Implementation Of Defence Schemes For Phishing Attacks On Mobile Devices

Aditya Kharat , Yogesh Komalwad , Ashvin Kokane , Rohan Gharde
Dept. of Computer Engineering
Maharashtra Institute Of Technology(MIT Pune)
Pune, Maharashtra

Abstract — *Phishing is a form of cybercrime where an attacker imitates a real person / institution by promoting them as an official person or entity through email or other communication mediums. In this type of cyber attack, the attacker sends malicious links or attachments through phishing emails that can perform various functions, including capturing the login credentials or account information of the victim. The standard way to specify page layouts is through the style sheet (CSS), the developed algorithm detects similarities in key elements related to CSS. Phishing detection includes approaches that uses profiles of trusted websites appearances to detect phishing.*

Index Terms — *Mobile phones; phishing attack; security; anti phishing; SLD(Second Level Domain).*

I. INTRODUCTION

Phishing is defined as the fraudulent acquisition of confiden-tial data by the intended recipients and the misuse of such data. Phishing attacks are often done through email. An example of Phishing; the email appears to be from a known web site like from a user's bank, credit card company, or Internet service provider. Generally, personal information such as credit card number or password is asked to update the accounts.

These emails contain a URL link that directs users to another website. This site is actually a fake or modified website. When users go to this site, they are asked to enter personal information to be forwarded to the phishing attacker.

A. Phishing Attacks

The aim is to steal sensitive data such as credit card and login information or to install malicious software on the victim's machine. Phishing is a common type of cyber attack that everyone must learn to protect them. Phishing starts with a fake email or other type of transmission designed to attract a victim. In this type of attack, the message appears to come from a trusted source.

In a phishing attack, attackers can use social engineering and other public information resources, including social networks like LinkedIn, Facebook and Twitter, to gather background information about the victim's personal and work history, interests and activities. With this prediscovery, attackers can identify potential victims' names, job titles and email ad-dresses, information about the names of key employees in their colleagues and organizations.

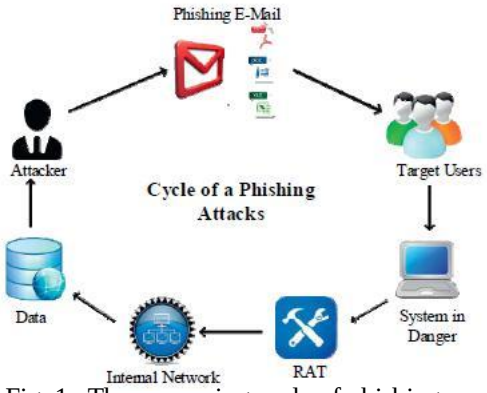

Fig. 1. The processing cycle of phishing attacks [1].

Phishing is also used to learn someone's password or credit card information. With the help of email prepared as if coming

from a bank or official institution, computer users are directed to fake sites. The common information that is stolen by a phishing attack is listed as follows:

User account number

User passwords and user name Credit card information

Internet banking information

B. Avoiding Phishing Attacks

A whitelist in the context of phishing detection is simply a list of trusted websites. For CSS detection to work properly, the list contains more than just the URL of the trusted website. Each entry in the whitelist database contains six strings: the URL of the trusted site, the domain of the site, the title of the site, the CSS fil ename, the CSS domain, and the CSS content of the file.

1) The URL of the trusted site: The URL of the trusted site is used to periodically update the CSS information in the database. The URL of the site, such as https: signin.ebay.com.

2) The domain of the site: The domain of the trusted site is the domain of the URL such as signin.ebay.com and is used to determine whether the current page displayed in the browser is on the whitelist or not.

3) The CSS filename: The CSS filename is the filename of the CSS file such as paypal.css and can also be used during CSS content detection to speed up detection by matching potential phishing site CSS filenames with filenames in the whitelist database.

4) The CSS domain: The CSS domain is the domain of the location of the CSS file such as secureinclude.ebaystatic.com. Often the domain is the same as the site domain, but in other cases such as eBay, the CSS file is hosted on a different domain. Storing the CSS domain is essential because if a match is found on a website not in the whitelist, then it is most likely a phishing site linking to the actual CSS file location of the legitimate site.

5) The CSS content of the file: The CSS content is the actual text contained in the CSS file that contains all of the style information. The CSS content is used to compare with the CSS content of a possible phishing site in order to determine if there is a match with a legitimate site.

C. OCR Techniques

The OCR technique is utilized to convert the screenshot into text. Screenshots can be used for the phishing detection in both web pages and applications. Since the source code of apps is not available, there is no way to acquire the content of an app login interface other than taking a screenshot. OCR has been utilized to extract text from simple text only logo. There is the believe that the OCR technique could achieve better performance on mobile phones because phones have a smaller screen size and a relatively higher pixel density.

II. LITERATURE REVIEW

A. MobiFish

It proposes an automated lightweight scheme for mobile phishing defense named MobiFish. MobiFish consists of three major components named WebFish, AppFish, and Account-Fish, which are designed to protect mobile web pages, appli-cations, and persistent accounts, respectively [2].

1) WebFish Schemes:: The defense scheme is initiated with URL loading. When a browser attempts to load a web page, WebFish first scans its URL to see whether the domain name is an IP address. WebFish obtains the HTML source code of the loading page and checks if there is any form in that page. Like legitimate login pages, phishing web pages also need a form with an input tag, which allows the user to enter (confidential) information and then submit. WebFish checks the existence of forms so that not every page has to go through the checking. If a form is found, WebFish starts the identity extraction and verification. On one hand, it extracts the SLD from the URL, which represents the actual identity of the site. Then, the SLD is indexed in the MWL(MobiFish White List) [2]. If it matches any of the SLDbrand name records in the MWL, the original SLD is replaced by the corresponding brand name. On the other hand, it calls the screencap native function to take a screenshot of the login page and extract the text with the OCR tool. The last step is to search the SLD in the text. If not found, it is marked as a phishing site.

2) AppFish Schemes:: AppFish monitors the possible paths for a phishing app to transmit data to remote server, which comprises of HttpGet / HttpPost, socket, SMS and email. This technique is based on the assumption that most well known enterprises use brand name as the SLD of their official websites which is also used, as an image, within their login forms.

3) AccountFish Scheme:: The account registry phishing attacks can be classified into three types, based on the iden-tities that the malicious app appears to be in the main menu and the account list. In the type A attack, the malicious app appears to be a different app to the target account (e.g., a game app registers a Twitter account). In the type B attack, the malicious app does not appear in the main menu at all. In the type C attack, the malicious app directly shows up as the target app (e.g., repackaged app), which means that they will have the same application name shown in the main menu as the account name that appears in the account list. The idea of the detection mechanism for the type A and type B account phishing attacks is to compare the app name in the main menu and the account label in the account list. As mentioned earlier, the malicious app can dynamically register and change the account information. AccountFish should be able to inspect the registration of accounts in runtime, which can only be accomplished by modifying the Android source code.

B. Non-content based approaches

Now a days phishing detection schemes can be usually divided into two categories:Heuristics-based schemes:-

Heuristics based schemes largely depend on features extracted from the URL and HTML source code, and then other techniques such as machine learning are used to determine the validity.

Blacklist based schemes:-

Blacklist-based schemes can only detect phishing sites that are in the blacklist, but cannot detect zero day phishing attacks that have appeared for days or even hours

[1]. It is possible that new phishing sites may have already stolen user credentials or even expired before being added into the blacklist.

C. Second Level Domain (SLD)

When the login form is detected then, the tool extracts second level domain (SLD) name from the URL [1]. The SLD is mapped to its brand name using the mapping list that is generated. Screen shot of a webpage is captured and converted to text using OCR tool. If there are some typical sensitive terms in the text, then the user will get the warning. If SLD is not present in the text that is extracted from the screenshot, then it is possibly a phishing web page.

D. Anti-phishing single sign on model using QR Code

This technique addresses the problem of phishing on single sign on authentication. Single sign on is an authentication process that permits users single username and password to access multiple applications or websites. The technique uses QR codes since they do not need mobile network data to read the data and it can store a large amount of information. There are two phases in this approach.

User Registration Phase

In User Registration the user receives a secret key which is later used in the verification phase to get access to the requested service.

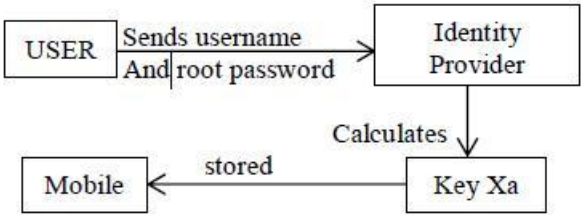

Fig. 2. User Registration Phase

User Verification Phase

In verification phase user requests a service from the service provider which sends the user identity to the identity provider.

E. Earths Mover Distance (EMD)

EMD is a metric of the similarity between two probability distributions over a region. The closer two images are, the smaller the EMD value will be about using the hash value of images in their work. The authors have also calculated pHash of a screenshot and evaluated the difference using hamming distance between two hash values. They presented experimental results illustrating that adding a small change to the original image will also lead a small increase in the hamming distance.

F. RRPHISH

Because of the maximum concurrent connections limitation of the same domain for browsers, brand sites usually run resource content on another domain, such as PayPal runs CSS, JS and image files on paypalobjects.com. RRPhish can automatically extend the blacklist which will be an effective complement to the blacklisting method. For different appli-cation scenarios, it can contain different algorithms with dif-

ferent complexity, such as heuristic rules or machine learning algorithms [3].

III. SYSTEM ARCHITECTURE

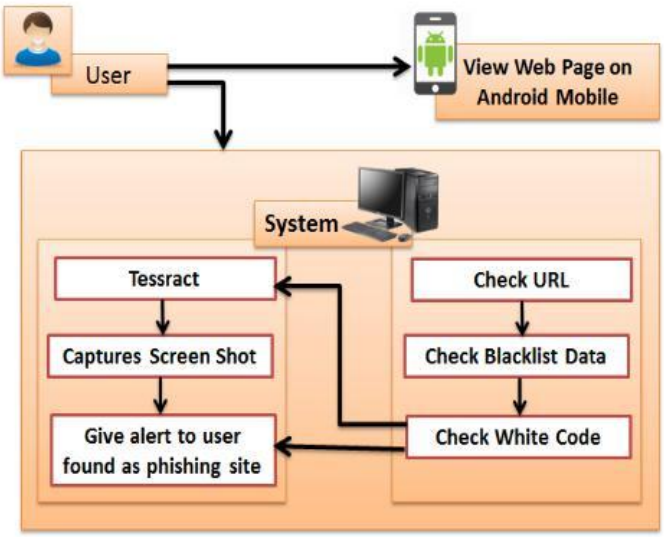

Fig. 3. System Architecture

As we know, the attackers use a number of methods to obfuscate the URL. So, it is complex to detect all that attacks, but the ObURL detection algorithm can detect the maximum number of URL obfuscation phishing attacks because follow-ing test cases are performed for checking the phishing site emails.

URL checking:

The system compares the web pages URL to original URL.

Blacklist checking :

Blacklist based schemes can detect only those phishing sites that are in the blacklist, and cannot detect zero day phishing attacks such as those that have only appeared for days or hours. It is possible that new phishing sites may have already stolen user credentials or have expired before being added to the blacklist.

White code checking

Here the URL is resolved to an IP address using DNS and the IP address is checked with the White List of IP addresses and if the IP of concerned site is not in white list then that site is considered as a threat. Authors have checked the IP and URL with whitelist to determine if the site is phished or not.

OCR using Tesseract:

It is a common method of digitizing printed texts so that they can be electronically edited, searched, stored more compactly, displayed online, and used in machine processes such as cognitive computing, machine transla-tion, (extracted) text-to-speech, key data and text mining. Tesserhact has supported output text formatting, hOCR positional information, and page layout analysis [1]. Support for a number of new image formats was added using the Leptonica library. Tesseract can detect whether text is mono spaced or proportionally spaced.

Capture screenshot and give alert :

If the site doesn't match with original web pages, then the system captures screenshot through Tesseract and gives alert to the user about Phishing attack.

IV. ALGORITHMS USED

ObURL Detection Algorithm

Input: Content of Email

Output: Prevent the user if URLs seems Counterfeit

Alert User: Possible Phishing

Safe User: No Phishing

DB: Database

If Input form found in E-mail Content
then Alert User;

End

For each iframe in E-mail
content do //get the
content of iframe
For each iframe in E-mail contents iframe
sources do If input form found then

Alert User;

End

For each hyperlink in E-mail contents iframe
sources do //perform the test 1 to 6

End

For each hyperlink found in E-mail content and iframe source

URL
do

Test 1: //DNS Test

If hypertext! = Anchortext

then

Alert User;

Test 2: // IP Address Test

If IP address found in hyperlink then

If IP address found in White list DB
then Safe User;

Else
Alert User;

// IP Address found in blacklist DB

Test 3: // Encoded Test

If hyperlink found encoded

then

Decode hyperlink;
Inform User;

Test 4: // Shorten URL Test

If URL is shorten

then Alert User;

Test 5: //hyperlink white list and blacklist test
If URL found in whitelist DB

then

Safe User;
Else

Alert User;
// URL Found in Blacklist DB

Test 6: // Pattern Matching Test

If hypertext and anchor text pattern is matching then

Alert User;

V. EXPERIMENTAL RESULTS

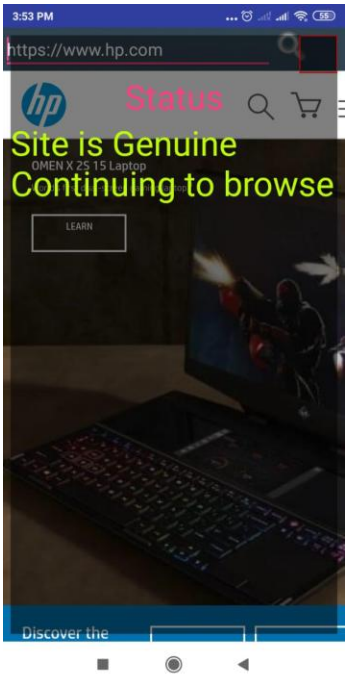

Fig. 4. Genuine Site Detected

Fig. 5. Blacklist Matching

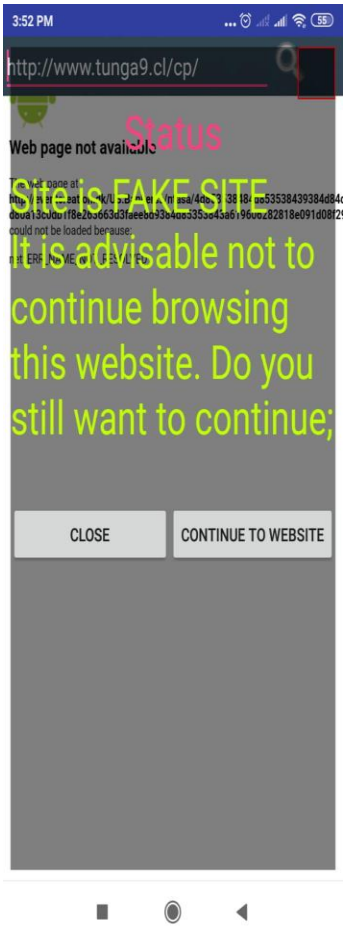

Fig. 6. Blacklisted Site Detected

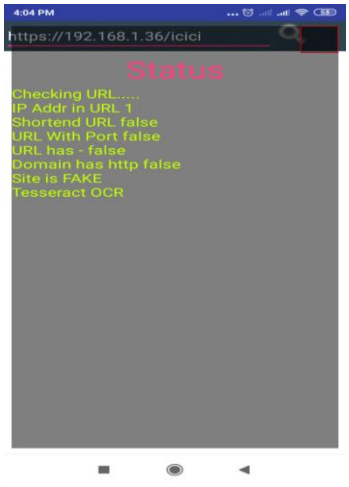

Fig. 7. CSS Checking

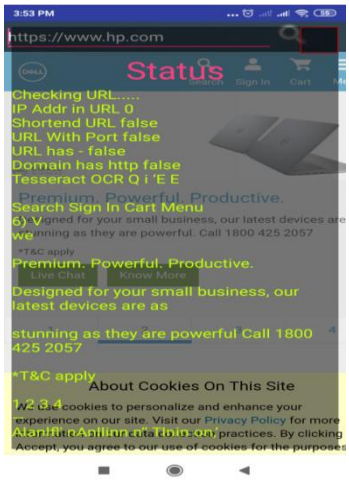

Fig. 8. Text Extraction

VI. ADVANTAGES

Protection from various types of phishing attacks. As we know, the attackers use a number of methods to obfuscate the URL. So, it is very complex to detect all the types of attacks, but the ObURL detection algorithm can detect the maximum number of URL obfuscation phishing attacks because it performs multiple tests.

It provides security against phishing site.

To prevent frauds on the internet. To prevent the hacking of private information like ATM card number, PIN etc.

Protect different application that runs on the internet.

VII. CONCLUSION

Phishing is an appalling threat in the web security domain. In this attack, the user inputs his/her personal information to a fake website which looks like a legitimate one. We have presented a scheme on phishing detection approaches based on visual similarity, white code and url checking. This scheme provides a better understanding of phishing website, various solution, and future scope in phishing detection.

REFERENCES

[1] Longfei Wu; Xiaojiang Du Senior Member IEEE ; Jie Wu, Effective Defense Schemes for Phishing Attacks on Mobile Computing Platform, 2015 IEEE International Con- ference on Consumer Electronics

[2] Longfei Wu; Xiaojiang Du Senior Member IEEE ; Jie Wu, MobiFish: A Lightweight Anti-Phishing Scheme for Mobile Phones., 2014 IEEE International Conference on Consumer Electronics

[3] Guang-Gang Geng ; Zhi-Wei Yan ; Yu Zeng ; Xiao-Bo Jin, RRPhish: Anti- phishing via mining brand resources request, 2018 IEEE Interna-tional Conference on Consumer Electronics

[4] Zuochao Dou ; Issa Khalil ; Abdallah Khreishah ; Ala Al-Fuqaha ; Mohsen Guizani, Systematization of Knowledge (SoK): A Systematic Review of Software- Based Web Phishing Detection, ommunications Surveys Tutorials.

[5] Guang-Gang Geng; Xiao-Dong Lee; Wei Wang; Shian-Shyong Tseng, Favicon - a clue to phishing sites detection, 2013 APWG eCrime Researchers Summit.

[6] Jhen-Hao Li ; Sheng-De Wang, PhishBox: An Approach for Phish-ing Vali- dation and Detection, 2017 IEEE 15th Intl Conf on De-pendable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Tech- nol-ogy(DASC/PiCom/DataCom/CyberSciTech).