웹페이지의 피싱 차단 탐지 기술에 대한 분석

김정태
목원대학교

Analyses of Detection and Protection for Phishing on Web page

Jung-Tae Kim
Mokwon University
E-mail : jtkim3050@mokwon.ac.kr

요 약

Phishing is a form of online identity theft that aims to steal sensitive information such as online banking passwords and credit card information from users. Phishing scams have been receiving extensive press coverage because such attacks have been escalating in number and sophistication. According to a study by Gartner, Many Internet users have identified the receipt of e-mail linked to phishing scams and about 2 million of them are estimated to have been tricked into giving away sensitive information. This paper presents a novel browser extension, AntiPhish, that aims to protect users against spoofed web site-based phishing attack

Ⅰ. Introduction

Phishing is a criminal trick of stealing victims' personal information by sending them spoofed e-mails urging them to visit a forged webpage that looks like a true one of a legitimate company and asks the recipients to enter personal information such as credit card number, password and etc. The victims may finally suffer losses of money or other kinds. According to the reports of Anti-Phishing Working Group. The Internet is playing an increasingly significant role in today's commerce and business activities. Unfortunately, poor security on the Internet and large financial gains provide a strong motivation for attackers to perpetrate such seemingly low risk, yet high-return online scams. Most of the phishing attacks are carried out by sending large volume of clearly crafted emails posing to originate from a legitimate business domain. These messages are intended for redirecting the recipients to a masqueraded website, which manifests the same behavior of a legitimate domain, for tricking the users to reveal their financial information. There are some characteristics of Phishing, such as huge harmfulness, powerful technicality, ulterior instrument and extensive influence. The most important harmfulness is that it will create "trust crises". The trust will be eroded gradually if without effective countermeasures to deal with the fraud, and all the sitter-in taking part in network transaction will be harmed in the end.

Ⅱ. Overview of Phishing

Generally speaking, Phishing is a method that exploits people's sympathy in the form of aid-seeking e-mails; the e-mail act as bait. These e-mails usually request their readers to visit a link that seemingly links to some charitable organization's website; but in truth links the readers to a website that will install a Trojan program into the reader's computer. Therefore, users should not forward unauthenticated charity mails, or click on unfamiliar links in an e-mail. Sometimes, the link could be a very familiar link or an often frequented website, but still, it would be safer if you'd type in the address yourself so as to avoid being linked to a fraudulent website. Phisher deludes people by using similar e-mails mailed by well-known enterprises or banks; these e-mails often asks users to provide personal information, or result in losing their personal rights; they usually contain a counterfeit URL which links to a website where the users can fill in the required information. People are often trapped by phishing due to inattention Besides, you must also be careful when using a search engine to search for donations and charitable organizations.

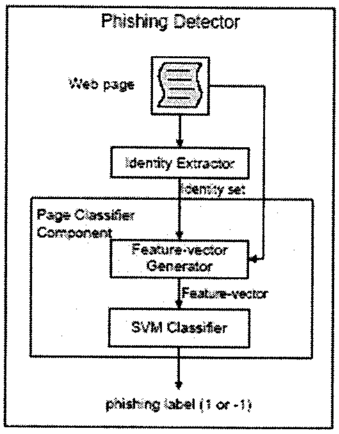

Figure 1. The architecture of the phishing detector

Evolving with the antiphishing techniques, various phishing techniques and more complicated and hard-to-detect methods are used by phishers. The most straightforward way for a phisher to scam people is to make the phishing Web pages similar to their targets. A phishing strategy includes both Web link obfuscation and Web page obfuscation. Web link obfuscation can be carried out in four basic ways:

1. adding a suffix to a domain name of the URL,
2. using an actual link different from the visible link,
3. utilizing system bugs in real Web sites to redirect the link to the phishing Web pages, and
4. using cousin domain names (e.g., replacing certain characters in the target URL with similar characters).

The Web page obfuscation can be carried out in three basic ways:

1. using the downloaded Web page from the real Web site to make the phishing Web page appear and react exactly the same as the real one does,
2. using a script or images to cover the address bar to scam users into believing they have entered the correct Web sites, and
3. using visual-based content (Image, Flash, JavaApplet, etc.) rather than HTML to avoid HTML-based phishing detection.

The processing modules for the true webpage and a suspicious webpage are similar in the steps. Both contain the following two:

a) Page Segmentation
b) Feature Extraction

In our approach, the visual features of the webpages are obtained based on the result of webpage segmentation. In both the True Webpage Processing Module and the Suspicious Webpage Processing Module, the webpage is first decomposed into a set of salient blocks which are visually (in terms of visual features) and semantically (in terms of content relevancy) consistent within each block but distinguishable between adjacent blocks. We simply employ the method in but do not group the salient blocks to higher levels.

III. Similarity Assessment

The Visual Similarity Assessment Module measures the visual similarity between two webpages in three aspects: block level similarity, layout similarity, and style similarity.

3.1 Block Level Similarity

We define block-level similarity as the weighted average of the visual similarities of all matched block pairs between two pages. The module first categorizes block content as either text or image and then extracts the features from the blocks. The block level similarity measures the visual similarity of two pages at the level of individual blocks. It is defined as the weighted average of the visual similarities of all matched block pairs between two pages. Web pages are of different types, the module sets their similarity to 0. Two blocks' total similarity is a weighted sum of the individual feature similarities. A feature's weight represents its importance to the total similarity, and we can assign the weights empirically.

3.2 Layout Similarity

Usually, it takes some efforts to make a brand new webpage mimicking a true webpage. A very convenient way is to copy the source file of the true one and modify it a little bit for this purpose. In this case, the main webpage structure is kept and the two webpages look very similar in their page layouts. Hence, we define the layout similarity as the ratio of the weighted number of matched blocks to the number of total blocks in the true webpage. In measuring layout similarity, we begin by finding several blocks with identical contents and then use the so-called neighborhood relationship model1 to match other blocks according to the spatial relations of all blocks on the page. We consider two blocks to be matched if both exhibit high visual similarity and satisfy the same position constraints

3.3 Overall Style Similarity

In addition to the webpage content, the style consistency is another important feature which can easily cheat the victims' eyes. Generally, all webpages owned by one company would keep the style consistent.

IV. Phishing: Attacks & Strategy

The increase in online services offered to consumers has naturally led to an increase in the exchange of personal information to access such services. This information is becoming ever more valuable due to the significant amount of money that could be stolen if someone's personal information got into the hands of a criminal. Phishing entails stealing someone's confidential information online for the explicit use of committing fraud. Phishing trends are on the rise: Phishing techniques range in complexity from deceptive attacks to various forms of malware or malicious software attacks. Deception attacks use social engineering to trick users into providing confidential information. The most common deceptive attacks come in the form of email in which, "a phisher sends deceptive email, in bulk, with a 'call to action' that demands the recipient click on a link,". In most scenarios the link leads to a fraudulent web site that prompts the user to enter their personal information. If the user enters and submits their personal information the attack has been successful. The phisher takes the stolen information and uses it to commit acts of fraud or sells the information to other criminals.

More advanced technical attacks move away from social engineering tactics and into the realm of malicious software. Malware attacks comprise the installation and execution of malicious software on a victim's personal computer. In a hybrid approach a phisher will use social engineering tactics to lure a user into opening or downloading a file that contains a malicious software installation. Security vulnerabilities are also exploited to install malicious software on an unsuspecting user's computer. Phishing attacks are on the rise and becoming increasingly complex. On the home front, consumers can protect themselves through education, awareness, and up-to-date security software. Consumers are not the only party that needs to be up to speed on phishing education. AntiPhish is an application that is integrated into the web browser. It keeps track of a user's sensitive information (e.g., a password) and prevents this information from being passed to a web site that is not considered "trusted" (i.e., "safe"). The development of AntiPhish was inspired by automated form-filler applications. Most browsers such as Mozilla or the Internet Explorer have integrated functionality that allows form contents to be stored and automatically inserted if the user desires. This content is protected by a master password. Once this password is entered by the user, a login form that has previously been saved, for example, will automatically be filled by the browser whenever it is accessed. Antiphish takes this common functionality one step further and tracks where this information is sent.

V. Conclusion

In this paper, we propose a novel approach to detect phishing webpages based on visual similarity. The visual similarity between two webpages is then evaluated in three metrics: block level similarity, layout similarity, and overall style similarity. A webpage is reported as a phishing suspect if any of them (with regards to the true one) is higher than its corresponding preset threshold.

References

[1] E. Levy, "Interface illusions," IEEE Security & Privacy, vol. 02, no. 6, pp. 66-69, Nov.-Dec. 2004.

[2] B. Krebs, "New industry helping banks fight back," Washington Post.com, Jan.4,200

[3] C. L. Schuba, "Analysis of a denial of service attack on TCP,"*IEEE Security and Privacy Conference*, 1997, pp. 208-223.

[4] http://www.antiphishing.org.

[5] G. Goth, "Phishing Attacks Rising, but Dollar Losses Down,"*IEEE Security and Privacy Magazine*, vol. 3, no. 1, January-February 2005, p. 8.