Imperial Journal of Interdisciplinary Research (IJIR)
Vol-3, Issue-2, 2017
ISSN: 2454-1362, http://www.onlinejournal.in

Mitigating Phishing Attack by Analysing Page Source

K. Siva Pavani[1], O. Prathyusha[2] & K. Ashlesha[3]
[1,2,3]Computer Science And Engineering, JNTUH

Abstract: *Attacker using the anonymity provided in the Internet, by which they are representing themselves as genuine users/companies and sending fake offers/messages to the users. Objective of this paper is to mitigate phishing attack by analyzing page contents. Most of the phishing attacks started from email from email messages, in this paper we discuss a new server-side émail add-on algorithm, by using the properties associated with the hyper-links, which present in emails. The methods we used here are assessment methods, which compare the similarities of the similarities of the links to the anchor text of that links. with this algorithm we find away to diminishing phishing attacks. we improved accuracy of finding phishing pages by inspection the URL links which are the main tools to lure the users to phishing site.*

1.Introcuction

The effectiveness and damage causing by the phishing attacks have been drastically increased. As the number of naive and non-it users using the internet more than compared to it aware users, making it easy for attacker to launch their attacks with ease. Limited protection and huge financial benefits, giving lead for attackers to perform low risk1, but high profit scams as cost incurred by criminals is pretty low and within a short duration attackers finish an attack and hide their identity. Attackers using the anonymity provided in the Internet, by which they are indicating themselves as legimate users/companies and sending fake offers/messages to the users. In phishing attack, attackers send a spoofed emails which are clearly crafted2, which intern looks like it came from áuthorized sources, which are used to lure users to open the contained URL's that lead them to a phishing website, intern tricking them to reveal sensitive information such as users' credentials, banking information, personal data etc.

Even though spam filtering techniques3 are present to protect from phishing emails, these are not effectively working as there are many number of tools that can bypass both rule based and spam filters. The success of attack stands in the äbility to craft the attack such that a non-it users are unable to identify the differences between the authorized and the spoofed messages. It is very difficult to identify whether a website is fake or not even by using good SSL mechanisms4. Industrial and academic research centres are focusing more on this phishing attack as it is now becoming top most attack on the internet. Even after proposing many protection mechanisms it is not 100% possible to mitigate these phishing attacks. Because, the percentage of non-IT users using these services óutweighs that of IT aware users In this paper, we analyze common process of phishing attacks and we will see a brief review of anti-phishing approaches. Our major focus is on email phishing. We firstly understand the common properties of hyper-links in e-mail messages. We found that the hyper-links posses one or more properties discussed as below 1) Actual link and visual link are different. 2) Phishers instead of direct DNS names, uses dotted decimal IP address notation. 3) Special processes are used to encode the links (actual or visual). 4) Phishers use fake DNS names which appear same as the under attack websites. We propose an email add-on algorithm which checks for phisher's identity, based upon the properties of the phishing links. As this email add-on using character-based algorithm, it is able to detect and prevent both known and unknown attacks. This émail add-on is light-weighted algorithm. The paper is divided in to the following sections. In Section 3, we explain about general procedure of a phishing attacks. And then we provide the available techniques to stop these attacks. Next analyze properties of the links and present email add-on algorithm in 4th section. Section 5 explains implementation of the email addon. In Section 6 conclude about this process and mentioned future scope of this paper.

2. Phishing Attack Procedure

1.In this paper, we presume that attackers use email messages as their main tool to implement these attacks. In general, attacker follow the below steps, first phishers establishes a fake website with similar designs5 used by the original site. In second step attacker send fake e-mails to users by representing himself as authorized personnel. When users receive this e-mail and click the fake hyper-links in e-mail. Attacker's fake site asks the user

required information. Attacker steals user information and performs tasks as legitimate user on original websites.

2.There are several ways to prevent phishing attacks, but none of those will completely stop these phishing attack, the best way to face them is to educate users to understand how these phishing attacks function.

The spoofed e-mails used by attackers can be said as one type of spam e-mails. Spam filters can be used to filter the phishing e-mails., the white-list, black-list and Bayesian filters having self learning skills, email stamps, keyword filters, etc., these can be used at client systems or email servers. Majority of the anti spam methods filter at receiver side by checking e-mail contents. Both black-list and white-list6 will not work, if the list's does not contain addresses in advance. Bayesian filters and Keyword filters can identify spam emails based on email-contents, these can detect any un-known spams. These can result in both false positives and false negatives. Moreover, these spam filters mainly designed only for general spams because these basically not taken properties of the phishing emails.

3. Defining Links in the Phishing Emails

A hyperlink references to data that the user can navigate by clicking on it. Hyperlink7 looks like <a href = "URL"> Visible text to users<\a>, 'URL' stands for universal resource locator, which gives information about the data user going to access, and 'Visible text' is the text visible to users.

URLs may have the following structure http://www. yahoo.com, ftp://60.90.1.2:2356, https://www.onlinebanking.com etc. 'Visible text' displays users a brief description about the URL and the contents he is going to visit in prior. Contents of the URL may not be same as visible text, attackers utilize this vulnerability to trick users. In this paper, we call URL as actual link and the visible text as visual link.

Attackers may use one of the following ways to define their hyper-links:

A) The actual link domain name does not match8 with the visual link. Consider an example, this hyper-link: http:// www.secure.onlinebaroda.in/login.php whichlooks like it is going to navigate to secure.onlinebaroda.in, which is the portal of a original site, instead it is pointing to a attacker website www.baroda.com.

B) Attacker uses dotted decimal IP address in the actual link and in visual link, instead original DNS name it displays another DNS name. www.onlinebaroda.com . The DNS of IP 89.85.74.85 is not onlinebaroda.

C) The actual link has been encoded9. This can be done in 2 ways: i) actual links transformed by encoding letters into their respective ANCI code; www.onlinebaroda.com . While the visual links are seems like pointing to www.onlinebaroda. com, but it is really pointing to http://4.34.195.41 ii) Special characters such as @,? are used. For example, the actual link looks like it is pointing to onlinebaroda, but really is pointing to IP address of 97.17.14.3 http:// www.onlinebaroda.com@97.17.14.3 iii) The following<ahref=http://www.onlinebaroda.com ?reDirect= http://97.17.14.3/"> Click on this link <a> It will redirect to attackers website 97.17.14.3.

D) Visual link does not display any link address, it simply displays some text. DNS name in actual link is identical to trusted company. In the following looks like it is redirecting VeriSign, but it actually not. Since VeriSign is actually owned by the attacker. Click here to pay.

An attacker can use any type of hyper-link style that he wants, that can belong to any one or many categories. An attacker normally uses more than one category in the same mail message to increase the probability10 of the success rate

4. Email Add-On Algorithm

Email Add-on algorithm analyzes the visual and actual links in the email message. The below pseudo code describes the email addon algorithm.

```
Vlink: visual link;

dvlink: decoded visual link;
alink: actual link;
dalink: decoded actual link;
vdns: visual DNS;
adns: actual DNS;
sdns: sender's DNS;
intEmailAddOn(Vlink, alink}
{
vdns = GetDNSName(vlink);
adns = GetDNSName(alink); returnPhishing_Mail;
If (adns is in dotted decimal format)
returnPhishing_Possible;
If(either alink or vlink in encoded form)
{
dvlink = decode(vlink);
dalink = decode(alink);
}
returnEmailAddOn(dvlink, dalink);
if(vdns is empty or not exsits)
returnDNSAnalysis(alink);
};
```

```
intDNSAnalysis (actual link)
{
if (adns in black-list)
returnPhishing_Mail;
if (adns in white-list)
returnSafe_Mail;
returnMatchPattern(alink);
};
intMatchPattern (alink)
{
if (sdns and adns are different)
returnPhishing_Possible;
for (each name pdns in seedset)
{
bvm = Similarity(pdns, alink);
if (bvm == true)
returnPhishing_Possible; }
returnSafe_Mail;
};
floatSimilarity_Index(strg, alink){
if (strg is part of alink)
return 1;
intmlen = max string length;
intmch = minimum transforms needed;
if (threshold < (mlen-mch) / mlen< l)
return 1
return 0;
};
```

The above pseudo code works as follows; it firstly we collect DNS information from the visual and actual link from email. We next check for the visual and actual DNS names. We say it is type 1 attack if both collected and actual dns names are not same. We say it is possible attack of type 3 if the dotted decimal notation11 has been used in any one fo the actual or visual links. We say it is type 2 attack if any one of the actual or visual links are represented in encoded format. If attacker uses encoded links, we 1st decode the encoded links, then we call recursively EmailAddOn procedure. If we are unable to find destination details in the visual link then we say it as type 4 attack, to analyze actual dns EmailAddOn calls DNSAnalysis procedure.

In DNSAnalysis, we call it phishing attack if the actual dns is present in blacklist. Similarly, we call it safe mail if the actual dns is present in whitelist. If actual dns is present in whitelist. If actual dns is neither present in blacklist nor in whitelist, we call Matchpattern procedure to find the unknown attacks, as this Matchpattern procedure is primarily designed to find and handle unknown attacks. The information12 we are having to deal the type 4 attack is the actual link from the hyperlink. To mitigate this attack, we have 2 methods; Firstly, we gather sender address details from the email message. Attacker mostly try to lure users by using authorized DNS's as sender address in email, we presume actual link dns and the dns present in sender email address are different. Secondly, in advance we gather user typed dns names while user surfing8 the internet in browsers and we assume these names are trustworthy because user manually typed them and we store them. Matchpattern checks firstly whether the original DNS of a hyper-link is not same as the DNS in the sender's email. We initiate the Similarity_index method if both the names are similar but not same with the stored list.

Similarity_index procedure analyzes similarity between actual dns and to the dns's in the stored data. The similarities of 2 dns's are calculated on the minimum changes (which may include adding, deleting, or replacing an element in the dns name) needed to convert one dns name to second dns name. We say 2 dns names are identical if the number of changes required is 0. We have they are sharing high similarity when the changes required are less, else low similarity13dns's. Consider example below, the similarity_index check of 'Google' and 'G00gle' is 4/6 because we have to replace the 2 'O's to make 'G00gle' as 'Google'. The similarity_index value of 'w3schools' , 'w3schools-ebox' are 10/15, because we have to remove last 5 chars from w3schools-ebox to make it w3schools and the similarity_index value of port value '5995', '59995' are 4/5, because we have to add '9' convert '5995' to '59995'. We say there is a possible phishing attack when the 2 DNS names are similar but not same

5. Implementation of EmailAddOn

We concentrated mainly on email phishing attacks because emails are the present common platform14 for the individuals and organizations to share information, most of the emails are auto generated, which makes the user to believe these are trustworthy15. The implementation this email addon is done by using query scripts, which are easy to design and cross-platform supported. There are mainly 2 components for this email addon, 1) Database for storing records 2) Server application to host the script file Whenever a mail is been received by the client, the email addon script checks the contents of the email template and alerts16 the user even before he opens the mail message. The script shows an alert whether the mail is safe or not. If the user himself found any mail, which a possible phishing mail he can send that mail for addition check for phishing activity.

We have placed a sample database consisting of both whitelist and black list links for the testing purpose. The program identified most of the phishing attacks with less false negative and false positive alerts, which intern increased the accuracy of the program.

6. Conclusion

Phishing is one of the most serious network security issues, causing finical lose to many companies and individuals. In this paper, we concentrated mainly on email phishing attacks because emails are the present common platform for the individuals and organizations to share information, most of the emails are auto generated, which makes the user to believe these are trustworthy. Attackers used this vulnerability to launch their attacks. In this paper, we did analyze the properties of the email hyper-links which are present in email body content. We designed server side email addon algorithm based on the gathered link properties. Since it is a property based algo-rithm, it is effective in detecting unknown attack as well.

In future we want to make it as browser addon, which can detect phishing attacks in all the WebPages which user visits. User need not to have any technical knowledge to use this addon as it doesn't need any technical configurations.

7. References

1.Anti-Phishing Working Group, Phishing activity trends report, 2008, Available from: http://www.antiphishing.org/ reports/apwg_report_Q4_2009.pdf, Oct.-Dec. 2009.

2.Ntoulas A, Najork M, Manasse M, Fetterly D. Detecting spam web pages through content analysis. Proceedings of the 15th International Conference on World Wide Web, Edinburgh, 2006.

3. Xiang G, Hong JI. A hybrid phish detection approach by identity discovery and keywords retrieval. Proceedings of the 18th International Conference on World Wide Web, Madrid, Spain, 2009.

4.Zhang Y, Hong J, Cranor L. Cantina: A content based approach to detecting phishing web sites. Proceedings of the 16th International conference on World Wide Web, New York, NY, USA, 2007.

5. Fette I, Sadeh N, Tomasic A. Learning to detect phish-ing emails. Proceedings of the 16th International Conference on World Wide Web, Banff, Alberta, Canada, 2007.

6. Almomani A, Gupta BB, Wan T-C, Altaher A, Manickam S. Phishing Dynamic Evolving Neural Fuzzy Framework for Online Detection "Zero-day" Phishing Email. Indian Journal of Science and Technology. 2013 Jan; 6(1). Doi: 10.17485/ijst/2013/v6i1/30571.

7. SadeghJavadi M, Meskarbashee A. New Approach to Congestion Mitigation Based on Incidence Matrix DCOPF. Indian Journal of Science and Technology. 2012 Feb; 5(2). Doi: 10.17485/ijst/2012/v5i2/30353.

8. John Livingston J, Umamakeswari A. Internet of Things Application using IP-enabled Sensor Node and Web Server. Indian Journal of Science and Technology. 2015 May; 8(S9). Doi: 10.17485/ijst/2015/v8iS9/65577.

9. Chandresekaran M, Chinchani R. Phoney: Mimicking user response to detect phishing attacks. 2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks, Buffalo-Niagara Falls, NY, (WoWMoM'06). 2006; 5.

10. Silnov DS. An Analysis of Modern Approaches to the Delivery of Unwanted Emails (spam). Indian Journal of
Science and Technology. 2016 Jan; 9(4). Doi: 10.17485/ ijst/2016/v9i4/84803.

11. Ma J, Saul LK, Savage S, Voelker GM. Beyond blacklists: learning to detect malicious web sites from suspicious URLs. In proceedings of the International Conference on Knowledge Discovery and Data Mining, Paris, France, 2009.

12. Ma J, Saul LK, Savage S, Voelker GM. Identifying suspicious urls: An application of large-scale online learning. Proceedings of the International Conference on Machine Learning, Montreal, Canada, 2009.

13. Pan Y, Ding XH. Anomaly Based Web Phishing Page Detection. Proceedings of the 22nd Annual Computer Security Applications Conference on Annual Computer Security Applications Conference, 2006. p. 381–92.

14. Kirda E, Kruegel C. Protecting Users against Phishing Attacks with AntiPhish. Proceedings of the 29th Annual International Computer Software and Applications Conference. 2005. p. 521–34.

15. Wu M, Miller RC, Garfinkel SL. Do Security Toolbars Actually Prevent Phishing Attacks? Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2006. p. 601–10.

16. Chandrashekaran M, Narayana K, Upadhyaya S. Phishing Email Detection Based on Structural Properties. Symposium on Information Assurance: Intrusion Detection and Prevention, New York, 2006.