AntiPhishing Against Phishing Attacks

PURTI
MCA, NET QUALIFIED,FEROZEPUR,PUNJAB

Abstract: ***Phishing is a type of network attack where the attacker creates a web page to fool users. For example by creating fake login page, fake emails for knowing user details. Phishing emails contain messages that ask the users to enter the personal information. Phishing scams have been receiving extensive press coverage because such attacks have been escalating in number and sophistication. To protect users from phishing attacks system designers and security professionals need to understand how users interact with those attacks. This paper gives brief information about phishing, its attacks, steps that users can take to safeguard their confidential information. To this end, Anti-Phish tracks the sensitive information of a user and generates warnings whenever the user attempts to give away this information to a web site that is considered un-trusted.***

Keywords: Phishing, Security, Anti-phishing

I. INTRODUCTION

Phishing attacks are becoming popular these days. The attacker will get a way to find the useful information from network. Phishing is basically performed by sending fake emails or link to replica web page asking user to fill in important details.

Phishing is a kind of passive network attack. Phishing emails contain links to the phishing webpage asking us to submit our information such as bank account details, security questions which we often use to recover our account password etc. Phishing collects data which further helps the attacker to perform some hit and trials and social engineering on our accounts. Phishing emails are sent to a large number of people asking them to submit important information like bank account details. Phishing pages are designed in such a way that we are not able identify whether it is fake or genuine. Also the URL's also sound similar like www.faecbook.com like this the attacker fools the victim. Some of the ethical hackers are working on anti-phishing so that these kinds of attacks can be prevented.

Rest of the paper is organized as follows, Section I contains the introduction to Phishing attacks, Section II contains stages of Phishing attack with structure of common attack tree method, Section III contain the some measures of Anti-phishing, Section IV contains categories of Anti-Phishing, section V explain the anti-phishing techniques, Section VI concludes research work with future directions

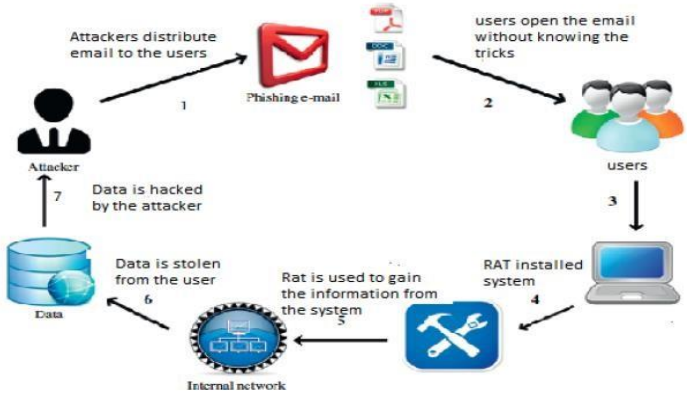

Fig 1: Phishing Attacks

II. Phishing Attack Stages

Phishing attacks involve several stages:

- The attacker takes a E-mail addresses for the intended victims. These could be guessed or obtained from Different sources.
- The attacker generates an E-mail that appears legal and requests the recipient to Perform some action.
- The attacker sends the E-mail to the intended victims in a way that appears legal and uncertain the true source.
- Depending on the content of the E-mail, the recipient opens a malicious attachment, Completes a form, or visits a web site.
- The attacker gathers the victim's sensitive information and may exploit it in the future.

There are numerous ways for the attacker to execute these steps. There are also countermeasures that intended victims can employ to prevent some of them. The attack trees below show the steps that the attacker (and victim) must take for a successful phishing attack. The trees also show ways that existing technology can be used to reduce vulnerability to phishing Attacks.

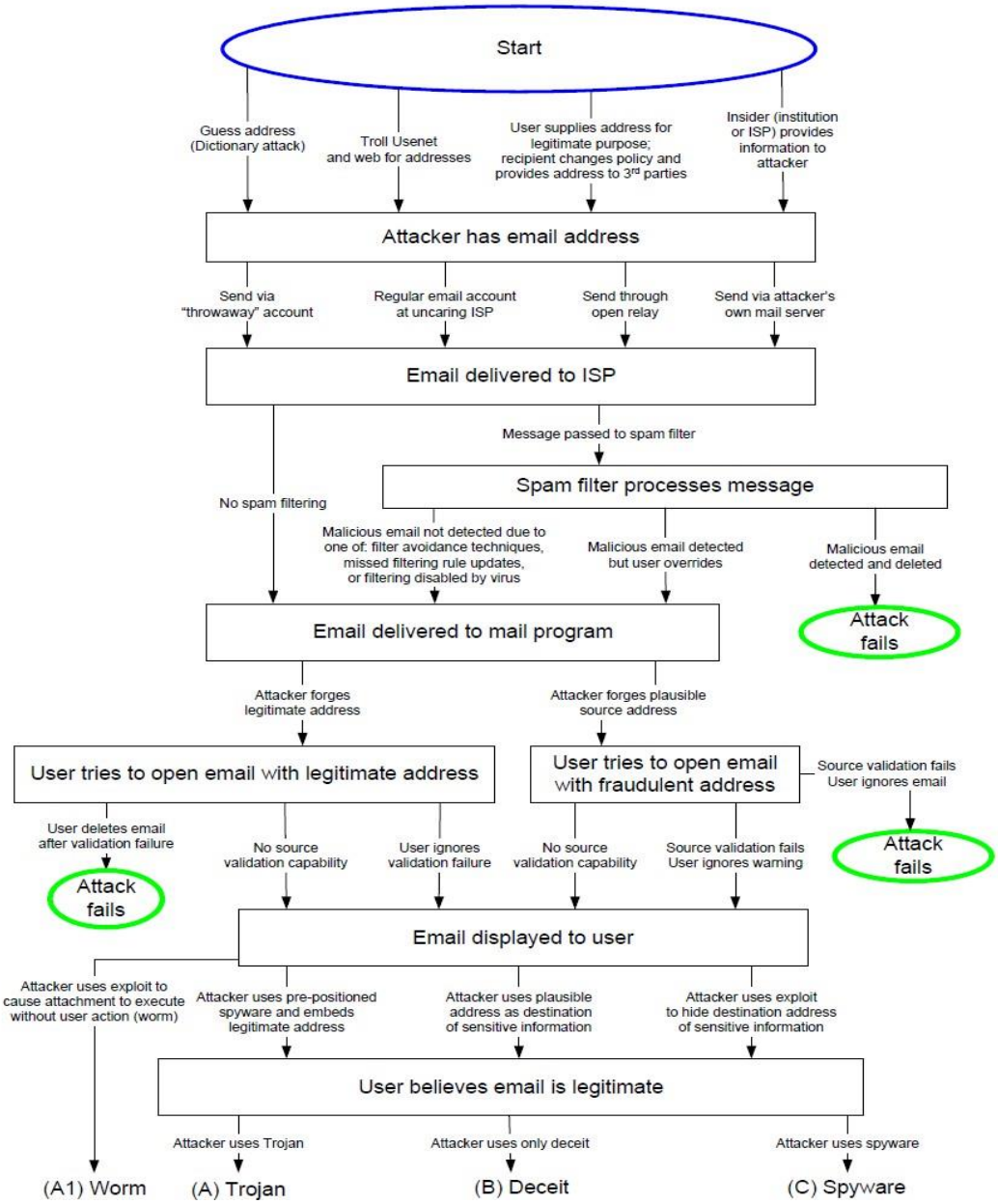

Fig 2: Common Attack Tree Methods

III. ANTI-PHISH

Anti-Phish is based on the premise that for inexperienced, technically unsophisticated users, it is better for an application to attempt to check the trustworthiness of a web site on behalf of the user. Unlike a user, an application will not be fooled by obfuscation tricks such as a similar sounding domain name.

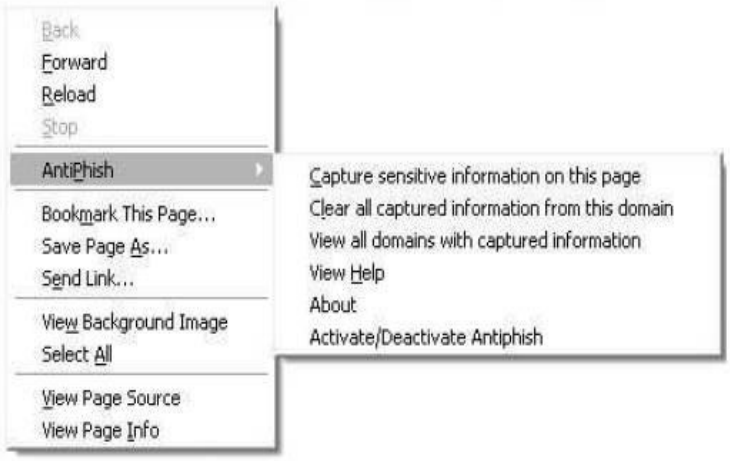

Fig 3: The Anti-Phish application menu integrated into the browser.

III.I. Main Functionality

Anti-Phish is an application that is integrated into the web browser. It keeps track of a user's sensitive information (e.g., a password) and prevents this information from being passed to a web site that is not considered "trusted" (i.e. "safe").

The development of Anti-Phish was inspired by automated form-filler applications. Most browsers such as Mozilla or the Internet Explorer have integrated functionality that allows form contents to be stored and automatically inserted if the user desires. This content is protected by a master password. Once this password is entered by the user, a login form that has previously been saved, for example, will automatically be filled by the browser whenever it is accessed. Anti-phish takes this common functionality one step further and tracks where this information is sent.

Figure 3 shows the right-click pop-up menu in the browser with the integrated Anti-Phish menu items. After Anti-Phish is installed, the browser prompts a request for a new master password when the user enters input into a form for the first time. After this password is entered, the Anti-Phish menu can be used to capture and store sensitive information. The master password is used to encrypt the sensitive information before it is stored. The symmetric DES algorithm is used for the encryption and decryption.

In our current implementation, user interaction is needed to tell Anti-Phish that a piece of information on a page is important and that it should be protected against phishing attempts. After the user enters sensitive information such as a password, the Anti-Phish menu is used to scan the page and to capture and store this information. Currently, the contents of all HTML text field elements of type password are captured and cached.

Besides storing the sensitive information, Anti-Phish also stores a mapping of where this information "belongs" to. That is, the domain of the web site where this information was originally entered is also stored. We use domains instead of web site addresses because some web sites are hosted on multiple servers with different addresses (e.g., the main web site might have the address www.ba-ca.com and based on load, the online banking service might be hosted on online1.baca. com and online2.ba-ca.com). Hence, if web server addresses or URLs are used instead of domains, false phishing alarms could be generated.

In our prototype, we provide simple dialogs for the management of stored sensitive information. The user can see a list of web site domains from which sensitive information has been captured and has the possibility of clearing this cached information[2].

If the user would like to use the same piece of sensitive information (e.g., the same password) on multiple web sites, this information has to be captured by AntiPhish for all sites where it is being used. This is typically done by first deactivating AntiPhish from the menu in order to prevent it from generating false phishing alerts.

III.II. Controlling the sensitive information flow

As far as AntiPhish is concerned, every page that contains a form is a potential phishing page. HTML form elements that can be used by the attacker to phish information from the user are text field elements of type text and password and the HTML text area element. Hence, whenever the user enters information into any of these form elements (e.g., the user presses a key or pastes text), AntiPhish checks the list of previously captured values (i.e., the "watch list"). For each value in this list that is identical to the one just entered by the user, the corresponding domain is determined. If the current site is not among these domains, a phishing attempt is assumed. The reason is that sensitive information is about to be transmitted to a site that is not explicitly listed as trusted. If AntiPhish detects, for example, that the user has typed his online banking password into a text field on a web site that is not in the online banking web site domain (i.e., an "un-trusted" web site), then it generates an alert and redirects to an information page about phishing attacks.

Interaction events that the user generates within the browser are used to intercept sensitive information flow to un-trusted web sites before the user can submit the information. AntiPhish is activated every time the user presses a key, loads a new page, clicks the mouse or has the current focus on a text element (i.e., text field or text area).

The flowchart in Fig. 4 depicts how the sensitive information flow is controlled by AntiPhish.

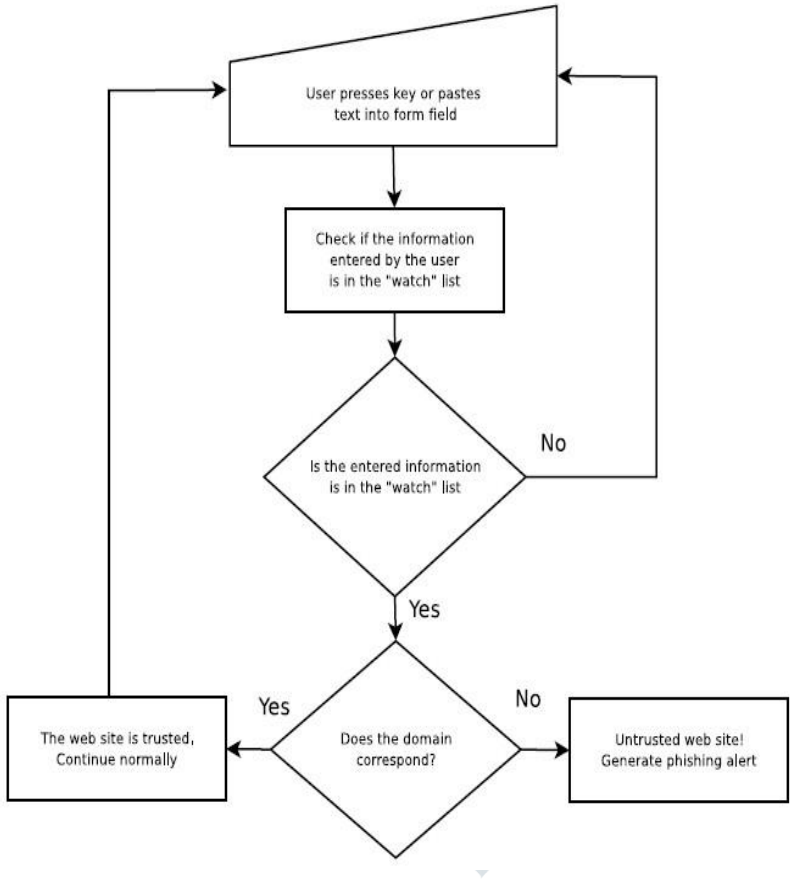

Fig. 4: Flowchart showing how the sensitive information flow is controlled by AntiPhish

IV. ANTI-PHISHING CATEGORIES

Anti-phishing refers to the method employed in order to detect and prevent phishing attacks. Anti-phishing protects users from phishing. A lot of work has been done on anti-phishing devising various anti-phishing techniques. Some techniques works on emails, some works on attributes of web sites and some on URL of the websites. Many of these techniques focus on enabling clients to recognize & filter various types of phishing attacks. In general anti-phishing techniques can be classified into following four categories [1].

Content Filtering- In this methodology Content/email are filtered as it enters in the victim's mail box using machine learning methods, such as Bayesian Additive Regression Trees (BART) or Support Vector Machines (SVM) [1].

Black Listing- Blacklist is collection of known phishing Web sites/addresses published by trusted entities like Google's and Microsoft's black list. It requires both a client & a server component. The client component is implemented as either an email or browser plug-in that interacts with a server component, which in this case is a public Web site that provides a list of known phishing sites [1].

Symptom-Based Prevention- Symptom-based prevention analyses the content of each Web page the user visits and generates phishing alerts according to the type and number of symptoms detected [1].

Domain Binding- It is a client's browser based techniques where sensitive information (e.g. name, password) is bind to particular domains. It warns the user when he visits a domain to which user credential is not bind.

V. ANTI-PHISHING TECHNIQUES

V.I. Attribute based anti-phishing technique

Attribute-based anti-phishing strategy implements both reactive and proactive anti-phishing defenses. This technique has been implemented in Phish-Bouncer [4] tool.

The Image Attribution check [4] does a comparison of images of visiting site and the sites already registered with phish- bouncer. The HTML Crosslink check looks at responses from nonregistered sites and counts the number of links the page has to any of the registered sites A high number of cross-links is indicative of a phishing site [4]. In false info feeder[4] check ,false information is input and if that information is accepted by site then it is probable that link is phished one. The Certificate Suspicious check validates site certificates presented during SSL handshake and extends the typical usage by looking for Certification Authority (CA) consistency over time.URL suspicious check uses characteristics of the URL to identify phishing sites.

Advantage: As attribute based anti-phishing considers a lot of checks so it is able to detect more phished sites than other approaches. It can detect known as well as unknown attacks.

Disadvantage: As multiple checks perform to authenticate site this could result in slow response time.

V.II. Genetic Algorithm Based Anti Phishing Techniques

It is an approach of detection of phishing web pages using genetic algorithm. Genetic algorithms can be used to evolve simple rules for preventing phishing attacks. These rules are used to differentiate normal website from anomalous website. These anomalous websites refer to events with probability of phishing attacks. The rules stored in the rule base are usually in the following form [5]:

if { condition } then { act }

For example, a rule can be defined as:

If {The IP address of the URL in the received e-mail finds any match in the Rule-set}
Then
{
Phishing e-mail
} [5]

This rule can be explained as: if there exists an IP address of the URL in e-mail and it does not match the defined Rule Set for White List then the received mail is a phishing mail [5].

Advantage: It provides the feature of malicious status notification before the user reads the mail. It also provides malicious web link detection in addition of phishing detection.

Disadvantage: Single rule for phishing detection like in case of URL is far from enough, so we need multiple rule set for only one type of URL based phishing detection. Likewise for other parameter we need to write other rule this leads to more complex algorithm.

V.III. Identity Based Anti Phishing Techniques

This technique follows mutual authentication methodology where both user and online entity validates each other's identity during handshake. It is an anti-phishing technique that integrates partial credentials sharing and client filtering technique to prevent Phishers from easily masquerading as legitimate online entities. As mutual authentication is followed, there would be no need for users to reenter their credentials. Therefore passwords are never exchanged between users and online entities except during the initial account setup process [1].

Advantage: It provide mutual authentication for server as well as client side. Using this technique user does not to reveal his credential password in whole session except for the first time when the session is initialized [1].

Disadvantage: In identity based anti-phishing if a hacker gain access to the client computer and disable the browser plug-in then method will be compromise against phishing detection [1].

V.IV. Character Based Anti Phishing Approach

Many time phishers tries to steal information of users by convincing them to click on the hyperlink that they embed into phishing email. A hyperlink has a structure as follows. <ahref="URI"> Anchor text <\a> [6]

where 'URI' (universal resource identifiers) provides the actual link where the user will be directed and 'Anchor text' is the text that will be displayed in user's Web browser and represents the visual link.

Character based anti-phishing technique uses characteristics of hyperlink in order to detect phishing links. Link-Guard [6] is a tool that implements this technique. After analyzing many phishing websites, the hyperlinks can be classified into various categories as shown in fig 6. For detection of phishing sites Link-Guard, first extracts the DNS names from the actual and the visual links and then compares the actual and visual DNS names, if these names are not the same, then it is phishing of category 1 If dotted decimal IP address is directly used in actual DNS, it is then a possible phishing attack of category 2[6]. If the actual link or the visual link is encoded (categories 3 and 4), then first the link is decoded and then analyzed. When there is no destination information (DNS name or dotted IP address) in the visual link then the hyperlink is analyzed. During analysis DNS name is searched in blacklist and white list .if it is present in white list then it is sure that the link is genuine and if link is present in blacklist then it is sure that link is phished one.

If the actual DNS is not contained in either white list or blacklist, Pattern Matching is done. During pattern matching first the sender email address is extracted and then it is searched in seed set where a list of address is maintained that are manually visited by the user. Similarity checks the maximum likelihood of actual DNS and the DNS names in seed-set.the similarity index between two strings are determined by calculating the minimal number of changes needed to transform a string to the other string.

Advantage: it cannot only detect known attacks, but also is effective to the unknown ones. Experiments showed that LinkGuard, can detect up to 96% unknown phishing attacks in real-time [6]. For phishing attacks of category 1, it is sure that there are no false positives or false negatives. LinkGuard handles categories 3 and 4 correctly since the encoded links are first decoded before further analysis [6].

Disadvantage: For category 2, LinkGuard may result in false positives, since using dotted decimal IP addresses instead of domain names may be desirable in some special circumstances [6].

V.V. Content Based Anti-Phishing Approach

GoldPhish [7] tool implements this technique and uses Google as its search engine This mechanism gives higher rank to well-established web sites. It has been observed that phishing web pages are active only for short period of time and therefore will acquire low rank during internet search and this becomes basis for content based anti-phishing approach[7]. The design approach can be broken down into three major steps. The first step is to capture an image of the current website in the user's web browser. The second step is to use optical character recognition techniques to convert the captured image into computer readable text. The third step is to input the converted text into a search engine to retrieve results and analyze the page rank.

Advantages: Generally GoldPhish does not result in false positive and provides zero day phishing [7].

Disadvantages: GoldPhish delays the rendering of a webpage. It is also vulnerable to attacks on Google's PageRank algorithm and Google's search service [7]

VI. CONCLUSION AND FUTURE WORK

In the above study we can conclude that most of the anti-phishing techniques focus on contents of web age, URL and email.

Character based anti-phishing approach may result in false positive but content based approach never results in false positive. Attribute based approach consider almost all major areas vulnerable to phishing so it can be best anti-phishing approach that can detect known as well as unknown phishing attack. Identity based anti-phishing approach may fails if phisher gets physical access to client's computer.

As a future work on phishing we can do more work on server side security. In the server side security policy we use dual level of authentication for user by which only authentic user can get the access of his account, and to educate the user about this policy will results in avoiding user to give his sensitive information to phished web site.

References:

[1] H. Tout, W. Hafner, "Phishpin: An identity-based anti-phishing approach", proceedings of international conference on computational science and engineering, Vancouver, BC, pp.347-352, 2009.

[2] M. Atighetchi, P. Pal, "*Attribute-based prevention of phishing attacks*", Eighth IEEE international symposium on network computing and application, 2009.

[3] V. Shreeram, M. Suban, P. Shanthi, K. Manjula, "Anti-phishing detection of phishing attacks using genetic algorithm", proceedings of Communication control and computing technology(ICCCCT), IEEE international conference, Ramanathapuram, India, pp.447-450, 2010.

[4] M. Bargadiya, V. Chaudhary, M. I. Khan, B. Verma, "*The web identity prevention: factors to considers in the anti-phishing design*", International journal of engineering science and technology, Vol. 2, No.7, 2010.

[5] The Antiphishing Working Group (2004) Home Page, http://www.anti-phishing.org

[6] Heise Security (2005) German Interior Minister Schily requests protection against online scams, http://www.heise.de/security/

[7] S. Kaur, S. Sharma, "*Performing Efficient Phishing Webpage Detection*", JCSE International Journal of Computer Sciences and Engineering, Vol. 3, Issue.7, pp.52-56, 2015.

[8] M. Shukla, S. Sharma, "*A Comparative Study of Existing Data Mining Techniques for Phishing Detection*", JCSE International Journal of Computer Sciences and Engineering, Vol. 5, Issue.5,2017.

[9] R. Dhamija, J. D. Tygar, M. Hearst, "Why Phishing works", Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, pp.581-590, 2006.

.