International Journal of Soft Computing 8 (3): 175-180, 2013
ISSN: 1816-9503

Online Identity Phish: Online Transaction Against Phishing Attacks

C. Emilin Shyni and S. Swamynathan
Anna University Chennai, Tamil Nadu, India

Abstract: Internet which is an appealing and fast way for distributing wide range of resources and services has become an economical way for dispensation of electronic information. The tremendous growth of internet in today's era has reduced the use of papers as electronic information is preferred over study distribution. Phishing is a delinquent technique of stealing victim's personal information by sending them spoofed emails urging them to visit a forged web page that looks like a true one. Here, researchers proposed a watermarking based approach and heuristics techniques and its implementation for extenuating phishing attacks a structure of web based identity stealing. The watermark message carries a secret message which is intended to be distinctive for each user. Phishing websites have their own methodological and community problems with each other and are very complicated. In this study, phishing detection and validation is implemented as a web service. Web services are independent layered module in a web application which process the user supplied URL as input and detect and prevent phishing.

Key words: Webpage, phishing, watermarking, forged, India

INTRODUCTION

Phishing is a alternative approach in association to additional forms of internet crimes, e.g., virus and hacking. Identifying theft is one of the most important security apprehensions in cyberspace as it has straight force on businesses, individuals and organizations. Phishing is a type of online individuality theft in which attackers sends fraudulent e-mails and use fake web sites that spoof a rightful business in order to attract innocent clients into allotting personal and important data such as bank account numbers, social security numbers, etc.

In many cases, phishing web pages targets in attacking most popular web pages and the system with these well-known web pages registered as the protected web pages can work well to detect these phishing web pages. In this study, researchers identified a phishing web page and discover its phishing target based on watermarking images. A phishing web page (Bargadiya *et al*., 2010) generally includes a few forward links to additional allied legitimate web pages but in no way to its target directly. The reason for this approach is as follows: Phishing is a social engineering attack whose success requires the active participation (Xiang and Hong, 2009) of the users. Thus, the co-operation of the users is required to mitigate (Xiang *et al*., 2011) the attacks to some extent. Secondly, researchers choose the image as a carrier for the watermark message since the client always be expecting to see an image on a web page.

The Anti Phishing Working Group (APWG) has mentioned that around 50.3% phishing attacks have happened in the commercial web sites alone (Xiang *et al*., 2011) because those web sites are money related ones. These statistical details are shown in Fig. 1.

LITERATURE REVIEW

APWG (Anti Phishing Working Group) describes the concepts of phishing attack, explores the attack vectors and published (Fu *et al*., 2006) examples of preventative best practice in a research on a web application. Phishing being a form of online crime that entice the people into giving up private or commercial information is a budding security threat that already costs victims billions of dollars each year.

The security toolbar (Wardman *et al*., 2009) is one of the known approaches to detect phishing which are usually located in peripheral area in the browser. But when compared to the web content, warning indicators in

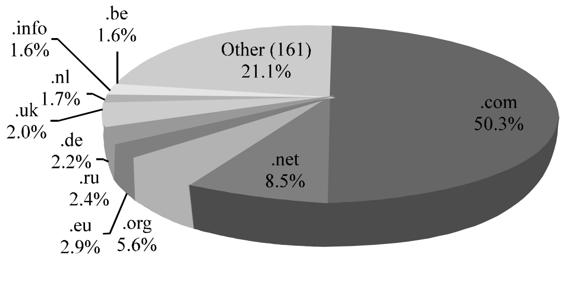


Fig. 1: Statistical view of phishing attack (Domain based)

Corresponding Author: C. Emilin Shyni, Anna University Chennai, Tamil Nadu, India toolbar became insignificant. The main disadvantage of toolbar is a reduced quantity of contextual data; secondly its incapacity to completely protect the user from the decision making process and finally toolbar is a simple cost argument.

The next tool that deals with automatic phishing detection mechanisms is the blacklist (Xiang and Hong, 2009) approach that attempts to inform clients of phishing sites either by pushing a modified list to the user plug-ins or having the users ensure with a server to demand data as it is coming up. These two specifications have their own drawbacks. These two approaches have their own difficulties. There occurs a definite latency issue when Blacklist server shows the modified set of phishing sites.

Another automatic approach available to detect phishing is heuristics (Werner and Courte, 2010) based technique identifies phishing sites in real time. But use of heuristics is found to be subjective and they produce large false positives. Reputation scoring yet another approach which is a fairly (Zhang *et al*., 2007) current improvement. This procedure involves ranking the phishing option of a given web page using reputation scores either collected from the certain web page or reported from the anti-phishing society. The trustworthiness of the reputation scoring algorithm is the vast test to this system. The web wallet (Cao *et al*., 2008) is a browser side bar which facilitates users to submit their sensitive information online.

This strategy of web wallet approach was successfully learned by Zhuge and Liu (2004) majority of the users to submit their login information. But the study also concluded that spoofing the web wallet interface itself was an effective attack.

SYSTEM ARCHITECTURE

The system architecture of is shown in Fig. 2. Here the client will enter the URL to view the required web page of a particular web server. That web service will have application server and web server. To increase the credibility of the web site of that particular web page, the client machine's current date and time will be displayed at the client browser. Usually, the phishing attack will occur that is the page may redirect during the money transaction.

When user registration takes place, a unique secret code will be generated for the user which will be encrypted and stored in the database of the server. When a user want to access a particular site, initially the URL address for the website will be entered in the phishing browser. The parser downloads the html content from the web site when the URL is fed to the browser. The parser analyses the form fields and the response fields. All the input is being gathered for the testing purpose. Form post is done after the random form filling. All the form post is recorded in heuristics. The heuristics recording is done till the end of URL. The heuristic matching is done; the legitimate and phishing web pages can be identified based on the heuristics.

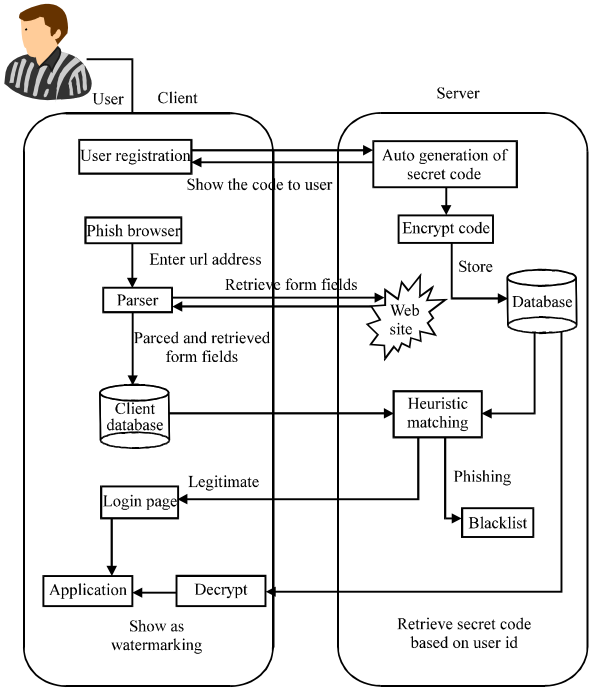


Fig. 2: System architecture of water marking

When the clients need to enter their personal details such as online banking password or ATM pin number the client needs to login. After logging in the client may not know whether he/she is in a correct page or not. Here is where the water marking technique plays a major role to give the highest credibility of that particular web page. After logging in and that is before giving the personal details, the user can check the credibility level of the web page. This credibility provision is only possible through water marking mechanism.

Online banking, online auction and share market come under this commercial domain. Usually, phishing attack happens when the user is entering his/her personal details such as bank transaction password or ATM pin number that he/she need to login. This is the place where the user is redirected to some other web page. Therefore, after logging in and before giving the sensitive information the user may be prompted to check the credibility of the web site using water marking techniques.

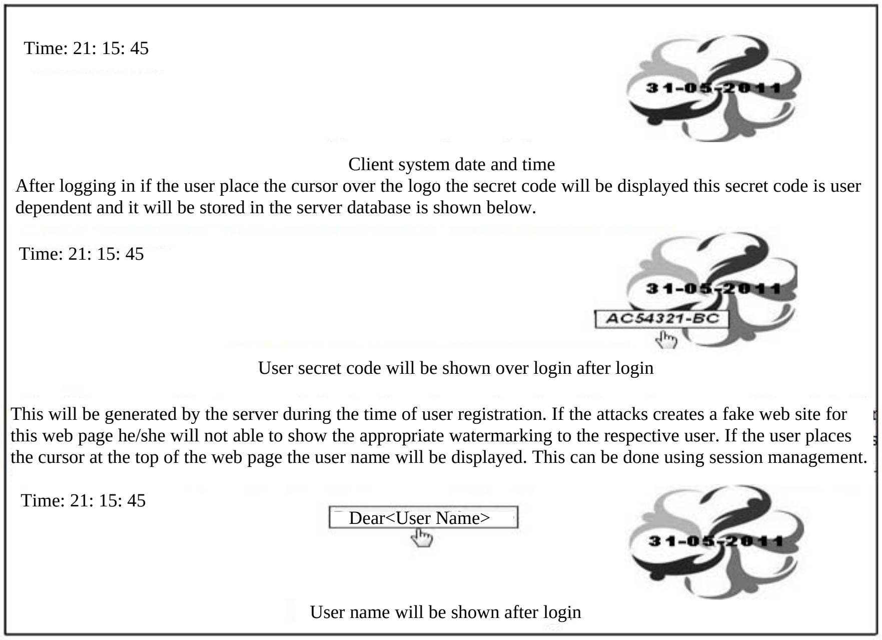
Before login into a commercial web site, the user can see his/her machine's date and time in the logo as shown in Fig. 3 which is initiated from the server. Sometimes the attacker may hack the server database to get the respective secret code of the user and may show the watermark in the fake website as a legitimate web site. To avoid this kind of problem, the server will encrypt the secret code before storing into the database. The following algorithm Fig. 4 is used to encrypt the secret string.

For example, let the secret code be AR12345678 and let X, Y and Z be the 3 keys having the values of keys are a, b and c, respectively. The ASCII equivalent of X is 97 and the ASCII equivalent of the first character of the secret code is 65. The new ASCII = ASCII equivalent of X + ASCII equivalent of A = 162. The binary value of 162 is = 100010 and the negation of binary = 01011101. The decimal conversion of the result is 93 the encrypted value of A is 93.

When the part of the image is in high level contrast for the human eye's attention, the MidInt (Vi) is the average value of the part of the image Vi and MidInt (Vi-Neighbouring) is the average value of all its surrounding selected images then the contrast value can be defined by the Eq. 1:

$$\text{QSimilarity}(\text{Wi}) = \text{MidInt}(\text{Vi})\ (1)\ \text{MidInt}(\text{Vi - Neighbouring}) \qquad (1)$$

This algorithm will convert the secret code into encrypted format which cannot be understood by human.


Time: 21: 15: 45

Client system date and time

After logging in if the user place the cursor over the logo the secret code will be displayed this secret code is user dependent and it will be stored in the server database is shown below.

Time: 21: 15: 45

User secret code will be shown over login after login

This will be generated by the server during the time of user registration. If the attacks creates a fake web site for this web page he/she will not able to show the appropriate watermarking to the respective user. If the user places the cursor at the top of the web page the user name will be displayed. This can be done using session management.

Time: 21: 15: 45

Dear<User Name>

User name will be shown after login

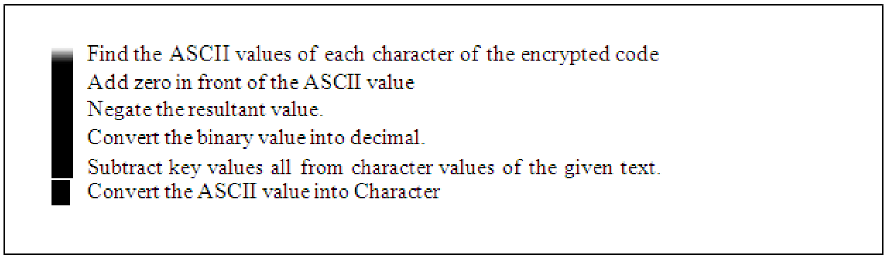
Fig. 3: System processing

- Take three key X, Y, Z and assign a character to each of them as X = a, Y = b and Z = c. (any character may be assigned)
- Find the ASCII value of X, Y and Z. Add the ASCII value of X to the ASCII value of first character that of Y to second character, Z to third character and alternatively add the ASCII values of X,Y, Z to consecutive characters.
- Convert each new ASCII value in to binary

  Negative the binary value and the end result increases the security.

  Convert the result into decimal values.
- The example is shown below:

Fig. 4: Encryption algorithm



Find the ASCII values of each character of the encrypted code
Add zero in front of the ASCII value
Negate the resultant value.
Convert the binary value into decimal.
Subtract key values all from character values of the given text.
Convert the ASCII value into Character

Fig. 5: Decryption algorithm

When the user logs in to this web site, the secret code of that particular user will be fetched from the database and decrypted into a human readable format using the following algorithm (Fig. 5).

Example: For example, let the encrypted secret code be 93, 75, 116, 108, 149, 104, 105, 103, 101, 102 and X, Y and Z be the three keys having the values a, b and c, respectively. The ASCII equivalent of X is 97 and the binary equivalent of the first character of the encrypted secret code is 1011101. Add zero in front of the encrypted secret code 01011101 and negate the value 10100010. The decimal equivalent of 10100010 is 162. The decimal equivalent 10100010 X-ASCII equivalent of A is 65. The character conversion of the result is A and the decrypted value of 93 is A. The position importance Q position of every image can be specified as the ratio of the number of pixels of the part of the image which are present in the middle position of the image to the total number of pixels in the specified image. This can be expressed as Eq. 2:

$$\text{QPosition (Vi)} = \text{middle (Vi) Sum (Vi)} \qquad (2)$$

Given below is the heuristic program behavior modeling to identify the phishing sites from the legitimate sites. The model gives the flexibility to detect phishing web sites that might steal information through an uncertain number of pages containing forms and employ various types of form generation techniques, (e.g., non XSS-based, XSS-based). The model differentiates a phishing site from a legitimate web site. It also applies offline analysis approach to navigate and download all the accessible pages by submitting random inputs and observe interesting responses.

In the Fig. 6, the state transformation from 0-1 is given by $<x_0, y_1>$ from 1-2 is given by $<x_1, y_2>$ and from 0-2 is given by $<x_0, y_2>$.

Experimental evaluation: This encrypted water marking mechanism is more secure than the previous since the date and time are initiated from the server and the secret code is displayed only after decryption. Even if the attacker hacks the server database he/she cannot understand the secret code. Some of the existing watermarking techniques are little more costly than this

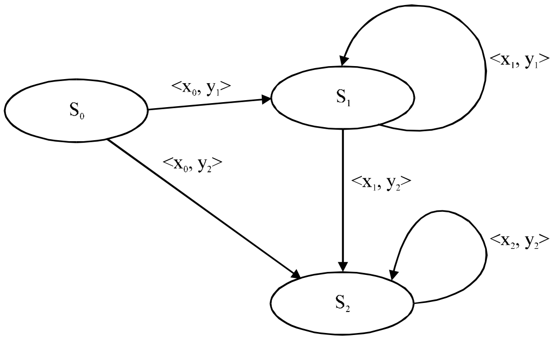


Fig. 6: FSM Model representing behaviors of phishing and legitimate web sitess

Table 1: Processing result for encryption based upon the ASCII value and binary value finds the final value

Characters	ASCII value	Key value	ASCII value adding the key	Binary value	Binary value a NOT operation	Final value
A	65	97	162	10100010	01011101	93
R	82	98	180	10110100	01001011	75
1	40	99	139	10001011	01110100	116
2	50	97	147	10010011	01101100	108
3	51	98	149	10010101	01101010	149
4	52	99	151	10010111	01101000	104
5	53	97	150	10010110	01101001	105
6	54	98	152	10011000	01100111	103
7	55	99	154	10011010	01100101	101
8	56	97	153	10011001	01100110	102

Table 2: Processing result for decryption based upon the decimal conversion and ASCII value finds the character

Encrypted value	Binary value	Binary value NOT operation	Decimal conversion	Key value	ASCII value	Final character
93	01011101	10100010	162	97	65	A
75	01001011	10110100	180	98	82	R
116	01110100	10001011	139	99	40	1
108	01101100	10010011	147	97	50	2
149	01101010	10010101	149	98	51	3
104	01101000	10010111	151	99	52	4
105	01101001	10010110	150	97	53	5
103	01100111	10011000	152	98	54	6
101	01100101	10011010	154	99	55	7
102	01100110	10011001	153	97	56	8

method since they need some additional software such as image magic. This is the simplest and easiest at the same time, efficient watermarking technique in order to prevent phishing attacks. Table 1 and 2 show the result for while processing the data.

When the user creates an account in a particular web application, the user secret code will be generated automatically and shown to the user. For security issues the secret code will be encrypted and stored in a database. The attackers having the web server to find the secret code of the user. To avoid the encryption technique is used. The encryption algorithm is given below with example the special feature of this algorithm is will gives more security to the secret code of the user by using key generation, binary conversion and taking not operation, similarly the decryption process also given below. Here, the main advantage is the secret code will be decrypted in the client side that is the server will send the encrypted secret code only to the client side using this researchers can prevents the man in the middle attacks also.

THE COMPARISON OF THIS METHOD AND THE EXISTING METHOD

Comparing the previous approach and the encrypted approach is in Table 3, the differences are tabulated above from which researchers can observe the following: the advantage of the approach is its credibility system, no usage of additional software and other additional techniques like encryption and decryption. A bar graph has been generated and shown below based on the above observation.

The credibility of the system before and after login, necessity of additional software and other advantage of the system are taken in the x-axis. The level of the system has been classified into high and low and taken as y-axis as shown in Fig. 7. Figure 8a and b shows the result of applying the algorithm. Figure 8a represents the key value and Fig. 8b shows after encryption. Figure 9 shows the time duration of the legitimate web sites.

Table 3: Comparison of previous approach and online identity phish approach

Previous approach	Online identity phish approach
No credibility before login	Gives the credibility before loginin terms of date and time
Gives the credibility after login using either username or secret code	Gives the high credibility after login using both username and code
Additional software is needed	No additional software is needed
No encryption and decryption technique	High encryption and decryption technique are used
No prevention of man in the middle attacks	Prevents man in the middle attacks

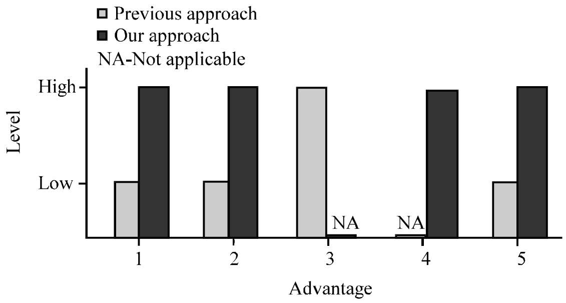


Fig. 7: Comparison of previous approach and online identity phish approach; 1) credibility before login (comparitivily); 2) credibility after login; 3) performance of additional software; 4) encryption and decryption; 5) prevention of man in the middle attack

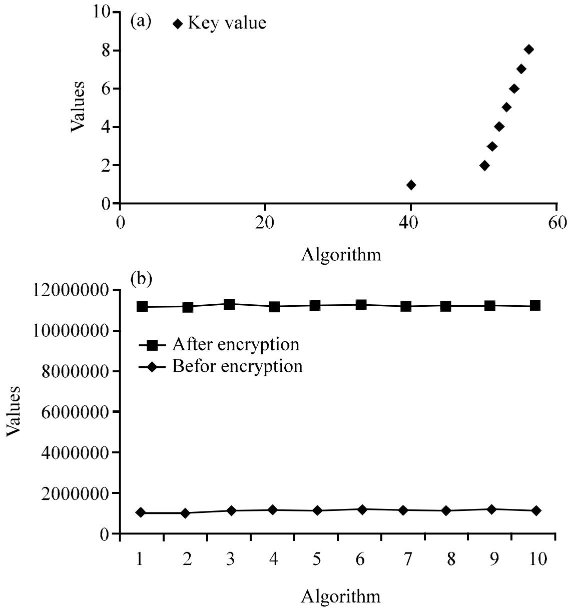


Fig. 8: After applying the algorithm; a) representing the key value; b) after encryption

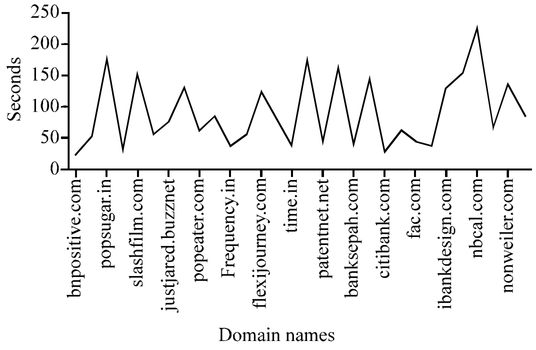


Fig. 9: Time duration for parsing legitimate web sites

CONCLUSION

Phishers can create a fake web site which is similar to the legitimate one and force the client to visit that fake web site and prompt them to enter the sensational information in order to steal the client's personal information. In this study, researchers propose the web service approach for detecting and preventing phishing attack by watermarking. Unlike the existing approach, the web service approach is platform independent and there is no load in the client side such as usage of additional tool. Researchers analyzed the web application with the developed web service and found the response time of the web application result set and error set. The response time without the approach doesn't give a major significance.

As part of future research, validations on the effectiveness of this approach are planned to be done. In addition to that the sturdy nature of the watermarking techniques can be mended by the use of high quality logo images in JPEG format or by spreading the message over all images in a web page.

REFERENCES

Bargadiya, M., V. Chaudhari, M.I. khan and B. Verma, 2010. Anti-phishing design using mutual authentication approach. Int. J. Comput. Sci. Inform. Technol., 1: 175-178.

Cao, Y., W. Han and Y. Le, 2008. Anti-phishing based on automated individual white-list. Proceedings of the 4th ACM Workshop on Digital Identity Management, October 27-31, 2008, Alexandria, VA., USA., pp: 51-60.

Fu, Y., L. Wenyin and X. Deng, 2006. EMD based visual similarity for detection of phishing webpages. Department of Computer Science, City University of Hong Kong. http://eople.sail.it.du/yf/ublication/WDA-antiphishing.pdf.

Wardman, B., G. Shukla and G. Warner, 2009. Identifying vulnerable websites by analysis of common strings in phishing URLs. Proceedings of the eCrime Researchers Summit, September 20-October 21, 2009, Tacoma, WA., pp: 1-13.

Werner, L.A. and J. Courte, 2010. Analysis of an anti-phishing lab activity. Inform. Sci. Educ. J., 8: 1-8.

Xiang, G. and J.I. Hong, 2009. A hybrid phish detection approach by identity discovery and keywords retrieval. Proceedings of the 18th International Conference on World Wide Web, Madrid, Spain, April 20-24, 2009, ACM, New York, USA., pp: 571-580.

Xiang, G., J.I. Hong, C.P. Rose and L.F. Cranor, 2011. CANTINA+: A feature-rich machine learning framework for detecting phishing web sites. ACM Trans. Inform. Syst. Secur., 14: 21-48.

Zhang, Y., S. Egelman, L. Cranor and J. Hong, 2007. Phinding Phish: Evaluating anti-phishing tools. Proceedings of the 14th Annual Network and Distributed System Security Symposium, February 28-March 2, 2007, Catamaran Resort Hotel, San Diego, CA., USA., pp: 1-16.

Zhuge, H. and J. Liu, 2004. A fuzzy collaborative assessment approach for knowledge grid. Future Gener. Comput. Syst., 20: 101-111.