A System to Detect Phishing Mail

Indrajeet kumar, Shipra shalvi, Rajan soni
Department of Computer Engineering,
Sinhgad Institute of Technology,
Lonavala, pune

***Abstract* - in order to get the sensitive user data like online banking user-id and passwords or credit card information, social networking site's data which may then be used by 'phishers' for their own personal gain or they may hire by firm or agency to do this is the primary objective of the phishing e-mails. Online activities has increased now a days exponentially so phishing scams which have now started achieving monstrous proportions. In this paper we present an Anti-Phishing application for the end user which keeps track of the url's and sites with which the user indulges in financial transactions, scans his e-mail account for mails which appear to have come from these institutions and detect him against suspected phishing e-mails, if the same are detected in his mailbox.**

***Keywords* - Phishing, Privacy, Security, Spam**

I.INTRODUCTION

PHISHING is a social engineering attack that aims at exploiting the weakness found in system processes as caused by system users. For example, a system can be technically secure enough against password theft, however unaware end users may leak their passwords if an attacker asked them to update their passwords via a given Hypertext Transfer Protocol (HTTP) link, which ultimately threatens the overall security of the system. Moreover, technical vulnerabilities (e.g. Domain Name Sys- tem (DNS) cache poisoning) can be used by attackers to construct far more persuading socially-engineered messages (i.e. use of legitimate, but spoofed, domain names can be far more persuading than using different domain names). This makes phishing attacks a layered problem, and an effective mitigation would require addressing issues at the technical and human layers. Since phishing attacks aim at exploiting weaknesses found in humans (i.e. system end-users), it is difficult to mitigate them. For example, as evaluated in [1], end-users failed to detect 29% of phishing attacks even when trained with the best performing user awareness program. On the other hand, software phishing detection techniques are evaluated against bulk phishing attacks, which makes their performance practically unknown with regards to targeted forms of phishing attacks. These limitations in phishing mitigation techniques have practically resulted in security breaches against several organizations including leading information security providers [2], [3]. Due to the broad nature of the phishing problem, this phishing detection survey begins by:

• Defining the phishing problem. It is important to note that the phishing definition in the literature is not consistent, and thus a comparison of a number of definitions is presented. • Categorizing anti-phishing solutions from the perspective of phishing campaign life-cycle. This presents the various anti-phishing solution categories such as detection. It is important to view the overall anti-phishing picture from a high-level perspective before diving into a particular technique, namely: phishing detection techniques (which is the scope of this survey). • Presenting evaluation metrics that are commonly used in the phishing domain to evaluate the performance of phishing detection techniques. This facilitates the com- parison between the various phishing detection tech- niques. • Presenting a literature survey of anti-phishing detection techniques, which incorporates software detection tech- niques as well as user-awareness techniques that enhance the detection process of phishing attacks. • Presenting a comparison of the various proposed phishing detection techniques in the literature.

A survey conducted by Gartner Inc., found that 3.6 million adults lost money due to phishing attacks during the period from Sep '06 to Aug '07, leading to a huge financial loss assumed to be of the tune of $3.2 billion in US alone [3] compared to $2 billion lost in the year 2006 [4]. This loss is not only due to the financial loss which is borne by the individuals and the financial institutions on account of the fraudulent transactions by the phishers. It is also due to the dent in the confidence and the resultant hesitancy of prospective clients of making use of web based businesses and services from the fear of being duped of their hard earned money.

There are a host of reasons responsible for the success of phishing attacks [5] a few of which are: lack of user's computer knowledge, use of obfuscated URLs, rapid technological advancements in the field of computer science, lack of awareness amongst internet users about elements phishing for their sensitive information etc.

In this paper we present a desktop based Anti-Phishing application for a naïve user against spoofed website based phishing attacks. The design of the application is based on the premise that a user is more susceptible to fall victim to a phishing e-mail which appears to have been sent from an institution like a bank, insurance company, investment company or an e-commerce site with which he has an existing relationship rather than from one with which he has no relationship. So if the user receives an e-mail claiming to be from, say Axis bank, but he does not have an account with them, then he is unlikely to forward any sensitive information to the phishing site. The application keeps track of names and URLs of websites with which the user has a relationship, scans the e-mail account of the user for e-mails which apparently have been sent by these institutions, looks for embedded URLs in these messages and generates a phishing warning for the mails which appear to be phishing e-mails.

II.PHISHING ATTACK TECHNIQUES

An attack technique may be defined as the path taken or the means adopted by a hacker/phisher to reach a destination computer [6]. In this section we present an overview of various attack vectors which are adopted by the phishers in order to try and lure the users to reach their phishing sites.

A number of methods have been devised by phishers to trick the users into doing what they want them to do. These methods can broadly be classified into two categories, those which rely on use of spoofed e-mails and websites (social engineering), and others which can be termed as exploit based phishing attacks. Exploit based phishing attacks are more sophisticated then the spoofing attacks and make use of certain inherent weaknesses in the web browsers, which are exploited by phishers to install certain malware in the user's machine, such as a key-logger or a screen-grabber, and use the same to steal information. The proposed design of the Anti Phishing Module revolves around the ways and means to mitigate the spoofing e-mails and web sites attacks.

A. Spoofing E-mails and Web Site

In their earlier days, the phishing attacks were e-mail based wherein the users were sent spam mails asking them to verify some form of their personal information via a reply e-mail. Today however it is very unlikely that any user will fall pray to such an attack (unlikely but not impossible). The primary reason being that users today understand that the financial institutions do not carry out sensitive transactions such as account verification through e-mail (it is, however, to be noted that there is a need to educate the users about safe online transactions in developing countries like India where Internet banking is a relatively recent phenomenon, and an average user is not aware about the unseen threats posed by the phishing community). Such organisations use their websites to provide interactive services to their clients which allow them to make use of encrypted web pages.

Many phishing attacks today, therefore make use of a combination of spoofed websites and e-mails to try and extract sensitive information from the users. These attacks generally employ some form of URL obfuscation techniques [7] to trick the user into visiting fake web sites which look and feel exactly similar to the original websites. More often than not the e-mails sent to the users ask them to verify their credentials with the organisation at the earliest (usually within 24 hrs) by clicking on an embedded URL. The weblink leads the user to an authentic looking but fake website of the organisation or even to the authentic website of the organisation but with obfuscated login and password dialog boxes using borderless windows. Some attacks also make use of hidden frames, images or Javascript code to control the way a webpage is rendered on the user's browser.

Lately, in addition to the use of e-mails, phisher community has started exploiting the Instant Message services to lure users into visiting the phishing sites [9]. During the chat sessions the phisher tries to convince the user to visit the spoofed website, the obfuscated URL for which the phisher forwards during the chat session itself, and divulge with his personal information.

B. Exploit Based Attacks

Exploit based attacks are more sophisticated when compared to the family of attacks described above. These attacks exploit some inherent weaknesses in the users' browsers or install some other malware such as a key-logger or a screen grabber which are programmed to keep tab of the user activity over his computer [8]. The data so gathered may be collected by the phisher through continuous streaming, local collection and batching of information which may then be uploaded on the phisher's server subsequently or by use of Trojan programs which allow the attacker to collect the user information as and when required. In addition, an attacker may use HTML or DHTML [8] to manipulate the display of information on the users' web browsers. These can be used to (a) Deliver additional content such as overriding page contents or graphics. (b) Executing screen grabbing / key logging observation code. (c) Provide a fake secure https wrapper for sites content, i.e., display a fake image of padlock at an appropriate location on the browser. (d) Hiding HTML code from the customers. (e) Loading images and HTML content in the background for later use by a malicious application.

Countering such attacks is beyond the scope of the Anti-Phishing application discussed in this paper. To counter these attacks what is required is that these security issues be taken up by the respective browser manufacturing teams who should try and fix these security bugs. The aim of our application is to try and mitigate spoofed e-mail and web site attacks by forewarning the users about presence of phishing e-mails in their e-mail account, thus reminding him to tread with caution when dealing with such mail messages.

III.DESIGN OF ANTI PHISHING APPLICATION

This anti phishing module is based on the following premise:-

- A user is more likely to fall victim to a phishing attack if the phishing e-mail received by him seemingly came from a financial/trading institution with which the user has a transaction relationship.
- For a naïve and inexperienced user, it would be better that the task of checking the authenticity of the URLs embedded in the e-mail be left to an application which cannot be fooled by the obfuscation techniques employed by the phishers.

C. Implementation Details

This work has been implemented in Java programming language using the Netbeans 6.0.1 Integrated Development Environment (IDE).

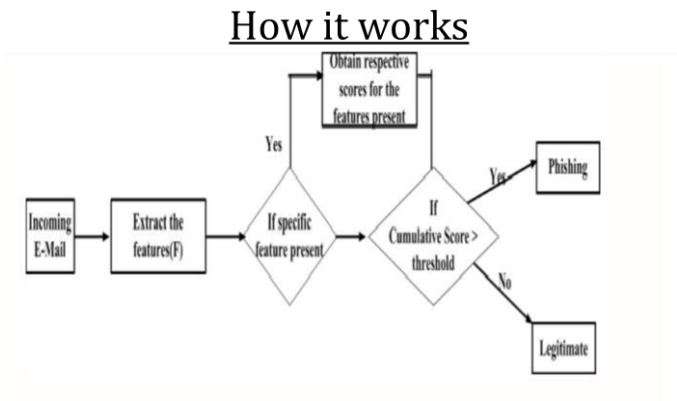

Fig 1

B Working Example

Suppose a naïve user is a customer of Axis Bank and has registered for their online banking services. Also suppose that the said user chooses to use our Anti Phishing application.

When the user runs the application for the first time, he is asked to enter the organisation's name and its secure URL address (as provide to the user by the bank). Accordingly he enters the bank's name as Axis bank and the URL as www.axisbank.com. The application now contacts the DNS and retrieves the corresponding IP address which in this case is 210.210.17.218. The MD5 value of this IP address is then calculated and is stored in the database.

In the next stage of the application, if the user wants to check his e-mail account, he is asked to provide his username and password to logon to his account. Once connected, the application shortlists the mails to be checked, i.e., either all the mails in the user's inbox (if it the first run of the application) or only those whose sent date is at the most 5 days less than the maximum sent date that was stored when the application was run the last time. The application looks for the substring "axis" in the 'From' header field of the short listed messages. Let there be a mail from onlineservice@alerts.axis.com with its subject being "IMPORTANT ALERT: Re-Confirm Your Net Banking Details, Update Your Account To Avoid Violation" (refer fig 1).

The message body of this e-mail is scanned for embedded URLs. It should be noted from the phishing e-mail shown in fig 1 that the phisher has tried to hide the identity of the destination URL behind a button titled "Update your account". The trick might fool a naïve or even an experienced internet user but the application's search returns the destination URL as *http://www.erainfo.es*. The corresponding IP address of this URL is fetched from the DNS and its MD5 value is matched against the value stored in the database provided by the user. A mismatch produces a warning against suspected phishing e-mails (as shown in fig 1 above).

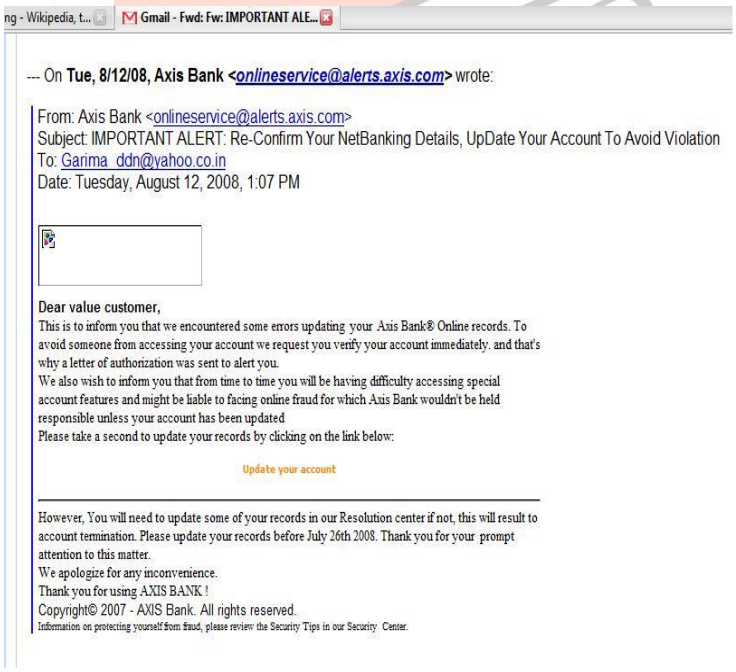

Fig 2

The user is thus warned against the existence of likely phishing e-mails in his account even before he physically opens his e-mail service. Forewarned about the same, he is unlikely to fall victim of the phisher's trap set for him.

IV. RESULTS AND DISCUSSIONS

To show the effectiveness of proposed system some experiments are conducted on java based windows machine. To measure the performance of the system we set the bench mark by selecting Gmail ID emails which may containing phishing URL with the Email body.

To determine the performance of the system, we examined how many relevant Phishing URL's are identified based on some phishing protocols. To measure this precision and recall are the best measuring techniques. So precision can be defined as the ratio of the number of Phishing URL's are identified to the total number of irrelevant and relevant phishing URL's identified. It is usually expressed as a percentage. This gives the information about the relative effectiveness of the system. Whereas Recall is the ratio of the number of relevant phishing URL's are identified to the total number of irrelevant phishing URL's identified. It is usually expressed as a percentage. This gives the information about the absolute accuracy of the system.

The advantage of having the two for measures like precision and recall is that one is more important than the other in many circumstances. In contrast, various professional searchers and intelligence analysts are very concerned with trying to get as high recall as possible, and will tolerate fairly low precision results in order to get it. Individuals searching their hard disks are also often interested in high recall searches. Nevertheless, the two quantities clearly trade off against one another.

For more clarity let we assign

- A = The number of relevant Phishing URL's identified,
- B = The number of relevant Phishing URL's not identified, and
- C = The number of irrelevant phishing URL's identified.

So, Precision = (A/ (A+ C))*100

And Recall = (A/ (A+ B))*100

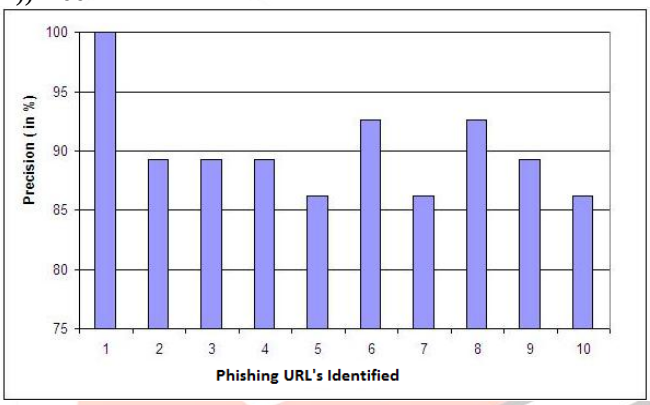

Fig. 1. Average precision of the proposed approach

In Fig. 1, we observe that the tendency of average precision for the identified Phishing URL's are high compared to other systems.

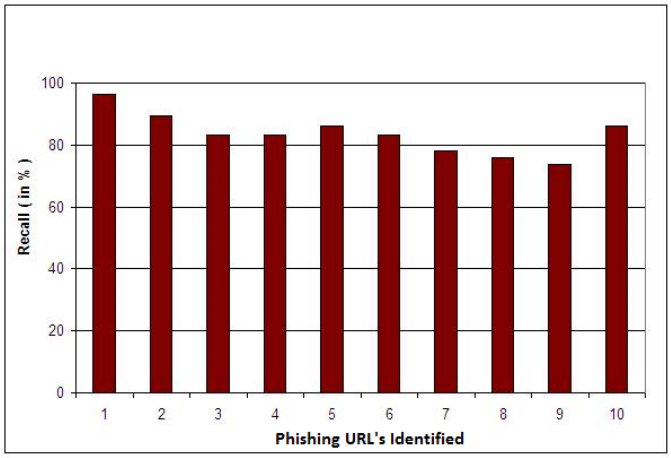

Fig. 2. Average Recall of the proposed approach

In Fig. 2, we observe that the tendency of average Recall for the identified Phishing URL's are high compared to other system. So this shows that our proposed system is achieving high accuracy than any other method.

V. REFERENCES

[1] Kapil Oberoi,Anil K. Sarje,anti-phishing algorithm for end user

[2] Anti-Phishing Working Group, http://www.antiphishing.org/index.html

[3] Rachna Dhamija, J.D.Tygar, "The Battle Against Phishing: Dynamic Security Skins" in Proceedings of the symposium on Usable privacy and security, pp. 77-88, yr 2005.

[4] detecting malicious url in an email –an implementation by dhanlakshmi rangayakulum

[5] intelligent rule based phishing website classification by rami m.mohammad , lee mcclucky