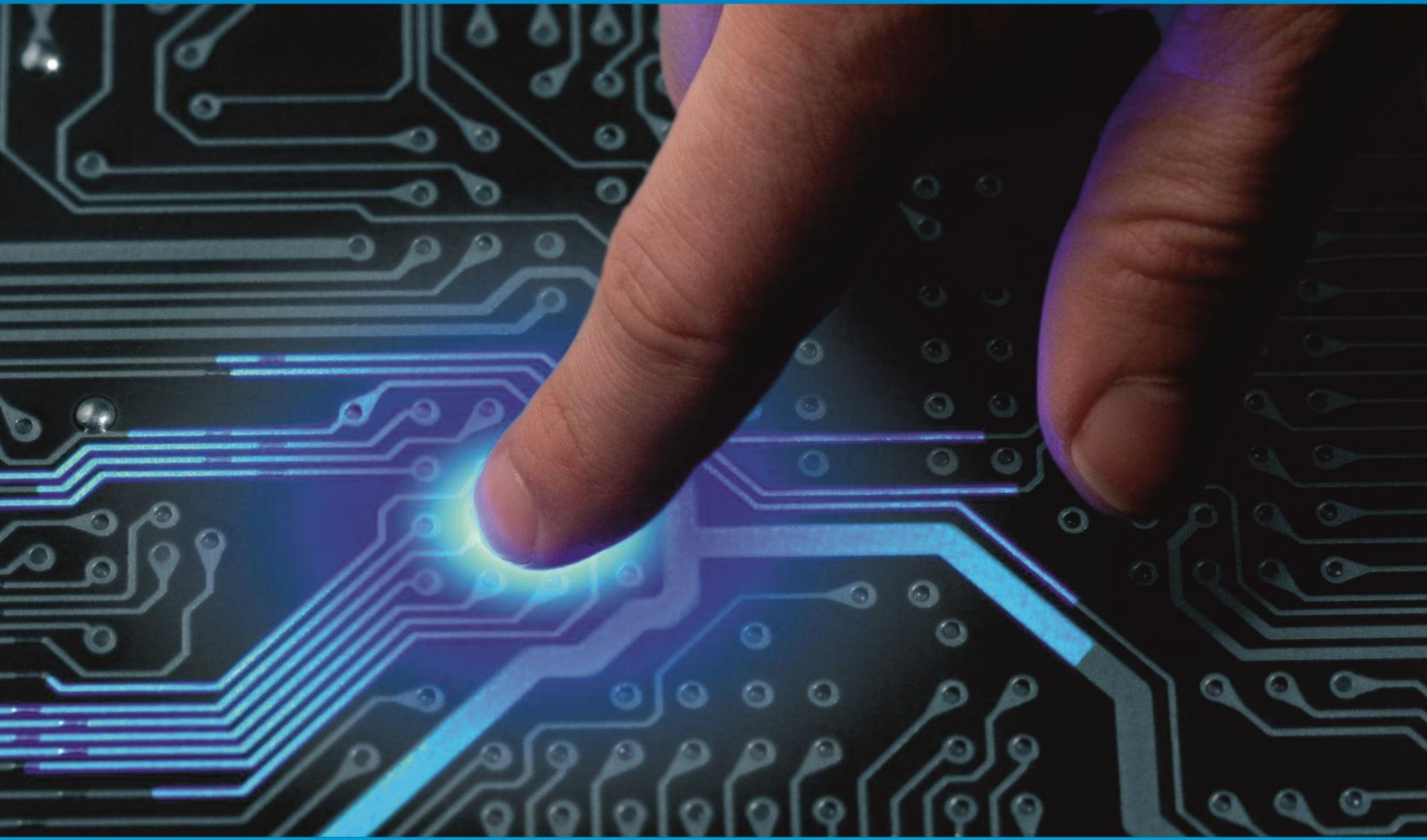
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 8, Issue 10, October 2020

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

Analyzed Study about Phishing& Detection Techniques Based on Visual Similarities

Umaima Sagir Pathan

Student, Dept. of I.T. , B.K.Birla College of Arts, Science, and Commerce(Autonomous), Kalyan , India

ABSTRACT: Phishing is like a weapon used to grab information like usernames, passwords & even personal data, by disguising oneself as a trustworthy entity in transmission (email spoofing, instant messaging, fraud websites, etc.). Phisher makes sure that the web site should have an equivalent look & feel of the legitimate website so that the user will easily get trapped by the attacker and may get the knowledge he requires. Phishing may be a part of social engineering which is employed by the attacker to trap the users. A phishing website may be a spoofed website which is an exact copy of the legitimate webpage. And actually, it's just a front from where the phishers can acquire the passwords or IDs of the user or the opposite confidential data. Then this information is employed by the attacker to use the accounts of the user or attack financially. So during this paper, we are getting to have an analyzed study of various sorts of phishing detection techniques supported visual similarity.

KEYWORDS : Phishing, History of phishing, Visual similarity approaches.

I. INTRODUCTION

Phishing is like a weapon used to grab information like usernames, passwords & even personal data, by disguising oneself as a trustworthy entity in transmission (email spoofing, instant messaging, fraud websites, etc.). Phisher makes sure that the web site should have an equivalent look & feel of the legitimate website so that the user will easily get trapped by the attacker and may get the knowledge he requires. Phishing may be a part of social engineering which is employed by the attacker to trap the users. Phishing was first introduced in the 1980s. Then also till today phishing is considered to be one of the most serious threats. After having such an advanced technological world still, the phishers find out the vulnerabilities of the solution to make the attack successful.

Phish is not only spelled like the word Fish but its concept is also similar to a fishing method like a baited hook is thrown in the form of fake site and it is expected you to get trapped.Phishing attempts are of various methods like email, SMS, websites, etc. Their mainly two types of phish webpages one is fake sites and another is scammed one.

A fake website is made in such a way that the user will not be able to understand that it is a fake duplicate of the legitimate website in appearance. To recognize such websites some technologies are introduced based on visual similarity to detect the phished site. In this paper following content will be explored:

- ➢ History of phishing.
- ➢ How phishing works.
- ➢ Visual similarity approaches.
- ➢ Conclusion.

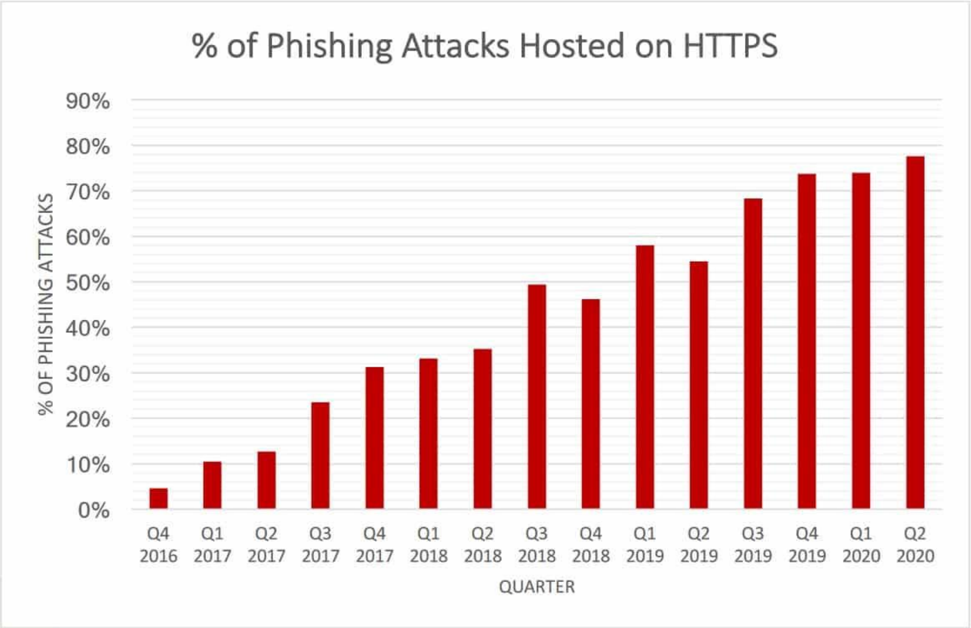
II.HISTORY OF PHISHING

Phishing scams are a kind of fraudulent attempt that mostly many of the users easily get trapped. It was firstly introduced in the year 1980s in a paper in detail and also the presentation was delivered in 1987 to International Users Group, Interex. And in the 1990s there was a very well scammed took placed at AOL known as “Early AOL Phishing”.

In which AOHell was designed to program and was associated with Warez through which the accounts of AOL users were hacked. And to get the sensitive information of the user they used to send the message to the user like “ verify your account" or " confirm billing information".

Once the user enters the password, the attacker can access their account for his / her malicious activities. And AOHell was programmed in early 1995.

In 2001 a payment system has affected the e-gold in June that was followed by "post 9/11" just after the attack on the world trade center on September 11. In Jan 2009, an attack was done which has followed into an unauthorized wire transfer of US$ 1.9 Million by Experi-metal's online banking. In Aug 2015, the Fancy bear has spoofed Electronic Frontier Foundation. In 2018, a company has developed the EOS.IO blockchain which was attacked by the sssphishing group to intercept the user's cryptocurrency wallet key.



Source: APWG

Fig 1: This is an analyzed status of the phishing attack done on HTTPS which is versusHTTP. It is reported by APWG.

III. **HOW PHISHING WORKS**

A phished email is sent to the victim with malicious attachments. One victim has open the mail it seems trustworthy to the victim and then he/she trust that mail assuming that it belongs to the bank or any authorized department. And then the user sends the sensitive data in reply. And he/she is thinking that the information is sent to bank authorization and it is safe. But in reality, the information is sent to the attacker. Then the attacker uses the information for his malicious activities like hacking the account and misusing it, even the user is financially robbed. Even fake websites are developed in such a way that the victim is not able to understand that is legitimate or fake. Thinking that it legitimates the victim to enter the ID and the password of any like Facebook, Instagram, etc. And that account is used by the attacker for their malicious activities.

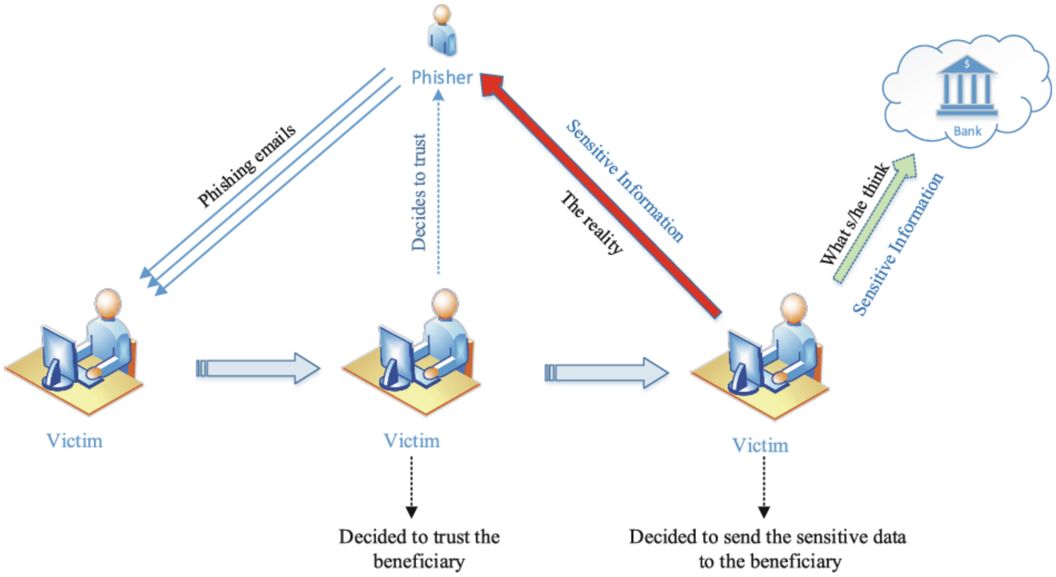
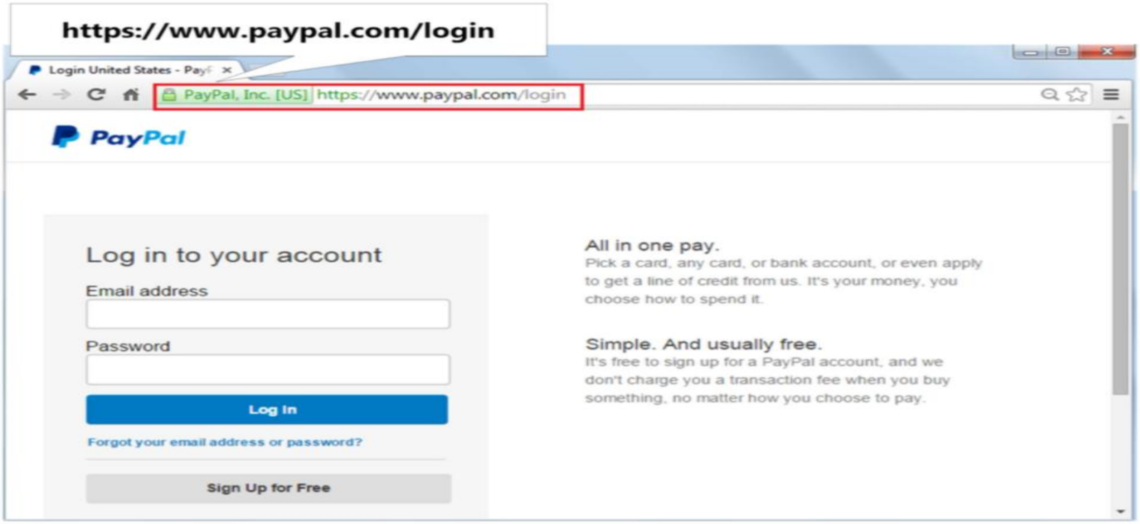


Fig 2: Structure of phishing attack.

Like emails, SMS, URL are phished the same way websites are also phished.A user can easily become a victim of the phishing attack by looking at the high visual resemblance of the phishing website with the targeted legitimate site, like page layouts, images, text content, font size, and font color. For example, the fake and the legitimate PayPal webpages, both pages have the same visual appearance but its URLs are different.



https://www.paypal.com/login

Login United States - PayP
PayPal, Inc. [US] https://www.paypal.com/login
PayPal
Log In
Sign Up for Free

Log in to your account

Email address

Password

Forgot your email address or password?

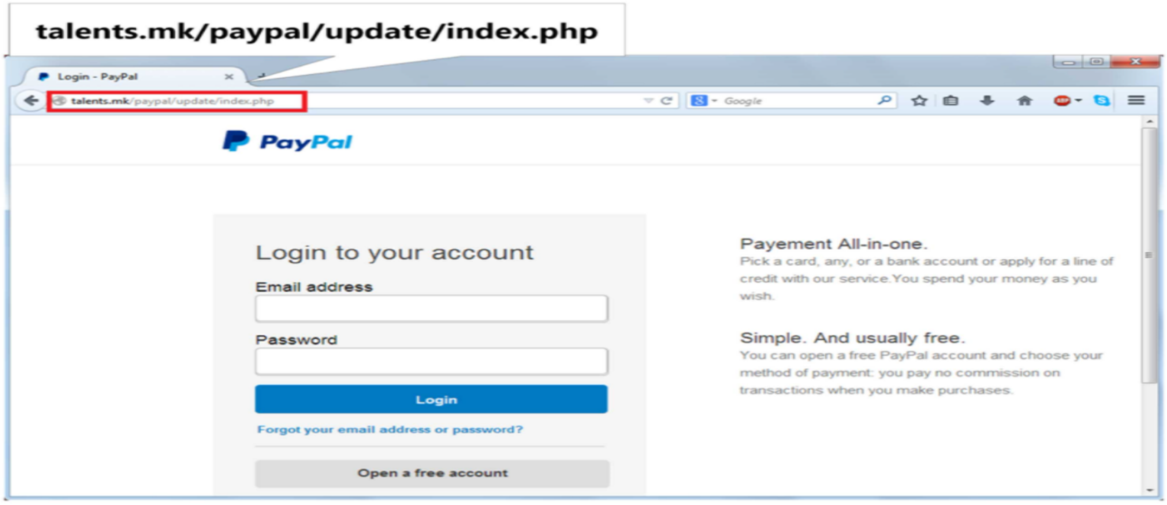
All in one pay.

Pick a card, any card, or bank account, or even apply to get a line of credit from us. It's your money, you choose how to spend it.

Simple. And usually free.

It's free to sign up for a PayPal account, and we don't charge you a transaction fee when you buy something, no matter how you choose to pay.

(a)


talents.mk/paypal/update/index.php

Payement All-in-one.

Pick a card, any, or a bank account or apply for a line of credit with our service.You spend your money as you wish.

Simple. And usually free.

You can open a free PayPal account and choose your method of payment: you pay no commission on transactions when you make purchases.

(b)

Fig3 : (a) is the legitimate PayPal webpage., (b) is the phishing PayPal webpage.

It is not always necessary that the people carefully notice on URL and SSL (Secure Socket Layer) certificate of internet sites. If an attacker doesn't copy the visual appearance of the targeted website well, then chances of inputting credentials by Internet users are very less.

An attacker can misguide the victim in the following ways :

1)Visual appearance: the fake website is a carbon copy of the legitimate website. With the help of the HTML source code of the genuine website the attacker generates the fake website.

2)Address Bar: with the help of script or image the attacker also covers the address or URL and make sure that the victim should believe that they are inputting their details on the safe website.

3)Embedded objects: attacker takes the help of embedded object like images or script to hide the textual content and the HTML coding to prevent the detection approaches.

4)Favicon similarity: favicon images are the signature image or symbol of the particular website. The attackercan copy the favicon image of that particular website.To increase the resemblance of the website.

A Taxonomy is presented for different types of detection techniques based on Visual similarity :

<u>**Name of the technique**</u>	<u>**Description**</u>	<u>**How it works**</u>	<u>**Conclusion**</u>
PhishZoo detection technique	This approach is the combination of whitelisting approach to identify the new targeted phishing website with a blacklisting & heuristic approach to alert the user. This method can also be combined with other whitelists, blacklist, and heuristic approaches to get a	The profile maker makes the profile of a real website and stores it. The stored profile contains data like SSL,URL, images, HTML contents & scripts. After that when the browser loads a website its contents are extracted. The SSL &URL are used to detect the whitelisted sites, if SSL & URL matches with the stored profile then it is a real site, if not then	This approach gives almost accurate result against sites that looks most likely to the legitimate one. It gives 96% accurate results & it can also detect the zero-hour phishing attack.

	more accurate result. It mainly works on text, SSL comprehension, and images.	he other content(images, HTML content, scripts) are compared with the stored profile, if the content doesn't matches then it is not a phishing site and if matches it will give a warning to the user that it can be a phishing site.	
BaitAlarm detection method	The main motive of the approach is to detect phishing pages based on the elemental feature of the phishing pages, that is, the similarity in-page visual layouts. To effectively attract victims, attackers usually try their best to form a phishing page look similar to the target page.	Phase 1: In this phase extraction of the CSS structure of the suspected website is done. Then a normalized model is prepared for the comparison-unit of the suspected webpage. Phase 2: After obtaining the normalized model, they match the 2 comparison-unit, to compute the similarity of the suspected webpage and victim webpage. Phase 3: If the similarity of the suspected page and victim page is beyond the pre-set threshold then both the pages are considered to be the same. And if they get more evidence like URLs of the two pages to have a different domain then the suspectedwebpage is concluded as a phishing webpage.	It has CSS based comparison.
Utilization of website logo approach	This approach mainly consists of two steps namely logo extraction and identity verification. With the help of machine learning, it extracts the logo and utilizes google image search.	Step1- Logo Extraction: In this process, the site is detected and the logo image from all downloaded image resources of the website is extracted. Then to identify the real logo image they use machine learning techniques. Step2-Identity Verification: Then the extracted logo image in this process takes on by the google image search to get the portrayed identity.	So it is concluded that a logo extraction process can improve the accuracy of the overall phishing detection.
The phishing detection technique is based on visual similarity without victim site information.	This approach is based on image processing. And doesn't require a dataset of the legitimate website.	Firstly URL is inserted as an input in the system to acquire the domain name of the URL and take the images mentioned in the web browser. Then with the help of imgSeek, the system searches for the image dataset with the image displayed in the web pages. If similar images are found from the image dataset then the threshold value of the image dataset, domain name, and name in the dataset of the image is compared.If the domain name corresponds to a name within the	This mechanism is based on the visual similarity of phishing sites that are the duplicate of the same victim site. And surprisingly it can be concluded as 224 distinct web page layouts duplicated by 2,262 phishing sites and gain a detection rate of over 80% by keeping the false-

		image database, we will consider the location of the inputURL to be the location registered within the image database.The proposed system outputs an equivalent result as a registered site. And if there are no similarimages whose similarity exceeds the threshold valuein the image database, the proposed system returns the result as unknown.	positive rate upto17.5 %.
Goldfish approach	The approach, called GoldPhish, uses a browser plug-in to detect and report phishing sites. We do this by using optical Character recognition (OCR) to read the text from an image of the page, for example, the company logo, grabbing the top-ranked domains from an inquiry engine and comparing them with the current web site.	This model is divided into three steps 1) To take the screenshot of the webpage that is active on the user's web browser. 2) Then with the help of an optical character recognition technique to convert the screenshot into computer-readable language. 3) Then in this step, the converted text is inserted as input in the search engine to get the result.	After testing the conclusion can be stated as on 100 legitimate sites and 100 phishing sites, the accuracy reported is 100% of legitimate sites and 98% of phishing sites.

IV.CONCLUSION

Phishing is a malicious cybercrime, through which victims' confidential data are gathered with the help of the victim only, by misguiding the victim that the victim is using trustworthy legitimate websites. In this paper, we studied what is phishing, its history, how it works, and five different types of phishing detection techniques based on visual similarity approaches. Here we conclude that to avoid phishing detection techniques, attackers usually insert images, Flash, ActiveX, and Java Applet in situ of HTML text. with the help of Visual similarity-based detection approaches, it is very easy and convenient to detect such embedded objects used in the phishing webpage.

Visual similarity-based techniques use a signature to spot phishing webpages. The signature is formed by taking common features from the whole website rather than one webpage. Therefore, one signature is sufficient to detect various targeted webpages of 1 website or different versions of an online site.

V.ACKNOWLEDGMENT

It gives us immense pleasure to express our deepest sense of gratitude and sincere thanks to our highly respected and esteemed thesis guide Prof. Swapna Augustine Nikale, Department of Information Technology ; B.K Birla College (Autonomous), Kalyan , for her valuable guidance, encouragement and help for completing this work. Her useful suggestions for the whole work and co-operative behavior are sincerely acknowledged.

REFERENCES

1. Gupta, B.B., Arachchilage, N.A.G. & Psannis, K.E. Defending against phishing attacks: a taxonomy of methods, current issues, and future directions. *Telecommun Syst* **67,** 247–267 (2018). https://doi.org/10.1007/s11235-017-0334-z.
2. Das, Avisha & Baki, Shahryar & El Aassal, Ayman & Verma, Rakesh & Dunbar, Arthur. (2019). SoK: A Comprehensive Reexamination of Phishing Research from the Security Perspective. IEEE Communications Surveys & Tutorials. PP. 1-1. 10.1109/COMST.2019.2957750.
3. A. El Aassal, S. Baki, A. Das, and R. M. Verma, "An In-Depth Benchmarking and Evaluation of Phishing Detection Research for Security Needs," in IEEE Access, vol. 8, pp. 22170-22192, 2020, DOI: 10.1109/ACCESS.2020.2969780.
4. Bhavsar, Vaishnavi & Kadlak, Aditya & Sharma, Shabnam. (2018). Study on Phishing Attacks. International Journal of Computer Applications. 182. 27-29. 10.5120/ijca2018918286.
5. Ferreira, A., & Teles, S. (2019). Persuasion: How phishing emails can influence users and bypass security measures. *International Journal of Human-Computer Studies*, *125*, 19–31. https://doi.org/10.1016/j.ijhcs.2018.12.004.
6. A. K. Mishra, A. K. Tripathy, and S. Swain, "Analysis and Prevention of Phishing Attacks in Cyber Space," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 430-434, DOI: 10.1109/ICSCCC.2018.8703343.
7. J. Kumar, A. Santhanavijayan, B. Janet, B. Rajendran, and B. S. Bindhumadhava, "Phishing Website Classification and Detection Using Machine Learning," 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020, pp. 1-6, DOI: 10.1109/ICCCI48352.2020.9104161.
8. Rao, R.S., Pais, A.R. Detection of phishing websites using an efficient feature-based machine learning framework. *Neural Comput & Applic* **31,** 3851–3873 (2019). https://doi.org/10.1007/s00521-017-3305-0.
9. A. Borkar, A. Donode, and A. Kumari, "A Survey on Intrusion Detection System (IDS) and Internal Intrusion Detection and protection system (IIDPS)," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, 2017, pp. 949-953, DOI: 10.1109/ICICI.2017.8365277.
10. Vasquez, M. H., & Barber, K. S. (2017). The Financial Crimes Management of Account Takeover Fraud. *The University of Texas at Austin December 2017*, 1–53. http://hdl.handle.net/2152/63762.
11. Xiang, G., Hong, J., Rose, C. P., & Cranor, L. (2011). CANTINA+: A feature-rich machine learning. Framework for detecting phishing websites. *ACM Transactions on Information and System Security*, *14*(2), 1–28. https://doi.org/10.1145/2019599.2019606.
12. Nawade, S., Wankhede, S., Bhaspale, D., & Sathwane, S. (2019). A phishing detection framework for websites based on data mining. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *5*(2), 651–656. https://doi.org/10.32628/CSEIT1726141.
13. Aneke, J., Ardito, C., & Desolda, G. (2020). Designing an Intelligent User Interface for Preventing Phishing Attacks. *Beyond Interactions*, 97–106. https://doi.org/10.1007/978-3-030-46540-7_10.
14. Varshney, G., Misra, M., & Atrey, P. K. (2016). A survey and classification of web phishing detection schemes. *Security and Communication Networks*, *9*(18), 6266–6284. https://doi.org/10.1002/sec.1674.
15. Jain, A., & Gupta, B. (2017, January 10). Phishing Detection: Analysis of Visual Similarity-Based Approaches. Retrieved November 15, 2020, from https://www.hindawi.com/journals/scn/2017/5421046/.
16. S. Afroz and R. Greenstadt, "PhishZoo: Detecting Phishing Websites by Looking at Them," 2011 IEEE Fifth International Conference on Semantic Computing, Palo Alto, CA, 2011, pp. 368-375, DOI: 10.1109/ICSC.2011.52.
17. J. Mao, P. Li, K. Li, T. Wei, and Z. Liang, "BaitAlarm: Detecting Phishing Sites Using Similarity in Fundamental Visual Features," 2013 5th International Conference on Intelligent Networking and Collaborative Systems, Xi'an, 2013, pp. 790-795, DOI: 10.1109/INCoS.2013.151.
18. K. L. Chiew, E. H. Chang, S. N. Sze, and W. K. Tiong, "Utilisation of website logo for phishing detection," *Computers & Security*, vol. 54, pp. 16–26, 2015.
19. M. Hara, A. Yamada, and Y. Miyake, "Visual similarity-based phishing detection without victim site information," 2009 IEEE Symposium on Computational Intelligence in Cyber Security, Nashville, TN, 2009, pp. 30-36, DOI: 10.1109/CICYBS.2009.4925087.

20. M. Dunlop, S. Groat, and D. Shelly, "GoldPhish: Using Images for Content-Based Phishing Analysis," 2010 Fifth International Conference on Internet Monitoring and Protection, Barcelona, 2010, pp. 123-128, DOI: 10.1109/ICIMP.2010.24.
21. Website Phishing Dataset .: https://www.kaggle.com/ahmednour/website-phishing-data-set
22. Phishing Website Detector.: https://www.kaggle.com/eswarchandt/phishing-website-detector
23. Detection of Phishing Website Using ML.:https://www.kaggle.com/shubham9696/dectecting-phishing-website-using-machine-learning
24. M. Khonji, Y. Iraqi and A. Jones, "Phishing Detection: A Literature Survey," in IEEE Communications Surveys & Tutorials, vol. 15, no. 4, pp. 2091-2121, Fourth Quarter 2013, DOI: 10.1109/SURV.2013.032213.00009.
25. A. C. Bahnsen, E. C. Bohorquez, S. Villegas, J. Vargas, and F. A. González, "Classifying phishing URLs using recurrent neural networks," 2017 APWG Symposium on Electronic Crime Research (eCrime), Scottsdale, AZ, 2017, pp. 1-8, DOI: 10.1109/ECRIME.2017.7945048.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

📱 9940 572 462 ✆ 6381 907 438 ✉ ijircce@gmail.com

www.ijircce.com

Scan to save the contact details