BAIT Alarm: Anti-Phishing Using Visual Similarities

Anuja Salve, Manisha Salgar, Akshata Sarode, Trushali Sardal, Prof. Archana Said

Abstract* - Phishing emails usually contain a message from a credible looking source requesting a user to click a link to a website where user is asked to enter a password or other personal information. Most phishing emails aim at withdrawing money from financial institutions or getting access to personal information. Phishing has increased hugely over the last years and is a serious threat to global security. Phishing has become the most popular practice among the criminals of the Web. URL, Domain and textual content analysis of email will results in a highly accurate anti phishing email classifier. We proposed a new antiphishing technique where we considered the advantages of whitelist, black list and heuristic technique for increasing accuracy and reducing false positive rate as well as true negative rate. We are using textual analysis, Domain analysis and URL analysis of e-mail in heuristic technique. Most of the phishing mails have similar contents, our proposed method Bait Alarm-antiphishing using visual similarities will increase the performance by analysing textual contents of email and lexical contains of URL analysis. It will detect legitimate mail if DNS is present in whitelist and if DNS is present in black list then it is considered as phishing mail. If it is not present in white list as well as blacklist then it is analyzed by using pattern matching with existing phishing DNS test and contents found in email and analysis of actual URL analysis. With the help of white list and black list we are avoiding detection time for phishing and authorized email. Simultaneously we are decreasing false positive rate by combining features of DNS, pattern matching ,textual content analysis of email and URL analysis.

Keyword: Hyperlink, iFrame, Network Security, Phishing, URLObfus-cation, Anti-phishing.

I. INTRODUCTION

Phishing is a type of attack where the attacker creates a replica of an existing Web page to fool users into submitting personal, unancial, or any other sensitive data like password data to what they think is their service provider’s Website. Phishing is the criminally fraud process of attempting to acquire sensitive information such as credit card details ,usernames, passwords by pretending as a legitimate trusted by customers in an electronic communication.

In simple word, the phishing means sending an e-mail to victim who contains some lured data that lead victim to spurious website and inquire about confidential data .The e-mail is socially engineered and the phisher try to convince the victim to divulge confidential information such as financial data, , bank account detail, credit card number and other credentials which can then be misuse by attacker. To detect and prevent the phishing attacks, different anti-phishing techniques used. Anti-phishing techniques are a protection scheme against the phishing attacks. It protects the user’s confidential data from the phishing attacks.

To robustly detecting phishing sites, we aim to use fundamental visual features of a web page’s appearance as the basis of detecting page similarities. In this paper, we propose a novel solution, Bait Alarm, to efficiently detect phishing web pages. Note that page layouts and contents are fundamental feature of web pages’ appearance. Since the standard way to specify page layouts is through the style sheet (CSS), we develop an algorithm to detect similarities in key elements related to CSS like the URL of trusted site, the domain and the title of the site.

II. RELATED WORK

Lots of research has been done on the internet security domain. To protects or alert the users against phishing attacks there are various techniques have been proposed that follows different strategies like client side and server side protection. To protect the users from URL obfuscation phishing attacks various tools are available which works on client side and the existing algorithms used to prevent this attacks which works on server side.

As the different methods used in URL Obfuscation phishing attacks any single tool are not capable to check all the types of methods. The tools like EarthLink, Netcraft anti-phishing and Spoof Guard toolbar produce very high false positive results. They all are relying only on blacklist and whitelist. This tools are not able to identify if the attacker have used shorten URL, IP address, encoding scheme or other methods . So, these different tools are not helpful now a day because there are number of methods used by phisher in URL obfuscation phishing attack.

The LinkGuard algorithm works on server side and it is the only one existing algorithm to check for the maximum methods used in URL obfuscation phishing attacks. LinkGuard algorithm is based on a) careful analysis of the characteristics of phishing hyperlink or URL and b) it produce very low false negative rate for the unknown phishing attacks. Basically, algorithm works for the hyperlink. When the hyperlink will be found in E-mail the algorithm performs the multiple tests such as

1) Visual and Actual DNS

2) Dotted decimal IP address.

3) Encoded scheme for URL.

4) Analyse domain name.

5) Blacklist and Whitelist.

6) Pattern matching of actual and visual DNS.

The proposed client side and server side security protection are not enough for the current phishing attacks. Because now days, the attacker uses the different methods such as input form in E-mail, shorten URL, iframe more than one hyperlink in E-mail, multilayer attacks. So, to protect and alert the users against these challenges we have implemented the ObURL detection algorithm which can use for the maximum detection of URL obfuscation attack.

III. MOTIVATION

Reading e-mails has become a dangerous activity. E-mails can carry dangerous viruses, worms which can be executed by merely opening e-mail or clicking on active link or picture in an e-mail. This e-mail phishing attacks are easily carried out by email spoofing. The two techniques known as e-mail forging and mass emailing made the task very easy for phishers to grab more number of victims. So we present a new solution, Bait Alarm, to detect different phishing attack using features that are hard to evade. The main intuition of our approach is that phishing pages need to preserve the visual appearance the target pages. So, We present an ObURL Detection algorithm to quantify the suspicious ratings of web pages based on similarity of visual appearance between the web pages. Since CSS is the standard technique to specify page layout, our new solution uses the CSS as the basis for detecting visual similarities among web pages.

IV. METHODOLOGY

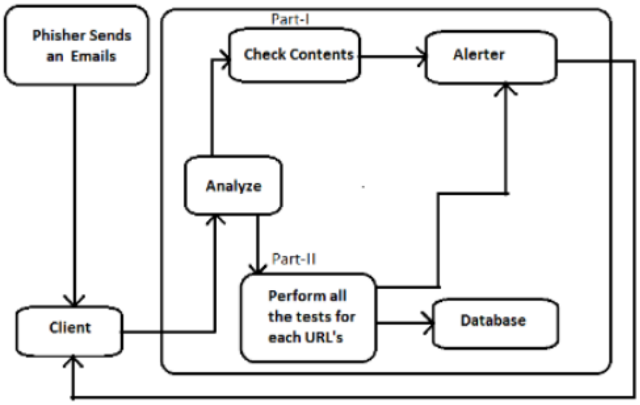

Fig 1: Workflow of Algorithm

The glory of Internet and its merits are being highly masked by the drawback associated with it. One of them the prime issue is Internet vulnerability, leading to data security issues. There are many phishing (fake) websites present on internet which collects private information of users.

We proposed a new algorithm, which we call Anti-phishing ObURL Detection Algorithm. The ObURL Detection Algorithm used to find out the URL Obfuscation Phishing attacks and it provides the multilayer security over the internet fraud. The Algorithm can detect the, content of hyperlink’s destination URL, hyperlink, input form, iFrame in email, input form in iFrame source URL, iFrame within iFrame, Black and White list Test, and after all that multiple tests will be perform such as DNS Test, URL Encode Test, IP address Test, Shorten URL Test, URL pattern matching Test On that collected email data. ObURL Detection Algorithm is effective to detect both known and unknown URL Obfuscation phishing attacks.

ObURL Detection Algorithm Contents
Input: Content of Email
Output: Prevent the user from phishing attack
Alert User: Possible Phishing
Safe User: No Phishing
DB: Database

Generally the Algorithm works in two parts:

1) Content check and

2) Perform all the tests for collected data.test cases are as following

1. DNS Test-

The DNS Test is first test. In this test, the ObURL Detection Algorithm will check the hyper text and anchor text. If both are different then it alerts the user with the message as both DNS are different, possible phishing, means the attacker is trying to hide something from the user.

2. IP Address Test-

The second test is IP Address Test. In the IP address test, attacker replace domain name by dotted decimal IP address, the ObURL detection algorithm will checks for the IP blacklist and IP whitelist sub tests.

3. URL Test-

The third test is URL Encode Test. In this test, actual URL encoding by attacker to hide from the user. The ObURL detection algorithm will find it, decode it and inform the user.

4. Shorten Test-

The fourth test is Shorten URL Test. The shortened URL is a combination of unique number or unique word and service provider site . In the shorten URL test, the ObURL detection algorithm detect if the URL is shorten by attacker.

5. Whitelist & Blacklist Test-

The fifth test is Whitelist and Blacklist Test. In this test, the algorithm will retrieve data from the database to check for the hyperlink domain in whitelist and blacklist database

6. Pattern Matching Test-

The sixth test is URL Pattern Matching Test. In this test, the algorithm will check collected data for the pattern matching. If all above tests are fails then algorithm will go to the pattern matching test.

VI. EXPERIMENTAL RESULT AND ANALYSIS

There are two other kinds of existing phishing detection approaches that are based on the similarity of web pages: based on page text and based on page image. Text-content-based methods find out the phishing pages according to the frequency of web pages , some sensitive words, keywoeds or the matching ratio between the suspicious page and the target page. Recent solutions check whether the contents of the page being visited is similar to other pages indexed by search engines. However, such solutions can be confused by attackers through embedding invisible contents. To capture the similarity in visual appearance, a few solutions are based on comparison of the image of a rendered page. However, this solution is not efficient. They can be affected by slight differences caused by different browser rendering engines. Moreover, if the target page cannot be indexed by search engines, such as a page that can be displayed only after a user login, the above solutions cannot be applied.

We will present a new proposed system, Bait Alarm, to detect phishing attack using features like DNS and CSS, our approach is that phishing pages need to protect from the visual appearance the target pages. so we have an algorithm to the suspicious ratings of web pages based on similarity of visual appearance between the web pages which we will be present. Since DNS and CSS are the standard technique to specify page layout as well as domain analysis, In our solution detecting visual similarities among web pages using CSS as basis for and DNS for domain information test. We present our approach as a Google Chrome extension, may in future scope we will use another search engine and used it to rate the suspiciousness of web pages & shows the correctness and accuracy of our approach with a relatively high performance overhead quantify.

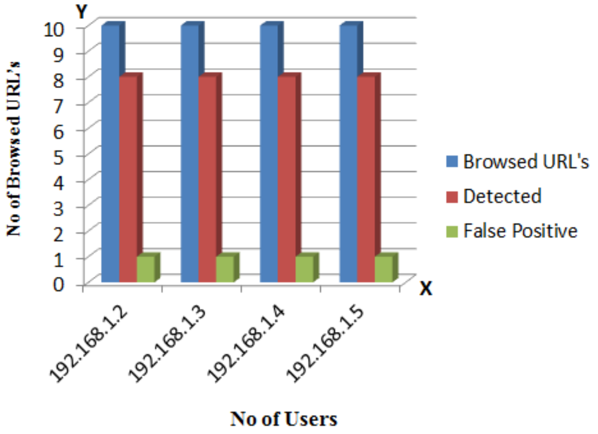

Fig 2: Detection Accuracy

The ObURL detection algorithm is a heuristic algorithm; it may cause false positive (i.e. non phishing site as phishing site) and false negative (phishing site as non phishing site) results. But this algorithm can detect both well know as well as unknown phishing attacks. False negative results are more dangerous than false positive results.

VII. CONCLUSION

We have studied different fishing attacks on email. We have described different phishing mail detection technique. So by analyzing DNS from the textual contents of mail, link and URL analysis we are trying to reduce false positive rate. At the same time we will take care of user by alerting user with possible phishing message due to possibility of phishing email.

A hybrid method has been proposed to detect phishing mail which is a combination of heuristic method, white list and blacklist. In heuristic detection technique we are considering textual analysis of email and lexical analysis of email for detection. This mechanism can effectively detect phishing mails as compared to the previous methods.

In this research, we have implemented and described ObURL Detection Algorithm against the URL obfuscation phishing attack. Our algorithm expects to analyze, detect and prevent the maximum obfuscated URLs.

The main goal of this research work is to provide the maximum security to the internet users against the phishing attacks. So, in future the researcher we will move towards the more secure algorithm for internet users .We are designing this software currently only for chrome extension further This software can be extended Firefox.

VIII. REFRENCES

[1] APWG, “Investigation report,” http://www.antiphishing.org/ reports/apwg trends report h2 2011.pdf, 2011.

[2] L. P., E. Jung, D. D., H. T.E., and H. J.P., “B-apt: Bayesian anti-phishing toolbar,” in Proceedings of IEEE InternationalConference on Communications, ICC’08. IEEE Press, May 2008.

[3] I. Fette, N. Sadeh, and A. Tomasic, “Learning to detect phishing emails,” in Proceedings of the International WorldWide Web Conference (WWW), May 2007

[4] Anti-phishing Working Group (APWG) Official site, http://www.apwg.org

[5] Phishing: The history of phishing attacks, URL: http://www.phishing.org/history-of-phishing.

[6] T. Ronda, S. Saroiu, and A. Wolman, “itrustpage: A userassisted anti-phishing tool,” in Proceedings of Eurosys’08. ACM, April 2008S

[7] C.Inc., “Could mark toolbar,” http://www.cloudmark.com/ desktop/ie-toolbar. [4] T. Ronda, S. Saroiu, and A. Wolman, “itrustpage: A userassisted anti-phishing tool,” in Proceedings of Eurosys’08. ACM, April 2008.