Review of Phishing Detection Techniques

Swati Gaikwad
Computer Engineering,
DACOE, Pune, India.
swatigaikwad0385@gmail.com

Abstract* — *Nowadays phishing attacks are increasing with burgeoning rate which is highly problematic for social and financial websites. Many anti-phishing mechanisms currently focused to verify whether a web site is genuine or not. In this paper we describe a novel approach for detecting phishing websites based on analysis of users' online behaviours—i.e., the websites users have visited, and the data users have submitted to those websites. Such user behaviours cannot be manipulated freely by attackers; detection based on those data can achieve high accuracy whilst being fundamentally resilient against changing deception methods. Various researches have been done for protecting the users from phishing attacks. They include blacklisting certain domains, spam filtering techniques, fake website detection, client side tool-bars and user education. Each of these existing techniques has some advantages and some disadvantages. The need to automatically discover a phishing target is an important problem for anti-phishing efforts. If we know the webpage which is considered as the target webpage, we can confirm which all are the phishing pages. It could help he owners to identify phishing attacks so that they can immediately take necessary counter measures. There are some techniques to identify the phishing website; we discuss some of those in this paper.

Keywords* — *Phishing websites detection, User protection, Identity theft, user binding relationship

I. INTRODUCTION

Phishing is the method used to steal personal information through spamming or other deceptive means. There are a number of different phishing techniques used to obtain personal information from users. As technology becomes more advanced, the phishing techniques being used are also more advanced.[1] To prevent internet phishing, should have knowledge of various types of phishing techniques and they should also be aware of anti-phishing techniques to protect themselves from getting phished. Let's look at some of these phishing techniques.

This paper reviews the phishing detection techniques and the advantages and dis-advantages of these techniques. Today all organizations are using the internet for sharing the message so the communication must be secure. Few unethical hackers are doing the cyber-criminal activity by committing fraud. Attacker sends the phishing message by the fake website as authentic site. This attack treated as a deceptive phishing attack which target to the financial organization. Attacker sends the phishing message by the fake website as authentic site. This attack treated as a deceptive phishing attack which target to the financial organization. In the phishing attack attacker will get the login and personal detail of the people who have account in the financial organization. Criminals complete their life cycle in very short period by the login and personal detail of the people. Phishing is the fraud emails, or spam, written to appear as if they have been sent by banks or other reputable organizations, with the intent of luring the recipient into revealing sensitive information such as usernames, passwords, account IDs, ATM PINs or credit card details. Typically, phishing attacks will direct the recipient to a web page designed to mimic a target organizations own visual identity and to harvest the user's personal information, often leaving the victim unaware of the attack. Obtaining this type of personal data is attractive to blackhearts because it allows an attacker to impersonate their victims and make fraudulent financial transactions. Victims often suffer significant financial losses or have their entire identity stolen, usually for criminal purposes. This thesis discusses the results of our research into phishing in a Dutch internet banking context. This approach aims to provide practical information on the phishing attack identification because identification of phishing website not easy task. When user will identify the phishing website then prevention will be done. Attackers are constantly innovating and advancing, and there are likely to be new phishing techniques already under development or in use today.

II. SOME SYSTEM USED FOR PHISHING DETECTION

A. ANOMALY BASED PHISHING DETECTION SYSTEM:

In this paper, the basic idea is every website has its identity. The fake website also gives some fake identity but while doing so phisher do some mistake. Like some of its original character remain with the website. The ANOMALY[2] compares the fake website with the legitimate website by using DOM objects and HTTP transactions.

This system consist two models- identity extractor and page classifier.

The identity extractor:

The keyword extraction algorithm in the information retrieval is used for Identity extraction. The webpage identity is determined by DOM objects in this method [3] As follows:-

I. Title: the title of one web page (namely, the text between the tag <title> and < \title>).
II. Description: the content property of the META whose name or http-equiv is "description".
III. Copyright: the content property of the META whose name or http–equiv is "copyright".
IV. ALT/title: the alt and title properties of the DOM objects such as IMG, AREA, INPUT, APPLET, OBJECT.

V. Body: the text in the main body or the images in a web page. There are many technologies to recognize texts from one image, such as Optical Character Recognition (OCR).

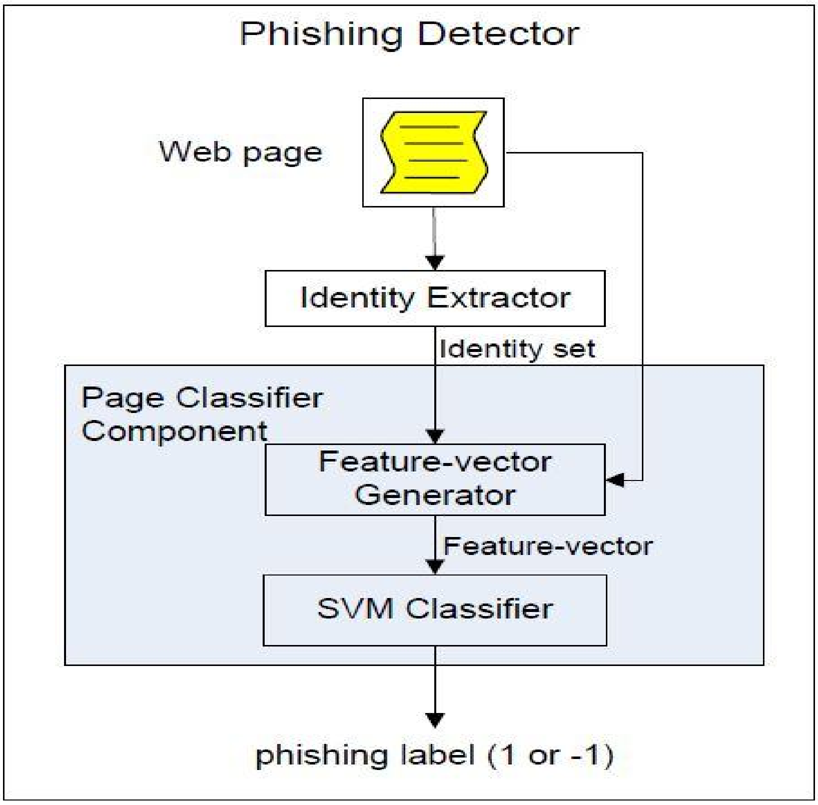

Fig 1: The architecture of phishing detection [1]

Page Classifier:

Another method of structural feature is HTTP transaction. The domain name in DNS record or cookie used for structural features. Thus the feature extracted by identity extraction model is not matched with its claimed identity. The Vapnik's Support Vector Machine is used as page classifier in this system. When user connects to the webpage, identity extractors collect the identity of the current webpage. It also extracts the structural features of the current web page. After that, the collected information is converted into the vectors. Those vectors are then sending to the SVM-page classifier, which returned the webpage tagged as phishing webpage or not [4].

Advantages of the system:

I. This technique is independent of the phishers attack methods.
II. It is independent on the third party.

Disadvantages of the system:

I. The network can be in an unprotected state as the system builds its profile.
II. If malicious activity looks like normal traffic to the system it will never send an alarm.
III. False positives can become cumbersome with an anomaly based setup. Normal usage such as checking e-mail after a meeting has the potential to signal an alarm

B. PHONEY: Mimicking User Response to Detect Phishing Attacks:

Madhusudhanan Chandrasekaran Ramkumar, Chinchani Shamburg Upadhyaya [5] The PHONEY system gives fake response to the user. It lies between the client and server to access each incoming mail. The client side is called as mail transfer agent (MTA) and the server side called as mail user agent (MUA). It provides the fake identity of user to the current website, until the website identity is proved as legitimate site. The diagram illustrates the architecture of PHONY system. The system hides the user identity from current website. It is deployed as client side tool between the mail server and the mail client. It detects the fake email based on phishing attack. The PHONEY system work as follows: First, the pre-processor probes the mail server for incoming messages. When the mail input, it parse the message body for embedded links and HTML forms. The emails which have the link and HTML forms are considered malicious.

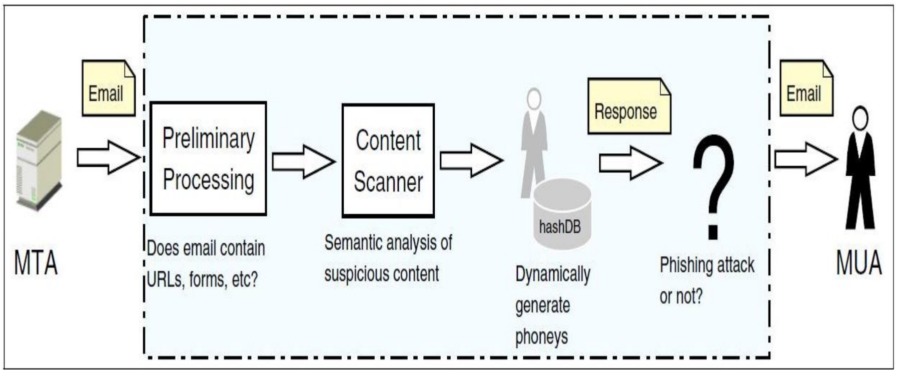
Email
MTA
Preliminary Processing
Does email contain URLs, forms, etc?
Content Scanner
Semantic analysis of suspicious content
hashDB
Dynamically generate phoneys
Response
Phishing attack or not?
Email
MUA

Fig 2: Block diagram of PHONEY architecture [5]

Then the control is passed to the content scanner. The content scanner then analyzes the webpage. The input forms are broken down for extracting input elements and its associated text. Then these extracted token compare with the HashDB. Suppose, the current page contain UName and in the HashDB it is USER NAME like wise. During this comparison, the field filled by fake response. The phantom user is used a virtually to creating fake identity of user. The honey token are created for responding to the current website. Then analyze the reply of the website and take appropriate action like commit or proceed further depending on the rule engine.

Advantages:

I. This technique provides the phishing webpage detection by hiding the users identity from phishers.

II. The computation time is very less.

Disadvantages:

I. As it provides fake identity of user, the phishers may find out the fake user by consistently analyzing fake response from user.

II. Phishers can use CAPTCHA (completely automated public Turing tests to tell computers and humans apart) to intact the response of the legitimate user.

C. DEFENDING THE WEAKEST LINK: PHISHING WEBSITES DETECTION BY ANALYSING USER BEHAVIOUR XUN DONG · JOHN A. CLARK · JEREMY L. JACOB[6]

The another technique is UBPD(User Behavior based Phishing Detection).[6] This method analyze the online behavior of user while sending credential information to the website. This method crate pair of the legitimate website and user's credentials with that websites (user name and password).This is called as user binding relationship. If this user binding relationship is break then this method generate an alert message for the user by, analyzing other criteria also like URL or domain of the website. We discuss the UBPD [5] system design. In this method there three modules.

I. User profile: In this module, it contain the user's binding relationship and whitelist.

II. Monitor: It collects data that user is sending to the webpage. And activate the detection engine.

III. Detection Engine: It works in two modes, first one is training mode and another is detection mode.

In training mode, it updates the user profile if necessary. In detection mode, it checks whether the webpage is phishing webpage or not. If so then send alert message to user. User's binding relationship is the pair of website. <Legitimate website, credential data >. The whitelist is the popular and mostly visited domain list in user's country. When UBPD is installed in the system, then it ask user for the websites with which user share his credential information. Mostly the websites related to the financial process. The UBPD automatically give user to the websites one by one and ask for the user credential related to that website. After accomplishing the login process, click on the TRAINUBPD button. This will create the binding relationship of user with that website. After that it will ask user to go next website or go further for next login process and so on. Also it will store the personal whitelist by analyzing user's browsing history. By analyzing 1000 webpage’s personal whitelist is constructed in user's country. This information is obtained from ALEXA, a company gives the network traffic ranking information.

P1: Temporary phishing score for domain1

P2: Temporary phishing score for domain2

Step One:

P1=2/3= 0.67 And P2= 1/3=0.33

P=biggest (p1,p2)=p1

The domain1 is the target of the phishing attack

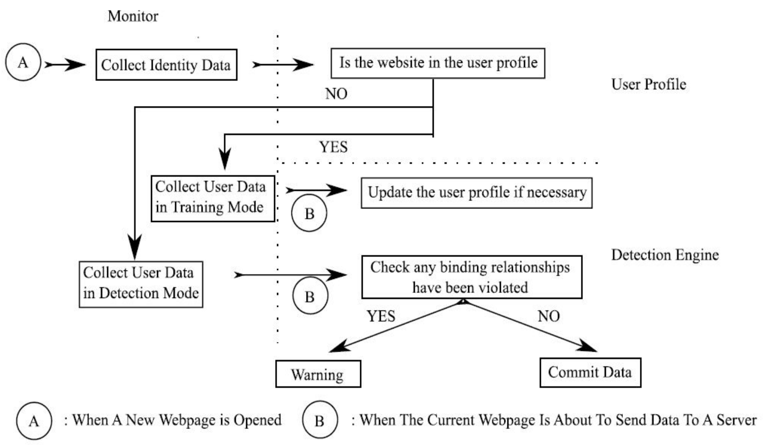
Monitor
A
Collect Identity Data
Is the website in the user profile
NO
User Profile
YES
Collect User Data in Training Mode
B
Update the user profile if necessary
Collect User Data in Detection Mode
B
Check any binding relationships have been violated
Detection Engine
YES
NO
Warning
Commit Data
A : When A New Webpage is Opened
B : When The Current Webpage Is About To Send Data To A Server

Fig3. Detection Process Workflow[6]

Phishing Score Calculation :

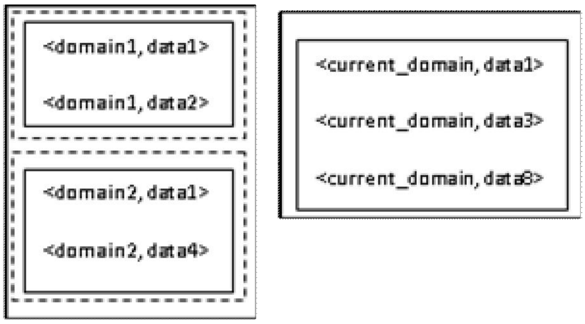

Fig. 4: An example of how the phishing score is calculated [6]

The insertion and fragmentation:

The insertion attack in which, the attacker asks for some extra information with the users credential information. the phishing score calculation method automatically solves the insertion attack problem.

Fragmentation: The phishers ask user credential in a part of information. This is solved by using phishing threshold. The threshold of the system is 0.6. If the current phishing score is above than the threshold then the current website is phishing website.

Reuse: It is possible that the user use the same username as a password to other site. And the user is not trained the website with which user have an account or in future user want to create an account. In such case, it generates false warning. Because of, break in user relationship pair. This is avoided by the use of whitelist. As that site is legitimate website. So it is appear in whitelist. Thus, it does not generate false warning.

Warning Dialogue box: The previous system also generates the warning dialog. But they are ignored by the user. The content of the warning are effective and short to prevent user by accessing phishing webpage. This system gives detailed warning dialog. It display the url name of current webpage and legitimate webpage, Ip address, registration time. So user can't ignore it.

Checking of domain identity: The checking of domain identity is done by comparing two websites domain. In big organization different domains have the websites. In this method, two IP addresses and two domain names are compare. As there are dynamic IP addresses also used. This system compares the net name, name of server and the countries where the IP address is registered. The WHOIS database is used to compare the websites.

User Profile's Privacy: This system also provides user profile privacy. As the user profile contain credential information. This security is provided by using SHA-1 algorithm. The domains are not possible to hashed, so they are encrypted using a secrete key. Thus this system provides high security to the user profile. As user profile contain confidential data.

Implementation of the system:

The UPBD is implemented in Firefox. The SHA1 algorithm is used to implement hashing function for user profile. The two fish is used as secrets key to encrypt and decrypt domain. The user can change the hash function and the encryption key also.

Advantages:

I. It does not depend on the phishing technique.

II. It can detect pharming attacks, which are undetectable by many existing systems.

III. Some system tries to detect phishing webpage. Some system detects phishing webpage when user opens new webpage. But the UBPD detect phishing web page when user sending the credential information to webpage. Thus UBPD can be used to fill the gap between these systems.

Disadvantages:

I. It can't handle the one time password between the user and the legitimate site.

II. The detection using WHOES database is time consuming method.

III. CONCLUSION

The anomaly gives accurate result but the algorithm for identity extraction is critical to give to successful detection. The PHONEY method has advantage that, the phishers can circumvent the given defence mechanism by replaying the response of the legitimate site for spurious inputs. The UBPD system is the best solution for the phishing detection system than the other as it works as compliment to other techniques. The system is easy to implement. It gives approximately accurate result. Also, the UBPD system has drawback it is used for only static data.

ACKNOWLEDGMENT

Head of the Department {Prof.Sangve S.M.} for their valuable guidance and for providing all the necessary facilities, which were indispensable in the completion of this project report. We are also thankful to all the staff members of Information Technology the Department for their valuable time, support, comments, suggestions and persuasion. We would also like to thank the institute for providing the required facilities, Internet access and important books.

REFERENCES

[1] citeseerx.ist.psu.edu

[2] Anomaly Based Web Phishing Page Detection Ying Pan, Xuhua Ding School of Information System, Singapore. Management University{ypan, xhding}@smu.edu.sg

[3] http://www.w3.org/tr/dom-level-2-html

[4] V. N. Vapnik. The nature of statistical learning theory. Springer, New York, 1995.

[5] PHONEY: Mimicking User Response to Detect Phishing Attacks –Madhusudhanan Chandrasekaran, Ramkumar Chinchani, Shambhu Upadhyaya , Department of Computer Science and Engineering, University at Buffalo 201, Bell Hall, Buffalo, NY-14260 {mc79, rc27, shambhu}@cse.buffalo.edu

[6] Defending the weakest link: phishing websites detection by analysing user behaviours