An Anti-phishing Strategy Based On Webpage Structuring

Rosali Pujapanda[1], Monalisha Parida[2], Ashis Kumar Mishra[3]

Abstract— With the growth of Internet use and e-commerce, the entire Internet Community is familiar with phishing attack. While spam is an annoyance in now-a-days, phishing attacks can cause the major financial disruptions for those victims. Lots of anti-phishing systems have been developed to fight against phishing attacks. Here, in our project, we have implemented an anti-phishing tool based on a strategy that makes visual similarity assessment by comparing the webpage structure of the pages and hence facilitating href comparison to detect phishing web pages. Our application is designed in such a way that it can be plugged into any email server, corporate intranet & email, anywhere to control phishing. This tool would detect web pages that try to act like original pages and phish critical user information.

Index Terms—Anti-phishing, Electronic Crime, Malicious Attack, Phishing website, Security Awareness, Web Security

1. INTRODUCTION

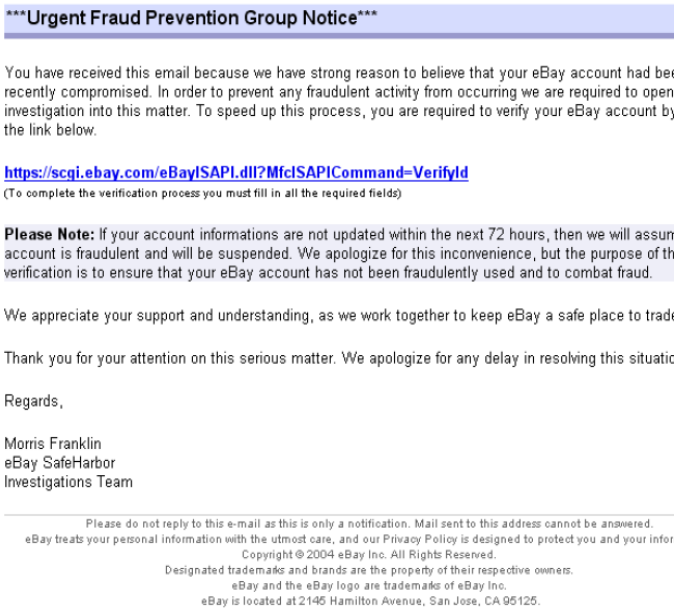
1.1 Origins of the word "PHISHING"

The term "phishing" ("FISH-ing) comes from the fact that internet scammers are using sophisticated lures as they "fish" for users' sensitive information such as user names, password, credit card details, by masquerading as trust worthy and well known enterprises such as PayPal and eBay. The "ph" comes from a common hacking term, "phreaking", which is the first type of hacking before the wide spread of use of the internet. In 1996, some American online accounts(AOL) were hacked and they were called as "phish". These phish were treated as a form of electronic currency where scammers could trade phish for hack software.[2]

1.2 How phishing attacks?

There are several type of phishing attacks. Thousands of phishing emails are sent out with a link to a phishing website, which solicit users' private information such as credit card data or password. When the form is submitted by the user, it sends the data to phishers while leaving the user on the real company's website so that they do not suspect anything. As the phishing emails and the websites look quite authentic, featuring corporate logos and format similar to ones used by legitimate enterprises, many victims do not know the phishing traps until they find their money have gone. In order to attack victims to submit their information to those phishing websites, the appearance of phishing emails and websites should be similar to legitimate ones. Copying the email and the webpage code from a major site is the most common ploy for phishing attacks. The original links in the legitimate email are replaced by links which redirect users to a replica page that appears to be a part of the company's website.

The figure is a screen capture of a phishing email in which it tries to gain trust from the users by pretending the real eBay email. The link leads to a replica webpage of eBay login page that was used to trick users into submitting their private information. On clicking the link, a pop up window is created and it masks its identity so that the address appeared legitimate.[3]



Urgent Fraud Prevention Group Notice

You have received this email because we have strong reason to believe that your eBay account had bee recently compromised. In order to prevent any fraudulent activity from occurring we are required to open investigation into this matter. To speed up this process, you are required to verify your eBay account by the link below.

https://scgi.ebay.com/eBayISAPI.dll?MfcISAPICommand=VerifyId
(To complete the verification process you must fill in all the required fields)

Please Note: If your account informations are not updated within the next 72 hours, then we will assun account is fraudulent and will be suspended. We apologize for this inconvenience, but the purpose of th verification is to ensure that your eBay account has not been fraudulently used and to combat fraud.

We appreciate your support and understanding, as we work together to keep eBay a safe place to trad

Thank you for your attention on this serious matter. We apologize for any delay in resolving this situatic

Regards,

Morris Franklin
eBay SafeHarbor
Investigations Team

Please do not reply to this e-mail as this is only a notification. Mail sent to this address cannot be answered.
eBay treats your personal information with the utmost care, and our Privacy Policy is designed to protect you and your infor
Copyright © 2004 eBay Inc. All Rights Reserved.
Designated trademarks and brands are the property of their respective owners.
eBay and the eBay logo are trademarks of eBay Inc.
eBay is located at 2145 Hamilton Avenue, San Jose, CA 95125.

1.3 Anti-phishing System

Nowadays, many anti-phishing systems have been implemented in many forms such as features embedded in browsers, as a part of website login procedures and as extensions or toolbars for browsers. These anti-phishing system help users to identify legitimate websites, alert users to phishing websites, eliminating phishing email, etc. Phishing Web pages generally use similar page layouts, styles (font families, sizes, and so on), key regions, and blocks to mimic genuine pages in an effort to convince Internet users to divulge personal information, such as bank account numbers and passwords. To confront those challenges, we, in this project, are developing an anti-phishing strategy that uses a visual approach to detect bogus Web pages. Given that most phishing attacks are initiated via email, our system is designed to run on mail servers and monitor and analyse both incoming and outgoing messages for potential phishing URLs.

2. LITERATURE REVIEW

2.1. Statistical Highlights

According to a description of phishing by APWG, the ways phishers steal consumers' personal information consist of social engineering and technical subterfuge. In technical-subterfuge schemes, phishers furtively plant crime-ware onto users' computers to intercept their online account user names and passwords, while in social-engineering schemes they send spoofed e-mails to consumers purporting to be from legitimate businesses and agencies, and then mislead consumers to counterfeit websites. In addition, according to a study by Gartner, 57 million US Internet users have received e-mails that linked to phishing scams and about 2 million of them claimed to have been tricked into leaking their sensitive information. A serious problem that consistently confuses ordinary Internet users is: Does the URL I have received by e-mail or other avenues link to a phishing page, if so, which website is the phishing target it attacks? Quite a few researchers have been engaged in anti-phishing research and a lot of solutions have been developed to detect whether a webpage is a phishing page or not. However, we have not seen any technical solution which can automatically find the phishing target. This is because it is very difficult for a machine to automatically discover the possible phishing target of any suspicious webpage, although it is easier for a human being.

According to Anti-Phishing Working Group(APWG)'s[4] estimation, 5% of all recipients of a particular phishing attack have provided personal data to the phishers. It means that if a single phishing email is spammed to 1,000,000 users , nearly 50000 of those recipients have submitted their bank, credit card account or other personal information to the phishers. Besides it is estimated that the phishing attack cost US$137 million to US$1.2 billion in one year. In 2008, HTTP port-80 continuous to be the most popular used of all phishing sites reported. Port-80 99.48%; Port-443 .20%.Financial services continue to be the most targeted sector during 2008 according to the reports. Financial service- 92.9%; Retail 1.4%; LSPS 1.4% . In 2008, United States remained the top country hosting sites due to a majority of attacks being targeted toward United States based companies. United States - 38.23%; Russia - 10.58%; France - 6.35% ; Germany - 4.71%;UK - 4.49%; India - 1%. As a result, how to effectively and efficiently discover the phishing target of a phishing webpage is a great challenge for anti-phishing, which will be addressed in this paper.

2.1.1. APWG-Phishing Activity Trends Summary (2009-2010)

The Anti-Phishing Working Group (APWG)[5] is the global pan industrial and law enforcement association focused on eliminating the fraud and identity theft that result from phishing, pharming and email spoofing of all types. Payment Services are the most targeted industry sector after

Financial Services held top position during 2009 as shown in the following figure. However, the category of ,Other rose from 13 percent to nearly 18 percent from Q4 2009 to Q1 2010, an increase of nearly 38 percent. The increase in the 'Other' category is attributed to the sharp increase in attacks against the online classifieds, social networking and gaming industries. The United States continued its position as the top country hosting phishing sites during the first quarter of 2010 with China maintaining a top three listing during the three month period.

2.2. How phishing attacks?

2.2.1 . PHISHING ATTACK ON DIFFERENT STAKEHOLDERS

Phishing involves many stakeholders, including consumers, financial institutions, online merchants, Internet Service Providers (ISPs), mail client and web browser vendors and law enforcement. Stakeholders are classified into the following categories: primary victims, infrastructure providers, for-profit protectors, and public protectors. The table below describes these stakeholders and their roles.

Primary victims

In most cases, consumers, organizations, financial institutions, and merchants are direct targets of phishing attacks. Each of them is negatively affected by phishing in a different way. For example, consumers who fall for phishing can potentially become victims of identity theft: they not only suffer monetary loss, but also psychological costs (e.g. fear, anxiety). Organizations such as the military and corporations worry that phishing may lead to further compromise of credentials that can be used to steal key intellectual property or conduct corporate espionage. Financial institutions lose money from fraud conducted with credentials acquired through phishing, and merchants lose money because these financial institutions eventually charge them for the fraudulent transactions. In general, these entities are most impacted by phishing, and have the strongest incentive to protect against phishing.

Categories	Example of Key stake holder	Roles
Consumer	----------	Primary Victims
Organization	Military, Universities, Corporations	
Financial Institution	Bank of America, Paypal	
Merchants	eBay, Amazon	
Registrars & Registries	GoDaddy, Verisign	Infrastructure Providers
ISP	AT & T, Comcast	
Email providers	Gmail, Yahoomail	
Browsers	Internet Explorer, Firefox	
Software Vendors	Symantec, RSA, Mark Monitor	For-Profit Protectors
Law Enforcement	FBI, Secret Service	Public Protectors
Computer Emergency Response Team	CERT-CC, CSIRTs	

Table 1 Different Stake holders with role

Infrastructure providers

Internet service providers, email providers, browsers, domain name registrars, and registries are infrastructure providers. In most cases, phishers do not go after these providers for their money; instead, they seek to gain access to the entities' infrastructures so that phishers may launch their attacks. For example, phishers register fake domain names with registrars. Phishers use compromised machines from Internet Service Providers as part of a botnet to launch phishing campaigns, sending emails to end user mailboxes or compromising mail provider accounts to send phishing email. These stake holders are important to study as they are in a better position than most victims to protect against phishing. However some infrastructure providers do not lose money from phishing, so they may not have sufficient incentives to devote resources to combating phishing.

For-profit protectors

Certain organizations actually benefit from phishing because it is an opportunity to develop and sell products to other stakeholders. These include companies that sell spam filters and anti-virus software, as well as companies that take down phishing websites. As they are the front-line defenders against phishing, we selected a few of our experts from these companies. However, as they make money from combating phishing, it could somewhat bias their recommendations.

Public protectors

In contrast to anti-virus vendors and spam filter companies who are for-profit protectors, law enforcement, computer emergency response teams (CERT), and academics are public protectors. There are some Para organizations such as the Anti- Phishing Working Group (APWG) and the Message Anti-Abuse Working Group (MAAWG) that aim to bring different stakeholders together to fight more effectively against phishing.

2.2.2. HUMAN FACTORS INSISTING FOR PHISHING ATTACK

2.2.2.1 Lack of Knowledge

a) Lack of computer system knowledge

Many users lack the underlying knowledge of how operating systems, applications, email and the web work and how to distinguish among these. Phishing sites exploit this lack of knowledge in several ways. For example, some users do not understand the meaning or the syntax of domain names and cannot distinguish legitimate versus fraudulent URLs (e.g., they may think www.ebaymembers-security.com belongs to www.ebay.com). Another attack strategy forges the email header; many users do not have the skills to distinguish forged from legitimate headers.

b) Lack of knowledge of security and security indicators

Many users do not understand security indicators. For example, many users do not know that a closed padlock icon in the browser indicates that the page they are viewing was delivered securely by SSL. Even if they understand the meaning of that icon, users can be fooled by its placement within the body of a web page(this confusion is not aided by the fact that competing browsers use different icons and place them in different parts of their display). More generally, users may not be aware that padlock icons appear in the browser "chrome" (the interface constructed by the browser around a web page, e.g., toolbars, windows, address bar, status bar) only under specific conditions (i.e., when SSL is used), while icons in the content of the web page can be placed there arbitrarily by designers (or by phishers) to induce trust. Attackers can also exploit users' lack of understanding of the verification process for SSL certificates. Most users do not know how to check SSL certificates in the browser or understand the information presented in a certificate. In one spoofing strategy, a rogue site displays a certificate authority's (CA) trust seal that links to a CA webpage. This webpage provides an English language description and verification of the legitimate site's certificate. Only the most informed and diligent users would know to check that the URL of the originating site and the legitimate site described by the CA match.

2.2.2.2 Visual Deception

Phishers use visual deception tricks to mimic legitimate text, images and windows. Even users with the knowledge described in (1) above may be deceived by these.

a) Visually deceptive text

Users may be fooled by the syntax of a domain name in "type jacking" attacks, which substitute letters that may go unnoticed (e.g. www.paypai.com uses a lowercase "i" which looks similar to the letter "l", and www.paypa1.com substitutes the number "1" for the letter "l"). Phishers have also taken advantage of nonprinting characters[6] and non-ASCII Unicode characters[7] in domain names.

b) Images masking underlying text

One common technique used by phishers is to use an image of a legitimate hyperlink. The image itself serves as a hyperlink to a different rogue site.

c) Images mimicking windows

Phishers use images in the content of a web page that mimic browser windows or dialog windows. Because the image looks exactly like a real window, a user can be fooled unless he tries to move or resize the image.

d) Windows masking underlying windows

A common phishing technique is to place an illegitimate browser window on top of, or next to, a legitimate window. If they have the same look and feel, users may mistakenly believe that both windows are from the same source, regardless of variations in address or security indicators. In

the worst case, a user may not even notice that a second window exists (browsers that allow borderless pop-up windows aggravate the problem).

e) Deceptive look and feel

If images and logos are copied perfectly, sometimes the only cues that are available to the user are the tone of the language, misspellings or other signs of unprofessional design. If the phishing site closely mimics the target site, the only cue to the user might be the type and quantity of requested personal information.

2.2.2.3 Bounded Attention

Even if users have the knowledge described in (1) above, and can detect visual deception described in (2) above they may still be deceived if they fail to notice security indicators (or their absence).

a) Lack of attention to security indicators

Security is often a secondary goal. When users are focused on their primary tasks, they may not notice security indicators or read warning messages. The image hyperlink spoof described in (2b) above would thwarted if user noticed the URL in the status bar did not match the hyperlink image, but this requires a high degree of attention. Users who know to look for an SSL closed-padlock icon may simply scan for the presence of a padlock icon regardless of position and thus be fooled by an icon appearing in the body of a web page.

b) Lack of attention to the absence of security indicators

Users do not reliably notice the absence of a security indicator. The Firefox browser shows SSL protected pages with four indicators. It shows none of these indicators for pages not protected by SSL. Many users do not notice the absence of an indicator, and it is sometimes possible to insert a spoofed image of an indicator where one does not exist.

In nutshell, in order to attract victims to submit their information to those phishing websites, the appearance of phishing emails and websites should be similar to legitimate ones . Copying the email and the webpage code from a major site is the most common ploy for phishing attacks. The original links in the legitimate email are replaced by links which redirect users to a replica page that appears to be a part of the company's website.

2.3 . SCENARIO ANALYSIS OF PHISHING ATTACK

Here are some examples of how phishing is used. In January 2009 Bryan Rutberg was tricked into providing the password to his Face book account. He was likely the victim of a spear phishing attack (Spear phishing is targeted communication toward employees or members of a certain organization or online group. Emails are customized with information publicly available on web sites like Face book or MySpace. The emails then direct people to a fake login page). Rutberg suspects that he responded to an email that asked him to click on a link to his Face book account. When he clicked on the link he was actually taken to a fake web page that looked like Face book where he entered his username and password. The attacker then took over Rutberg's account and sent messages telling his friends that he had been robbed and asking them to send money to Western Union's branch in London. Thinking Rutberg was in need of cash, his friend sent the money. Rutberg's friend was an indirect victim of phishing and a direct victim of a scam similar to the "Nigerian" or 419 scam.

These scams are directed to "reliable and trustworthy" people. Computer users often use the same user name and password to access a number of websites, including banking, credit card and PayPal accounts. In a variation of the example above, using the same username and password, the attacker is able to access and transfer money from the user's bank account to an intermediary's account - a money mule - who forwards the funds out of their own account to the attacker who is located in another country. The money mule gets to keep a percentage of the money as a commission. The money mule performs this money laundering task either unwittingly or as an accomplice. In another attack, thousands of bogus subpoenas from the U.S. District Court in San Diego were "served" by email in corporate executives. The email contained an image of the official seal from the court and contained a link, supposedly to download a copy of the entire subpoena. However, when a recipient clicked on the link, key-logging software was installed on the user's computer instead. This is called "whaling" attack (Whaling is phishing that is targeted at corporate executives, affluent people and other "big phish.", Like spear phishing, whaling emails often are customized with information directed to the recipient (name and other personal information) and sent to a relatively small group of people). The Increasing Complexity of Malware Phishing is increasingly perpetrated with the use of specially designed malicious code— "malware." This custom code comes in the form of worms, viruses, Trojan horses, spyware, key loggers

and other routines that are designed to perform a variety of tasks. Some phishing malware propagates as viruses (code that spreads itself by infecting other programs) or as worms (self-spreading computer programs). These programs create an army of "zombie computers" that are centrally controlled as part of a "botnet" with the goal of "monetizing" the control over the infected systems - to turn such control into a source of revenue for the phishers. Another means of attack is to compromise the web server and provide malicious code that is delivered via the legitimate (although compromised) server itself. This is referred to as cross-site scripting (XSS). When the victim visits the page, he or she is presented with content that has been 'injected' into the page through XSS. The script runs on the client machine and sends personal data to the attacker.

The top 5 phishing target sites are [16]

The Top 5 Phishing Target Sites

Sites	Number of Records	CR	FNR	FPR
eBay	701	96.8%	0.0%	0.1%
PayPal	632	97.7%	0.0%	0.1%
Marshall & Ilsley Bank	138	97.7%	0.0%	0.1%
Charter One Bank	116	98.0%	0.0%	0.1%
Bank of America	51	95.4%	2.0%	2.1%

Total Number of Phishing Target Pages: 300 pages in 74 sites.
CR: Correct Rate; FNR: False Negative Rate; FPR: False Positive Rate

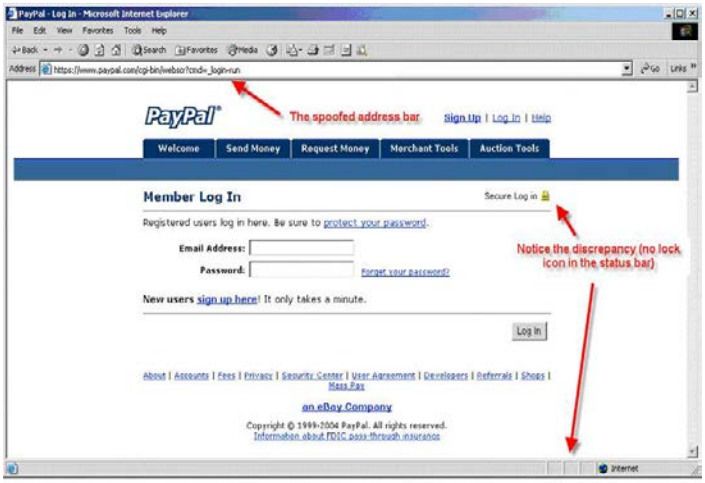
2.4. PHISHING ATTACK STRATEGIES

2.4.1 Phishing Attack Using Internet Access

Most employees browse the web for personal reasons, such as on-line shopping or research, at some time. Personal browsing may bring employees, and therefore the company computer systems, into contact with generic social engineers who will then use the staff in an effort to gain access to the company resources. The two most common methods of enticing a user to click a button inside a dialog box are by warning of a problem, such as displaying a realistic operating system or application error message, or by offering additional services.

The following Figure shows how a hyperlink appears to link to a secure PayPal website (https), while the status bar does not show anything that indicates for sure that it will take the user to a hacker's site. A hacker can suppress or reformat the status bar information.

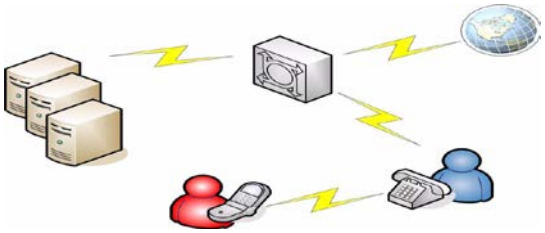
Figure: Web Page Phishing hyperlink



2.4.2 Phishing Attack using Phone Access

The telephone offers a unique attack vector for social engineering hackers. It is a familiar medium, but it is also impersonal, because the target cannot see the hacker. Phone phishing hacking is not considered to be a major threat. However, as more businesses embrace this technology, phone phishing is set to become as widespread as e-mail and website phishing is now.

The most common approach is for the hacker to pretend to be the IT supervisor or outsource IT support engineer, requesting in a hurry all passwords and authenticated credentials to analyze and resolve the claimed problems reported to him, as shown in the following Figure .



Requests for information or access over the telephone are a relatively risk-free form of attack. If the target becomes suspicious or refuses to comply with a request, the hacker can simply hang up. But it should be noted that such attacks are more sophisticated than a hacker simply calling a company and asking for a user ID and password. The hacker usually presents a scenario, asking for or offering help, before the request for personal or business information is made (Business Security Guidance, 2006).

2.5. Phishing Techniques

In a typical attack, the phisher sends a large number of spoofed (i.e. fake) e-mails to random Internet users that seem to be coming from a legitimate and well-known

business organization (e.g. financial institutions, credit card companies, etc).

A. Basic URL Obfuscation Ref :

URL obfuscation [9] misleads the victims into thinking that a link and/or web site displayed in their web browser or HTML capable email client is that of a trusted site. These methods tend to be technically simple yet highly effective, and are still used to some extent in phishing emails today.

1. Simple HTML redirection

One of the simplest techniques for obscuring the actual destination of a hyperlink is to use a legitimate URL within an anchor element but have its href attribute point to a malicious site. Thus clicking on a legitimate-looking URL actually sends the user to a phishing site.

2. Use of JPEG images

Electronic mail rendered in HTML format is becoming more prevalent. Phishers are taking advantage of this by constructing phishing emails that contain a single image in JPEG format. When displayed, this image appears to be legitimate email from an online bank or merchant site. The image often includes official logos and text to add to the deception. However, when users click on this image, they are directed to a phishing site.

3. Use of alternate encoding schemes

Hostnames and IP addresses can be represented in alternate formats that are less likely to be recognizable to most people. Alphanumeric characters can be changed to their hexadecimal representations.

4. Registration of similar domain names

At initial glance, users may attempt to verify that the address displayed in the address or status bar of their web browser is the one for a legitimate site. Phishers often register domain names that contain the name of their target institution to trick customers who are satisfied by just seeing a legitimate name appear in a URL.A widely implemented version of this attack uses parts of a legitimate URL to form a new domain name as demonstrated below:

Legitimate URL http://login.example.com

Malicious URL http://login-example.com

B. Web Browser Spoofing Vulnerabilities

Over the past two years, several vulnerabilities in web browsers have provided phishers with the ability to obfuscate URLs and /or install malware on victim machines.

1. International Domain Names (IDN) Abuse:

International Domain Names in Applications (IDNA) is a mechanism by which domain names with Unicode characters can be supported in the ASCII format used by the existing DNS infrastructure. IDNA uses an encoding syntax called puny code to represent Unicode characters in ASCII format. A web browser that supports IDNA would interpret this syntax to display the Unicode characters when appropriate. Users of web browsers that support IDNA could be susceptible to phishing via homograph attacks, where an attacker could register a domain that contains a Unicode character that appears identical to an ASCII character in a legitimate site (for example, a site containing the word “bank” that uses the Cyrillic character “a” instead of the ASCII “a”).

2. Web Browser Cross-Zone Vulnerabilities:

Most web browsers implement the concept of security zones, where the security settings of a web browser can vary based on the location of the web page being viewed. We have observed phishing emails that attempt to lure users to a web site attempting to install spyware and/or malware onto the victim’s computer. These web sites usually rely on vulnerabilities in web browsers to install and execute programs on a victim’s computer, even when these sites are located in a security zone that is not trusted and normally would not allow those actions.

C. Specialized Malware

Over the past two years, there has been an emergence of malware being used for criminal activity against users of online banking and commerce sites. This type of specialized malware (which can be considered a class of spyware) greatly increases the potential return on investment for criminals, providing them with the ability to target information for as many or as few sites as they wish. One benefit for criminals is that most malware can easily be recognise to change targeted sites and add new ones. Malware also provides several mechanisms for stealing data that improve the potential for successfully compromising sensitive information.

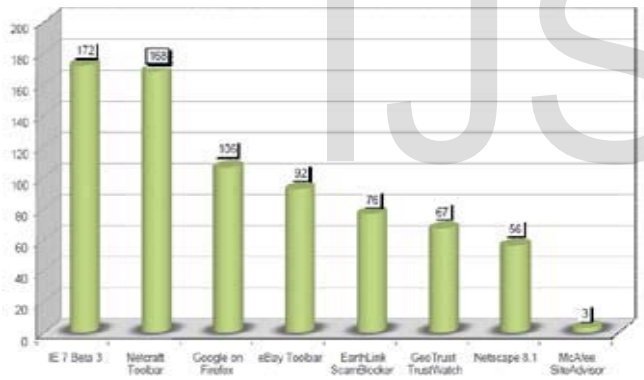
3.6 Anti-Phishing Techniques:

A. Email-Level Approach

It includes authentication and content filtering. The email filtering techniques,[10]commonly used to prevent phishing. These are quite popular in anti-spam solutions because they try to stop email scams from reaching target users by analyzing email contents. Phishing messages are usually sent as spoofed emails; therefore, researchers have proposed numerous path-based verification methods. Current mechanisms, such as Microsoft's Sender ID or Yahoo's Domain Key, are designed by looking up mail sources in DNS tables. The challenge in designing such techniques lies in how to construct efficient filter rules and simultaneously reduce the probability of false alarms.

B. Browser Integrated Tool Approach

A browser-integrated tool [11][12] usually relies on a blacklist containing the URLs of malicious sites to determine whether a URL corresponds to a phishing page. In Microsoft Internet Explorer (IE) 7, for example, the address bar turns red when a malicious page loads. Fig: Composite Accuracy Score Result:



A blacklist's effectiveness is strongly influenced by its coverage, credibility, and update frequency .Currently, the most well-known blacklists are those Google and Microsoft maintain for the popular browsers Mozilla Firefox and IE, respectively. Fig. 4 shows the accuracy of various toolbars. However, experiments show that neither database can achieve a correct detection rate greater than 90 percent, and the worst-case scenario can be less than 60 percent.

C. Webpage Content Analysis

It analyzes a Web page's content [12], such as the HTML code, text, input fields, forms, links, and images. In the past, such content-based approaches proved effective in detecting phishing pages. Phishers responded by compiling pages with non-HTML components, such as images, Flash objects, and Java applets. A phisher might design a fake page composed entirely of images, even if the original page contains only text information. In this case, content-based anti-phishing tools can't analyze the suspect page because its HTML code contains nothing but HTML <img/> elements.

D. Visual similarity based analysis

This solution[13][14] involves detecting phishing pages based on the similarity between the phishing and authentic pages at the visual appearance level, rather than using text-based analysis. An important feature of a phishing webpage is its visual similarity to its target (true) webpage. Hence, a legitimate webpage owner or its agent can detect suspicious URLs and compare the corresponding WebPages with the true one in visual aspects. If the visual similarity of a webpage to the true webpage is high, the owner will be alerted and can then take whatever actions to immediately prevent potential phishing attacks and hence protect its brand and reputation. This module extracts the Web pages' features and measures the similarity to the true pages according to three metrics: block-level, layout, and style. If the visual similarity is higher than the corresponding threshold, the system issues a phishing report to the customer. However, this approach is susceptible to significant changes in the Web page's aspect ratio and important colours used.

2.6 Some well-developed anti-phishing systems

2.6.1 Microsoft Phishing Filter

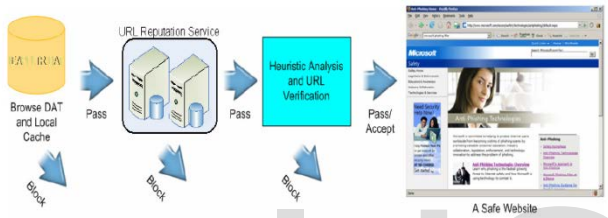
Microsoft Corporation has introduced the Microsoft ® Phishing Filter to block websites and caution users about reputed and suspected phishing websites.[15] Currently, the phishing filter is embedded in the Microsoft products only. Internet Explorer 7, MSN ® Search Toolbar and MSN ® Hotmail are now protected by the Microsoft Phishing Filter. It uses several patent-pending technologies designed to warn or block users from potentially harmful websites:

Several Layers of phishing detections in the phishing filter Reporting system for users to report and for collecting phishing attacks Jeremy Dailman, an IE program manager, expressed that "it was the only one that consistently caught

more than 60% of phishing sites while having the lowest possible rates of incorrect ratings (otherwise known as false positives)" [16]. According to Microsoft, over 1,000,000 suspected and phishing websites are blocked each week, and over 10,000 phishing websites are added to blacklists every week in which it can provide quick response in blocking the known phishing websites.

2.6.2 How to Detect Phishing Attack?
According to a report written by Jefferson, Microsoft Phishing Filter detects phishing attacks in 3 processes.[17] The following Figure shows the phishing detections in the Microsoft Phishing Filter. Each process has different strategy to detect and filter specific type of phishing attacks.

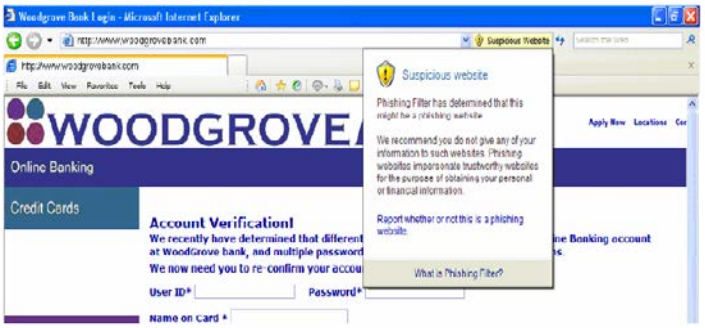
Phishing detections in Microsoft Phishing Filter



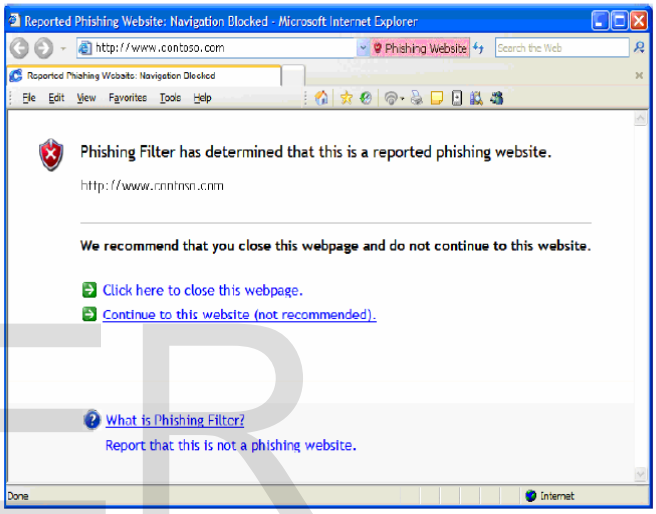
Process 1 : Browser DAT and URL Local Cache

The Microsoft Phishing Filter first checks the .DAT file and cache in the local machine to find a match of a previously rated phishing site when a user enters a website using Browser Technology embedded with Phishing Filter such as Internet Explorer 7 and Windows Live Toolbar. As it is the first checking, it simply checks the domain and path of the URL (i.e. http//domain.com/path) with the query string data removed. If the URL matches the rated phishing site in the .DAT file and cache, a yellow warning or a red warning will be raised depending on the phishing rating of the URL to alert the user about phishing attacks. A yellow warning means that the user has entered a suspected phishing website and it is recommended to avoid entering any personal information on the website. Figure 3.2 shows the yellow warning raised by the phishing filter in the Internet Explorer 7. A red warning means that the user has entered a confirmed phishing website. A threat level warning page is raised to navigate the user to a new page. This warning page offers users the option to close the webpage immediately or proceed at their own risk to the website. Figure 3.3 shows the red warning raised by the phishing filter in the Internet Explorer 7. If the Browser DAT file and URL Local cache do not contain the phishing ratings of the URL, the phishing detection will move on to the URL Reputation Web Server.

Yellow Warning raised in Internet Explorer 7:



Red Warning raised in Internet Explorer 7:



Process 2: URL Reputation Web Server
When the first process, Browser DAT file and URL Local Cache, cannot judge for the phishing attacks, the Phishing Filter in the local machine tries to match a rated phishing list of URLs stored on the URL Reputation Web Server hosted by MSN. If the URL matches the list, a yellow or red warning will be raised depending on rating of the website. Figure shows the yellow warning and red warning raised by the phishing filter in MSN Search Toolbar. If the match is not identified, the phishing detection will move on the next detection process, the Heuristic Analysis and URL Verification.

Yellow warning raised in MSN search toolbar:

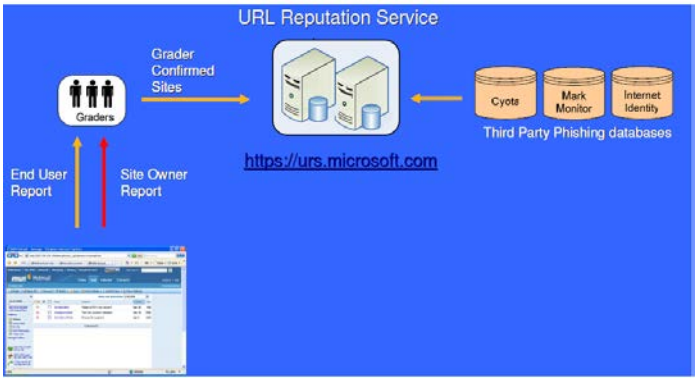
Red warning raised in MSN search Toolbar:

Process 3: Heuristic Analysis and Process Verification:

If URL Reputation Web Server still cannot judge for the URL, the webpage is subjected to a heuristic analysis by the Phishing Filter. Based on the heuristic analysis and the URL verification, the Phishing Filter will calculate the phishing rating for the URL. If the rating exceeds a limit, a yellow or red warning will be raised to alert user depending on the rating of the URL. Otherwise, the webpage continues as normal.

2.6.1.2 How to Collect Phishing Data?

The Microsoft Phishing Filter collect phishing data by 2 ways: A team of Graders from Microsoft confirms the phishing reports from end users and site owners and inserts these data to the URL reputation web server; Phishing database from third parties also provide phishing data to the URL reputation web server.[18] The following figure shows the collection of phishing data.



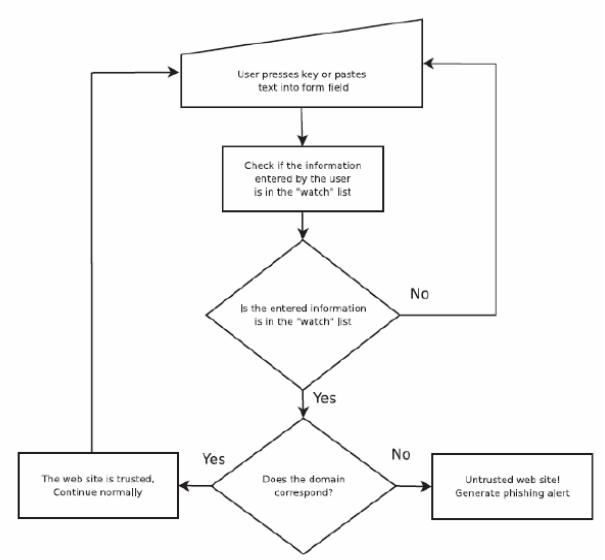
2.6.2 Mozilla AntiPhish

Mozilla Firefox 2.0 has introduced the AntiPhish, a Firefox extension for anti-phishing support. This AntiPhish simply keeps track of users' sensitive information and prevents this information from being passed to a website that cannot be trusted.[19] Although AntiPhish only supports Firefox 2.0, the performance in fighting against phishing attacks is outstanding. A recent study [20] commissioned by Mozilla in 2006 claimed that the anti-phish in Firefox 2.0 had a higher accuracy in detecting phishing websites than Microsoft Internet Explorer 7. The study found that Firefox in its most secure configuration blocked 81.5% of all phishing websites. Internet Explorer 7 blocked just 66.35% of phishing websites. And also, Firefox 2.0 failed to flag a phishing websites in 117 instances when Internet Explorer 7 caught it, while the Internet Explorer 7 let 243 URLs slop through that Firefox stopped.

2.6.2.1 How to Detect Phishing Attack?

Instead of checking the characteristics of phishing attacks, AntiPhish detects use another approach to detect phishing. It uses JavaScript to check the HTML of the webpage where information is going to be submitted to determine if the sensitive information "belongs" to the domain of the website. Figure 3.7 shows how the sensitive information flow is controlled by AntiPhish [19].

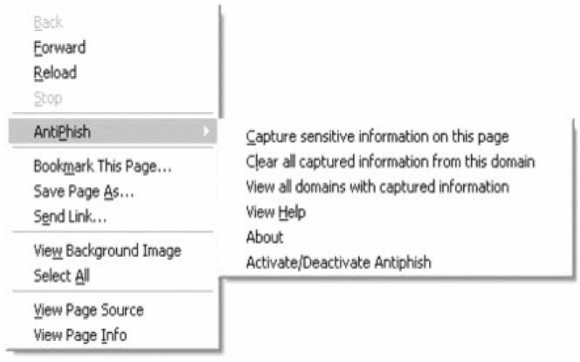
Flow Chart of how anti-phish controls the sensitive information:



1. Before Detection :

When the AntiPhish is first installed in the Firefox 2.0, a browser prompts a request for a master password when a user enters input into a form for the first time. The master password is used to encrypt the sensitive information before it is saved for the auto form-filler. The encryption and decryption are done by the symmetric DES algorithm.

2. Store Sensitive Information into the "Watch List" After the user fills in the sensitive information and submits the form, the AntiPhish scans the webpage and to capture this information. This scanning is done by manually in the current version of AntiPhish. The AntiPhish then stores the sensitive information and a mapping of where this information "belongs" to. The information and the domain name will be encrypted and stored into a "watch" list. Figure 3.8 shows the AntiPhish application menu integrated into Firefox 2.0.



The AntiPhish application menu integrated into the Firefox 2.0

2.6.2.2 Control the Sensitive Information Flow

As HTML form elements such as text field of type text and password and the HTML text area are the most likely to be used to phish information, AntiPhish starts checking the potential phishing attacks when a website contains a form and the form elements mentioned above. When the user enters any information into any of these form elements, AntiPhish checks the 'Watch List". For each value in the list that is identical to the one just entered by the user, the corresponding domain is determined. If the current website is not among these domains, AntiPhish generates an alert and redirects to an information page about phishing attacks. Figure 3.9 shows the phishing alert message box in Firefox 2.0.

Phishing alert message in the Firefox 2.0

2.6.2.3 How to Collect Phishing Data? Since the AntiPhish is designed to avoid user to submit sensitive information to "phishing" websites rather than blocking the "phishing" website directly, it is no need to store phishing information in a server for users to download.

2.6.3 Google Safe Browsing

"Responsible disclosure allows companies like Google to keep users safe by fixing vulnerabilities and resolving security concerns before they are brought to the attention of the bad guys." [21] As one of the most famous search engine in the world, Google play an important role in anti-phishing. Google safe browsing is designed to identify phishing websites in the Google Toolbar. Originally Google Safe Browsing was a Firefox extension, but it has been integrated into the Google Toolbar in the later version.

2.6.3.1 How to Detect Phishing Attack?

Among the anti-phishing system studied in this paper, Google Safe Browsing use the simplest way to detect phishing attack — the blacklist-based detection. There are 2 types of blacklist for the Google Safe Browsing: Blacklist in

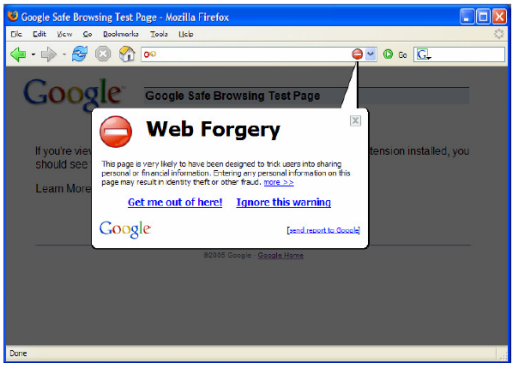
local machine and Blacklist in the Lockup Server. Both blacklists will be updated periodically to make sure that the blacklisting is effective and most updated. The advantages of using blacklist are that it is simple and easy to deploy. The Google Safe Browsing provides an "enhanced protection" option for users to select if they would like to be protected by local blacklist or remote blacklist in the look up server.

1. Phishing Detection on Local Blacklist

If enhanced protection is disabled, the user keeps a local blacklist of phishing URLs in the local machine. When the user visit websites through Firefox, every URL the browser requests will be checked if it is in the local blacklist. The list is updated timely from the Lookup Server in which it sends diffs from the user's current blacklist or a new full blacklist.

2. Phishing Detection on Lookup Server

If enhanced protection is enabled, Google Safe Browsing looks up URLs in a remote blacklist hosted on a Lookup Server. Enhanced protection is a recommended option to protect from being phished as it provides better coverage because the remote blacklist is updated every minute, which is more updated than the local blacklist. When the user visits websites through Firefox, every browser request of the URL will be checked in the remote blacklist. When a blacklisted page is loaded in the user's machine, the page is disabled and a warning will be shown to the user. The user can choose to navigate away or to continue to visit the page. The following figure shows the alert message raised by the Google Safe Browsing in the Firefox.

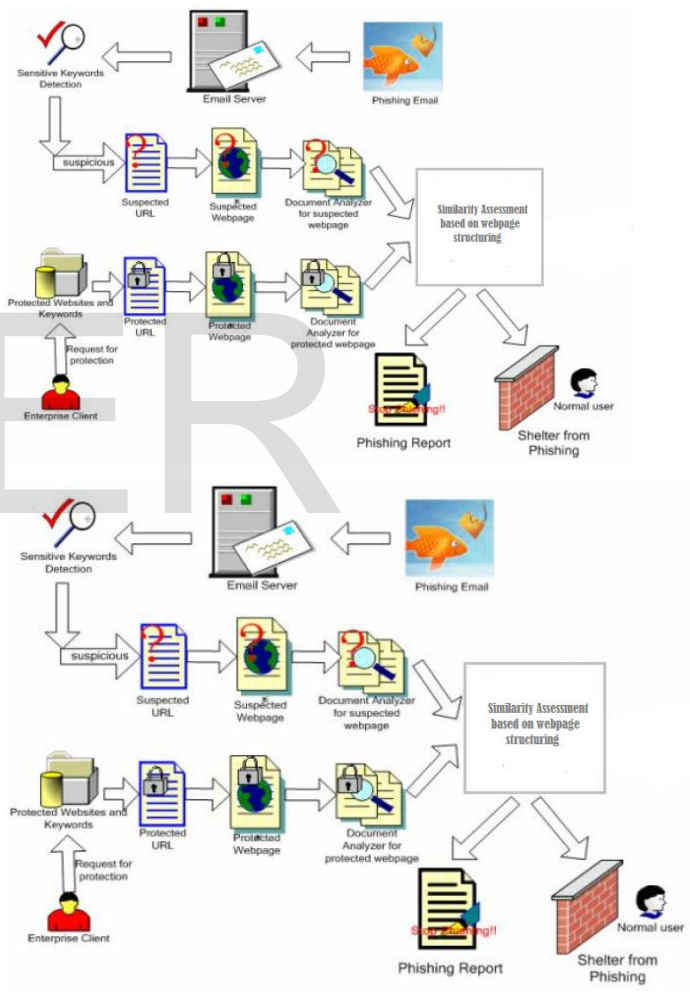


Alert message raised by Google Safe Browsing in the Firefox 2.0

2.6.3.2 How to Collection Phishing Data?

It is not known how URLs are added to the remote blacklist in the Lookup Server. According to the Google Safe Browsing for Firefox Official website, the toolbar combines "advanced algorithms with reports about misleading pages from a number of sources.[22][23] Google Safe Browsing uses blacklists as well as heuristics is the same as the phishing analysis in the Microsoft Phishing Filter. Although we do not know how Google Safe Browsing gets the phishing websites, it can be predicted that end user reporting is one way in growing the blacklist. For end user reporting, it allows users to report both false positive and false negatives. False positive means "it is a not a phishing website but it reports as phishing". False negative means "it is a phishing website but it does not report as phishing".

3. PROPOSED MODEL



Steps:-

1. A phisher sends a phishing email with an url link to an user of our email server. 2. Similar to spam checker component, we have a Anti Phishing component in Mail Server which spans a separate thread for each user.
3. As soon as a user receives a new email in his Inbox, this Anti Phishing component is triggered automatically. It scans

the new message for any links by searching either for href attribute of a tag or http or www. in the email content.
4. If a link is found it downloads content of the link which is nothing but a set of html tags (an web page)
5. Now it extracts/parses the structure of the HTML tags, by removing the contents of these tag, for that particular link and checks the protected site database which has similar structure for all protected websites like eBay. In.
6. If the structure of this link matches with HTML structure of any website in protected list then definitely this link is a copy of that protected site, and hence the href of the website in the protected list whose structure matches to that of the suspected website is compared with the href of the suspected website. If it does not match then definitely this link is trying to phish user's data and hence this components adds this link to blacklist database.
7. Next time when user accesses this link after accessing his Inbox, Since it is present in blacklist url database, user will get an error message about Phishing and will be prevented access to this website.

System Specification-

1. Hardware Requirement:

- Intel(R) Core(TM) i7-2670QM CPU @ 2.20GHz
- 6.00 GB of RAM

Note that the hardware requirements mentioned above are not minimum requirement. They are only the resources 2.provided for the project.

Software Requirement:

Our project requires

- Microsoft Visual Studio 2010 with .NET Framework with support of MYSQL.

3. Programming Language:

- With Microsoft Visual Studio 2010, the server is implemented using ASP.NET under Microsoft development environment. And also, ADO.NET provides consistent access to data sources like MYSQL database server.

Database Lists Used:

- Protected list- stores url that would be registered and protected and are safe to be accessed by the user. These urls are called protected urls. The keywords of the urls would also be stored (i.e. www.paypal.com contains keywords "PayPal" or "pay pal" or "pay" or "pal"). Apart from this, the html structures of these urls are also stored.
- Black List- stores "known" phishing url till date. These urls are called Black urls.

Current Work:

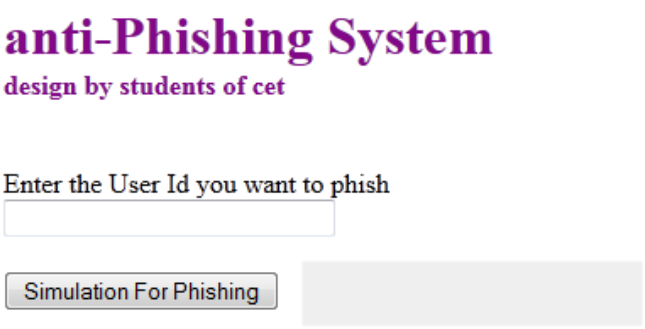
We have designed an email server named as "an anti-phishing system". It serves within an intranet network consisting of number of users connected to it in such a way that any link that the user tries to connect will have to be linked via the anti-phishing system created by so that it can check and differentiate the phishing websites. As a part of our simulation, we have created the necessary database and the anti-phishing work is under progress. However, to just have a snapshot of what the result will be like, we have implemented the anti-phishing by comparing only the href of the suspected and protected URL.

Simulation:

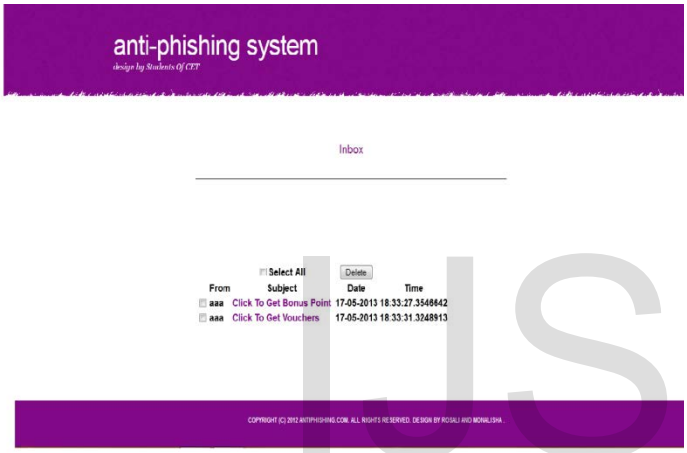
We have designed the following mail server.

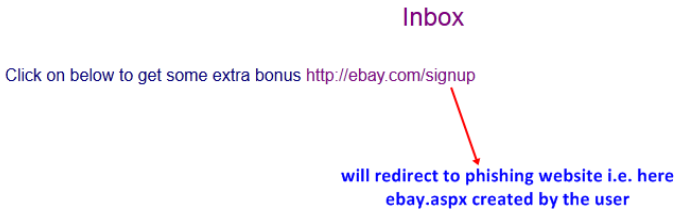
In order to show how phishing exactly occurs,as per our model, we have included the following by taking the example of ebay site. The simulation screen for showing

phishing is as

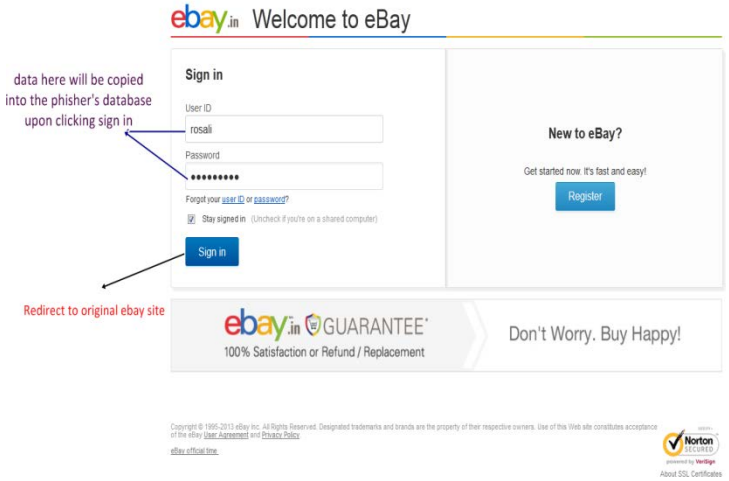


In the textbox, the user(the one who is trying to phish other users in its mail server) has to enter the user id of the user he wants to phish. Upon clicking the button 'Simulation For Phishing', an email is automatically send to the victim's inbox. The victim's inbox will have the following mails:

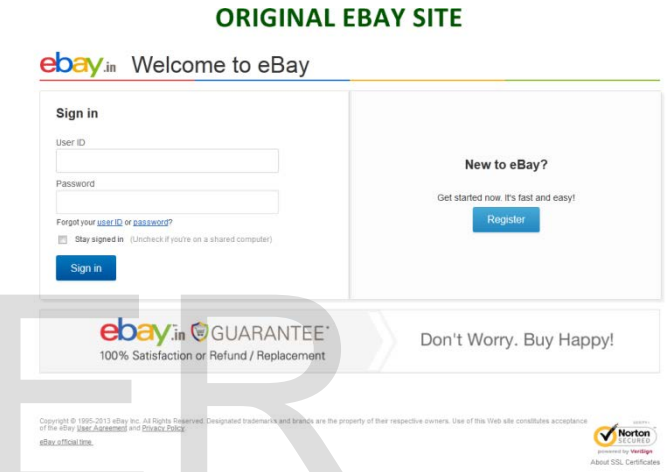




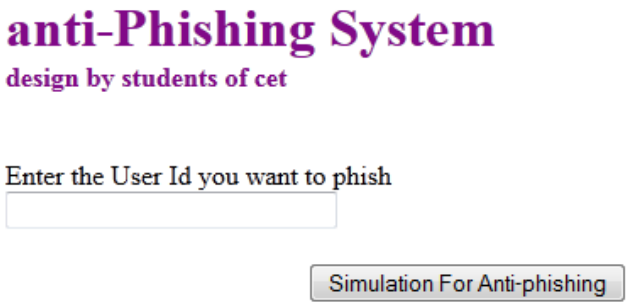
Upon clicking the subject i.e. ''Click to Get Bonus Point" the following message is displayed which is be having a link, say http://ebay.com/signup. On clicking this link, rather than redirecting to original eBay site, it gets redirected to a webpage similar to that of the original eBay, generated by the phisher to fool the user. Upon clicking 'sign up' button, after entering the user name and password, the relevant data is copied to the phisher's database and the user is now redirected to the original eBay site. Thus the phisher gets the user name and password of the victim and hence would misuse this. The snapshot of the phishing site after clicking http://ebay.com/signup will be as



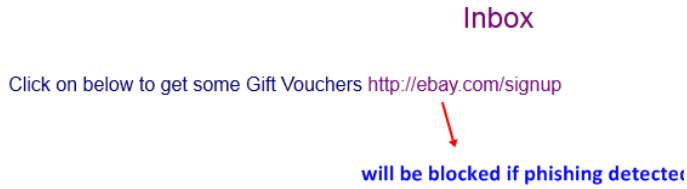
The original eBay site is as:



This is how phishing occurs. In order to check this, we have designed an anti-phish system. The simulation screen for anti-phishing system is as:



As described above, an email is send to the victim upon clicking 'Simulation For Anti-phishing' button. The email is as:



On clicking the link, the html structure of the webpage is automatically extracted and compared to those stored in our protected database. If it matches to any of the webpage in the protected list, then it might be so that someone might be trying to phish. Hence the href of both the suspicious link and the protected link are compared, if it matches then it is the original site and is safe to be accessed. If phishing is detected then the victim gets an alert message and is prohibited from accessing that very phishing site and the url of this link is copied into the black list for future detection so that on encountering this same link again, the site will be automatically blocked before even comparing the html structure. The alert message is as:

CONCLUSION AND FUTURE WORK:

In this paper, our main contributions include a new way of discovering the phishing target of a given phishing webpage, which is more significant than only identifying a given suspicious webpage as phishing or not in previous work.

This project design is subjected to certain drawbacks which we are trying to eradicate. This design may experience Denial-of-Service (DoS) attack in future. DoS is a computer crimes which violates the Internet proper user policy. DoS floods the network and prevent valid user to access the server. Again our system is designed to detect phishing only in html code level. It cannot detect phishing when the websites are coded in other languages like flash, java applet. However this drawback can be overcome by implementing the same technique using html parsing in java.

Acknowledgements

We would like to thank our teacher for his great efforts of supervising and leading us, to accomplish this fine work. To our friends and families, they were a great source of support and encouragement. We thank every person who gave us something to light our pathway; we thank them for believing in us.

---

[1]*Rosali Pujapanda is currently pursuing bachelor's degree program (2009-2013) in computer science and engineering in College of Engineering and Technology (BPUT), Odisha, India. E-mail: rpujapanda2@gmail.com.*

[2]*Monalisha Parida is currently pursuing bachelor's degree program (2009-2013) in computer science and engineering in College of Engineering and Technology (BPUT), Odisha, India. E-mail: jolly92mona@gmail.com*

[3]*Ashis Kumar Mishra has completed his post graduate from KIIT University (2011) and under graduate from Biju Pattnaik University of Technology (2007). He is currently a faculty member in Department of CSE in College of Engineering and Technology, Bhubaneswar, Odisha, India. E-mail: ashiskumar.misra@gmail.com. He has published 9 numbers of International Journals and a National Conference in fields of Cryptography, Real Time Systems, Evolutionary algorithm, Cloud Computing and machine learning.*

References

[1].Research Paper- An Antiphishing Strategy Based on Visual Similarity Assessment, Wenyin Liu, Xiaotie Deng, Guanglin Huang, and Anthony Y. Fu, City University of Hong Kong

[2]. An Anti-phishing working group. Origins of the word "phishing"
available http://www.antiphishing.org/word_phish.html

[3].Phishing email
example, http://code.jenseng.com/createPopup/email.html

[4]. Anti-Phishing Working Group (APWG),
2008 http://www.antiphishing.org

[5]. The Anti-Phishing Working Group ,"APWG Phishing Trends Reports, [Online] Available
: www.antiphishing.org/phishReports Archive.html

[6]. Secunia, Internet Explorer URL Spoofing vulnerability (2004)

[7]. Secunia, Multiple Browsers Vulnerable to the IDN Spoofing Vulnerability(2005)

[8]. "Phishing Statistics", Secure Computing ,2007, http://www.ciphertrust.com/resources/statistics/phishing.php

[9]. Jason Milletary, "Technical Trends in Phishing Attacks", Carnegie Mellon University,2005

[10]. Sumit Siddharth, "Anti Spamming Techniques.pdf"

[11]. P. Robichaux, D.L. Ganger, "Gone Phishing: Evaluating Antiphishing Tools for Windows," 3Sharp Project Report, Sept. 2006; [Online] Available : www.3sharp.com/ projects/antiphishing/.

[12] C. Ludl et al., "On the Effectiveness of Techniques to Detect Phishing Sites", Proc. Detection of Intrusions and Malware, and Vulnerability Assessment, LNCS 4579, Springer, 2007, pp. 20–39

[13] W. Liu et al., "An Antiphishing Strategy Based on Visual Similarity Assessment", IEEE Internet Computing, vol. 10, no. 2, 2006, pp. 58–65

http://www.betanews.com/article/Microsoft_Expands IE7_Phishing_Filter/1170804311

[17].Jefferson W. (2006). Microsoft Phishing Filter Feature in Internet Explorer7 and Windows Live Toolbar -Privacy Assessment Report. Available: http://www.jeffersonwells.com/client_audit_reports/Microsoft_PF_IE7_IEToolbarFeature_Privacy_Audit20060728.pdf

[18]. Steve L., (2006) Internet Explorer7 Security Features. Available: http://download.microsoft.com/documents/uk/technet/learning/downloads/security/InternetExplorer7SecurityFeatures9Nov06.ppt

[19]. Engin, K. and Christopher, K. (2005) Protecting Users against Phishing Attacks. Oxford Journals on Computer 49-554-561

[14] L. Wenyin et al., "Detection of Phishing Webpages Based on Visual Similarity," Proc. World Wide Web Conf. (special interest tracks and posters), A. Ellis and T. Hagino, eds., ACM Press, 2005, pp. 1060–1061.

[15].Microsoft Corporation (2005) Microsoft Phishing filter : a new approach to building trust in e-commerce content [Online]. Available: http://www.microsoft.com/downloads/details.aspx?FamilyId=B4022C6699BC-4A30-9ECC-8BDEFCF0501D&displaylang=en[2005, 20 September]

[16]. Nate M. 2007. Microsoft Expands IE Phishing Filter. BetaNews [Online] Available:

[20].Mozilla (2006) Firefox 2 Phishing Protection effectiveness testing. [Online] Available: http://www.mozilla.org/security/phishing-test.html

[21]. Michael, K. (2007) Google on Security Alert. Intranet Journal [Online] Available: http://www.intranetjournal.com/articles/200701/ij_01_04_07b.html

[22]. Google Safe Browsing for Firefox, http://www.google.com/tools/firefox/safebrowsing

[23]. Lorrie, C., Serge, E., Jason, H. and Yue, Z. (2006). Phinding Phish: An Evaluation of Anti-phishing Toolbars. [Online] Available: http://www.cylab.cmu.edu/files/cmucylab06018.pdf