Detection of Phishing e-mail

Jasveer Singh
Dept. Of CS/IT, Graphic Era University, Dehradun, Uttrakhand, India

Abstract
Phishing has increased enormously over the last years and is a serious threat to global security and economy. Criminals are trying to convince unsuspecting online users to reveal sensitive information, e.g., passwords, account numbers, social security numbers or other personal information. Majority of the present day phishing attacks employ e-mail as their primary carrier, in order to allure unsuspecting victims to visitthe masqueraded website. In this paper describe the procedure of phishing attacks and a number of features that are particularly well-suited to identify phishing emails .Finally describe the detection of the phishing email .

Key words
Phishing, Phishing email, attacks, detection.

I. Introduction

As people increasingly rely on the Internet for business, personal finance and investment, Internet fraud becomes a greater and greater threat. Internet fraud takes many forms, from phony items offered for sale on eBay, to scurrilous rumors that manipulate stock prices, to scams that promise great riches if the victim will help a foreign financial transaction through his own bank account. One interesting species of Internet fraud is phishing. Phishing attacks use email messages and web sites designed to look as if they come from a known and legitimate organization, in order to deceive users into disclosing personal, financial, or computer account information. Phishing emails usually contain a message from a credible looking source requesting a user to click a link to a website where she/he is asked to enter a password or other confidential information. Most phishing emails aim at withdrawing money from financial institutions or getting access to private information. Phishing has been growing really fast, as shown in Fig. 1. According to the Anti-Phishing Working Group, in the past two years, the number of unique reported phishing attacks per month has increased more than 160 times and the number of unique reported phishing sites per month has increased about 16 times. In June 2006, 130 legitimate brands have been attacked. [1]

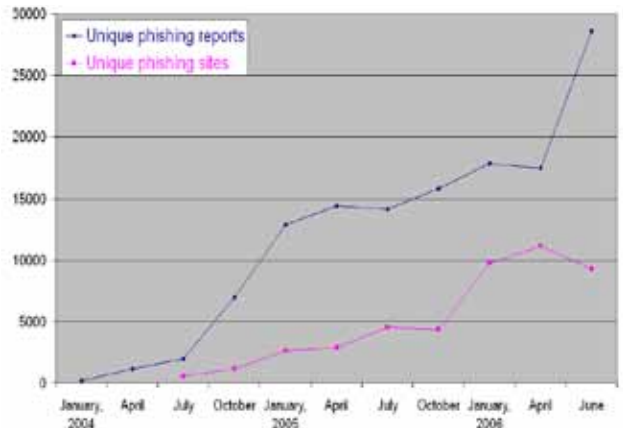

Fig. 1: Records of phishing in 2004-2006

II. The Procedure of Phishing Attacks

In general, phishing attacks are performed with the follow- ing four steps see Fig.2:

1. Phishers set up a counterfeited Web site which looks exactly like the legitimate Web site, including setting up the web server, applying the DNS server name, and creating the web pages similar to the destination Website, etc.
2. Send large amount of spoofed e-mails to target users in the name of those legitimate companies and organizations, trying to convince the potential victims to visit their Web sites.
3. Receivers receive the e-mail, open it, click the spoofed hyperlink in the e-mail, and input the required information.
4. Phishers steal the personal information and perform their fraud such as transferring money from the victims'.

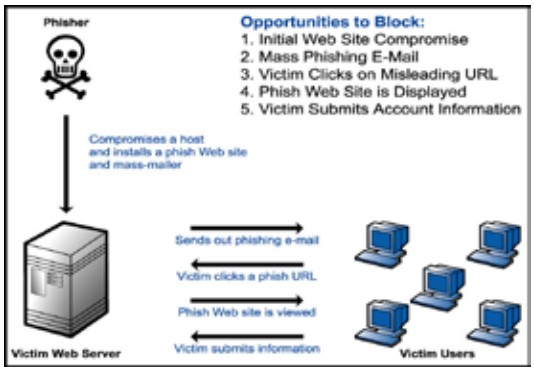

Fig. 2 : Phishing attack

A common phishing attack is (for a phisher) to obtain a victim's authentication information corresponding to one website (that is corrupted by the attacker) and then use this at another site. This is a meaningful attack given that many computer users reuse passwords – whether in verbatim or with only slight modifications.

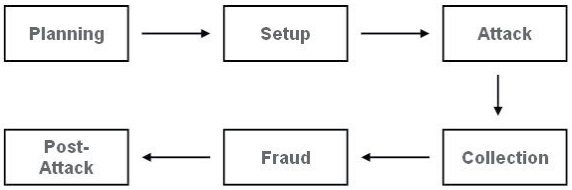

Fig. 3: Phishing attack lifecycle

The phishing attack lifecycle can be decomposed in (see Fig.3) : Planning, Setup, Attack, Collection, Fraud and Post-Attack Actions.

III. Types of Phishing Attacks

- **Deceptive Phishing.** The term "phishing" originally referred to account theft using instant messaging but the most common broadcast method today is a deceptive email message. Messages about the need to verify account information, system failure requiring users to re-enter their information, fictitious account charges, undesirable account changes, new free services requiring quick action, and many other scams are broadcast to a wide group of recipients with the hope that the unwary will respond by clicking a link to or signing onto a bogus site where their confidential information can

be collected [3].

- **Malware-Based Phishing** refers to scams that involve running malicious software on users' PCs. Malware can be introduced as an email attachment, as a downloadable file from a web site [3].
- **Data Theft.** Unsecured PCs often contain subsets of sensitive information stored elsewhere on secured servers. Certainly PCs are used to access such servers and can be more easily compromised.
- **Keyloggers and Screenloggers** are particular varieties of malware that track keyboard input and send relevant information to the hacker via thebrowsers as small utility programs known as helper objects that run automatically when the browser is started as well as into system files as device drivers or screen monitors
- **Session Hijacking** describes an attack where users' activities are monitored until they sign in to a target account or transaction and establish their bona fide credentials. At that point the malicious software takes over and can undertake unauthorized actions, such as transferring funds, without the user's knowledge.
- **Web Trojans** pop up invisibly when users are attempting to log in. They collect the user's credentials locally and transmit them to the phisher [3].
- **Hosts File Poisoning.** When a user types a URL to visit a website it must first be translated into an IP address before it's transmitted over the Internet. The majority of SMB users' PCs running a Microsoft Windows operating system first look up these "host names" in their "hosts" file before undertaking a Domain Name System (DNS) lookup. By "poisoning" the hosts file, hackers have a bogus address transmitted,taking the user unwittingly to a fake "look alike" website where their information can be stolen.
- **DNS-Based Phishing** ("Pharming"). Pharming is the term given to hosts file modification or Domain Name System (DNS)-based phishing. With a pharming scheme, hackers tamper with a company's hosts files or domain name system so that requests for URLs or name service return a bogus address and subsequent communications are directed to a fake site.
- **Man-in-the-Middle Phishing** is harder to detect than many other forms of phishing. In these attacks hackers position themselves between the user and the legitimate website or system. They record the information being entered but continue to pass it on so that users' transactions are not affected. Later they can sell or use the information or credentials collected when the user is not active on the system.
- **Search Engine Phishing** occurs when phishers create websites with attractive (often too attractive) sounding offers and have them indexed legitimately with search engines. Users find the sites in the normal course of searching for products or services and are fooled into giving up their information.

IV. Features for Email Data

To train the classifier appropriate features must be extracted from the emails. We group the features into three groups: basic features, dynamic Markov chain features, and latent topic model features[9]. After providing these features we furthermore perform feature processing (scaling and nor- malization) and feature selection (i.e., automatic elimination of the most irrelevant features).Basic features are features that can directly be extracted from the email without much further processing. Note that we only use features that can be derived directly from the email itself. In particular, we do not use features that require information about specific linked-to websites. We use the following basic features [4]:

- **Structural features:** Reflect the body part structure of an email, i.e., information about the number of body parts, discrete and composite body parts, and alternative body parts.
- **Link features:** Reflect various properties of links contained in an email, i.e., information about the total number of links, internal and external links, links with IP-numbers, deceptive links (links where the URL visible to the user is different from the URL the link is pointing to), links behind an image, etc.
- **Element features:** Reflect what kinds of web technologies are used in an email, i.e., information about whether HTML, scripting and in particular JavaScript, and forms are used.
- **Word list features:** We use a positive word list, i.e., a list of words hinting at the possibility of phishing.

V. Detection of phishing email

The act of sending an e-mail to a user falsely claiming to be an established legitimate enterprise in an attempt to scam the user into surrendering private information that will be used for identity theft. The e-mail directs the user to visit a Web site where they are asked to update personal information, such as passwords and credit card, social security, bank account numbers, that the legitimate organization already has.

Fig. 4: Module for detecting of phishing attack

To detect phishing attacks module (see fig. 4) will:

- Employ filtering model that screens out good email from phishing email.
- Suspicious ones are stored for other modules use.
- Employs PLSA, a natural language processing technique.

Some Countermeasures of Detection of phishing email:-

1) The people behind OpenDNS have recently launched a site called PhishTank [5]. On this site, users can report phishing attacks. Similar Alerts can be found at FraudWatch International.
2) Anti-phishing toolbars are small utilities that can be quite useful in protecting a user from a known Phishing attack. Some of these toolbars that have been catching my attention include Anti-phishing Toolbar from Netcraft , Phishing Filter from Microsoft , Scamblocker from Earthlink, Trustwatch from Geotrust and Anti-fraud Toolbar from Cloudmark. SpoofGuard, EarthLink and Netcraft, were able to identify over 75% of the phishing sites tested. However, some of the toolbars were not able to identify even half the phishing sites tested [6].
3) Phishing Awareness is extremely significant for internetusers. The majority of the problem lies due to the indifference of the clients in the area. The consumer advice can be taken from the Anti Phishing Working Group (APWG) .

4) Some common recommendation for general practices & behavior [8].

VI. Conclusion

Phishing is a form of online identity theft that aims to steal sensitive information from users such as online banking passwords and credit card information. The last years have brought a dramatic increase in the number and sophistication of such attacks. Although phishing scams have received extensive press coverage, phishing attacks are still successful because of many inexperienced and unsophisticated Internet users. Attackers are employing a large number of technical spoofing tricks such as URL obfuscation and hidden elements to make a phishing web site look authentic to the victims.In this paper present how the phishing attacks are done and the main focus of the paper is on the phishing e-mail and how detect these phishing attacks and their countermeasures.

References

[1] Anti-Phishing Working Group. "Phishing activity trends report", June 2006. [Online] Available : http://antiphishing.org/reports/apwg report june 2006.pdf,.

[2] "Identity Theft: What to Do if It Happens to"

[3] A. Emigh, "Phishing attacks: Information flow and chokepoints, in: Phishing, Countermeasures", M. Jakobsson, S. Myers, eds, Wiley, 2007, pp. 31–64.

[4] A. Bergholz, J.-H. Chang, G. Paaß, F. Reichartz, S. Strobel, "Improved phishing detection using model-based features", in: Proceedings of the Conference on Email and Anti-Spam (CEAS),Mountain View, CA, USA, 2008.

[5] "Database for information on phishing sites reported by the public PhishTank". [Online] Available : http://www.phishtank.com/.

[6] Lorrie Cranor, Serge Egelman, Jason Hong, Yue Zhang "Phinding phish: An evaluation of anti-phishing toolbars". The 14th Annual Network & Distributed System Security (NDSS) Symposium 2007 – San Diego, CA - 28th February - 2nd March., in press.

[7] Mitesh Bargadiya, Vijay Chaudhari, Mohd. Ilyas Khan, Bhupendra Verma "The Web Identity Prevention: Factors to consider in the anti- phishing design", Vol. 2(7), 2010, 2807-2812, pp. 2811, ISSN: 0975- 5462.

[8] I. Fette, N. Sadeh, A. Tomasic, "Learning to detect phishing emails", in: Proceedings of the International World Wide Web Conference (WWW), Banff, Canada, 2007, pp. 649–656.

I completed my M.Sc.(IT) from Graphic Era University Dehrdun and also pursuing M.Tech also from Graphic Era University. My area of Research is Wireless Network Security and Networking.