Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

An Inexpensive Way to Prevent Phishing using OCR Technology

Bharti Sharma
Department of Computer Science & Engineering
Sunderdeep Engineering College
Ghaziabad , India

Ashutosh Kumar Rao
Department of Computer Science & Engineering
Sunderdeep Engineering College
Ghaziabad , India

***Abstract :-*Phishing is one of the main thrusts of the country and world , e-commerce is also predominant in the growing mobile services abroad mainly from the growing mobile services, although many consumers are tired with phishing. Fishing is also a major part of the mainly unfortunate attacks, from URL addresses to suspicious routes. Most browser vendors predominantly adopt blacklists and heuristic based although both have limited limitations.This paper is proposed to protect against the novel method of image recognition of phishing attacks using the OCR technique of phishing attacks. Can be realistically different from the website by people reading on the website and comparing people with the website. index terms was the highest score performance to implement site URL prototype for the detection accuracy of experimental successful tests.**

Keywords- Phishing, OCR , phishing prevention

I. INTRODUCTION

The Internet has become an important element of our life. In daily life it continues to provide important resources for people such as playing online games for education and other important role. It plays a role in the role of mobile devices. Mobile services are the only important financial transactions. Our life is a new initiative. All transactions are processed by our finger. In the current figures of British Bankers Association and Ernst & Young (EY) in 2017, there are rapid changes in consumer financial transactions being carried out by mobile banking and Paytm, phone pay etc. This number has been increasing year by year as shown in the picture one.

Meanwhile, there is a very challenging problem among mobile consumers which is mobile security. Apple and Google They both have different application platforms. There are still difficulties to prevent phishing for the security of the applications. In the survey shown in Mcafree, in 2015, about 97% of consumers were unable to identify phishing emails correctly. In addition to using QR-code, QR phishing uses new and convenient methods every time. It is intelligent and supported to detect but there is a lack of security in different browsers in laptop computers and mobiles, so the user cannot easily identify the symptoms of phishing attacks, so identify each user URL as fake email and Unfortunate URLs are made invisible to QR codes and QR codes Most user URLs Do not check. Check the URL before accessing. The survey of wandair is visible as it appears that every 20 seconds a new fishing site is created in seconds.

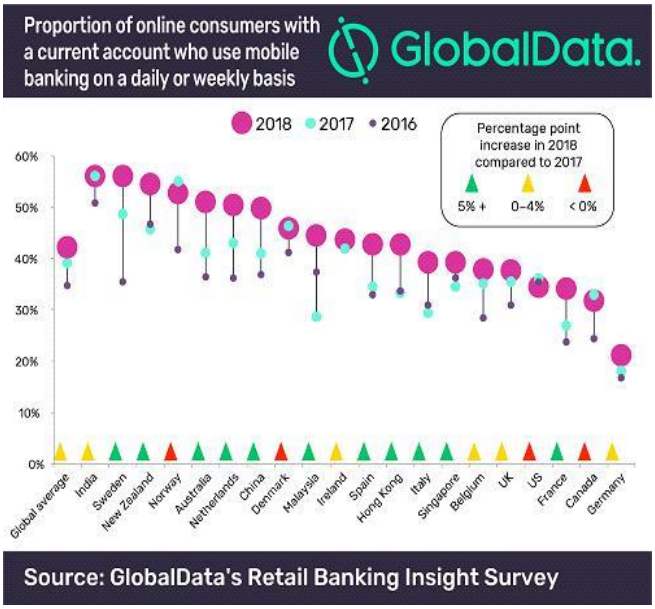

Fig. 1. Mobile banking with dealy weakly basis.

It is effective to isolate suspicious sites separately because these sites are not detected and are not even registered in real time. Kaspersky lab anti-phishing investigation More than 247 million ebanking consumers access different types of new pages It is revealed that 54% of the individuals have access to the financial website by the account holders in 2016 vs 47%. This is the first time that More than 50% of phishing has financial fissuring accounts. So with this growing human trend, Phishing Defense can accelerate better in the mobile platform or the PC platform, hence we present a novel method by letter. In some demonstrations to avoid phishing attacks, the remainder of the letter is used to increase the rate of accuracy through evaluation experiments. But the following searches were used in section 2 n 3. We describe the methodology that experimental tests have been evaluated and related limitations are also analyzed. At the end of section 4, section 5 covers this work in detail and includes All Our conclusions are included.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

II. RELATED WORK

All the reviews we conducted have a special focus on three aspects: strategies to detect phishing, phishing prevention and safety protection on mobile and PC platforms. In felts or wagner [2], the risk of phishing on mobiles and the risk of platform phishing and the risk of being platformed and platform has been assessed in the transfer control concept introduced in mobile browsers' websites as standard common origin policy Follow Origin Policy 1 because it is all potentially unreliable that aggregates individual websites aggregating domains into one another DOM No Se must be separated from Domino Se from Domino Se from Domino, although neither Android nor Android nor Apple restricts access to mobile websites, so phishing always occurs during control transfers such as trusted applications. Link that can be unfortunate to fusing QR code in collaboration with the emerging technology of QR code as a phishing vector As shown because it is easy to produce and the attacker alters the entire QR code gives the transforms the code replaces the code returns the transforms or the attacker returns some QRs Modifies the code. Pixel of QR code. Pixel of vector is used as vector attack. Encode lte malicious. An work was diagnosed with the unfortunate and malicious cases that use that use does redirect to phishing showed existing data QR-code in September March 2011!The fight against phishing will continue and it is plausible. Strategies are offered to detect the current prevention and fitting. Usually most browser vendors adopt two strategies, adopting the strategy Blacklisting and holistic techniques against phishing. In this method, the URL is first checked by checking the status before accessing the URL. Listing is a stable right way. Phishing cannot be stopped completely, the only reason is that all the black lists require people to report them in the phishing URL. Some way to physically improve the black list. Not enough, therefore, a new fishing website will not be able to be found in another adequate method through phishing. Website content such Eshtaan etc. went took some went from the machine learning approach!

The learning method used by most machines is lexical and host-based to describe the features adjective to the adjective example for some reached in [8], [12], [13], in URLs The properties of the analysis are considered to be related to the current electrification that is analyzed by the properties of the analysis for WHOIS2, such as IP, such as registration information and so on. Due to this, there may be difficult major problems in the future. WHOIS ban is being pursued and despite being disputed, it continues to generate very private information. Most of the information was erased from the WHOIS to comply with the European Union's General Data Protection Act Act (GDPR). Not [14] | Followed by all the phishing security hints the resource is mature on the browser. An active warning is considered more effective than a passive alert by the probe such that often users ignore popup messages.Although the prevention of phishing is more on the mobile platform, the mobile platform is more complex than expected. Also the problems are laptop and computer. Both of them keep on facing the additional challenges. Phishing links come from most of the first phishing. This email and mobile Mobile users do not support knowing exactly how to identify the security of the platform It is unable to determine whether it can be accessed by the URL address and whether it is safe, especially when there is a lack of security. Users have a special lack of security. For example: google crome provides a lot of security compared to other browsers. It is highly secure. It provides warnings on URL cards but does not provide this feature on the Google Chrome mobile platform if the user checks the correct address of the URL If you are not a victim of any type of scam, then some scams are done by java script scrool tools. This is done to hide what replaces a fake URL within the page of the page. This is done to confuse users although the actual address is often below the page, although this risk was only recently fixed. Although this old risk is a recent risk platform.

Quick processing in other URLs poses a risk of phishing. In mobile platforms, the code remains QR code encoded, so it is unread by humans and it inserts specific QR code by dumping. Needs to be done thus the QR code has the possibility to trigger in the trigger The manipulated QR code [17], [18]Apart from the 31 QR scanner applications that have been used in 19M, the Security Alert is the only two apps to perform that feature, but they also measure false negativity errors in a higher proportion besides having 2 sources (crome browser , fasttrack). It has been recommended to improve the prevention and accuracy of phishing but the static limits black list is not far-fetched, so in this paper we present a novel approach using OCR3 technology to prevent phishing. Dynamic detection method can be deployed, it will not only blacklist the content of the web but will also check the latest with accuracy. There are boundary issues but this restriction is appropriate in the works of the issues and these restrictive issues can be avoided by using machine learning from the respective WHOIS.

III.METHODLOGY

A. Main objectives of the research.

The purpose of this research is only to determine the purpose of the legislation using the website, the recognition technique using the background and image of the people. Then by checking all the aspects, we identify the activities

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

of the phishing on the base website. Its divided into four parts. Its steps are as shown below.

1) *Remove Image*

2) Describe all the contents of the image

3) Check and confirm the based URL

4) Check and verify the accessed URL

B. Preliminary Requirements

In the initial workup, a python program was written to locate the images to verify the feasibility in. Find a website based on it which is a regular part of the task to register some open source such as APIs as follows. Main requirement

- **google (OCR) related APIs.**
- **Google serch related APIs. C. Related procedure**

The following steeps applied for 40 URLs taken the phishing and more Analize with proof or concept.

- **Extract image**: To confirm the objectives, the first step for us website was to use a web crawler. There are usually two ways to extract images related to it. Most of the websites have a logo image or background link. HTML or .css files. Code In the first nasty angle, the image of the people in the angle is transferred and the HTML code is usually inserted into the code under the image tag such as google.com microsoft.com etc.!

In the second approach the image of the people is added by means of .css and under their characteristic features is the background background image, such as logo. When we extract the image, all these situations should be considered in situations that are complex and due to the variety, the background is able to prevent from phishing.

2). Describe the image by content: In this stage it is used by us Google Optical Character Recognition (OCR). It takes image-based content information to be detected because it only works on the API itself. Specific text in the image is determined to identify the text. Content images such as .css called symbol changed called it Is and the description image is ignored by the <html> <img> tag. After that manipulation gives its efficiency. We also check other like by Microsoft query to try to be the result went the material removed is no longer as good as expected mentioned will be in the next section

3). Based on the URL check and confirm the confirmation by checking: By checking the details we can identify our region above the content of the image. In Google search, the search of the keyboard is achieved using the search of the keyboard search of the URL based on the corresponding keyword by api keyword

The address can be confirmed and we only put Title 3 in it, the list of results we gave in our initial use, their results. Generally, the first url can be a url based on Website: The second link may be a wiki link to the website and the third link may be news related to this website. This link may be news related to this website or other branches of this website such as for example, phishing paypal text for one use. The paypal website and related URL were confirmed by this text. I was identified by using the Google search engine as shown in Figure 3.

```
The related official url is:
{'url': 'https://www.paypal.com/us/home', 'Organization': 'PayPal, Inc.'}
{'url': 'https://itunes.apple.com/us/app/paypal-mobile-cash/id283646709?mt=8',
'Organization': 'Apple Inc.'}
{'url': 'https://www.paypal.com/login', 'Organization': 'PayPal, Inc.'}
```

Fig. 3. Search key words to PayPal, and related results

4) Fully verify the accessed URL: Just like we have already done in most of the website, we also do SSL (secure sockets layer) to further improve security Transmission of all data due to encrypted gender And ensures that it is private and integral under an established channel website survey and browser browser. [20] We therefore verify the SSL certificate Compare to compare and to access went urls safety information.

First we tried to retrieve the SSL and the hash associated with them and determined the results of the thumb print URL and the based URL to confirm it to maintain and maintain the status quo. However, the results were not so good. As expected by us, hash values may correspond to different domain names and branches of the same company. For example the SSL certificate hash value shown in Figure 4 https://www.google.com.uk / So finally we used all the organizations that identify the certificates by SSL to verify these websites that belong to the same company and also verify the security. Also served to register statement such as has been against using the URL of other features to enhance the confidence to confirm the ownership of the website in order to SSL, issuing method by our current.

```
{'url': 'www.google.com', 'hash': '8512c2a42dac995fa6ca65843ec38fd9'}
{'url': 'www.google.co.uk', 'hash': 'b3c781e6c93a646f70e3f26e5c831bfe'}
```

- And the text of the URL must be known, otherwise the API will not be there

- these logo must be English as an OCR API, s serves in English as a prototype simple

In the initial test, we have collected 40 different fish tanks in the best way. We have found the fishing URLs from the fish tank most of the time. We have blacklisted all the fishing URLs, it was about financials. 75.5% (29/40) All these sites with accounting finance sites such as paypal, alibaba, amarican express, netWest etc, are shown in the figure

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

number 5, all of them in the best ratio of 72.5%. paypal was 79% (23/2 9) which is shown on figure 5.

IV. EVALUATION

A. Electricity and their results together:

Many URLs are considered to be malicious and url It was discovered that these URLs are created to meet the conditions:

• All steps are to go online to locate these URLs

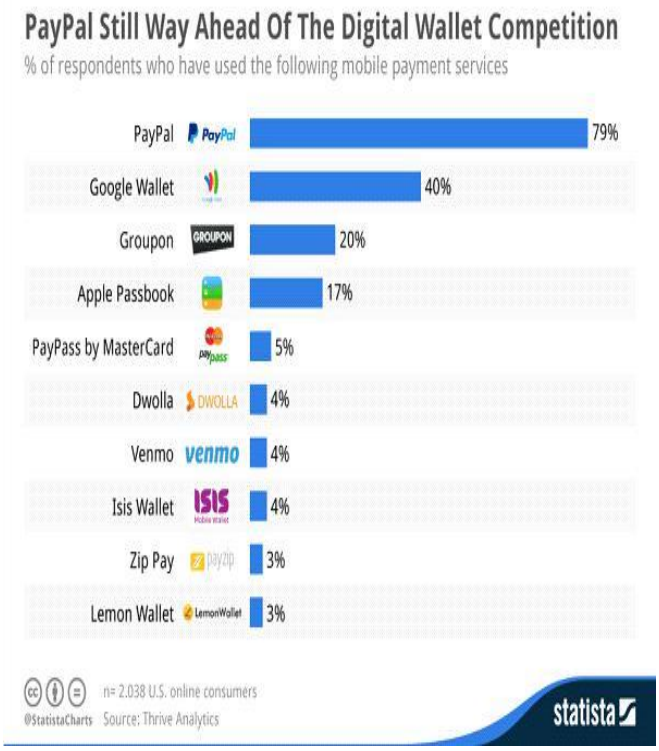

Fig. 6. PayPal was the highest proportion of 40 financial accounting sites at 79%.

90% (36/40) of our approach had so many phishing URLs that were successfully identified and only 4% of the URLs we missed analyzed all of these failures by us and we knew the main reason and we Knowing the main reason: Knowing the main reason in which the logo came to the conclusion that means the purpose of the website's purpose would be wrong in step 2, if the logo image was extracted which is clear it

The method would be wrong because of the specificity of the results in step 1 which we conducted the financial analysis.

• Difficult to clearly identify what the web crawler presented in the text rather than the logo image

• So the whole page is defined in an image ocr API for response, s more details are presented

B. challenges of the limitations.

Of course, technology has contributed significantly to OCR. The end result of this method is subject to accuracy. For the sake of recognition, we first try to use the microsoft OCR API in our approach, although this has been the result of testing expected. According to this it is not good because it will not work in some important cases like the color of the logo with the working form background Also only supports jpeg.png, png, and bhp format. Some websites have made special use of SVG files.The special format of the logo image is thus switched to google ocr api, our approach is achieved by crossing all possible limits by which the recognized results are also affected by the web crawler because the useless results recognized logo Whose image is wrong, so the boundaries of all the issues are the same and the challenges are as follows

• **Extraction of logo's image accuracy:** Phishing can be done in a different way by phishing in many types of places. By this method it is difficult to accurately locate the logo image for example on the image of the page but on the image Is.

• **Determining the cost of the OCR API, s:** There are different APIs that can be used to identify the image to be used, however they are not independent. The copy of a free quote is recognized, 1000 rounds, but an internal check Need for which to be charged thus if we want to make a big response then we should consider the same level of OCR API, s

• **For efficiency of systems:** This method can complicate the phishing website and the url can be very archived images or

.css if we have stored stored in it the computer response can be very long to identify the image either It is used to identify the response of a web crawler image

• **For Single Identification:** Phishing is a diverse threat. It can only provide personal information about the victim, but the victim can execute to infect and keep using wictim to create an event and also stop the method. Transplant virus is not possible in phishing site

V. CONCLUSION AND WORK TO BE DONE IN FUTURE.

Fitting attacks are way easier to produce when deploying and most vendors use phishing to prevent phishing from various perspectives although we cannot keep this solution with the constant integrality of phishing websites. Fishing We have reviewed related literature about the logo and we have taken an approach to prevent and identify it. Has been researched on the entire phishing website using ocr technology. We have exceeded all the limitations of the approach. All the current methods have been researched and resolved to make it dynamic and provide privacy of all the methods. And restrictions can also be avoided in issues like results of machine learning for WHOS even though

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

phishing has been compromised though There may be some limitations of the technology that can be identified. These can be further improved. Some of these detection results appear promising against the accuracy rate. Our aim is to implement security approaches in mobile platforms. Mobile platforms in future Will face a lot of challenges:

.Repeated detection of limited resources may affect network bandwidth either in power or in network mobile.

.Whether it is possible to use minimal data or a network to detect phishing.

. Apply all manipulations to the survey to insert and reduce all functionality in the current browser.

Resource consumption on the mobile user side | So we keep trying to overcome all the current limitations and are doing this and for the security of mobiles we are deploying a more suitable solution and trying to overcome the challenges successfully.

REFERENCES:

[1] A. Aouad, Mobile banking is risk of the in the UK – Business Insider, 2017. [Online] Available:http://uk.businessinsider.com/mobilebanki ng- is-risk-onthe-in-the-uk-2017-6. [Accessed: 17-Nov-2018].

[2] A. P. Felt and D. Wagner, Phishing on Mobile Devices. in Web 2.0Security and Privacy Oakland, California, 2011.

[3] S. Cook, 50+ Phishing Statistics, Facts and Trends 2017-2018 — Comparitech,2018. [Online]. Available:https://www.comparitech.com/blog/ vpn-privacy/phishing-statistics-facts/#gref. [Accessed: 17-Nov-2018].

[4] T. Vidas, E. Owusu, S. Wang, C. Zeng, L. F. Cranor, and N. Christin, QRishing: The Susceptibility of Smartphone Users to QR Code Phishing Attacks, Springer, Berlin, Heidelberg, 2013, pp. 5269.

[5] B. Regls, Phishing News: QR Code Phishing Scheme, International Journal of Computational Intelligence and Information Security, May 2012 Vol.3,No.5, 2016.

[6] Wandera, Mobile Phishing Report 2018,2017.[Online]. Available: http://go.wandera.com/rs/988-EGM 040/images/Phishing%20%282%29.pdf

[7] Kaspersky, Financial phishing accounts for over 50% of all phishing attacks fo..., 2018. [Online]. Available: https://www.finextra.com/pressarticle/72837/financi al-phishingaccounts- for-over-50-of-all-phishing-attacks-for-the-first-time.[Accessed: 17-Nov-2018].

[8] M K. Krombholz, P. Frhwirt, P. Kieseberg, I. Kapsalis, M Huber, and E. Weippl, QR Code Security: A Survey of Attacks and Challenges for Usable Security, Springer, Cham, 2014, pp. 7990.

[9] V. Sharma, A Study of Malicious QR Codes. .

[10] N. Mazher, I. Ashraf, and A. Altaf, Which web browser work best for detecting phishing, in 2013 5th International Conference on Informationand Communication Technologies, 2013, pp. 15.

[11] Y. Zhang, J. Hong, and to L. Cranor, CANTINA: A Content-BasedApproach to Detecting Phishing Web Sites.

[12] M. S. I. Mamun, as M. A. Rathore, A. H. Lashkari, N Stakhanova, andA. A. Ghorbani, and Detecting Malicious URLs Using Lexical Analysis,Springer, Cham, 2016, pp. 467482.

[13] J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, Beyond blacklists: learning to detect malicious web sites from suspicious URLs, in Proceedings of the 15th ACM SIGKDD international conference on Knowledge to discovery and data mining - KDD 09, 2009, p. 1245.

[14] BBC, GDPR risks making it harder to catch hackers - BBC News, 2018. [Online]. Available: https://www.bbc.co.uk/news/technology-44290019. [Accessed: 17-Nov-2018].

[15] S. Egelman, L. F. Cranor, and J. Hong, Youve been the warned, in Proceeding of the twenty-sixth annual CHI conference on Human of factors in computing systems - CHI 08, 2008, p. 1065.

[16] Y. Niu, F. Hsu, and the H. Chen, iPhish: Phishing of Vulnerabilities on Consumer Electronics.

[17] K. Peng, H. Sanabria D. Wu, and C. Zhu, Security Overview of QR Codes. - Student project in the MIT course 6.S57, ’14.

[18] P. Kieseberg et al., QR Security Code . Proceedings to the 8th International Conference on Advances in Mobile phone Computing and Multimedia, November 08-10, 2010, Paris, France.

[19] H. Yao and D. Shin, Towards preventing QR code based attacks on android mobile using security warnings, in Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communicationssecurity -- ASIA CCS 13, 2013, p. 341.

[20] S. Chugh, Why Google is Forcing You To Have SSL Certificate onYour Websites 2018. [Online]. Available: https://serverguy.com/security/ google-forcing-ssl-certificate-websites/. [Accessed: 17-Nov-2018].

[21] Wikipedia, Same Origin Policy. W3C.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)