RESEARCH ARTICLE OPEN ACCESS

Anti-Phishing Technique to Detect URL Obfuscation

Jigar Rathod, Prof. Debalina Nandy
M. Tech (CE) Researcher Scholar, RK University, India.
Dept. Of Computer Engineering, RK University, India.

Abstract
Phishing is a criminal scheme to steal the user's personal data and other credential information. It is a fraud that acquires victim's confidential information such as password, bank account detail, credit card number, financial username and password etc. and later it can be misuse by attacker. In this research paper, we proposed a new anti-phishing algorithm, which we call ObURL Detection Algorithm. The ObURL Detection Algorithm used to detect the URL Obfuscation Phishing attacks and it provides the multilayer security over the internet fraud. The ObURL Detection Algorithm can detect the hyperlink, content of hyperlink's destination URL, iFrame in email, input form, input form in iFrame source URL, iFrame within iFrame, and after all that multiple tests will be perform such as DNS Test, IP address Test, URL Encode Test, Shorten URL Test, Black and Whitelist Test, URL pattern matching Test On that collected data. Our experiments verified that ObURL Detection Algorithm is effective to detect both known and unknown URL Obfuscation phishing attacks.

Keywords— Anti-phishing, Hyperlink, iFrame, Network Security, Phishing, URL Obfuscation.

I. INTRODUCTION

Internet security is a tree branch of network security specifically related to the internet. Its objective is to constitute rules and measures to protect the confidential data against attacks over the internet. Phishing is a criminal scheme to steal the user's personal data and other credential information. It is an illicit mechanism utilizing both social engineering and technical stratagem to steal victim's personal data, financial account data and other credential information [1].

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 [2]. The e-mail is socially engineered and the phisher try to convince the victim to divulge confidential information such as financial data, credit card number, bank account detail and other credentials which can then be misuse by attacker [1]. To detect and prevent the phishing attacks, anti-phishing techniques used. Anti-phishing is a protection scheme against the phishing attacks. It protects the user's confidential data from the phishing attacks. There is diverse anti-phishing techniques have been developed that follows different strategies like client side and server side protection against the phishing attacks [3].

As the use of internet service is continuously increasing, the phishing attacks are also increasing to steal user's confidential information. So, it is major challenge to protect the user's confidential data over the internet by detecting and preventing the phishing attacks. Now a day most of the phishing attacks are done using URL Obfuscation phishing attack. Because of the different methods used in this attack the existing algorithm cannot detect all the obfuscated URLs. The methods includes such as shorten URL, iframe in webpage, iframe in e-mail, IP address instead of domain name, encoded URL, input form in e-mail and other method which cannot be detect by the existing algorithm. So, in my research work I proposed to implement ObURL Detection Algorithm to detect obfuscated URL with the solution of current issues and will also check some other factors such as iFrame, source URL of iFrame, content of iFrame's source URL, input form and shorten URL maximum detection of URL obfuscation phishing attack.

In short, URL Obfuscation is a one type of phishing attacks in which the message recipient follows a hyperlink without realizing that they have been duped. Unfortunately, phisher have access to an increasingly large arsenal of methods for obfuscating the URL [4].

1.1 General steps of URL Obfuscation phishing attacks:

In general the phishing attacks perform with the following four steps as shown in figure 1 [5]:

1. First of all, the phisher have to create a counterfeit website to lure the victim which seems as legitimate one.
2. Then, the URL of that website is attached to spoofed e-mail and send to the lots of users. The content of e-mail will convince the victim to click on that URL.

3. If victim clicks on that obfuscated URL and visit that website, it convinces the victim to enter some confidential information.
4. Phisher then acquire some entered data and later it can be misuse by phisher.

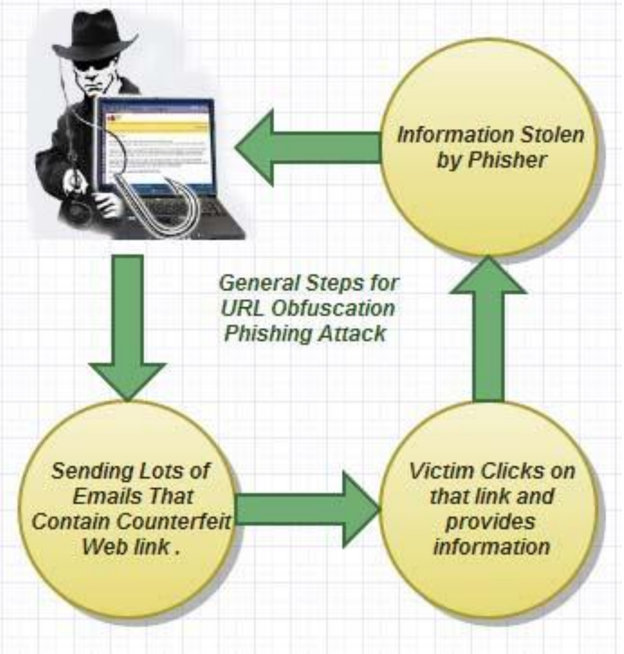

Information Stolen by Phisher
General Steps for URL Obfuscation Phishing Attack
Sending Lots of Emails That Contain Counterfeit Web link .
Victim Clicks on that link and provides information

Figure 1: General Steps for URL Obfuscation Attack

1.2 Methods used in URL Obfuscation Phishing Attack:

There are several methods for obfuscating the URL. Most common methods includes for URL obfuscation is:

A. Bad domain name or Misspelled domain name

One of the obfuscation method is uses the bad domain name or misspelled domain name to dupe the victim. Now a days it is easy to register the own domain name at minimal cost. So, this is the most common method used by attacker [6].

Consider the financial site Genuine Bank has registered domain name http://www.genuinebank.com and associated customer transactional site http://www.netbanking.genuinebank.com. The attacker can setup any of the following domain names to obfuscate the real destination host [7].

- http://www.genuinebanks.com
- http://www.netbank.genuinebank.com
- http://www.netbanking.genuinebanks.com

Phishing by Top level Domain (TLDs)

Anti-phishing working group (APWG) phishing activity trends report for 1st quarter of 2013 have analyse the phishing attacks using Top-level domain as shown in Figure 2 [6]:

B. Host Name Obfuscation or Using IP Address.

Host name obfuscation method uses the IP address instead of domain name. It obfuscates the host name by replacing it with the IP address.

As we have considered the domain name of financial site Genuine Bank is http://www.genuinebank.com, suppose the IP address for this domain name is 173.193.212.4. The phisher may use the IP address as a part of URL to obfuscate the host name or hide the destination from the end user [5].

For example: The following URL http://www.genuinebank.com , Could be obfuscated such as, http://173.193.212.4.

There are the other formats also available for the IP address such as dot less IP address in decimal, dotted IP address in octal, dotted IP address in hexadecimal, dot less IP address in hexadecimal [8].

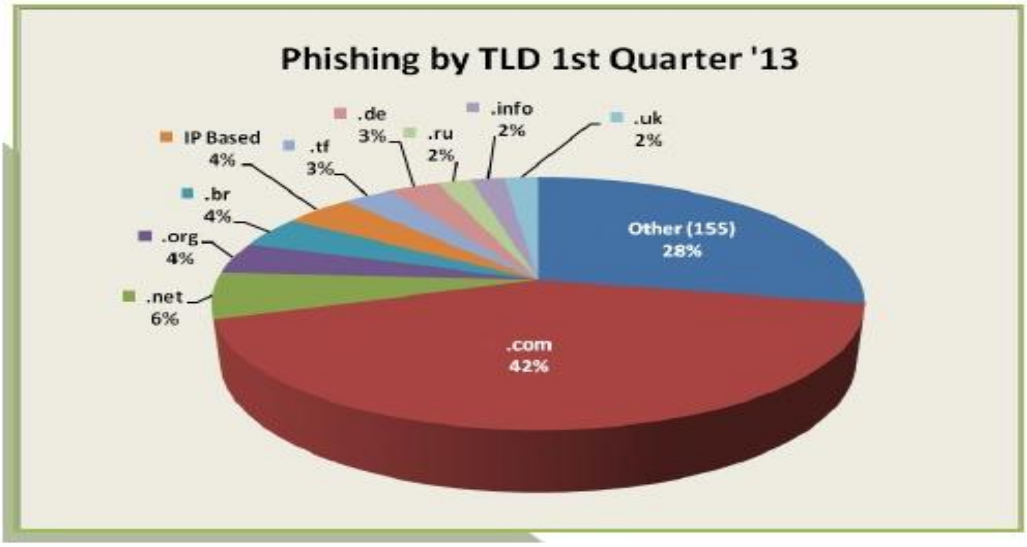


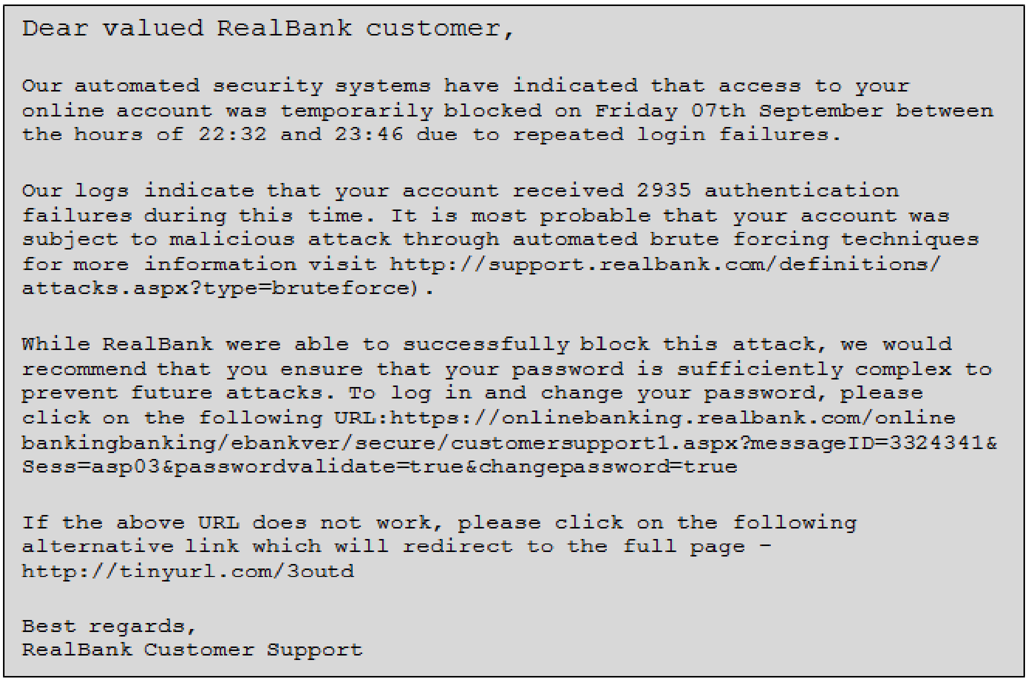
Figure 2: Phishing Attacks Trends using TLDs.

C. Shortened URLs

The length and complexity of web based application URLs is minimize using shortened URL method. The shortened URL is a combination of service provider site and unique number or unique word [9]. This navigates the user to the original destination. There are several service providers which can be used by phishing attacker to obfuscate the destination URL, such as http://tinyurl.com. http://bitly.com, http://goo.gl.

For example, the shorten URL for http://en.wikipedia.org/w/index.php?search=shortened+url+ &title=Special%3ASearch&go=Go URL is in goo.gl service provider produce http://goo.gl/VmwBNh, and in bitly.com it is http://bit.ly/1kfh1P0, and in tinyurl.com it is http://tinyurl.com/lbnj3un. Where, goo.gl, bit.ly and tinyurl.com is a service provider site and 'VmwBNh', '1kfh1P0' and 'lbnj3un' are the unique words. Because of the shorten URL are not relative to destination URL, the user can not judge the destination URL [10]. The attacker sends this shortened URL through the e-mail and the content convince the user to click on that link. If the user gets click on that URL, it is directed to the counterfeit webpage.

The message will write as an official message. For Example: as shown in below figure 3 [4]:



Dear valued RealBank customer,

Our automated security systems have indicated that access to your online account was temporarily blocked on Friday 07th September between the hours of 22:32 and 23:46 due to repeated login failures.

Our logs indicate that your account received 2935 authentication failures during this time. It is most probable that your account was subject to malicious attack through automated brute forcing techniques for more information visit http://support.realbank.com/definitions/ attacks.aspx?type=bruteforce).

While RealBank were able to successfully block this attack, we would recommend that you ensure that your password is sufficiently complex to prevent future attacks. To log in and change your password, please click on the following URL:https://onlinebanking.realbank.com/online bankingbanking/ebankver/secure/customersupport1.aspx?messageID=3324341& Sess=asp03&passwordvalidate=true&changepassword=true

If the above URL does not work, please click on the following alternative link which will redirect to the full page - http://tinyurl.com/3outd

Best regards,
RealBank Customer Support

Figure 3: Shortened URL E-mail Example

D. Encoded URL Obfuscation

In this method, the attacker obfuscates the URL using the encoded scheme. This encoding scheme tends to be supported by most of the web browser and can be interpreted in different way by web browser and their custom applications.

II. OBURL DETECTION ALGORITHM IMPLEMENTATION

We have implemented ObURL Detection Algorithm to detect the maximum URL Obfuscation Phishing Attacks. ObURL Detection Algorithm stands for Obfuscated URL Detection Algorithm.

The ObURL detection algorithm provides the multilayer security. Because of the use of internet service is continuously increasing, the phishing attacks are also increasing. So, the ObURL detection algorithm will secure the data against the phishing attacks over the internet.

As we know, the attacker uses the number of methods to obfuscate the URL. So, it is complex to detect all that attacks but the ObURL detection algorithm can detect the maximum number of URL obfuscation phishing attacks because of it performs the multiple tests such as DNS Test, IP address Test, URL encode Test, Shorten URL Test, Whitelist and Blacklist Test, Pattern matching Test and also check the iFrame, source URL of iFrame, content of iFrame's source URL, input form in email.

3.1 ObURL Detection Algorithm:

Input: Content of Email
Output: Prevent the user if URLs seems Counterfeit
Alert User: Possible Phishing
Safe User: No Phishing

```
DB: Database

If Input form found in E-mail Content then
    Alert User;
End
For each iframe in E-mail content do
    //get the content of iframe
    For each iframe in E-mail content’s iframe source do
        If input form found then
            Alert User;
    End
    For each hyperlink in E-mail content’s iframe source do
        // perform the test 1 to 6
End
For each hyperlink found in E-mail content and iframe source URL do
    Test 1: //DNS Test
        If hypertext! = Anchortext then
            Alert User;
    Test 2: // IP Address Test
        If IP address found in hyperlink then
            If IP address found in Whitelist DB then
                Safe User;
            Else
                Alert User;
                // IP Address found in blacklist DB
    Test 3: // Encoded Test
        If hyperlink found encoded then
            Decode hyperlink;
            Inform User;
    Test 4: // Shorten URL Test
        If URL is shorten then
            Alert User;
    Test 5://hyperlink whitelist and blacklist test
        If URL found in whitelist DB then
            Safe User;
        Else
            Alert User;
            // URL Found in Blacklist DB
    Test 6: // Pattern Matching Test
        If hypertext and anchor text pattern is matching then
            Alert User;
```

3.2 Description of ObURL Detection Algorithm

Generally the ObURL Detection Algorithm works in two parts: 1) content check and 2) perform all the tests for collected data. The ObURL Detection Algorithm works as follows. In its main routine ObURL detection algorithm, first check for the input form in the E-mail. In the recent phishing era the attacker sends an input form in the e-mail to users. So, this algorithm will first check for input form in E-mail content and if it is presented then algorithm will alert the user with the message: input form found, do not enter any confidential data.

After that, the iframe operation will start. In this process, the ObURL detection algorithm will check for the iframe in E-mail content. This algorithm can also work for multiple iframes. If there are multiple iframes in E-mail content then, ObURL detection algorithm will check for all and collect all the iframes. Then to provide the multilayer security, this algorithm will collect the content of iframes source and check for the further iframe. When all the iframes are collected the ObURL detection algorithm will start to check for the input form in each iframes. If the input found in any iframe, this algorithm will alert the user.

After the iframe operation is complete, the ObURL Detection Algorithm will work for the hyperlink. This algorithm will collect all the hyperlinks from E-mail content and from the iframes source URL content. When all the hyperlinks are collected, the ObURL detection algorithm will perform all the six tests for those hyperlinks. One by one all the hyperlinks will operate by the ObURL Detection Algorithm.

All the six tests will use against all possible methods to make the obfuscated URL. The ObURL detection algorithm includes the tests such as DNS Test, IP Test, URL Encode Test, Shorten URL Test, Whitelist Test & Blacklist Test, and Pattern Matching Test.

The performance of all the tests is described below:

Test 1: The first test is DNS Test. In the DNS test, the ObURL Detection Algorithm will first extract the hyper text and anchor text. If both are not same then it alerts the user with the message as both DNS are different, possible phishing. The DNS test believes on that if the attacker is not writing the same anchor text in hypertext. It means the attacker is trying to hide something from the user.

Test 2: The second test is IP Address Test. In the IP address test, if the attacker uses the dotted decimal IP address instead of domain name, the ObURL detection algorithm will checks for the IP blacklist and IP whitelist sub tests. For example: if the IP address found by the algorithm, it checks in the whitelist and blacklist IP address database. If the IP address found in blacklist IP database then the algorithm will alert the user with the message as blacklisted IP, do not enter. And if the IP found in whitelist IP database then the algorithm will consider user is safe. If the IP not found in whitelist and blacklist then this test is useful to alert the user as not found in whitelist, be careful.

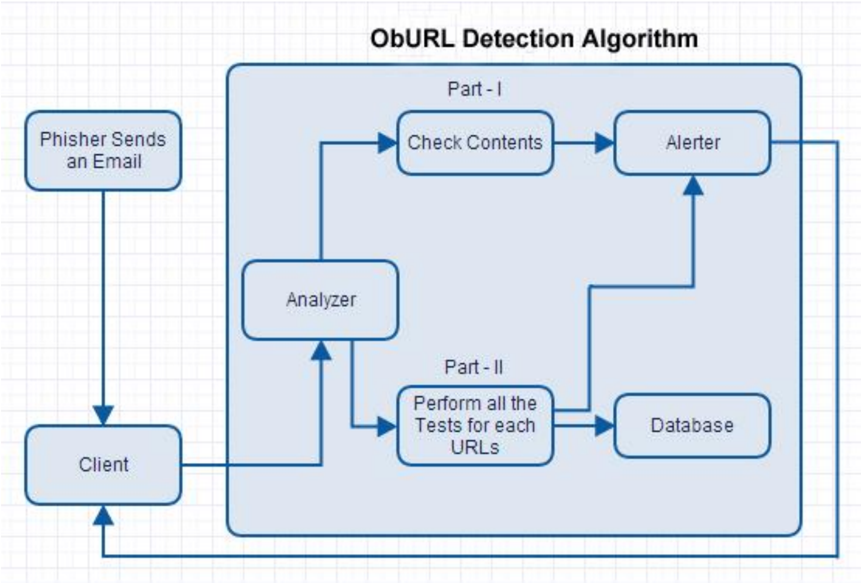


Figure 4: Block Diagram ObURL Detection Algorithm

Test 3: The third test is URL Encode Test. In the URL encode test, if the attacker encoding the actual URL to hide from the user. The ObURL detection algorithm will detect it, decode it and inform the user.

Test 4: The fourth test is Shorten URL Test. In the shorten URL test, the ObURL detection algorithm detect if the URL is shorten by attacker. Because of the shorten URL widely used in phishing attacks now a days, it is very important to alert the user with the message that URL is shorten.

Test 5: The fifth test is Whitelist and Blacklist Test. In this test, the algorithm will access the database to check for the hyperlink domain in whitelist and blacklist database. If the domain name in hyperlink found in whitelist then the algorithm will make the user safe and if it found in blacklist database then the algorithm will alert the user and prevent from that attack.

Test 6: The sixth test is URL Pattern Matching Test. In this test, the algorithm will check for the pattern matching. The pattern matching will be occurring between the hyper text and the anchor text. If both hyper and anchor text are identical but not the same. It means the attacker wants that the user will think it is the legitimate one.

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 know as well as unknown phishing attacks. False negative results are more harmful than false positive results.

III. Results and Discussion

We have implemented the ObURL Detection algorithm. And our experiments verified that ObURL detection algorithm is effective to detect and prevent both known and unknown phishing attacks with minimal false positive results.

We have done experiments of ObURL detection algorithm using the 303 URLs. In which 47 was shorten URLs (30 legitimate and 17 phishing URLs), 145 was blacklisted URLs, 57 was white listed URLs, 31 was Blacklisted IP address and 23 was white listed IP address.

The results of ObURL Detection Algorithm improve the security of user's confidential data against the URL obfuscation phishing attacks. The tests such as DNS test, URL encode test, whitelist and blacklist test and URL pattern matching test are produce the results same to the existing Link Guard algorithm. Because of there did no need to improvement into those tests. We have no need to compare the results of Link Guard Algorithm and the ObURL Detection Algorithm for DNS test, URL encode test, whitelist and blacklist test and URL pattern matching test because the both algorithm produces the same results.

Ob URL Detection Algorithm IP address test reduces the false positive results. Where, in Link Guard algorithm the IP address test was produce the false positive results. The IP address test in Link Guard algorithm performs only to check whether the IP address used instead of domain name or not?

Without knowing the IP address is blacklisted or whitelisted. So, for all the IP address either it is legitimate or counterfeit the LinkGuard will produce positive results. In experiments we have used 54 IP address in which 31 was blacklisted and 23 was white listed IP address. So, for all this 54 IP address LinkGuard produce positive results. But in ObURL Detection Algorithm the false positive results is reduced as shown in below figure:

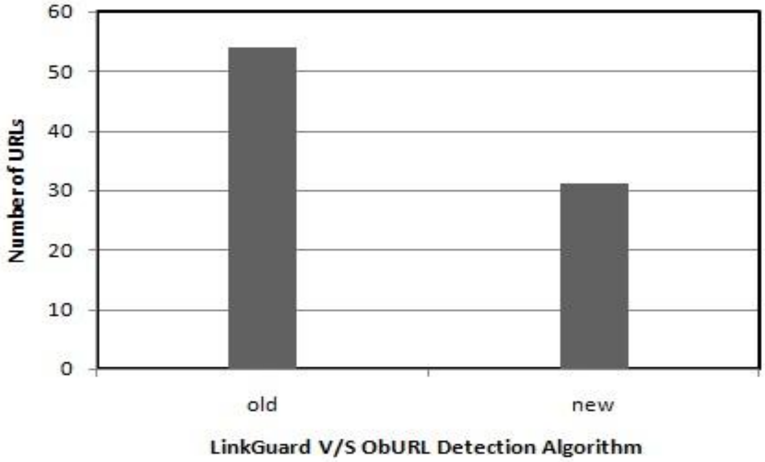


Figure 5: LinkGuard V/S ObURL Detection Algorithm IP Address Test

Our algorithm is also used in the detection and prevention of newly invented URL obfuscation phishing methods such as input form in E-mail content, input form in iframe source, iframe in E-mail content, iframe in E-mail contents iframe source. The attacker provides the input form in E-mail and the content of E-mail convince the user to submit the data. The attacker may also provide iframe in E-mail so the user can be diverted to counterfeit URLs.

To provide the security against the newly invented methods the ObURL Detection Algorithm is only works. Our experiments on iframe and input form in the various categories may produce the positive results. In our experiments we have used some URLs who contain input form or iframes and for this type of tests the ObURL detection algorithm provides multilayer security. If iframe is presented in email contents iframe source then our algorithm will find all the iframes and checks for the further iframe and input form in that iframe. So the experiments prove that the ObURL detection algorithm provides multilayer security. The LinkGuard algorithm is not helpful for the detection of these newly invented methods. The scope of LinkGuard is only for hyperlinks.

The ObURL Detection Algorithm performs the new test named Shorten URL Test. In this test the algorithm detect the URL if it is shorten. We have cover this test into the algorithm because of the Shorten URL is a new method used by attacker to obfuscate the URL. When the attacker uses the shorten URL method to obfuscate the URL and sends that shorten URL to the victim, the ObURL detection algorithm can detect that shorten URL and may prevent the phishing attack. Where, LinkGuard algorithm is not capable to test such a thing. The shorten URL Test may produce the false positive results (i.e. non-Phishing URL as Phishing URL). It is because of the ObURL detection algorithm works to check for, is the URL is shorten or not? So, for all the shorten URL the ObURL detection algorithm alerts the user. We have experiment on 47 shorten URL, 30 of them were legitimate and 17 were shortening of phishing URL. So, it produce the results as for all the shorten URL it alerts the user that the URL is shorted, be careful. So, after the complete experiments of ObURL detection algorithm we can say that the maximum false positive results are produce in the shorten URL test.

IV. RELATED WORK

Lots of research has been done on the internet security domain. To protects the users against phishing attacks there are various techniques have been proposed that follows different strategies like client side and server side protection [2]. 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 Toolbar, Netcraft anti-phishing toolbar, SpoofGuard toolbar produce very high false positive results. They all are relying only on blacklist and whitelist. They are not capable to identify if the attacker have used IP address, shorten URL, encoding scheme or other methods [11]. So, these 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 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, LinkGuard algorithm works for the hyperlink. So, when the hyperlink will be found in E-mail the algorithm performs the multiple tests such as 1) Visual and Actual DNS (comparing the visual and actual DNS). 2) Dotted decimal IP address (produced the false positive results, when the IP address is found instead of domain name it produce the results as phishing. So, when any legitimate site is using IP address instead of domain names it produces the false positive results). 3) Encoded scheme for URL (works if the actual or the visual URL is encoded). 4) Analyse domain name (when there is no destination information, LinkGuard call the analyse DNS to analyse the actual DNS). 5) Blacklist and Whitelist (it checks for the DNS in blacklist and whitelist database). 6) Pattern matching of actual and visual DNS (it designed to handle the unknown phishing attacks. It checks for the similarity between actual and visual DNS) [12].

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

V. Future Work

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 can move toward the more secure algorithm for internet users. The ObURL detection algorithm may produce the very minimal false positive results and false negative results. It may produce the false positive results in IP address test and false negative results in shorten URL test. When the URL is shorten, the ObURL detection algorithm only alerts the user, not converts to the original domain name. URL pattern matching test also produces the false positive results. So, the future work may include making an ObURL Detection Algorithm more solid with the solution of new upcoming methods in URL Obfuscation phishing attack.

VI. Conclusion

In this research, we have implemented and described ObURL Detection Algorithm against the URL obfuscation phishing attack. Our algorithm expects to detect and prevent the maximum obfuscated URLs. Our experiment on number of URLs proves that the ObURL Detection Algorithm is the one who provides the maximum security with compare to LinkGuard Algorithm and also provides the multilayer protection to prevent the URL obfuscation phishing attack. The ObURL Detection Algorithm can works for both know as well as unknown phishing attacks. Our algorithm provides the multilayer security with the minimal false positive results. The ObURL Detection Algorithm increases the level of legitimacy and security to protect the user's confidential data over the internet.

References

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

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

[3]. Gaurav, Madhuresh Mishra, Anurag Jain, (March-April 2012) "Anti-phishing techniques: A Review" International Journal of Engineering Research and Application (IJERA), ISSN: 2248-9622, Volume.2 Issue.2, Pages: 350-355.

[4]. The Phishing Guide, Understanding & Preventing phishing attacks. By: Gunter Ollmann, Director of Security Strategy IBM Internet Security System.

[5]. Jigar Rathod, Prof. Debalina Nandy "URL Obfuscation Phishing and Anti-Phishing: A Review" International Journal of Engineering Research and Application (IJERA), ISSN: 2248-9622, Volume.4, Issue.1 (Version 1), January 2014, Pages: 338-342.

[6]. Anti-Phishing Working Group (APWG) Phishing Activity Trends Report 1st Quarter 2013. Published on July 23, 2013 URL: http://docs.apwg.org/reports/apwg_trends_report_q1_2013.pdf.

[7]. The ocean is full of phish. By: Todd Fitzgerald. URL: http://www.infosectoday.com/Articles/Phishing.htm.

[8]. P.Malathi, Dr.P.Vivekanandan (November-2012) ―An efficient framework for internet banking International Journal of Engineering Science and Research Technology (IJESRT), ISSN: 2277-9655, Pages: 545-551.

[9]. Soojin yoon, Jeongeun Park, Changkuk Choi, Seungjoo Kim (July 24-26 2013). “SHRT – New method of URL shortening including relative word of target URL” Symposium on usable privacy and security (SOUPS) 2013, Newcastle, UK.

[10]. Dr.Marthie Grobler (August 2010) “Phishing for Fortune” Information Security for South Africa Conference, Sandton, South Africa. 2-4 August 2010, Pages: 8-15.

[11]. Lorrie Cranor, Serge Egelman, Jason Hong, and Yue Zhang (November 2006) “Phinding Phish: Evaluating Anti-phishing Tools”, CyLab Carnegie Mellon University.

[12]. U.Naresh, U.Vidya Sagar, C.V.Madhusudan Reddy (Sep-Oct 2013) “Intelligent Phishing Website Detection and Prevention System by Using LInkGuard Algorithm” International Journal for Scientific Research (IOSR), e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume.14 Issue.3, Pages: 28-36.