An Anti-Phishing Application for the End User

Praveen Kumar Medapati, UCID: 31365623

Abstract:

Seeking sensitive user data in the form of online banking user_id and passwords or credit card information, which may then be used by 'phishers' for their own personal gain is the primary objective of the phishing e-mails. With the increase in the online trading activities, there has been a phenomenal increase in the 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 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 warns him against suspected phishing e-mails, if the same are detected in his mailbox.

I. INTRODUCTION

Committing a crime in complete anonymity, having gained someone else's identity is a dream come true for any criminal. Phishing is a form of online identity theft which employs both social engineering and technical subterfuge to steal consumers' personal identity data and financial account credentials.

Prior to the advent of Internet such efforts on part of criminals were limited to isolated individuals, who were lured into divulging their personal information through social engineering. However, with the rapid growth of internetworking around the globe and the phenomenal popularity gained by Internet since the early days of 1990's, the going has never been easier for such criminals or 'phishers' as this community has better come to be known as.

The preferred strategy of the phishers today is to send out millions of spam mails to potential targets around the globe, masquerading as if these came from original institutions such as banks, insurance companies etc. These mails urge the recipients to click on the embedded URLs which lead them to fraudulent but apparently official looking phishing websites where the gullible users are made to divulge with their personal information such as passwords, account numbers and such. These are then collected by the phishers from the server side using web tools such as key loggers and used for their personal gains.

Since it first occurred in the mid 1990's by attacking America Online, phishing has become a profound threat to online services provided by financial organizations, ISPs, retailers and governments. On an estimate, almost 5% recipients of phishing e-mails give away their personal information to these phishing sites while they are in operation. 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 compared to $2 billion lost in the year 2006. 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 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.

The paper is structured as follows: In the next section we talk about various types of phishing attacks. Section 3 describes the design and working of our application along with a live example of how a warning about a phishing e-mail is generated. In Section 4 we discuss the related work. Section 5 talks about the future work followed by Section 6 wherein we conclude our paper.

II. PHISHING ATTACK VECTORS

An attack vector may be defined as the path taken or the means adopted by a hacker/phisher to reach a destination computer. 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, such as use of SSL technology.

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 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. 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. 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 DHTMLto 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) 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.

(b) 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.

A. ***Main Functionality***

As brought out above, this application works on the premise that a user is more worried about the authenticity of the e-mails which appear to have come from institutions with which he has a relationship, rather than e-mails received from all and sundry. As an example, consider a user who has an account with the ICICI Bank and not with Axis Bank. This user then is more likely to fall prey to a phishing mail which claims to be from ICICI Bank rather than a mail which asks her to verify her details with Axis Bank.

To ascertain the websites which are of interest to a user, there is thus a need of user interaction with the application. Accordingly, the application initially asks the user to enter the name and the trusted URLs of such institutions/websites where he sends his login details, i.e., username and password. The application fetches the IP Address corresponding to the URL from the DNS server, calculates its message digest (using MD5) and stores this data in a table within the database.

On its subsequent run, the application connects to the e-mail service provider of the user (in this case Gmail from Google) and scans for URLs in the message bodies of only those messages which appear/claim to have been sent by the websites of interest to the user. IP Addresses of URLs so located are fetched by the application and their MD5 values calculated. These values are compared against the values stored in the database for that institution. A warning message which includes subject fields of all the e-mails which report a mismatch in the values of message digest are then displayed, cautioning the user that these mails are suspected to be phishing mails, as shown in figure 1 below.

B. ***Connecting to E-Mail Server***

POP3 (Post Office Protocol version 3) and IMAP4 (Internet Message Access Protocol) are the two most prevalent Internet standard protocols used by the local e-mail clients for e-mail retrieval from a remote server over a TCP/IP connection. Although IMAP4 is more user friendly and offers a host of facilities to the users, it is not supported by most of the ISPs (e.g. Yahoo mail does not support IMAP4). The wide popularity of the POP3 protocol is largely due to its appeal to ISPs, not to the users. Using the POP3 protocol, ISPs can elect not to allow the user to leave a copy of the mail on the Mail Server, thus minimizing hard drive storage space.

The present prototype of the Anti Phishing Module uses POP3 to connect to the Gmail account of the user to retrieve the desired information (Gmail incidentally supports both POP3 and IMAP4). POP3 works over a TCP/IP connection using TCP on network port 110. Gmail however uses the deprecated alternate-port method, which uses TCP port 995.

One of the major disadvantages of using POP3 for mail retrieval is that it does not distinguish between a new message and a message which has already been fetched. As a result every call made to run the application results in it checking the e-mail inbox folder from the very beginning. Although this does not seem to be much of a problem in case there are a limited number of messages in the user's inbox, the time taken to complete run of the application increases manifold in case there are say a couple of thousand messages in the inbox folder (in our trial runs over an inbox containing about 300 messages, the application took an average of 6-7 minutes to give the result).

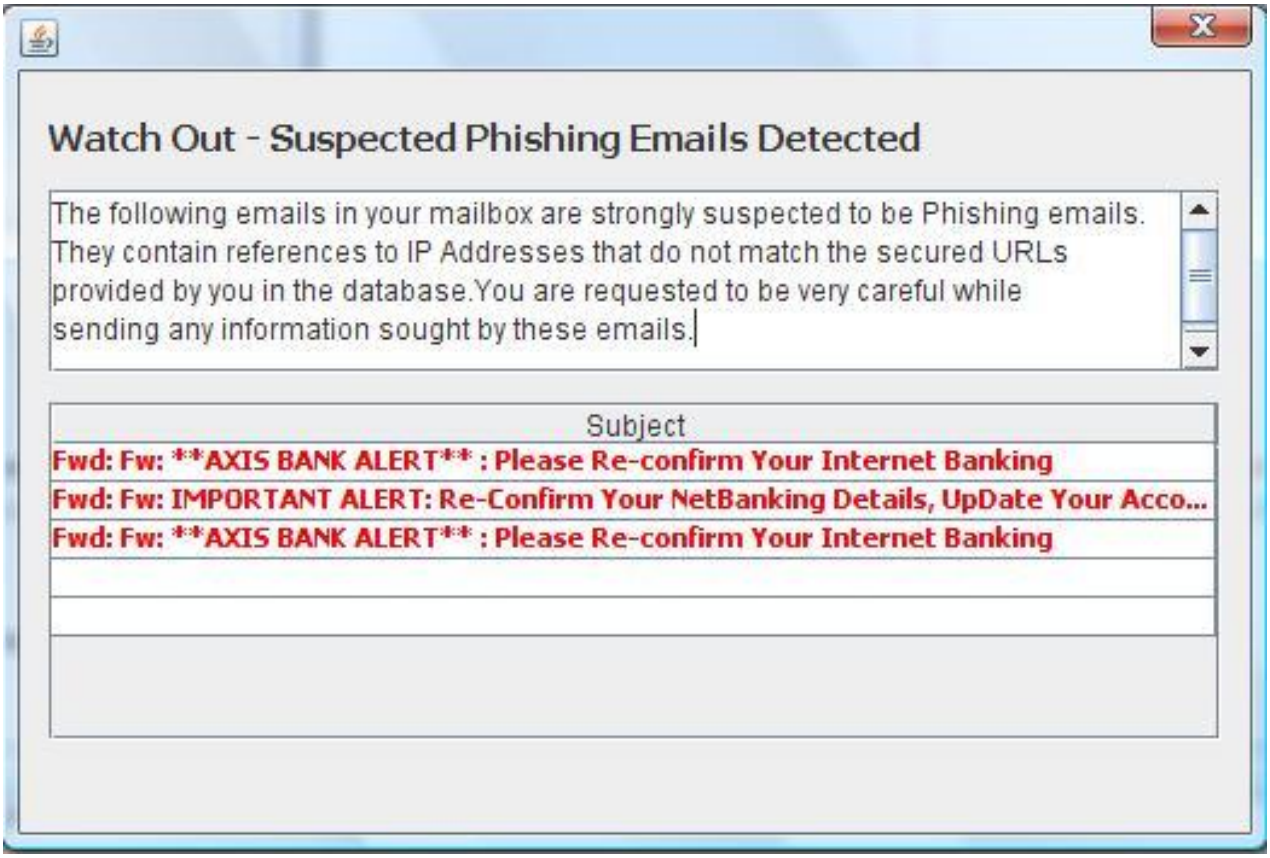

Figure 1: Warning message to the user

To get over the problem, this application makes use of the Sent Date field of the e-mail header (POP3 does not support Received Date field). The maximum value of the Sent Date field of all the messages checked is saved by the application and in the next run is starts checking the messages whose Sent Date is 5 days behind this maximum value. This is done because:

- It might so happen that due to some problem, the mails sent to user's e-mail server get delayed and are not delivered by the time when the application is run.
- The average life time of a phishing site is 5-6 days. Thus it may safely be assumed that if a mail is sent 5 days back and is yet to be delivered to the user it would have been rendered harmless by the time it arrives in his mailbox.

The workflow chart of the application is given in Fig 2.

C. Implementation Details

This work has been implemented in Java programming language using the Netbeans 6.0.1 Integrated Development Environment (IDE). The IDE is available as a free download from http://www. netbeans.org/community/releases/60/.

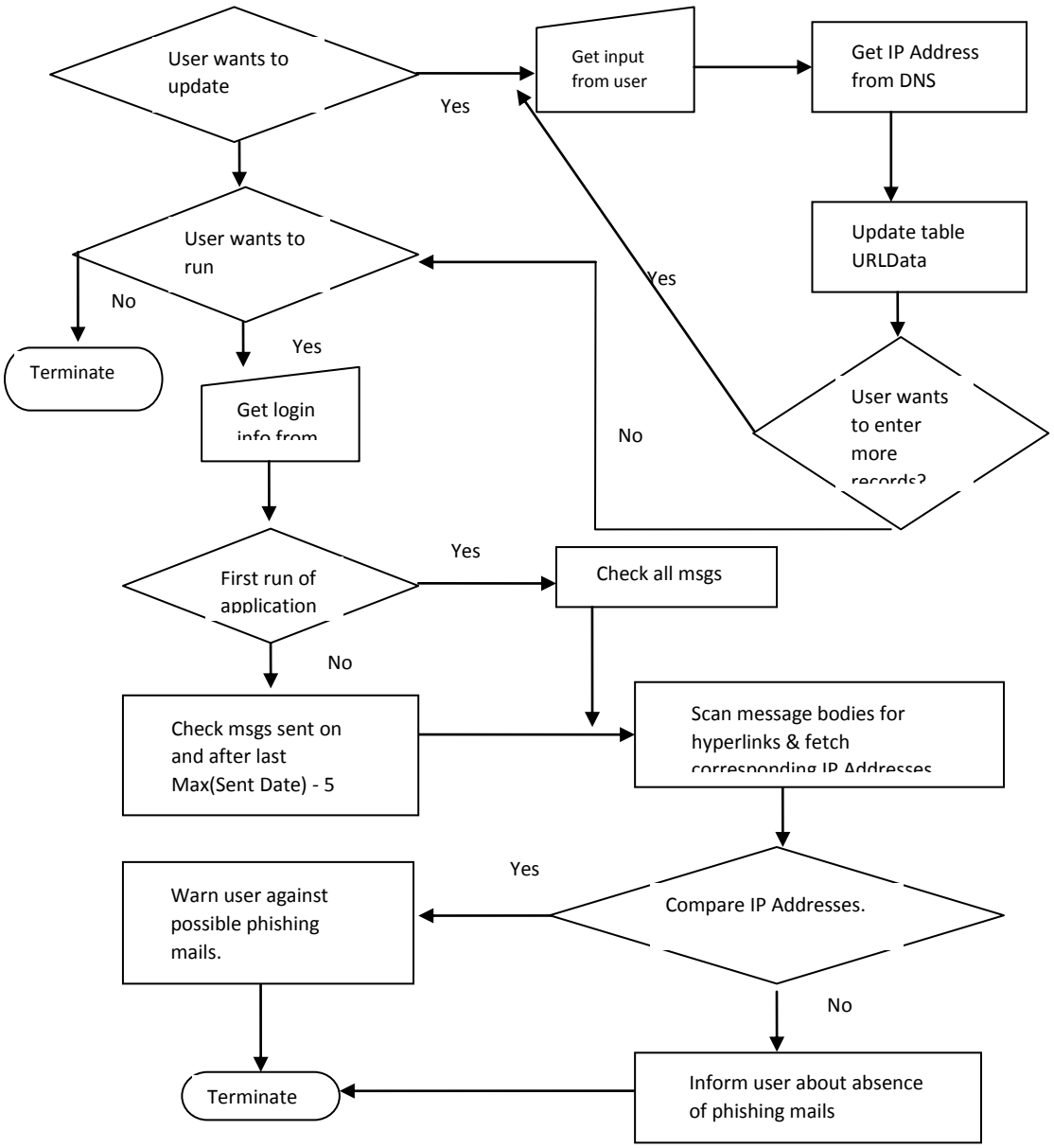

Figure 2. Workflow chart of the Application

Besides the standard Java APIs, JavaMail API has been extensively used in the module. This API is an optional package which is used for reading, composing and sending e-mails. It provides protocol independent access for sending and receiving messages. To use this API, there is a requirement to download JavaMail implementation, unbundle the javamail-[version].zip file, and add the mail.jar file to the project. JavaMail API also requires the JavaBeans Activation Framework, which is also to be downloaded. The framework adds support for typing arbitrary blocks of data and handling it accordingly. After downloading the framework, it is required that we unbundle the jaf-[version].zip file, and add the activation.jar file to the project's CLASSPATH.

For the purpose of handling the data fed in by the user and that fetched from the mail server Microsoft Access Relational Database Management System (RDBMS) has been employed.

D. ***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 3).

The message body of this e-mail is scanned for embedded URLs. It should be noted from the phishing e-mail shown in fig 3 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).

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.

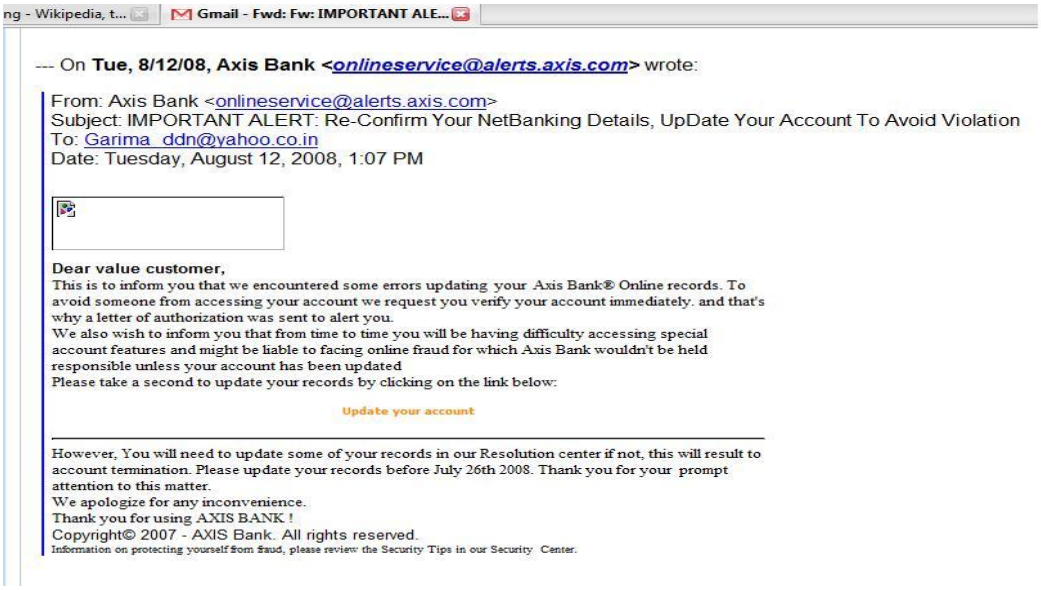

Figure 3. An example of a Phishing E-Mail apparently from Axis Ban

V. FUTURE WORK

This application is just a small step towards tackling the phishing menace an a lot more needs to be done. Some of the features that may be added to this application in order to make it more useful are:

- It is likely that large institutions use multiple servers/mirror sites in order to manage the web traffic on their websites. These mirror sites will have different IP addresses and thus there is a chance that the application may yield false positives when it is run. To mitigate this problem, work is currently underway to fetch and store the digital signature of the login pages corresponding to the authentic URL as provided by the users. Since the source code of the login page of an institution remains the same on all the mirror sites, a mismatch in the same would point to a suspicious website.
- Work is also underway to try and receive alerts as and when the login pages of the institutions change. As and when such an alert is received, the corresponding signature in the database will be amended.
- The present application is a single user application. Making it a multi user application will enhance its utility.
- Integrating the application with web browsers will make it more useful.
- Anti Phishing Working Group (APWG), a pioneer institution formed with the aim to tackle the phishing problem, maintains a black list of sites which are reported upon and confirmed to be phishing sites. Integrating this list with the application so that a passive search of the websites being visited by the user may be carried out and a warning issued in case the site is listed in the black list.
- Automated mechanism to report a phishing site to the authorities so that steps may be initiated to shut them down.
- The email service providers may be approached to integrate the application with their product so that the dependence of the user to access his emails from a single machine, in order to make use of the application, can be done away with. This would also speed up the application significantly since then each incoming mail would be checked for its authenticity. The suspected emails can then be segregated and kept in a separate folder, as is done in the case of spam emails presently.

VI. CONCLUSION

The specter of online identity threat was never so real as it is today primarily due to rapid growth of the Internet and increase in online trading activities which offer a cost effective method to service providers, such as banks, retailers etc., to reach out to their customers via this medium. This has also provided the phishing community an excellent tool to try and fool the netizans into divulging sensitive information about their banking accounts, credit cards details, etc. Recent years have witnessed a host of phishing scams with each doing the other in terms of reach to the users and the level of sophistication.

Indeed the best measure available against such scams is user awareness. Users should be trained against following blind links and the tendency to part with sensitive information over e-mails which may later cause heavy loss to them. However with the ever increasing reach of the Internet, this in itself is a Herculean task. There is thus a need to develop tools which may be of some assistance to the users in dealing with the menace of phishing. This work to try and develop an Anti Phishing application for the end user is a small attempt in this direction.