ISSN:0975-9646

An Hybrid Anti-Phishing Approach Using MAC Verifier, Pattern Similarity Index (PSI) and SLIQ Decision System

Anil Kumar Goutam[1], Tejalal Choudhary[2]
Department of Computer Science and Engineering,
SD Bansal College of Technology Indore, India

***Abstract*— In this era of technology, internet and its application is accepted in a wide range, the main of these applications are providing ease in transition, convey messages and data from one end of the world to another conclusion. These methods are consuming various sensitive information for their own record. But due to theoretical design of web applications this information can be stoned by attackers by which they are drawing or creating problems in real life. The primary aim of this presented work is to study about various phishing techniques and their effects on our daily life additionally finding some acceptable and/ or adoptable detection and prevention techniques by which system automatically detects a phishing web URL uses data mining techniques. Along with the studying the work had also identified the problems associated with the current detection. Here the traditional system seems to provide detection with high false positive rates with static rules for pattern collections. This work proposes a hybrid anti-phishing approach using some of the well known phishing detection factors like MAC address of web pages. Laos the approach holds the pattern similarity index for analyzing the most relativity of the entered pattern with the phished information. For getting an accurate classification and decision, the SLIQ based mining algorithm is used. At the initial level of work the approach seems to provide effective results in near future.**

***Keywords*— Web Security, Phishing, Pattern Analysis, MAC, Pattern Similarity Index (PSI), Decision System, SLIQ;**

I. INTRODUCTION

Phishing is attempting to get information (and sometimes, indirectly, money) such as usernames, passwords, and credit card details by impersonating as a trustworthy entity in an electronic communication. Communications maintaining to be from popular social web sites, auction websites, online payment processors or IT administrators are commonly used to lure the unsuspecting public. Phishing emails may contain links to web sites that are infected with malware. Phishing is typically carried out by e-mail spoofing 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 a model of social engineering techniques used to deceive users and exploits the poor usability of current web security technologies. Attempts to deal with the rising number of reported phishing incidents include legislation, user training, public awareness, and technological security measures.

II. BACKGROUND

Phishing is a form of social engineering in which an attacker, also known as a pulsar, attempts to fraudulently retrieve legitimate users' confidential or sensitive credentials by mimicking electronic communications from a trustworthy or public governance in an automated fashion. A complete phishing attack involves three roles of phishers. Firstly, mailers sends out a large number of fraudulent emails (usually through botnets), which direct users to fraudulent websites. Secondly, collectors set up fraudulent websites (commonly hosted on compromised machines), which actively prompt users to supply confidential information. Finally, cashers use the confidential information to achieve a payout. Monetary exchanges often occur between those fishers. Phishing has spread beyond email to include VOIP, SMS, instant messaging, social networking sites, and even multiplayer games. Below are some major categories of phishing.

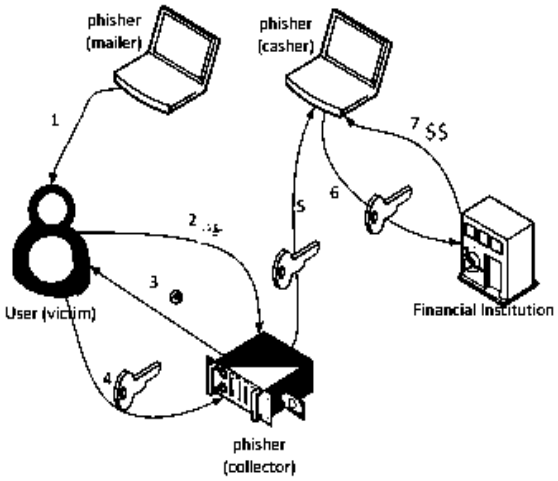

Figure 1 shows the phishing

- **Spear Phishing**

Phishing attempts directed at specific individuals or companies have been termed spear phishing. [3]. Attackers may gather personal information about their target to increase their probability of success.

- **Clone Phishing**

A type of phishing attack whereby a legitimate, and previously delivered, E-mail containing an attachment or link has had its content and recipient addresses taken and

used to create an almost identical or cloned E-mail, The attachment or Link within the email is replaced with a malicious version and then sent from an email address spoofed to appear to come from the original sender. It may claim to be a re-send of the original or an updated version to the original.

This technique could be used to pivot (indirectly) from a previously infected machine and gain a foothold on another machine, by exploiting the social trust associated with the inferred connection due to both parties receiving the original E-mail.

- **Whaling**

Many phishing attacks have been coordinated particularly at senior executives and other prominent focuses inside businesses, and the term whaling has been begat for these sorts of attacks.

- **Link Manipulation**

Most strategies for phishing utilize some manifestation of technical deception intended to make a connection in an E-mail (and the spoofed website it prompts) seem to fit in with the spoofed organization. Incorrectly spelled URLs or the utilization of sub-domains are common tricks utilized by phishers. In the accompanying example URL, http://www.yourbank.example.com/, it shows up as if the URL will take you to the example segment of the your bank website; really this URL focuses to the "your bank" (i.e. Phishing) area of the example website. An alternate common trick is to make the displayed text for links (the text between the <A> tags) recommend a reliable destination, when the connection really goes to the phishers' site. The accompanying example join,/en.wikipedia.org/wiki/Genuine, seems to direct the client to an article titled "Genuine"; clicking on it will truth be told bring the client to the clause entitled "Deception". In the lower left hand corner of most browsers clients can preview and verify where the connection is going to take them. Floating your cursor over the connection for several seconds may do a comparable thing, however this can at present be set by the phisher.

A further problem with URLs has been found in the handling of internationalized domain names (IDN) in web browsers that may permit visually identical web addresses to prompt different, perhaps malicious, websites. Notwithstanding the attention surrounding the flaw, known as IDN spoofing. on the other hand homograph assault. phishers have taken preference of a comparative danger, utilizing open URL redirectors on the websites of trusted organizations to disguise malicious URLs with a trusted domain. Even digital certificates don't tackle this problem on the grounds that it is very feasible for a phisher to purchase a substantial certificate and subsequently change content to spoof a genuine website.

- **Filter Evasion**

Phishers have used images instead of text to make it harder for anti-phishing filters to detect text commonly used in phishing e-mails.

- **Website Forgery**

When a victim visits the phishing website, the deception is not over. Some phishing scams use JavaScript commands so as to adjust the address bar.This is carried out either by placing a picture of a legitimate URL over the address bar, or by closing the original address bar and opening another one with the legitimate URL.

An attacker can even use flaws in a trusted website's own particular scripts against the victim. These types of attacks (known as cross-site scripting) are especially problematic, because they direct the user to sign in at their bank or service's own particular page, where everything from the web address to the security certificates appears right. Really, the connection to the website is crafted to do the attack, making it extremely hard to spot without specialist knowledge. Just such a flaw was used in 2006 against PayPal.

A Universal Man-in-the-middle (MITM) Phishing Kit, discovered in 2007, provides a simple-to-use interface that allows a phisher to convincingly reproduce websites and catch log-in details entered at the fake site.

To stay away from anti-phishing techniques that scan websites for phishing-related text, phishers have started to use Flash-based websites. These look a ton like the genuine website, yet conceal the text in a multimedia object.

- **Phone Phishing**

Not all phishing attacks oblige a fake website. Messages that claimed to be from a bank advised users to dial a phone number regarding problems with their bank accounts. Once the phone number (possessed by the phisher, and provided by a Voice over IP service) was dialed, prompts advised users to enter their record numbers and PIN(Personal Identification Number). Vishing (voice phishing) sometimes uses fake caller-ID information to give the appearance that calls come from a trusted organization.

- **Other Technique**

Another approach used successfully is to forward the customer to a bank's legitimate website, then to place a popup window requesting credentials on top of the website in a way that it appears the bank is requesting this sensitive information.

One of the latest phishing techniques is tab nabbing. It takes advantage of the multiple tabs that users use and silently redirects a user to the affected site.

Evil twins are a phishing technique that is hard to detect. A phished creates a fake wireless network that looks similar to a legitimate public network that may be found in public places such as airports, hotels or cafes. Whenever someone logs on to the bogus network, fraudsters attempt to capture their passwords and/or credit card information.

III. OBJECTIVE

This proposed study includes the following area of study, is given as:

(i) Study of different phishing attack: in this part of the proposed work, various kinds of phishing attacks are evaluated and reviewing their properties and resources that are required to phish any victim.

(ii) Finding the types of phishing attacks: classify the attacks according to their behaviour and identification techniques.
(iii) Identifying the problem domain and relevant solution : finding the area of the problem and offer the optimum solution for that problem
(iv) Implement the proposed technique using attractive GUI and simulate real world problem and solutions
(v) Justify the proposed solution using performance analysis and comparison with the traditional techniques

IV. PROBLEM DEFINITION

The current usage of SSL/TLS by browsers, still allows web spoofing, i.e. misleading users by impersonation or misrepresentation of identity or of credentials. Indeed, there is an alarming increase in the amount of real-life web-spoofing attacks, usually using simple techniques. Often, the swindlers lure the user to the spoofed web site, e.g. impersonating as financial institution, by sending her spoofed E-mail messages that link into the spoofed web-sites; this is often called a phishing attack. The goal of the attackers is often to obtain user-ID's (Identity), passwords/PIN (Personal Identification Number) and other personal and financial information, and abuse it e.g. for identity theft. Thus, the significant improvement in detection of spoofed sites is required.

V. PROPOSED SOLUTION

As described in the problem domain, this work suggested the solution as an browser extension i.e. plug in. It also includes usability experiments, to measure and compare the effectiveness of the approach to sites identification indicators. The following work provides the given solutions:

(i) Provide prevention from the spoofing and phishing done by using a URL.
(ii) Allow to open, but the authentic web pages.
(iii) Provides the authenticity rating to the web pages.
(iv) Designs the certificates providing authenticity.
(v) Maintain the data base according to which authenticity rating will be provided.

To design and implement a secure toolbar for the web servers, the phishing control mechanism must be embedded with the browser. The control on the browser must be defined in such a way that it should contains or preserves the authenticity of the genuine web pages and must block the access of the modified web pages. These alternate web pages seems like original once is termed as phishing web. The work had also found that for effectively analysing the phishing web pages following facts are very must important. It is required to analyze the URL patterns and checks the phish reporting online database along with the feedbacks. After conclude we are design a new system that is following the above given directives.

To provide the optimum solution for the Anti-phishing we proposed the below given system architecture. To properly understand and implement the complete model some modules are designed, their description and working is given as:

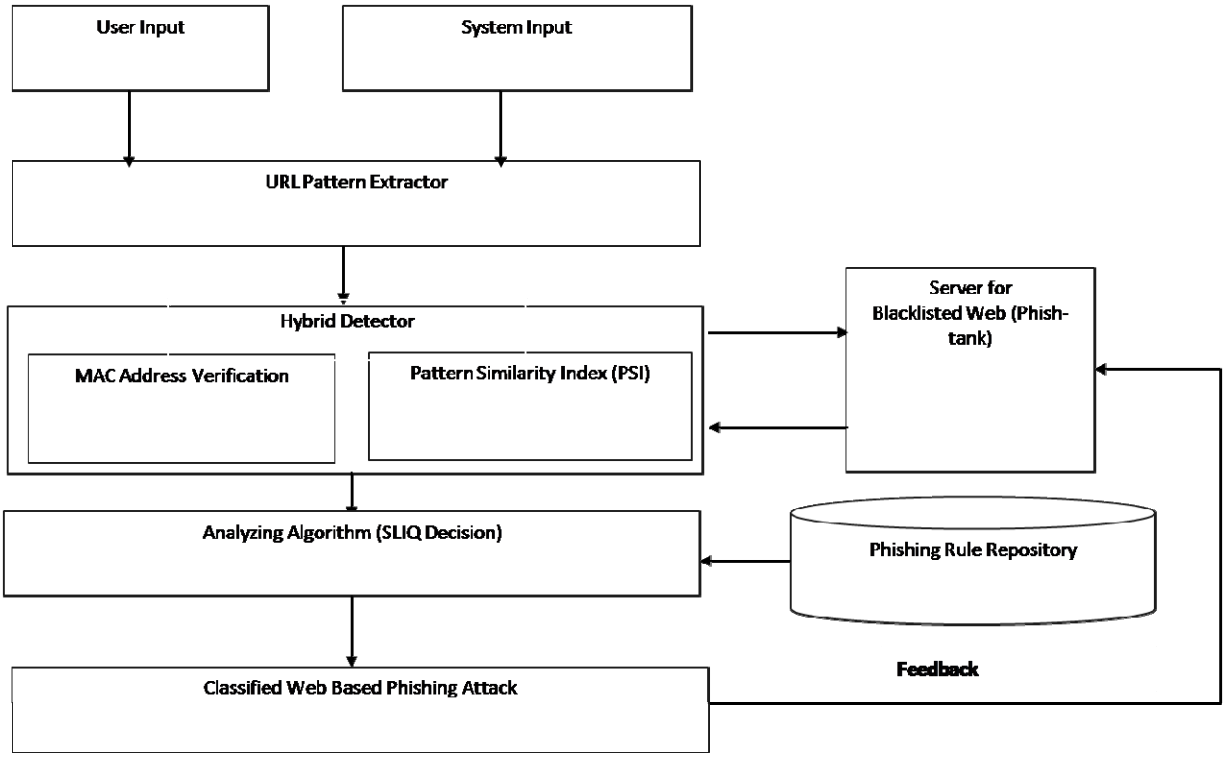

Figure 2: Proposed Hybrid Anti-Phishing Approach

Component of proposed system

a) **Inputs:** Here the system is having two different types of input. The fist one is user generated in the form of an entered URL in the browser. Second is the system generated input where the browser generates the query about the web page authenticity.

b) **URL Pattern Extractor:** This modules works as a separator for the URL into multiple categories. Here the extractor works as parser for analyzing the information available with the entered URL.

c) **Hybrid Detector:** This modules works towards detecting the probability of web page as a phished page. Here the attacker works towards making the changes somewhere in the web or just made a duplicate copy with alternative information. Now the module has two important phases: MAC address verifier and Pattern Similarity Index (PSI). The MAC address verifier fetches the original address of the web page. The PSI module calculates the relativity of the pattern inserted by the user with the blacklisted web based phish tank.

d) **Server for Blacklisted Web (Phish tank):** that is an updated database where the entire phish reported web URLs is stored, proposed system contains a relational data table which store these web URL patterns and used to build data model form algorithm selected.

e) **Phishing Rule Repository:** This component is work as database repository which holds the rules for phished pages detection. Mainly it handles the various query based patterns which verifies the authenticity of the pages. **T**his database in common for all guests who use proposed tool, this database contains user feedback about URLs.

f) **Analyzing Algorithm SLIQ:** This module contains the algorithm for analyzing the web page as a fake page using a well known mining algorithm SLIQ. It is used to mine the correct and most relative information and pattern. It works as decision support system for malicious web. It consumes phish-tank database along with the navigational model. This module also had data model is a decision tree which is grown using a phish tank database and used to analysis the URL pattern which is found in the database. After analysis of web URL decision is reached.

g) **Feedback:** a user interface provided in the proposed model to submit feedback for a URL if required to report and this is taken in both databases. It is a user interface where user's navigated URL information is stored and provides the various user experiences about the navigated page.

VI. EVALUATION PARAMETERS

To demonstrate the results the work had started with the conventional definition of different keywords used in the evaluation. The performance parameters that are required to simulate is given as:

- **System Accuracy**

To evaluate the performance of the proposed system is calculated using the n cross validation method. The overall classification accuracy is given below and evaluated using the below given formula, the listed accuracy of the system are the best performance during different experiments.

$$accuracy = \frac{totalvalues}{Totalcorrectlyclasstfied} classifyX100$$

- **Error Rate**

The error rate is a scale which provides the information how much amount of objects are incorrectly classified during experiments. Due to experiments we found that as the size of database increase the error rate increases, which is judged using the contributed pattern.

$$error = \frac{TotalIncorrectlyclasstfied}{TotalNumberofobjects} X100$$

VII. EXPECTED OUTCOME

In this proposed work finding a way to produce and implement a plug in, a browser extension or an individual tool for improved secure identification indicators. Users can specify a name/logo to a secure website. This name/logo is presented by plug in, when the browser presents that secure site; otherwise, Trust Bar presents the certified site's owner name, and the name/logo of the Certificate Authority who identified the owner. Some of these ideas are already adopted by browsers, following is my work.

Describe usability experiments, which measure, and prove the effectiveness, of plug in improved security and identification indicators. Therefore, try to derive general secure-usability principles from my experiments with the plug in.

VIII. CONCLUSION

In this proposed work we are search for an advance option for detecting and preventing phishing attacks by evaluating the structure of URL navigated using web browsers. The proposed methods we motivated with phishing detection and preventions of using phishing pre-URL pattern and post URL pattern detection methodology. In this concept we make some small modifications on detection and prevention technique. Two kinds of databases are utilized first working locally and second is global database. In local database Phish tank web URL are stored and in second database user feedbacks and experiences are available. After evaluation of performance that's user choice to use any kind of data model implemented in system. For preparing data set we break the URLs in small chunks and treated as the attribute for creating data model (decision tree). any web URL is verified more than one process, first by using Phish tank database, secondly with decision tree data model and finally with the user experience.

REFERENCES

[1] Guang Xiang, Jason I. Hong by "A Hybrid Phish Detection Approach by Identity Discovery and Keywords Retrieval" published in ACM 978-1-60558-487-4/09/04 in Apr 2009.

[2] Aaron Emigh, Radix Labs "Anti-Phishing Technology" published in report was produced in conjunction with the United States Secret Service San Francisco Electronic Crimes Task Force in January 19, 2005.

[3] Aanchal Jain, Prof. Vineet Richariya " Implementing a Web Browser with Phishing Detection Techniques " published in World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 1, No. 7, 289-291, 2011.

[4] Yue Zhang, Jason Hong, Lorrie Cranor "CANTINA: A Content-Based Approach to detecting Phishing Web Sites" published in ACM 978-1-59593-654-7/07/0005, 2007

[5] Brad Wardman, Tommy Stallings, Gary Warner, Anthony Skjellum "High-Performance Content-Based Phishing Attack Detection" UAB Department of Computer & Information Sciences, the UAB Department of Justice Sciences in 2009.

[6] Angelo P. E. Rosiello,Engin Kirda,Christopher Kruegel, Fabrizio Ferrandi"A Layout-Similarity-Based Approach for Detecting Phishing Pages" in 2008.

[7] Madhusudhanan Chandrasekaran, Ramkumar Chinchani, Shambhu Upadhyaya "PHONEY: Mimicking User Response to Detect Phishing Attacks" Proceedings of the 2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks .

[8] Andre Bergholz, Gerhard Paa , Frank Reichartz, Siehyun Strobel "Improved Phishing Detection using Model-Based Features" in 2009.

[9] Ge Wang, He Liu, Sebastian Becerra, Kai Wang, Serge Belongie, Hovav Shacham, and Stefan Savage "Verilogo: Proactive Phishing Detection via Logo Recognition" in 2011.

[10] S.Arun, D.Anandan, T.Selvaprabhu, B.Sivakumar, P.Revathi, H.Shine "Detecting Phishing Attacks In Purchasing Process Through Proactive Approach" An International Journal (ACIJ), Vol.3, No.3, May 2012.

[11] Jonathan Zdziarski Weilai Yang Paul Judge" Approaches To Phishing Identification Using Match And Probabilistic Digital Fingerprinting Techniques" in 2004.

[12] Sadia Afroz and Rachel Greenstadt "PhishZoo: An Automated Web Phishing Detection Approach Based on Profiling and Fuzzy Matching" in 2009.

[13] Fernando Sanchez and Zhenhai Duan "URL Obfuscation Phishing and Anti-Phishing: A Review " published in ISSN : 2248-9622, Vol. 4, Issue 1(Version 1), January 2014.

[14] Srishti Gupta, Ponnurangam Kumaraguru "Emerging Phishing Trends and Effectiveness of the Anti-Phishing Landing Page" in 2014.

[15] Shaun Cooley, William E- sobel" Anti-Phishing Early Warning System Other Publications Based On End User Data Submission Statistics" in 2012.

[16] Daisuke Miyamoto, Yuzo Taenaka, Toshiyuki Miyachi, Hiroaki Hazeyama "PhishCage: Reproduction of Fraudulent Websites in the Emulated Internet" in march 2013