



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

www.iasir.net

Token Based Security for Prevention of Phishing Attack at Client Side

Rashmi Gupta

M.E (CSE) Student, Punjabi University Regional Centre of IT and Management, Mohali, Punjab
India

Nivit Gill

Assistant Professor (CSE), Punjabi University Regional Centre of IT and Management, Mohali, Punjab
India

Abstract: Phishing is an electronic identity theft in which the attacker uses a combination of social engineering techniques and web spoofing techniques to deceive a user into revealing sensitive information. The literature addresses this issue extensively and presents a number of solutions, which are either client based or server based. Generally client based solutions have an upper hand over the server based solutions. In this paper, a client side anti phishing technique has been proposed that employs hash algorithm for the verification of web sites at URL level. A database of tokens (hash codes) of legitimate web sites' URLs is stored on the client machine. When a user enters a site address in the browser, the system computes its integrity code and matches it with the stored integrity code (token). For the validation of the proposed system, it has been compared with the existing 'whitelist' based anti phishing method in terms of false positive and false negative metrics. Both the systems have been implemented using MATLAB®, and simulated using the test data. From the simulation results, it has been observed that the proposed system is better than the existing system ('whitelist') for the detection of phishing sites at client side.

Keywords: phishing, MD5 hashing, authentication, Tokens.

I. Introduction

Organizations are investing vast amount of money in creating stronger resistance to their information resources via firewalls, anti-virus software and biometrics. These actions have made the business world more efficient by blocking pressure from the outside and made it gradually more complicated for hackers or viruses to break in systems [1]. But still there are some threats that put such organizations at risk. This paper analyzes the main threat in organizations that are phishing attacks.

Phishing is a fraudulent attempt to acquire information such as usernames, passwords, and credit card details by masked as a trustworthy entity in an electronic communication [2]. Communications claimed to be from popular social web sites, any auction sites, and online payment processors. Phishing emails may contain a link to website that is infected with malware. Phishing is typically carried out by e-mail spoofing or instant messaging and often it directs users to enter personal details at a fake website whose look and feel are identical to the original one. Phishing is an instance of social engineering techniques used to deceive users, and it exploits the usability of current web security technologies [2]. The Attempts to deal with these growing number of phishing incidents include user training, and public awareness, and technical security measures.

A lot of work has been done by researchers for the detection of phishing attacks. In this paper, phishing attacks are detected at website level with low false positive rate. The general approach to detection phishing sites is that a web browser relies on a white list of some known web sites in which verification is done only on the basis of matching URL. The current study is generating a client based token system in which information is extracted to detect phishing sites using hash method.

A. Organization of paper

The current section provides a glimpse of the paper with brief introduction of phishing attacks. The second section presents the review of the previous work that has been referred to carry out this research work. The design of the proposed system is discussed in section 3. This section covers the whole process of generating a token based client system. The implementation details of the proposed system, including the data set and performance metrics are given in the fourth section. The simulation results and their comparison with the existing system are also presented in fifth section. Finally, research work has been concluded in the sixth section.

II. Related Work

Anti-phishing techniques have been proposed by various authors to detect and prevent the phishing attacks. The author Madhuri *et al.* proposed a phishing detection scheme based on the URL domain identity and webpage image matching. It was found that this combined approach performed better than other tools [3].

Juan Chen *et al.* explained a new algorithm called link guard which utilized the general characteristics of the hyperlinks in phishing attacks. The proposed algorithm first extracted the DNS name from the actual and visual link, and then compared the actual and visual DNS names [4]. Mallikka *et al.* proposed an effective image based anti-phishing scheme in [5], which uses distinct features of web pages for differentiating between legitimate and phishing sites. This method determined the similarity of two images by taking out a vector of main features from each image of web page, and then computing the distance between the vectors. The author Y. Zhang *et al.* explained the various anti-phishing tools like callingid toolbar, earthlink toolbar, cloudmarkanti-fraudtoolbar, eBay toolbar, firefox2, netcrafttoolbar. These extensions warn users when these users are browsing some suspected phishing site [6].

EnginKirda *et al.* explained a novel browser extension; Anti-phish tracks the confidential information of a user and thus generates warnings whenever the user attempts to steal this information to a web [7]. R. Dhanalakshmi *et al.* presented the system in which when the client requests the website for a transaction process, this phishing detection system validates the website with four features such as WHOIS registration, IP address, and domain and inter domain values of the corresponding requested website. The phisher login into the website means they cannot get that session key so they enter a wrong session key then the system denies the access [8]. V. Prasad *et al.* presented two anti phishing approaches wherein one inhibits the user from entering into a spoofing site and the other defends the user's password [9].

III. Proposed System

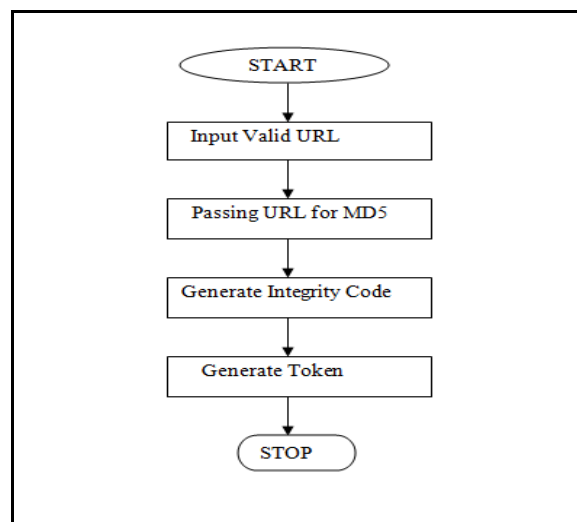
The aim of this research is to design a token based technique at client side for automated detection of phishing sites. From the literature review, it has been found that some anti phishing techniques use password based generator functions. From these techniques, the idea has been extended for URL authentication, which is very important as it may lead to a phishing site. The proposed system attempts to provide a method to check whether the site is a phishing site or not, with a low false positive rate.

B. Design Components

The design of the proposed system has three components, which are explained below:

Token/ Stored Digested Code: Token is a valid certificate which contains integrity code that is verified on request. It resides in the client machine and a phishing engine is implemented on client machine to verify the link. First the client enters the valid URL, and this URL is supplied to a routine that generates the integrity code using a hashing algorithm, MD5. The flow chart depicting the working of Token Generation System is given in Figure 1.

Figure 1: Flow chart depicting the process of Token Generation



The MD5 message-digest algorithm is a widely used cryptographic hash function that produces a 128-bit (16-byte) hash value. Specified in RFC 1321, MD5 has been developed in a large variety of security applications, and it is also used to check data integrity. MD5 was designed by Ron Rivest in 1991 to replace an earlier hash function, MD4. An MD5 hash value is normally expressed as hexadecimal number, 32 digits long [6].

Service: This component works in the background. It initiates the trigger when the user opens a site in the browser. The service will retrieve the URL that the user enters and then sends it to verification engine for further processing.

Verification Engine: Verification Engine takes input as URL from service and tokenizes it into page variable and 'actual URL'. Then it generates the digest integrity code for the retrieved 'actual URL' and compares the computed code with the stored token of the URL. If verified, then the 'actual URL' is okay, i.e. not a phishing web site, and the user is allowed to surf, otherwise it will report the phishing attack to the user.

These components of the proposed system, when come into play, result in a complete system for the client side detection of phishing sites.

C. Working of Proposed System

The motive of the proposed system is to determine whether the asked site is a phishing site or not. The three components of the proposed system, namely Stored Digested Code (Token), Service and Verification Engine follow a specific procedure for detecting phishing site.

The working of the proposed system is given in following steps:

Step 1: Read web phishing URL. The input URL is read by the system and get it to the code extractor where it will decompose the URL and separate page variable. The browser will check whether the token exists on PC or not. If it does not exist on PC then first it will get a valid token and this token will be stored in client machine. But if it exists on PC then it will take input URL to phishing engine.

Step 2: Code extractor separates the URL and page variable and pass it to service generator for further processing, which is done by service generator.

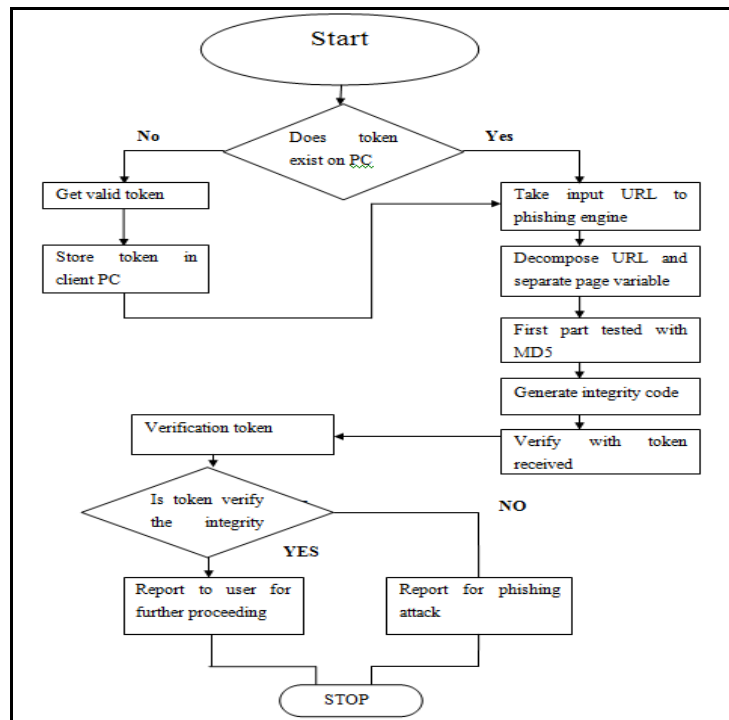
Step 3: Service generates the token, based on input string of URL using MD5 hashing technique. The first part will be tested with MD5 and it will generate the integrity code. Next it will be verified with the token received.

Step 4: Now the generated code will be verified with the token received. Now the system will check whether the token verifies the integrity code or not. Verification Engine verifies the code generated by service and the token code available. The result is true, if they match, and false otherwise.

Step 5: Based on the conditions true or false obtained from previous step, Trigger block, allow or deny for this particular site. If the verification is okay, then it will allow the user for safe surfing else it will alarm the user.

This procedure has been illustrated with the help of a flow chart shown in Figure 2.

Figure 2: Flow chart diagram for the proposed system



In order to verify the proposed method, it has been implemented and simulated with test data and its performance has been evaluated based on few metrics.

IV. Simulation

The implemented system is checked under different attack conditions and its performance is evaluated. The performance of the proposed system is compared with the existing white list based phishing detection system. For implementation, MATLAB® tool has been used. The performance of the proposed system can be assessed using various measures such as false positive rate and false negative rate. The performance of any anti phishing system depends on its rate of detection.

D. Data Sets

The performance of the proposed system has been determined by testing its implementation with test data. This test data has been created for fifty well known web sites. The tokens for these web sites have been stored on the client machine on which this implemented system is simulated. A data base of these sites has also been stored in the form of a white list for evaluating the performance of the existing anti phishing system. The information like URL and domain are stored in the database. Table 1 lists few of the sites from the database.

Table 1: Data set for proposed system and existing system

S. No	URL link
1.	www.icici.com
2.	www.google.com
3.	www.icici.com
4.	www.hdfc.com
5.	www.sbi.co.in
6.	www.yesbank.com
7.	www.sbp.com
8.	www.gmail.com
9.	www.pnb.com
10.	www.obc.com

The system retrieves the user supplied URL and verifies with the stored token. The implementation of the proposed system and existing system has been evaluated for this test data and their performance has been compared based on the chosen metrics.

E. Performance Metrics

To evaluate the performance of proposed system against the existing system, following metrics have been employed [9]:

False negative (Fn): The probability the system fails to recognize a phishing attack. The formula for calculation of false negatives is:

$$Fn = \frac{\text{The number of phishing websites that are not detected}}{\text{Total number of websites that are used in the test}}$$

False positive (Fp): The probability the system recognizes a legitimate website as a phishing website. The formula for calculation of false positives is:

$$Fp = \frac{\text{The number of legitimate websites that are identified as phishing website}}{\text{Total number of websites that are used in the test}}$$

Lower the two rates, more efficient is the phishing detection.

V. Results and Discussion

The purpose of this experiment was to determine what percentage of phishing and legitimate URLs would be detected using the proposed system. Both the systems, proposed as well as existing, have been implemented and simulated for the test data. Table 2 presents the results of the simulation of the proposed system for ten runs. The table presents the results in form of TP (True Positive), TN (True Negative), FP (False positive) and FN (False Negative).

Table 2: Results for the Proposed System

S. No.	URL link	Attack	FP/FN
1.	www.icici.com	No	TN
2.	www.google.com	No	TN
3.	www.icicii.com	Yes	TP
4.	www.gooogle.com	Yes	TP
5.	www.sbi.co.in	No	TN
6.	www.sbii.co.in	Yes	TP
7.	www.gmial.com	Yes	TP
8.	www.gmail.com	No	TN
9.	www.yaaho.com	Yes	TP
10.	www.yahoo.com	No	TN

In order to evaluate the performance of the proposed system, the existing system of white list based detection of phishing web sites has been implemented and simulated for the test data. The results of the existing system for ten simulations are tabulated in Table 3 in terms of TP, TN, FP and FN.

Table 3: Results for Existing System

S. No.	URL link	Attack	FP/FN
1.	www.icici.com	No	TN
2.	www.google.com	Yes	TP
3.	www.icicii.com	Yes	TP
4.	www.gooogle.com	Yes	TP
5.	www.sbi.co.in	Yes	TP
6.	www.sbii.co.in	No	TN
7.	www.gmial.com	Yes	TP
8.	www.gmail.com	No	TN
9.	www.yaaho.com	No	TN
10.	www.yahoo.com	Yes	TP

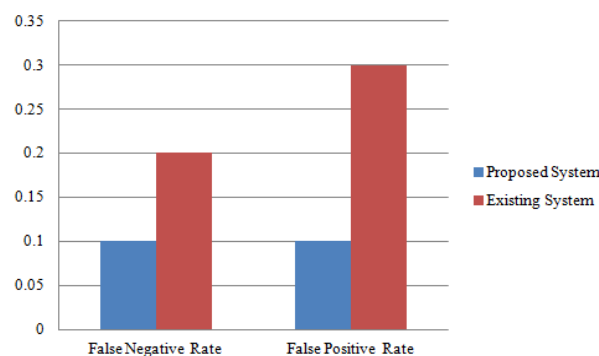
The results of the simulations of the proposed as well as existing system have been analyzed and compared. For the simulation results shown in Table 2 and Table 3, false negative and false positive rates have been calculated for the proposed system and existing system using the Equation 1 and Equation 2. Table 4 shows the result for calculated false positive and false negative rate for the proposed and existing system.

Table 4: False Negative and False Positive Rate for Proposed and Existing System

	False Negative Rate	False Positive Rate
Proposed System	0.1	0.1
Existing System	0.2	0.3

Figure 3 presents a bar graph showing the comparison of false negative rate and false positive rate for the two systems.

Figure 3: False Negative and False Positive Rate of Proposed System and Existing System



From the results, it has been observed that the proposed system is better than the existing system on the basis of false positive and false negative rate.

VI. Results and Discussion

Phishing detection methods are consistently being revised so as to keep up with new techniques used by the phishers. Combating phishing is an ongoing battle that will probably never end, much like the ongoing battle with spam emails. The literature review related to phishing reveals that many solutions have been devised for the detection of phishing attacks. Generally, these solutions are either client based or server based. Even though server side solutions have many benefits, but client side solutions are preferred to them [3], and thus, client side protection is considered as an efficient technique for detecting phishing sites.

In this paper, a client side token based solution for phishing attacks has been proposed that verifies the site at URL level. The token (integrity code) has been generated and stored for the well known websites using MD5 hashing algorithm. When the user requests for a particular website, then the integrity code for the requested web site is computed and verified with the stored token. To check the effectiveness of the proposed system, its performance has been compared with the existing “whitelist” based system, in terms of false positive and false negative metrics. The existing system which is based on ‘whitelist’, performs verification on the basis of a string matching algorithm. Both the systems have been implemented using MATLAB® and tested using the test data. The simulations results show that the values of false positive rate and false negative rate for the proposed system are less as compared to the existing system. In other words, the proposed system is better than the existing system.

The present study concludes that the proposed system which is based on tokens (hash codes of legitimate web sites) is better than the existing system based on ‘whitelist’ (URLs of legitimate web sites). It has also been observed that the proposed system is effective only if the token for a legitimate target site is already stored in the database.

References

- [1] Roger Needham and Butler Lampson, "Network Attack and Defense," in *Security Engineering: A Guide to Building Dependable Distributed Systems.*: Anderson, 2001, ch. 18, pp. 367-390.
- [2] Yue Zhang, Jason I. Hong, and Lorrie Faith Cranor, "Cantina: a content-based approach to detecting phishing web sites," in *In Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 639-648.
- [3] Madhuri S. Arade and P. C. Bhaskar, "Antiphishing Model with URL & Image based Webpage Matching," *International Journal of Computer Science and Technology (IJCSST)*, vol. 2, no. 2, pp. 282-286, June 2011.
- [4] uan Chen and Chuanxiong Guo, "Online Detection and Prevention of Phishing Attacks (Invited Paper)," pp. 1-7. [Online]. [http://national3.com/o/online-detection-and-prevention-of-phishing-attacks-\(invited-paper\)-w6294.html](http://national3.com/o/online-detection-and-prevention-of-phishing-attacks-(invited-paper)-w6294.html) accessed on 5-3-2013.
- [5] Malikka Rajalingam, Saleh Ali Alomari, and Putra Sumari, "Prevention of Phishing Attacks Based on Discriminative Key Point Features of WebPages," *International Journal of Computer Science and Security (IJCSS)*, vol. 6, no. 1, pp. 1-18, 2012.
- [6] Yue Zhang, Serge Egelman, Lorrie Cranor, and Jason Hong, "Phishing Phish: Evaluating Anti-Phishing Tools," in *In Proceedings of the 14th Annual Network and Distributed System Security Symposium*, 2007.
- [7] Engin Kirda and Christopher Kruegel, "Protecting Users Against Phishing Attacks with AntiPhish," in *in Proceedings of 29th Annual International Computer Software and Applications Conference (COMPSAC)*, vol. 2, 26-28 July, 2005, pp. 517-524.
- [8] R. Dhanalakshmi, C. Prabhu, and C. Chellapan, "Detection Of Phishing Websites And Secure Transactions," *International Journal Communication & Network Security (IJCNS)*, vol. 1, no. 2, pp. 15-21, 2011.
- [9] Venkata Prasad Reddy, V. Radha, and Manik Jindal, "Client Side protection from Phishing attack," *International Journal of Advanced Engineering Sciences and Technologies (IJAEST)*, vol. 3, no. 1, pp. 39-45, 2011.
- [10] Roger Needham and Butler Lampson, "Network Attack and Defense," in *Security Engineering: A Guide to Building Dependable Distributed Systems.*: Anderson, 2001, ch. 18, pp. 367-390.
- [11] Tyler Moore and Richard Clayton, "An Empirical Analysis of the Current State of Phishing Attack and Defence," in *In Proceedings of the 2007 Workshop on the Economics of Information Security (WEIS)*, May 11, 2007, pp. 1-20.
- [12] Chuan Yue and Haining Wang, "Anti-Phishing in Offense and Defense," in *in Proceedings of Annual Computer Applications Conference-ACSAC*, 2008, pp. 345-354. D. Kornack and P. Rakic, "Cell Proliferation without Neurogenesis in Adult Primate Neocortex," *Science*, vol. 294, Dec. 2001, pp. 2127-2130, doi:10.1126/science.1065467.