Detection Of Phishing Websites And Secure Transactions

R. Dhanalakshmi, C. Prabhu, C. Chellapan
Department of Computer Science and Engg.
Anna University , Chennai- 60025, TamilNadu ,India
E-mail : drcc@annauniv.edu, dhanalakshmisai@gmail.com

ABSTRACT :Phishing is an electronic online identity theft in which the attackers use a combination of social engineering and web site spoofing techniques to trick a user into revealing confidential information. It steals the user's personal identity data and financial credentials. Most of the phishing attacks emerge as spoofed E-Mails appearing as legitimate ones which makes the users to trust and divulge into them by clicking the link provided in the E-Mail. To detect a Phishing website, human experts compare the claimed identity of a website with features in the website. For example, human experts often compare the domain name in the URL against the claimed identity. Most legitimate websites have domain names that match their identities, while Phishing websites usually have less relevance between their domain names and their claimed (fake) identities. In addition to blacklists, white lists, heuristics, and classifications used in the state-of-the-art systems, we propose to consider websites' identity claims. To enable secure transactions ,Password hashing has been done with MD5 hashing algorithms that strengthens web password authentication. It is also shown that getting original password from hashed form is not an easy task due to addition of salt value. If the user is valid, get a session key via mobile, through which further access can be done.

Keywords: Phishing,MD5 hashing, authentication, spoofing, Tokens

I. Introduction

In the field of computer security, phishing is the criminally fraudulent process of attempting to acquire sensitive information such as usernames, passwords and credit card details by masquerading as a trustworthy entity in an electronic communication. A phishing website is a broadly launched social engineering attack that attempts to defraud people of their personal information including credit card number, bank account information, social security number and their personal credentials in order to use these details fraudulently against them. Phishing has a huge negative impact on organizations' revenues, customer relationships, marketing efforts and overall corporate image. Communications purporting to be commonly used to lure the unsuspecting public. Phishing is typically carried out by e-mail or instant messaging, and it often directs users to enter details at a fake website whose look and feel are almost identical to the legitimate one. Phishing is an example of social engineering techniques used to fool users, and exploits the poor usability of current web security technologies. Attempts to deal with the growing number of reported phishing incidents include legislation, user training, public awareness, and technical security measures. As per Symantec Message Labs Security report of March 2011, the following statistics were observed

- Spam – 79.3% in March
- Viruses – One in 208.9 emails in March contained malware (an increase of 0.13 percentage points since February 2011)
- Phishing – One in 252.5 emails comprised a phishing attack
- Malicious websites – 2,973 web sites blocked per day

II. RELATED WORKS

A number of anti-phishing solutions have been proposed to date which are attempted at E-Mail level with validation of sites and Content analysis techniques(e.g Naïve Bayes Filter)[7]. Dhamija et al. [12] involves the use of a so-called dynamic security skin on the user's browser.The technique allows a remote server to prove its identity in a way that is easy for humans to verify, but difficult for phishers to spoof. First, browser extension provides a trusted window in the browser dedicated to username and password entry. Second, scheme allows the remote server to generate a unique abstract image for each user and each transaction. This image creates a "skin" that automatically customizes the browser window or the user interface elements in the content of a remote web page.Their extension provides the user with a trusted password window. This is a dedicated window for the user to enter usernames and passwords and for the browser to display security information. They present a technique to establish a trusted path between the user and this window that requires the user to recognize a photographic image.The remote server generates an abstract image that is unique for each user and each transaction.

They implement the Secure Remote Password Protocol (SRP), a verifier-based protocol developed by Tom Wu, to achieve mutual authentication of the user and the server.In general, two factor user authentication schemes serve to protect the server from fraud, rather

than protecting the user from phishing attacks if they do not provide a mechanism for the user to authenticate the server. This ignores the "limited human skills" property.In a later study [6], Dhamija et al. report that more than 20% of the users do not take visual cues into consideration when surfing and that visual deception

attacks can fool even the most sophisticated users. They showed 22 participants 20 web sites and asked them to determine which ones were fraudulent, and why. Their key findings are good phishing websites fooled 90% of participants, existing anti-phishing browsing cues are ineffective, and 23% of participants in their study did not look at the address bar, status bar, or the security indicators. Lui et al. [30] analyze and compare legitimate and phishing web pages to define metrics that can be used to detect a phishing page. A web page is classified as a phishing page if its visual similarity value is above a pre-defined threshold.

CANTINA[17] examines the content of a web page to determine whether it is legitimate or not, in contrast to other approaches that look at surface characteristics of a web page, for example the URL and its domain name. CANTINA makes use of the well-known TF-IDF (term frequency / inverse document frequency) algorithm used in information retrieval and more specifically, the Robust Hyperlinks algorithm for overcoming broken hyperlinks. Their results show that CANTINA is quite good at detecting phishing sites, detecting 94-97% of phishing sites. They also show that they can use CANTINA [17] in conjunction with heuristics used by other tools to reduce false positives.

III. PROPOSED SYSTEM

A. Aim

To distinguish the phishing websites from the legitimate websites and ensure secure transactions us provide the session key to improve the security.

B. Existing Approaches

There are many existing techniques used to detect the phishing website but it has some false positive values. Hence there is a possibility to identify the legitimate website as a phishing website. The proposed system checks the four features to identify phishing website such as WHO IS, URL, domain and inter domain features.

1. The methods used to identify the phishing websites are,
    i. Black list
    ii. White list
    iii. Heuristic based
    iv. Classification
2. The existing anti-Phishing technique, password Hash technique is ineffective against some attack. Phishers stole the database entries and then tries hashes and after an exhaustive search they get the password.
3. During implementation ,SHA-1 is more secure but slow in execution as SHA-1 includes more rounds comparatively.

C. Proposed Methodology

1. This system checks the following features for a given website
    i. WHOIS features
    ii. IP address
    iii. Domain and Inter domain and it is depicted in Fig 1.
2. Password hashing has been described with MD5 hashing algorithms that strengthens web password authentication. It is also shown that getting original password from hashed form is not an easy task due to addition of salt value. If the user is valid get a session key via mobile, through which further access can be done.
3. If the user unfortunately login into the Phishing site an automatic alert will be displayed.
4. Each and every time a user enters the username and password to the website a random number is sent to the users mobile phone. Even the Phishers get the secret number they can't perform any transaction without the random number.

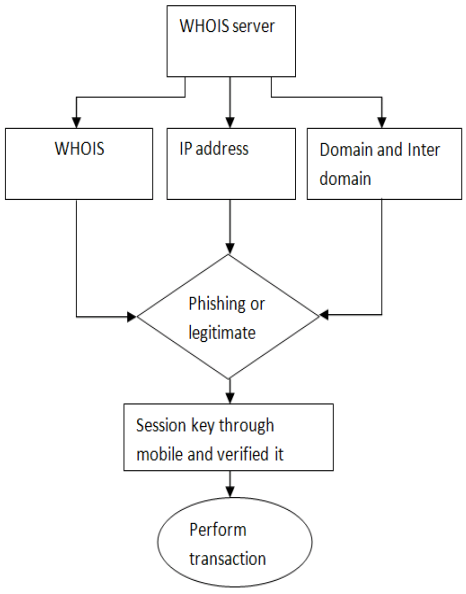

Fig 1 Phishing website detection mechanism

D. List of Modules

1. Phishing Data base Construction
2. Validation of Website by WHOIS server
3. User login
4. Password Salting and Hashing Using MD5 and authentication

E. Phishing data base construction

The phishing data base construction is carried using periodic monitoring of the phishing attacks and IP address along with their WHOIS, Domain name. This information is updated in the DNS server for further validation which is useful for further screening of the phishing website. If any web link of this data base is requested by the user without knowing it is phishing web site, immediately the user will be intimated with an alert that this site is phishing website. Hence the user will be not giving any sensitive data to that phishing website.

F. Phishing bank site

The Phishing site resembles the original site which steals the user's confidential data and stores the data in their database. Later they can apply those details into original bank site and perform the transactions.

FBlock diagram for phishing site interaction

G. Validation of website and domain by WHOIS server

The main DNS server is having all the information of Original & Phishing Web sites. Each web site has who is information along with the IP address. WHOIS is all about the website registration; owner name of the web site is registered, along with the company details. Every Web site has an IP address, which will be used for authentication. Phishing Database is always updated with the Phishing Website's details for verification.

H. Authentication server

Whenever the client accesses any site the site is validated in the WHOIS server. The WHOIS is a database in an authentication server. Phishing websites tend to have poorly managed WHOIS records. For example, some of the basic records including registrant and the date of registration might be missing or WHOIS lookup might not even succeed .It is given in Fig 2 For a given website, server sends WHOIS query for its domain and extracts three kinds of features below. Among these, Registrant becomes a Common Name candidate for this website identity.

• Registrant: who (what) owns this domain?

• Dates: what are the dates of registration, update, and expiration?

• Name server: how many name servers exist? Does this domain have its own name servers?

If the site is registered then only the server allows the user to access that site otherwise it will intimate the user by displaying some alert messages.

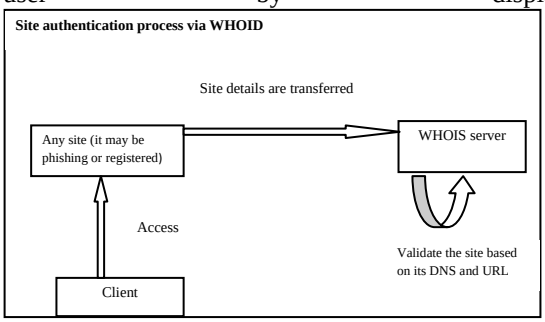

Fig 2 : Querying the WHOIS server

Validation of website by IP address and Inter domain

The DNS server will also have the complete details regarding the Domain Name & Inter domain in the web address. Each & every web site will have Domain name (.Com, .Co.in, .Edu,. Tech, .Co.uk, .in, Org & etc). Inter domain is all about any two domain names in the same link, www.123.com/456.com. Phishing Database is always updated with the Phishing Website's details for verification.

USER LOGIN

The client application is designed to get the data from the platform. Here the client sends user name and password for getting authentication. The authentication for client access is given, if and only if both the user name and password matches to the details in database. Else access is denied. After authentication, client get session via mobile using which the client can perform further transaction.

IV. PASSWORD SALTING AND HASHING USING MD5 AND AUTHENTICATION

Using the concept of hashing passwords secured from unauthorized users. User just type simple password e.g. "test" but it will be stored in the form of hash in database using MD5 algorithm. Here salt is used to add special characters with the hashed

password, date and time with the help of salting. This method can help in storing encrypted passwords in the database and improve security, as most of the high security data is kept in a form not usable for unauthentic users.

This module is used to generate session for the login client. If the user is an authenticated person, he gets his session key via mobile using which the client can perform further transaction and is given in Fig 3.

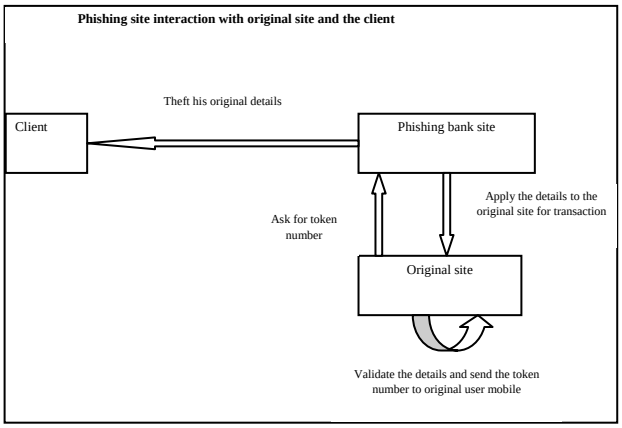

Fig 3 : Client Transactions

Thus a hacker cannot know or access transaction since he cannot know the session. This generation of the password is achieved by Real-time Mobile connected with the Bank server. The mobile number of the user is obtained from the bank database. Once the user gives the user name and password the password hashing and salting process is preceded. This value is send as SMS to the mobile number of the user for further Authentication to avoid any further attacks using IP Spoofing.

The DNS server will first verify whether the requested web site is Genuine or not. If the DNS server identifies that the requested website is of Phishing website, the server will intimate the user that the requested web site is Phishing, so that user will not give any sensitive data to the phishing website. If the website requested is Genuine, any way further verification is also carried using Salting and Hashing using MD5 algorithm, this is sent as SMS to the Legitimates mobile number. Once the Legitimate User gives the input of the Session key which is sent as SMS and if it is authenticated, only then the user is allowed for further transaction process.

MD5-MESSAGE DIGEST ALGORITHM

A. Logic

The algorithm takes as input a message of arbitrary length and produces as output a 128-bit message digest. The input is processed in 512-bit blocks.

A. Steps

Step 1: Appending padding bits

The message is padded so that its length in bits is 512 bits. Length of the padding is in the range of 1 to 512 bits. The padding consists of a single 1-bit followed by the necessary number of 0-bits.

Step 2: Append length

A 64-bit representation of the length in bits of the original message(before padding) is appended to the result off step 1.

Step 3: Initialize MD buffer

A 128-bit buffer is used to hold intermediate and final result of the hash function. The buffer can be represented as force 32-bit registers (A, B, C, and D). These registers are initialized to the following 32 0-bit integer (Hexadecimal values).

Step 4: Process message in 512 bits+ (16 words) blocks

The algorithm consists of four rounds of processing. These four rounds have a similar structure but using different primitive logical function.

Each round takes input the current 512 block being processed and 128-bit buffer value updates the contents of the buffer.

The output of fourth round is added to the input to the first round to produce the next round. The addition is done by independently for each of the four words in the buffer with each of the four words in the buffer with each of the corresponding word in current step, using addition.

Step 5: Output

After all 512-bits blocks have been processed, the output from that stage is 128-bit message output.

Operation

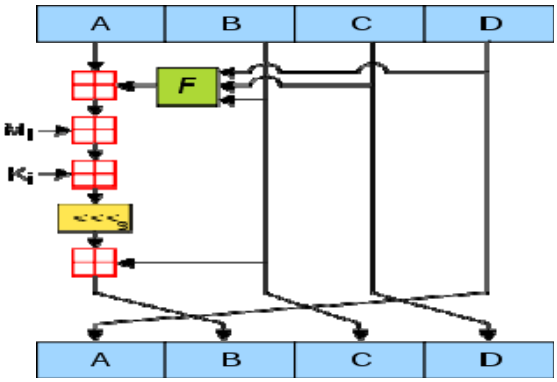

Fig 4 : one MD5 operation

There are four possible functions *F*; a different one is used in each round:

$$F(X,Y,Z) = (X \wedge Y) \vee (\neg X \wedge Z)$$
$$G(X,Y,Z) = (X \wedge Z) \vee (Y \wedge \neg Z)$$
$$H(X,Y,Z) = X \oplus Y \oplus Z$$
$$I(X,Y,Z) = Y \oplus (X \vee \neg Z)$$

$\oplus, \wedge, \vee, \neg$ denote the XOR, AND, OR and NOT operations respectively.

MD5 hashes

The 128-bit (16-byte) MD5 hashes (also termed message digests) are typically represented as a sequence of 32 hexadecimal digits. The following demonstrates a 43-byte ASCII input and the corresponding MD5 hash:

MD5 ("The quick brown fox jumps over the lazy dog")
= 9e107d9d372bb6826bd81d3542a419d6.

V. RESULTS

After querying the whois database the results are given in Fig 5. The table that contains the information about the websites. The information like URL, domain and inter-domain and these are stored in the database.

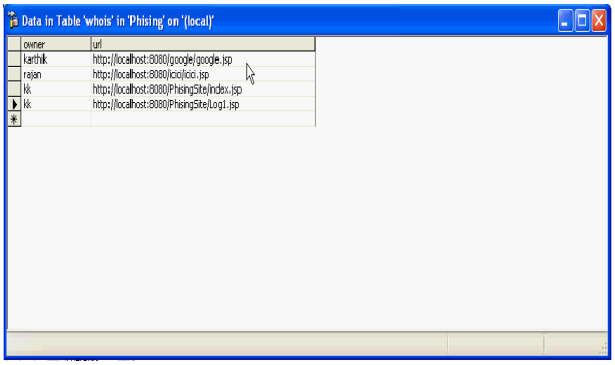

Fig 5 WHOIS database.

Fig 6. Shows the login screen for the user login. The user enters into his/her account by giving the username and password. The server allocates the access rights to the user

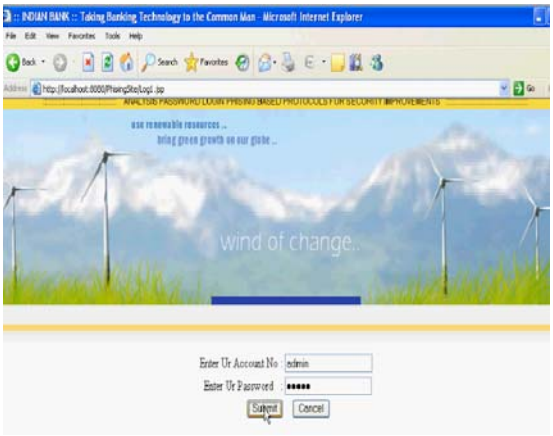

Fig 6 : Login Page for banking website

The system retrieves all the user details and it will stored in the database.For Secure transactions we use session key verification. For each and every transaction random key is generated and send to the user's mobile. The user want to enter that key, after that the verification is done and is depicted in Fig 7.

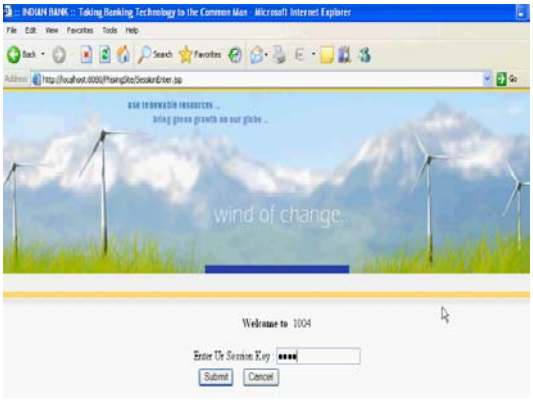

Fig 7 : Secured Transaction via session keys

V. Conclusion

Phishing detection system is used in online banking & E-business. In the banking applications, the online transaction is very useful for all the customers of the corresponding bank. The client request the website for transaction process then this phishing detection system validate the website with four features such as WHOIS registration, IP address, domain and inter domain values of the corresponding requested website. The website related information's are stored in a WHOIS database for verification process. If the website is phishing website then this system will alert the user. If unfortunately the client enter into the phishing website then the phisher will theft all the information's entered by the user and apply those information into the original website.

In the original website the administrator will create a new account and if already created account means user login into that and precede the transaction process. The client enter the account number and password for login into that website that password stored securely in database by using a MD5 algorithm. After login into the website this system generates the session key and send through the user's mobile. The phisher login into the original website means they can't get the session key so they enter a wrong session key then this system deny the access. The original user means the transaction is performed successfully.

VI. FUTURE ENHANCEMENT

More experiments with larger datasets will also be preformed to make AIWL (Automated Individual White-List) more efficient. The change rate of IP will be a problem in AIWL, longer time-span need to be used to gather the web sites' IP and analyze.

ACKNOWLEDGMENT

This work is supported by the NTRO, Government of India. NTRO provides the fund for collaborative project "Smart and Secure Environment" and this paper is modeled for this project.Authors would like to thanks the project coordinators and the NTRO members

References

1. ***Anti-Phishing Working Group (APWG). APWG Homepage. http://www.antiphishing.org/, 2007.***
2. ***Blake Ross, Collin Jackson, Nicholas Miyake, Dan Boneh, and John C. Mitchell.Stronger Password Authentication Using Browser Extensions. In 14th Usenix Security Symposium, 2005.***
3. ***Engin Kirda and Christopher Kruegel. Protecting Users against Phishing AttacksThe Computer Journal, 2006.***
4. ***Fritz Schneider, Niels Provos, Raphael Moll, Monica Chew, and Brian Rakowski. Phishing Protection Design Documentation. http://wiki.mozilla.org/Phishing Protection: Design Documentation, 2007.***
5. ***J. Ma, L. Saul, S. Savage, and G. Voel, Beyond blacklists: Learning to detect malicious web sites from suspicious urls, In The 15th ACM SIGKDD ker Conference On Knowledge Discovery and Date Mining, 2009.***
6. ***Liu Wenyin, Guanglin Huang, Liu Xiaoyue, Zhang Min, and Xiaotie Deng. Detection of phishing webpages based on visual similarity. In 14th International on World Wide Web (WWW): Special Interest Tracks and Posters,2005.***
7. ***Lobato DH, Lobato JM (2008). Bayes Machines for binary classification. Pattern Recognition Letters. Elsevier, 29: 1466-1473.***
8. ***Maher Aburrous a,*, M.A. Hossain a, Keshav Dahal a, Fadi Thabtah b, "Intelligent phishing detection system for e-banking using fuzzy data mining",Expert Systems with Applications,Elsiever ,Volume 37,2010, pp. 7913-792.***
9. ***Neil Chou, Robert Ledesma, Yuka Teraguchi, Dan Boneh, and John Mitchell Client-side defense against web-based identity theft. In 11th Annual Network and Distributed System Security Symposium (NDSS '04), San Diego, 2005.***
10. ***Niels Provos. Phishing Protection: Server Spec: Lookup Requests. http://wiki.mozilla.org/Phishing Protection: Server Spec#Lookup Requests, 2007.***
11. ***Phishtank. Phishtank feed: validated and online. http://data.phishtank.com/data/onlinevalid/index.xml, 2007.***
12. ***Rachna Dhamija and J. D. Tygar. The battle against phishing: Dynamic security skins. In Proceedings of the 2005 symposium on Usable privacy and security, New York, NY, pages 77–88. ACM Press, 2005.***

13. *Rachna Dhamija, J. D. Tygar, and Marti Hearst. Why Phishing Works. In Proceedings of the Conference on Human Factors In Computing Systems (CHI) 2006, Montreal, Canada. ACM Press, 2006.*
14. *S. Garera , N. Provos, M. Chew, and A. D. Rubin, A framework for detection and measurement of phishing attacks, In Proceedings of the 2007 ACM workshop on Recurring malcode, 2007.*
15. *Yahoo. Yahoo! AntiSpam Resource Center. http://antispam.yahoo.com/domainkeys, 2007.*
16. *Yue Zhang, Serge Egelman, Lorrie Cranor, and Jason Hong. Phinding Phish:Evaluating Anti-Phishing Tools. In Network and IT Security Conference: NDSS*
17. *Y. Zhang, J. Hong, and L. Cranor, Cantina: A content-based approach to detecting phishing web sites, In the 16th International Conference on World Wide Web, May 2007.*