ONLINE DETECTION AND PREVENTION OF PHISHING

NAME : MUTHUKUMAR SELVARASU

DATE : 21th May, 2010.

Contents

Page No.

ABSTRACT

Phishing is a new kind of network attack in which the attacker creates a copy of an existing Web page to fool users (e.g. by using special e-mails or instant messages) to submit personal details, password and important details. In this paper, the researcher propose a new end-host based anti-phishing algorithm, called LinkGuard is through the use of the generic characteristics of the hyperlinks in phishing attacks. These properties are by the analysis of data phishing archive of the Anti-Phishing Working Group (APWG) which derived. Because it is based on the generic characteristics of phishing attacks that can LinkGuard detects known and unknown phishing attacks. The Researcher personally tested this solution into his Windows XP machine. It is basically proof of concept that recognize effective link Guard and to prevent both known and unknown phishing attacks with minimal false negatives. Link Guard recognizes successfully 185 of 200 phishing attacks. In his experiments also showed that LinkGuard is lightweight and can detect and prevent phishing attacks in real time.

I. INTRODUCTION

The word "phishing" first emerged in the 1990s. The early hackers often use 'ph replace "f" to produce new words in the hacker community, as they chop the data by mobile phones in general. Phishing is a new word pronounced as" fishing".

The commonly used method is sending e-mails to potential victims, who appeared to be sent by banks, online organizations or ISPs. In this e-mail, they will make some causes, such as the password of user's credit card has been mis-entered many times, or they provide services modernization, to seduce, visit their website to match, or change your account number and password by using the hyperlink provided in the e-mail. User will then be linked to a fake Web site by clicking these links. The style, the functions carried out, sometimes even the URL of the fake sites are similar to the real site. It is very difficult for user to know that user actually visit a malicious Web site. If the account number and password input, then the attacker successfully gather information on the server side and is able to carry out their next step activities with that information stand out (such as transfer money from user bank account).

In this paper the Researcher examine the common methods of phishing attacks, and reviewed the possible anti-phishing approaches. We then focus on end-host-based anti-phishing approach. First analyze the common characteristics of the links in phishing e-mails. The analysis shows that the phishing links one or more characteristics as listed below: 1) the visual link and the actual link are not the same; 2) the attackers often use dotted decimal IP address instead of the DNS name, 3) special tricks used to encode the hyperlinks in bad faith, 4) the attackers often use fake DNS names that are similar (but not identical) to the target site. The Researcher proposes an end-host based anti-phishing algorithm that called LinkGuard, based on the characteristics of phishing hyperlink. Since LinkGuard link is character-based, it can detect and prevent not only known phishing attacks but also unknown.

The rest of this paper is organized as follows. In Section II, The Researcher gave the general procedure of a phishing attack and offers the available methods to prevent phishing attacks. Then analyze the properties of the hyperlinks in phishing attacks and present the algorithm in Section III. Section IV describes implementation of the system and gives LinkGuard the experimental results. Section V concludes and lesson learned this paper.

II. PHISHING ATTACK PROCEDURE AND PREVENTION METHODS

In this paper, we assume that phishers use to send e-mail as an important method to conduct user for phishing attacks. Nevertheless, the analysis and algorithm uses for the attacks such as instant messaging.

A. THE METHOD OF PHISHING ATTACKS

In general, phishing attacks with the follow out four steps:

1) Phishers set up a fake website that looks exactly like the legitimate site, such as setting up the Web server, the application of the DNS server name, and the creation of web pages similar to the desired site, etc.

2) Send large quantity of fake e-mails to users in the names of legitimate businesses and organization and to convince the potential victim, visit their web sites.

3) The recipient received the e-mail, open it, click on the link in the fake e-mail, and enter the required information.

4) Phishers steal personal information and perform their fraud as a transfer from the victims account when the data entered into their site.

B. APPROACHES TO PREVENT PHISHING ATTACKS

There are several (technical or non-technical) ways to prevent phishing attacks:

1) By educate the users to understand phishing attacks, when phishing e-mails are received

2) To punish the Phishing attackers by legal methods

3) Use technical methods to stop phishing attackers.

In this paper the Researcher only focus on the third step.

Technically, if we cut one or more of the steps required by a phishing attack, then we have successfully prevent an attack that. Below we briefly review these approaches.

1) Block the phishing Web sites in time, if we can detect the phishing Web sites in time, can block the Web sites and phishing attacks to be prevented. It is relatively easy (manual), if a site is a phishing site or not, but it is difficult to find these phishing sites in time. Here we list two methods for detecting phishing site. i) The web master periodically scans root DNS for suspicious Web sites. ii) Since the phisher must duplicate the contents of the destination site, he has to use tools (automatic) Download the Web pages from the target site. It is therefore possible to detect this kind of download on the Web server and trace the phisher.

2) Increasing the security of Web sites: The Business Web sites such as the Web sites of banks, new methods to ensure the security of the personal data of users. One method is to increase the security, to use hardware devices. For example, Barclays Bank offers hand-card to their customers. Before shopping the web, users must insert their card into the card reader and their Input (Personal Identification Number) PIN code, then the card will produce a unique password security, user transactions can only be carried out after the correct password is received (John Leyden, 2005). Another method is to use the biometric (eg. voice, fingerprint, iris, etc.) for user authentication. For example, replace the Paypal single password verification by speech recognition to enhance the security of the site. With these methods cannot fulfill their tasks, phishers and after being part of the victims information received. However, all these techniques additional hardware to implement the authentication between the user and the Web sites, so the cost increase and bring some inconveniences.

3) Block the phishing e-mails from various spam filters: Phishers typically use e-mails as bait to lure "potential victims. SMTP (Simple Mail Transfer Protocol) (Anti-Phishing Group, 2010) is the protocol to send e-mails on the Internet. It's a very simple protocol, simply authentication mechanisms is required to setup. Information related to the sender, such as name and e-mail address of the sender of the message route, etc., can be falsified in SMTP.

4) Install online anti-phishing software on user's computer: Despite all these efforts it is still possible for users to visit fake Web sites. As a last defense, the user can install anti-phishing

tools into their computers. The anti-phishing tools in use today can be divided into two categories: Blacklist / Whitelist based and rule-based.

III. LINKGUARD

CLASSIFICATION OF THE HYPERLINKS IN THE PHISHING E-MAILS

In order to gain (illegal) useful information from potential victims, phishers usually try to convince the user to click the hyperlink embedded in the phishing e-mail. A hyperlink has a structure as follows. <a href="URI"> Anchor text <\ a>

1) The link DNS domain name has in the anchor text, but the target DNS names in the visible link does not match the actual in the link. For example, the following hyperlink: <a href ="http://www.profusenet.net/checksession.php"> eSecure banking login </ a> seems to secure.regionset.com that the portal is linked to a bank, but it is actually a phishing site linked www.profusenet.net.

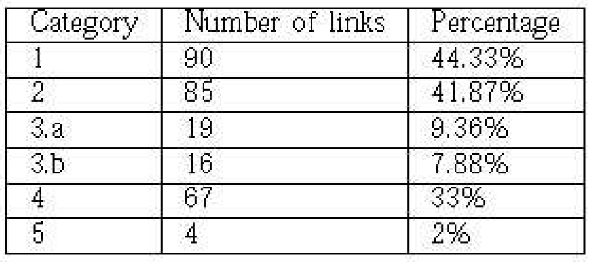
2) Dotted decimal IP address is used directly in the URI or the anchor text instead of DNS names. See below for an example.
<a href= "http://61.129.33.105/secured site/www.skyfi.com/
index.html?MfcISAPICommand=SignInFPP& UsingSSL=1"> Login </ a>

3) The link is forged in bad faith by the use of specific encoding schemes. There are two cases:
a) The compound is formed by encoding the alphabet into the corresponding ASCII codes. See below for such a link. <A href = "http://% 34% 2E% 33% 34% 2E% 31% 39% 35% 2E% 34% 31:% 34% 39% 30% 33 /% 6C /% 69% 6E% 64 % 65% 78% 2E% 68% 74% 6D "> www.citibank.com </ a>, as you pointed out this link www.citibank.com seemed, it really points http://4.34.195.41:34/ 1 / index.htm.

b) Special characters (eg @ link in the visible) are used to mislead users to believe that the e-mail from a trusted sender. For example, the following link to Amazon seems, is connected, but it is actually the IP address 69.10.142.34 connected.
http://www.amazon.com:fvthsgbljhfcs83infoupdate @ 69.10.142.34.

4) The link does not contain information about travel destinations in its anchor text and uses DNS names in its URI. The DNS name in the URI is usually similar to a famous company or organization. For example, the link seems to be sent by PayPal, but it is not. Since PayPal-cgi is actually registered by the phishers, so that the users believe that there something wrong with paypal click <a href = "http://www.paypal-cgi.us/webscr.php hat? cmd = login "> here to confirm your account </ a>

5) The attacker exploit the vulnerability of the target launch site to redirect users to phishing sites or CSS (Cross Site Scripting) attacks. For example, the following link <a href="http://usa.visa.com/track/dyredir.jsp?rDirl= http://200.251.251.10/.verified/"> Click here <a> Once clicked, the user to the phishing site redirect



Category	Number of links	Percentage
1	90	44.33%
2	85	41.87%
3.a	19	9.36%
3.b	16	7.88%
4	67	33%
5	4	2%

Table 1.

Table 1 summarizes the number of hyperlinks and their percentages for all categories. It is observed that most phishing e-mails using DNS name (Category 1, 44.33%) or dotted decimal IP addresses (Category 2, 41.87% are fake). Encoding tricks are often used (Category 3 and 3, 17.24%). And phishing attackers often try to users through the creation of DNS names that start very similar to the real e-sites or by providing information on the destination is not in the anchor text (category 4 are fools). Phishing attacks use the fact that the vulnerability of Web sites (category 5) are small number (2%) and we let this kind of study for future attacks. (Anti-Phishing Group, 2006)

After the characteristics of phishing hyperlinks understood, then able to design anti-phishing algorithms that detect known and unknown phishing attacks and in real time. The Researcher have presented Link Guard algorithm in the next section.

IV. LINKGUARD ALGORITHM

The LinkGuard works by analyzing the classification between the visual link and the actual link. It also counts the similarities of a URI with a known trusted site. The algorithm is illustrated in here

A. BASIC ALGORITHM

```
v_link: visual link;
a_link: actual_link;
v_dns: visual DNS name;
a_dns: actual DNS name;
sender_dns: sender's DNS name.
int LinkGuard(v_link, a_link} {
    v_dns = GetDNSName(v_link);
    a_dns = GetDNSName(a_link);
    if ((v_dns and a_dns are not
    empty) and (v_dns != a_dns))
        return PHISHING;
    if (a_dns is dotted decimal)
        return POSSIBLE_PHISHING;
    if(a_link or v_link is encoded)
    {
        v_link2 = decode (v_link);
        a_link2 = decode (a_link);
        return LinkGuard(v_link2, a_link2);
    }
    /* analyze the domain name for
    possible phishing */
    if(v_dns is NULL)
        return AnalyzeDNS(a_link);
}
```

```
int AnalyzeDNS (actual_link) {
/* Analyze the actual DNS name according
to the blacklist and whitelist*/
if (actual_dns in blacklist)
return PHISHING;
if (actual_dns in whitelist)
return NOTPHISHING;
return PatternMatching(actual_link);
}
int PatternMatching(actual_link){
if (sender_dns and actual_dns are different)
return POSSIBLE_PHISHING;
for (each item prev_dns in seed_set)
{
bv = Similarity(prev_dns, actual_link);
if (bv == true)
return POSSIBLE_PHISHING;
}
return NO_PHISHING;
}
float Similarity (str, actual_link) {
if (str is part of actual_link)
return true;
int maxlen = the maximum string
lengths of str and actual_dns;
int minchange = the minimum number of
changes needed to transform str
to actual_dns (or vice verse);
if (thresh<(maxlen-minchange)/maxlen<1)
return true
41 return false;
}
```

As per Table.1 (Categories 3 and 4) we have to decipher the links first, and then recursively call LinkGuard (lines 8-13 back). If there is no destination information (DNS name or dotted IP address) in the visual link (Category 5), LinkGuard calls on the actual DNS (lines 16 and 17 to analyze).

Link Guard assumes all five categories of phishing attacks. DNS name is included on the blacklist and then we are sure that there is a phishing attack (lines 18 and 19). Similarly, if the actual DNS is contained in the Whitelist, so it's not a phishing attack (lines 20 and 21). If the actual DNS is not in the blacklist or whitelist contain then pattern-matching called (line 22).

Pattern Matching is designed to unknown attacks (blacklist / whitelist is useless to deal with in this case). For category 5 of phishing attacks, which is all the information that we have the actual link from the link (as the visual link contains no DNS or IP address of the destination Web site) that provide very little information for further analysis . To solve this problem, we try two methods: first, we extract the email address from the e-mail. Since phishers generally for the users, by trying to fool (fake) legal DNS name in the sender's email address and we expect that the DNS name in the sender's address different from the actual link. Secondly, we set to collect proactively DNS names manually input by the user surfs the Internet and if they save the name in a seed, and as these names are input by the user by hand, we assume that these names are trustworthy. Pattern matching then checks if the actual DNS name of a hyperlink is different from the DNS name in the address of the sender (lines 23 and 24), and if it quite similar (not identical) with one or more names in the seed the similarity is set by the call (lines 25-30) processes.

B. IMPLEMENTATION AND VERIFICATION OF LINKGUARD

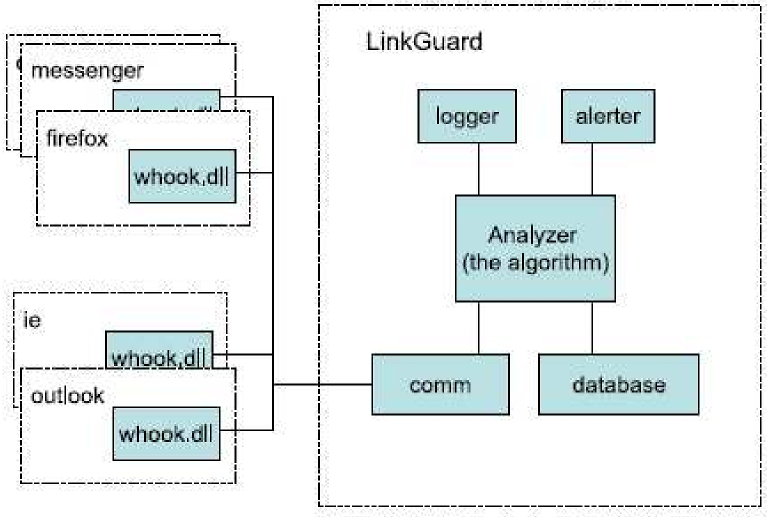


Figure 1

The Researcher implemented the algorithm in Windows XP machine. There are two parts: a dynamic library (whook.dll) LinkGuard executive. The structure of the implementation is shown in Fig. 1. Whook is a DLL used by LinkGuard, it is dynamically loaded into the address space of the execution of processes by the operating system. Whook is responsible for the collection of data, such as the compou nds mentioned and visual relationships, the user input URLs. More specifically whook.dll is used:

1) Install a BHO (Browser Helper Object) for IE, Firefox to monitor user input URLs,

2) Install an event with the Whook SetWinEventHook provided by the Windows operating system to collect relevant information

3) Retrieve data from analyze sender which is from E-mail address from Outlook

4) Analyze and filter the received content

CONCLUSION

Phishing has become a serious problem of network security, which makes finical lose billions of dollars to both consumers and e-commerce companies. And perhaps more fundamentally, Phishing, E-Commerce distrusted and less attractive to ordinary consumers. In this paper, The Researcher has studied the characteristics of the links that were embedded in phishing e-mails investigated. Then determined an anti-phishing algorithm on the basis of the derived features as explained. Since Phishig Guard-based characteristic, it can detect not only known attacks, but also effective, to the unknown. The Researcher have learned how to implement in the algorithm in Windows XP machine using Whook.dll, this experiment showed that light-weight LinkGuard and recognizes up to 96% unknown phishing attacks in real time.

The Researcher believe that Link Guard not only useful for detecting phishing attacks, but can also detect the malicious or unwanted links in Web pages and instant messages to protect.

The Researcher wishes to thank the University to make the report of the investigation and the opportunity to learn this work, also thankful to Dr. Toh, for motivating this work and contributing helpful comments.

REFERENCES

Anti-Phishing Group, (2010), The Anti-phishing working group. Retrieved on May 15 from http://www.antiphishing.org/

David Geer, (2005), Security Technologies Go Phishing. Retrieved on May 16, 2010 from http://www.geercom.com/r6018.pdf

EarthLink , (2008), EarthLink. ScamBlocker. Retrieved on May 16, 2010 from http://www.earthlink.net/software/free/toolbar/

Georgina Stanley. (2006), Internet Security -Gone phishing. Retrieved on May 18, 2010 from http://www.cyota.com/news.asp?id=114

I. Androutsopoulos, J. Koutsias, K.V. Chandrinos, (2000), An Experimental Comparison of Naive Bayesian and Keyword-Based Anti-Spam Filtering with Encrypted Personal E-mail Message. Retrieved on May 15, 2010 from http://research.microsoft.com/en-us/um/people/joshuago/spambibliography.mht

John Leyden, (2005), Trusted search software labels fraud site as 'safe'. Retrieved on May 17, 2010 from http://www.theregister.co.uk/2005/09/27/untrusted search/

Meng Weng Wong, (2008), Phishing Prevention methods. Retrieved on May 18, 2010 from http://www.openspf.org/whitepaper.pdf.

Neil Chou, Robert Ledesma, Yuka Teraguchi, Dan Boneh, and John C. Mitchell, (2004), Client-side defense against web-based identity theft. Retrieved on May 15 from http://www.cse.msu.edu/~alexliu/courses/825Spring2008/lectures/presentation.ppt

PhishGuard.com, (2009), Protect Against Internet Phishing Scams . Retrieved on May 16, 2010 from http://www.phishguard.com/