

Phishing Attack Detection by Using a Reputable Search Engine

Robert Ma
Electrical and Computer Engineering Department
University of Toronto
robertma@eecg.toronto.edu

ABSTRACT

Phishing attack is one of the most critical issues on the Internet today as it poses serious threats on millions of end users and commercial institutions resulting large amounts of financial damages. Many approaches have attempted to solve the problem; however, most solutions have been proven ineffective as they lack the necessary interaction with users. In this paper, a novel anti-phishing approach, one that bases on a search engine, is presented. It aims to provide an effective and yet lightweight solution that addresses most common problems found in other current anti-phishing solutions. Our solution exploits search results made with unique keywords extracted from a suspicious web site on a reputable search engine, Google. Then, these search results are analyzed to determine whether a website is genuine or counterfeit. Detailed design and implementation issues will be presented in this paper as well as accuracy and performance issues will be addressed.

1. INTRODUCTION

In recent years, Internet has become an integral part of our lives and over 1 billion of active users are browsing web sites every day worldwide [8]. Being such an enormous but yet still developing medium, it is almost the perfect environment to nurture frauds and scams. Phishing is one of the common techniques used on the Internet today to commit scams which legitimate-looking but fake websites or emails are set up by the attackers to deceive victims in divulging their personal information. In a study made by the Anti-Phishing Working Group (APWG) in year 2006, within the month of July, 14191 unique phishing sites were found and 23670 cases of phishing were reported

[2]. In year 2004 alone, an estimate of 20 million phishing emails were sent out, resulting nearly 10 billion dollars in damage [4]. These numbers clearly show that phishing attack has become one of the most serious threats to Internet users of today. It is therefore critical to prevent phishing from spreading immediately.

Many researches were done in developing ways to prevent Internet users from falling into phishing traps. Most common anti-phishing solutions use web browser toolbars that provide users with information like the “real” domain name of the website or whether Secure Sockets Layer (SSL) is used on this website. With these information displayed in toolbars, the hope was that users could identify counterfeit websites easily. However, the fact is these solutions are ineffective against phishing attacks, and these are some of the reasons [10]:

- i) Toolbars are usually located in a peripheral area in the browser hence warning indicators in them become insignificant when compared to the web content.
- ii) Security is rarely user’s main focus when browsing web sites. It’s unlikely that user will continuously pay attention to these indicators.
- iii) Indicators usually show something is wrong and advise user not to proceed, but they do not suggest good alternatives. This may encourage users to risk submitting their information anyway, since they don’t see any other way to accomplish their goal.

While the growth of phishing is become out of control, on the other hand, the number of websites also increases dramatically at the same time. Many search engines, as a result, have developed as a by-product of the Internet expansion. Several

reputable search engines are commonly used by Internet users today, they include: Google, Yahoo, AltaVista, and etc. Among these popular search engines, Google is known to have the cleanest design, fastest search results, and its unique PageRank technology [6].

In this paper, we present a novel anti-phishing solution that aims to tackle the abovementioned common problems found in other anti-phishing products. In our approach, a web browser (Mozilla Firefox) plug-in will be developed to performance phishing website detections by leveraging the Google search engine. When analyzing whether a website is authentic, our plug-in extracts unique keywords from the website and make a search query in the Google search engine. The search results of the unique keywords are then used to compare with the website that the users are currently visiting and to determine whether the website is a phishing website. Since most phishing sites are short-lived; they would have much less visitors compared to the legitimate sites they imitate. Hence, phishing sites will never show up as the top search results. Hence when there is a mismatch in the domain names of the websites in the top search results and the suspicious website, it can be derived that the suspicious web site is unlikely to be the one that the users are expecting they are visiting.

2. RELATED WORK

There are many anti-phishing solutions available in the market, and most of them take either the back-end or the front-end approaches to this problem.

An example of back-end approaches would be to stop phishing at the email level [1]. This approach works because most current phishing attacks use broadcast emails to lure victims to a phishing website.

Another similar solution, PHONEY [3] tries to mimic user response by providing fake information to the suspicious websites that request critical information. Then, the websites' responses are forwards to the decision engine for further analysis.

Alternatively, most front-end approaches utilize web browser toolbars, providing information to users which hints whether a website is legitimate or not.

SpoofStick [9] displays the website's real domain name, in order to expose phishing sites that obscure their domain name. An attack might use a legitimate-looking domain name as a sub-domain, e.g. `www.ebay.com.ww2.us` to fool users; SpoofStick would display this domain as `ww2.us`.

TrustBar [7] makes secure web connections (SSL) more visible by displaying the logos of the website and is certificate authority (CA). This is useful against phishing because many legitimate websites use SSL to encrypt the user's sensitive data transmission, but most phishing sites do not.

Google Safe Browsing for Firefox [5] pops up an alert when a user is on a web that Google determines to be illegitimate. It uses several techniques to determine whether a page is genuine, including the use of a blacklist containing pages that have been identified as suspicious and/or misleading based on automated detection or user reports. Furthermore, this approach also examines pages' content and structure in order to catch potentially misleading pages. This solution is the most similar to our approach of all other related work. However, in our approach we leverage the Google search engine as our knowledge database whereas in Google Safe Browsing for Firefox relies on an existing blacklist.

3. DESIGN

In this project, a web browser plug-in is developed to detect phishing website and warn users when they are identified. Mozilla Firefox is chosen to be the web browser that this project will build on top of because of its popularity, multi-platform interoperability, and its robust plug-in platform which enables developer freely contribute new tools to the browser. The design of our plug-in will try to be clean and simple but at the same time address the common problem found in existing anti-phishing solutions.

To minimize any effect can be made to users' web browsing experience, this plug-in will be lightweight in processing and should take up as little space as possible, so that the size of the main web window is not affected during web browsing.

Many common anti-phishing products provide warning indicators that do not draw users' attention when they should beware of. In case of phishing website is detected, our solution will pop up a dialog, which interrupts the current task the user is trying to accomplish. As a result, it would be impossible that users would miss a potential warning and carelessly fall into traps of phishing websites. Furthermore, since our approach bases on Google search results to determine a web site's authenticity, in the event a phishing web site is identified, our tool can easily suggest one of the top search results as an alternative which users can proceed with instead of the suspicious phishing web site. This feature addresses the issue that some other anti-phishing tool only provides a warning but no alternatives to users and hence users take risks in proceeding to the phishing website as they has no other options to work with.

REFERENCES

- [1] Adida, B., Hohenberger, S., Rivest, R., Lightweight Encryption for Email. USENIX Steps to Reducing Unwanted Traffic on the Internet Workshop (SRUTI), 2005.
- [2] Anti-Phishing Working Group. Phishing Activity Trends Report, July 2006. http://www.antiphishing.org/reports/apwg_report_july_2006.pdf
- [3] Chandrasekaran, M., Chinchani, R. PHONEY: Mimicking User Response to Detect Phishing Attacks. *Proceedings of the 2006 International Symposium on a World of Wireless, Mobile, and Multimedia networks (WoWMoM'06)*, 2006.
- [4] D. Illett. Phishing attacks skyrocket in 2004. 2004. http://news.com.com/21007349_35479145.html
- [5] Google Safe Browsing for Firefox BETA. <http://www.google.com/tools/firefox/safebrowsing/>
- [6] Google Technology. <http://www.google.com/technology/>
- [7] Herzberg, A., Gbara, A. TrustBar: Protecting (even Naïve) Web Users from Spoofing and Phishing Attacks. 2004. <http://cs.biu.ac.il/~herzbea/Papers/ecommerce/spoofing.htm>
- [8] Internet World Stats. Top 20 Countries with the highest number of Internet users. 2006. <http://www.internetworldstats.com/top20.htm>
- [9] SpooftStick. 2004. <http://www.spooftstick.com/>.
- [10] Wu, M., Miller, R.C., Little, G. Web Wallet: Preventing Phishing Attacks by Revealing User Intentions. *Symposium On Usable Privacy and Security (SOUPS) 2006*. Pittsburgh, PA, USA.