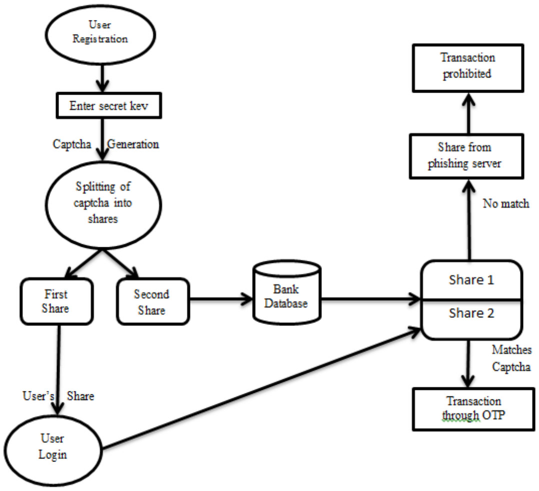

Fig. 2. Architecture diagram

B. *Module description*

1) *Registration with Secret Code:* The user enters personal details like name, age, address etc., along with a secret key (8 digits). The key may contain any alphanumeric character. This secret key is to be concatenated with the key from server's side.
2) *Image captcha Generation:* In captcha generation phase, the key string from the user is combined with the key string from the server's side. The 16-digit string is converted into black and white pixels using linear programming approach and the generated image is stored in the server.
3) *Shares Creation:* Visual Cryptography (VC) is a method of encrypting a secret image to shares, such that stacking a sufficient number of shares reveals the secret image. Original image captcha is split 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. Any one share cannot be used to regenerate the original captcha.
4) *Login Phase:* The user enters login credentials and also uploads the share of the image stored in the system. The share from bank's side is also overlaid. After decrypting, the user is allowed to access the transaction page, if the captcha matches the original captcha. If not, the share probably originated from a phishing website and the user should refrain from further communication with the site. The user is provided with a OTP (One Time Password) for additional security during transactions.

C. *Advantages of proposed system*

The proposed system can achieve real-time response and stable performance to detect phishing URLs.

- It prevents password and other confidential information from being stolen by the phishing websites.
- The proposed system provides three levels of security for the user as well as the organization giving services.
- This system utilizes OTP which provides the maximum level of security in terms of confidentiality, authentication, reliability and privacy.

5. Results and discussion

The three most commonly used anti-phishing techniques are blacklist based method, heuristic based and similarity based method. Their disadvantages are overcome in the proposed system which has better accuracy, response time and efficiency. The accuracy of the three techniques is represented in the graph below.

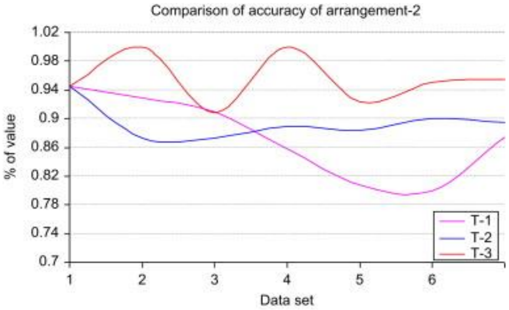

Fig. 3. Accuracy comparison chart

In the above chart, T-1 represents Blacklist technique, T-2 represents Heuristic based technique and T-3 stands for Similarity based technique.

In Blacklist based technique, the websites which are suspicious are matched with a list of websites that are classified as phishing websites. If there is a match, then such a website is classified as a phishing site and prevented from the user's access. Blacklist-based strategy has low false alarm likelihood, yet it can't distinguish the sites that are not in the marked database. Since the existence cycle of phishing sites is excessively short and the foundation of blacklist has a long slack time, the exactness of the method isn't excessively high. The proposed system overcomes these drawbacks by avoiding the use of blacklists to classify phishing sites.

In Heuristic-based technique, the heuristics of suspicious sites matches with the list of capabilities, which are commonly found in phishing sites. This method has a high likelihood of false alert, and it is simple for the assailant to utilize specialized techniques to evade such prevention method. The proposed system is difficult to break as the captcha has to be generated from the legitimate server's site to match with the user's share of the captcha.

Visual similarity-based approaches compare the visual appearance of a suspicious website and its corresponding legitimate site. Visual similarity-based techniques use features set like text content, HTML Tags, Cascading Style Sheet (CSS), image processing, etc., to make decision. This method needs lengthy timespan to ascertain a couple of pages, so utilizing the strategy to distinguish phishing sites on the customer terminal isn't reasonable. Also, there is low precision rate for this technique relies upon numerous elements, for example, the content, pictures, and similitude estimation. The proposed system provides a quicker classification of sites.

6. Conclusion

A phishing assault can have annihilating outcomes. For people, this incorporates unapproved buys, the taking of assets, or recognizes robbery. In addition, phishing is frequently used to pick up a solid footing in corporate or legislative systems as a piece of a bigger assault, for example, a progressed diligent danger (APT) occasion. In this last situation, workers are imperiled so as to sidestep security edges, disperse malware inside a shut domain, or increase advantaged access to anchored information. An association surrendering to such an assault ordinarily supports serious money related misfortunes notwithstanding declining piece of the overall industry, notoriety, and purchaser trust. Contingent upon degree, a phishing endeavor may grow into a security episode from which a business will have a troublesome time recuperating. It is imperative to design systems so as to prevent users from being mishandled.

7. Future Enhancement

The proposed system involves three stage verification. The first stage involves captcha generation based on the secret code entered by the user during registration. The captcha is shared between the user and the bank server. Each and every time the user has to upload the captcha in order to transact. Instead the process can be automated. This way the system will be more user-friendly.

References

[1] Mohammad, Rami, McCluskey, T.L. and Thabtah, Fadi Abdeljaber, "Intelligent Rule based Phishing Websites Classification*", IET Information Security*, pp. 153-160.

[2] Yicong Zhou, Zhongyun Hua, Chi-Man Pun and C. L. Philip Chen, "Cascade Chaotic System with Applications", *IEEE Transactions On Cybernetics"*, vol. 45, No. 9, September 2015.

[3] Chuan Pham , Luong A. T. Nguyen ,Nguyen H. Tran , Eui-Nam Huh , Choong Seon Hong, "Phishing-Aware: A Neuro-Fuzzy Approach for Anti-Phishing on Fog Networks", *IEEE transactions on network and service management*, vol. 15, no. 3, September 2018.

[4] Aburrous, M., Hossain, M.A., Dahal, K., Thabtah, F. "Intelligent phishing detection system for e-banking using fuzzy data mining", *Journal of Expert Systems with Applications*, vol. 40, pp. 4697-4706 September, 2013.

[5] Yu-Guang Yang, Qing-Xiang Pan, Si-Jia Sun and Peng Xu, D, "Novel Image Encryption Based On quantum Walks", *Scientific Reports*, Article number: 7784, 2015.

[6] W. S. Chen, K. H. Chih, S. W. Shih, and C. M. Hsieh, "Personal identification technique based on human IRIS recognition with wavelet transform", *International Conference on Acoustics, Speech and Signal Processing(ICASSP)*, vol. 2, pp. 949-952, 2005.

[7] Yicong Zhou, Karen Panetta, Sos Agaian, and C. L. Philip Chen," (n, k, p)-Gray Code for Image Systems", *IEEE transactions on cybernetics*, vol. 43, no. 2, April 2013.

[8] B. B. Zhu, Y. Chun, W. Yidong, and L. Shipeng, “Scalable protection for MPEG-4 fine granularity scalability,” *IEEE Transactions on Multimedia,* vol. 7, no. 2, April 2005.

[9] R. Bose and S. Pathak, “A novel compression and encryption scheme using variable model arithmetic coding and coupled chaotic system,” *IEEE Transactions on Circuits and Systems,* vol. 53, no. 4, 2006.

[10] X. Liao, S. Lai, and Q. Zhou, “A novel image encryption algorithm based on self-adaptive wave transmission,” *Signal Process*, vol. 74 issue 3, pp 781-811, February, 2015.