International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 04 | Apr 2019 www.irjet.net p-ISSN: 2395-0072

Phishing Website Detection System

Shaswat Kumar[1], Musquan Karovalia[2], Bhagyashree Prajapati[3], Shraddha Kadam[4], Aabha Patil[5]

[1,2,3,4]*Student* [5]*Assistant Proffesor*
[1,2,3,4,5]*Department of Computer Engineering, Shree L.R Tiwari College of Engineering,Mumbai University, India*

Abstract - *Phishing is an unlawful activity wherein people are misled into the wrong sites by using various fraudulent methods. The aim of these phishing websites is to confiscate personal information or other financial details for personal benefits or misuse. As technology advances, the phishing approaches used need to get progressed and there is a dire need for better security and better mechanisms to prevent as well as detect these phishing approaches. The primary focus of this paper is to put forth a model as a solution to detect using chrome extension. The chrome extension will help to keep the user secured as well as allows to give an alert message to the user. There are two major phases using chrome extension such as comparing the url with the phishtank database and comparison with the Indexed DB to give better results.*

***Key Words*: Malicious URL's, Phishing, Web Security, Indexed DB, Chrome Extension .**

1. INTRODUCTION

Computer security is a vast field consisting of many different methodologies and technologies. It defends against a variety of attack vectors, both internal and external. Phishing is a significant security threat to the internet, it is an electronic online identity theft in which the attackers use spoofing techniques like fake websites that mimic legal websites to trick users into revealing into their private information. As this global impact of phishing attacks will continue to be on the increase and thus requires more efficient phishing detection techniques to curb the menace. Over the years, phishing attacks grew in number and intensity too. Phishing attacks now target users of online banking, payment services such as PayPal, and online e-commerce sites.

A Detection of Phishing Website System has been developed. Our system uses a chrome extension which fetches the url provided by the user and compares it with the database which will perform a quick scan for existing URLs. If it concludes that the received url is malicious, it will change the color to red. If the tool concludes that the url is legit. It will let the user to open the webpage.

This paper is organized into five sections. The first section gives a brief introduction about the system. The second section is the study of related existing systems. The third section details out the implementation of the system. The fourth section provides the results obtained. Finally the conclusion gives the summary and future scope about the system.

2. LITERATURE REVIEW

In Literature review, we discuss about the various aspects of the project by taking reference of the existing projects that are similar to the makers of this current project.

[7]Hossein Shirazi, Kyle Haefner, Indrakshi Ray developed a framework, called "Fresh-Phish", for creating current machine learning data for phishing websites. Using 30 different website features using python, they built a large labeled dataset and analyzed several machine learning classifiers against this dataset to determine which is the most accurate.

[8]Tara Baniya, Dipesh Gautam and Yoohwan Kim have performed this paper that proposes a cyberspace has opened a new platform for criminal activities because people have to provide the most private information such as user name, password, credit card information, social security number. Paper was surveyed on various methods of cyber attacks, attempts to mitigate them, their strength and weakness. Also discussed on generic architecture of URL blacklisting that are being used by various malware detection system.

[9]Ankit Kumar Jain and B. B. Gupta Visual similarities based techniques are very useful for detecting phishing websites looks very similar in appearance to its corresponding legitimate website to deceive users into believing that they are browsing the correct website. This utilise the feature set like text content, text format, HTML tags, Cascading Style Sheet (CSS), image, and so forth, to make the decision. These approaches compare the suspicious website with the corresponding legitimate website by using various features and if the similarity is greater than the predefined threshold value then it is declared phishing. Paper presents a comprehensive analysis of phishing attacks some of the recent visual similarity based approaches for phishing detection. Survey provides a better understanding of the problem, current solution space, and scope of future research to deal with phishing attacks efficiently using visual similarity based approaches.

[1]A detection technique for phishing websites is proposed by Abdulghani Ali Ahmed et al which examines Uniform Resources Locators (URLs) of suspected web pages as per five extracted features. Phishtank and Yahoo directory datasets are used to assess the accuracy of the results given by proposed solution. The final report thus establishes that the detection mechanism can detect various types of

phishing attacks without fail. However, there are still chances of receiving false alarms.

3. DATABASE.

3.1 PHISHTANK DATABASE.

Phishtank Database is stored on the server side.Phishtank Database consists of all top phishy sites. As Phishtank shares its database by providing API key we have downloaded its data. We update the database frequently. We use this database to compare it with the URL fetched.

3.2 INDEXED DB

IndexedDB is an asynchronous, transactional, key-value object store. Asynchronous means that IndexedDB won't block the user interface. Transactional means that operations in IndexedDB are all-or-nothing. IndexedDB is stored on the client's side,. A key value object store means that each record is an object, as opposed to a row. In a key-value object store, each record is a self-contained object. It may, but usually doesn't have a relationship to records in another object store. We created this database by monitoring the activity of the user with the frequently used websites and stored it in this database. The Indexed DB data is stored on the client side.

4. PROPOSED SYSTEM

The Phishing Website Detection system notifies the user whether the site he or she is using is a phish or malicious site or not depending upon the match found in our database.

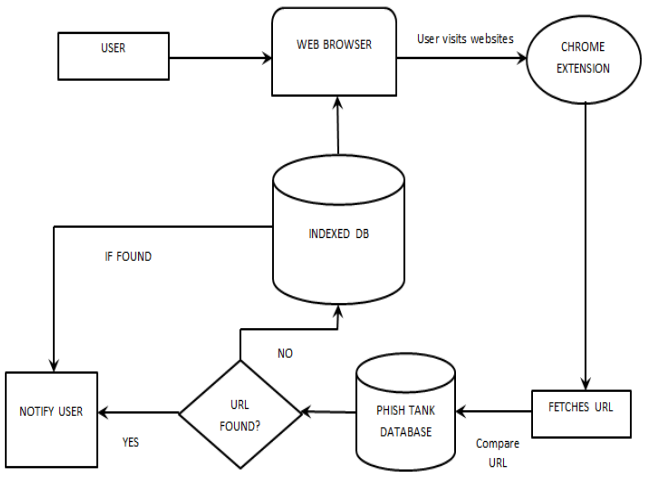


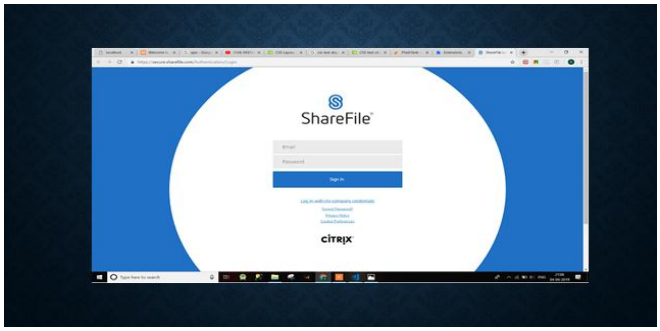
Fig -1: System Architecture

5. IMPLEMENTATION OF THE SYSTEM

Here we will discuss about how we implemented our system and is represented in a flowchart manner in Figure 1.

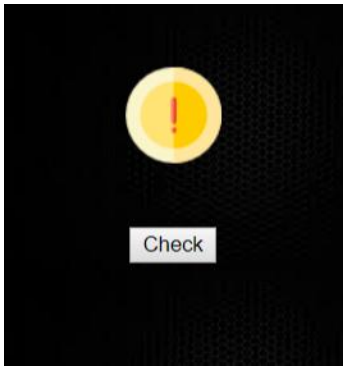
Step 1: USER OPENS WEB BROWSER ANS VISITS WEBSITE

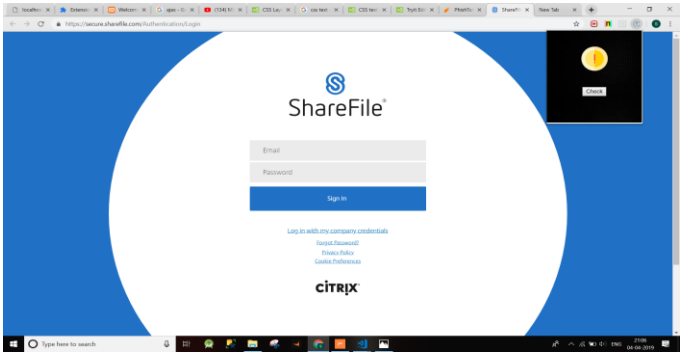
In Fig. 1we can see that first the user opens web browser and visits any website.



Step 2: CHROME EXTENSION

We have created a chrome extension which will start its functioning after the user clicks on check button on the chrome extension.





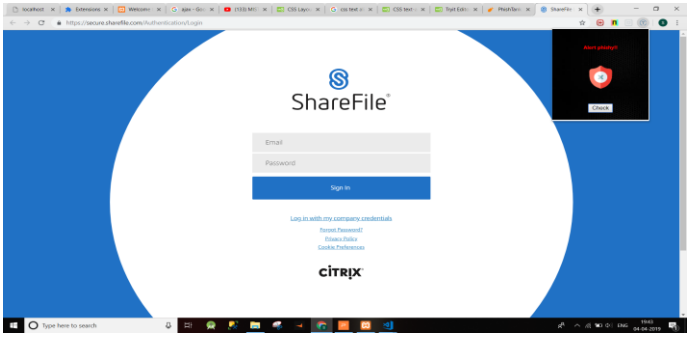
If the URL is a malicious URL, the chrome extension will change its color from yellow to red and if the URL is not a malicious URL, the chrome extension will change its color from yellow to green.

Step 3: FETCHES URL:

When the user visits the website the chrome extension will fetch the url.

Step 4: PHISH TANK DATABASE

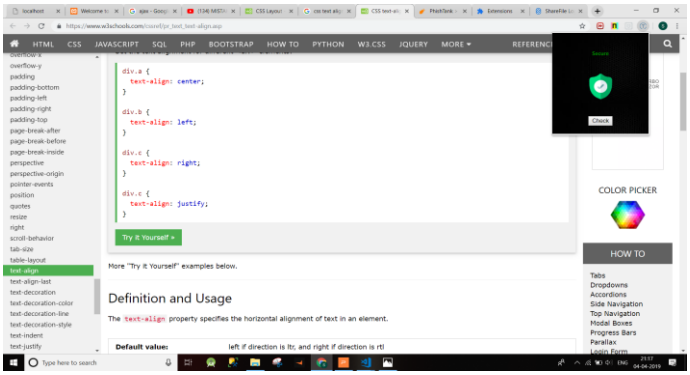
After the chrome extension fetches URL, it will be compared with the phish tank database. This database contains URL, creation date, etc of the phish or malicious websites. If the URL is found then the user will be notified by the chrome extension by change in the color from yellow to red.



If the URL is not found it will be directed to Step 5.

Step 5: INDEXED DB

Indexed DB is a database stored at client side. We have created this database by monitoring the activity of the user. The sites which the user frequently uses is stored in this database. If the URL is not found in step 4,it will come to step 5 and check the URL here. If the match is found it will notify the user.



6. RESULTS

Following is the list of few Phishy URL's which we tested

URL	SYSTEM RESULT	NETCRAFT RESULT
https://sites.google.com/site/freehabbocoinsgbbo00/	Detected	Not Detected
http://themedsmarket.ru/	Detected	Detected
http://www.tunga9.cl/cp/	Detected	Detected
http://princestudio.net/tienda/libraries/joomla/BT/index.php	Not Detected	Detected
http://www.tunga9.cl/cp/	Detected	Detected
http://www.gkjx168.com/images/	Detected	Detected
http://paypal.com.cg-bin.login.submited.elkklkdgh54m54k565vdflgqdfaziou456789.begiu.com/	Detected	Not Detected
http://jppost-afu.com:81/pp.html	Detected	Detected

The above table contains only few URL's that we have tested. We tested total 53 samples and the accuracy and the error is as follows:

Total samples tested =53
correctly classified samples = 49
Incorrect classified samples = 4

Accuracy = correctly classified samples/Total samples
= 49/53
= 92.45%

Error = Incorrectly classified samples/Total samples
= 4/53
= 7.54%

7. CONCLUSION

Phishing is the one the major cyber threats nowdays . Through this system we have developed a chrome extension which will fetch the URL and compare it with the existing database of system .The main aim of this system is to detect the phishing website and alert the user by changing the color from yellow to red if it is a phishy website or changing color from yellow to green if it is not a phishy website so as to prevent the user from sharing or providing their sensitive information to the attackers.

REFERENCES

[1] P. Hayati and V. Potdar, "Toward Spam 2.0: An Evaluation of Web2.0 Anti-Spam Methods "In 7th IEEE International Conference onIndustrial Informatics Cardiff, Wales, 2009.

[2] P. Hayati, K. Chai, V. Potdar, and A. Talevski,"HoneySpam 2.0:Profiling Web Spambot Behaviour," In 12th International Conference on Principles of Practise in Multi-Agent Systems, Nagoya, Japan, 2009, pp. 335-344.

[3] P. Szor. "The art of computer virus research and defence." Addison-Wesley Professional, 2005.

[4] P.N. Tan and V. Kumar, "Discovery of Web Robot Sessions Based on their Navigational Patterns", Data Mining and Knowledge Discovery, vol. 6, pp. 9-35, 2002

[1] M. Khonji, Y. Iraqi, and A. Jones, "Phishing detection: literature survey," IEEE Communications Surveys & Tutorials, vol. 15, no. 4, pp. 2091–2121, 2013.

[2] R. Islam and J. Abawajy, "A multi-tier phishing detection and filtering approach," Journal of Network and Computer Applications, vol. 36, no. 1, pp. 324–335, 2013.

[3] A. K. Jain and B. B. Gupta, "Comparative analysis of features based machine learning approaches for phishing detection," in Proceedings of the 10th INDIA-COM, New Delhi, India, 2016.

[4] G. Weaver, A. Furr, and R. Norton, Deception of Phishing: Studying the Techniques of Social Engineering by Analyzing Modern-Day Phishing Attacks on Universities, 2016.

[5] Kaspersky Lab, "Spam in January 2012 love, politics an sport,"2013,http://www.kaspersky.com/about/news/spam / 2012/Spam in January 2012 Love Politics and Sport.

[6] APWG Q1-Q3 Report, 2015, http://docs.apwg.org/reports/ apwg trends report q1-q3 2015.pdf.

[7] Hossein Shirazi, Kyle Haefner, Indrakshi Ray, "Fresh-Phish: A Framework for Auto-Detection of Phishing Websites", 2017 IEEE International Conference on Information Reuse and Integration 978-1-5386-1562-1/17 $31.00 © 2017 IEEE DOI 10.1109/IRI.2017.40

[8] Tara Baniya, Dipesh Gautam, Yoohwan Kim, "Safeguarding Web Surfing with URL Blacklisting", 2015 12th International Conference on Information Technology - New Generations 978-1-4799-8828-0/15 $31.00 © 2015 IEEE DOI 10.1109/ITNG.2015.30

[9] Eric Medvet, Engin Kirda, Christopher Kruegel, "Visual-Similarity-Based Phishing Detection", Copyright 2008 ACM ISBN # 978-1-60558-241-2 ...$5.00

[10] Ahmed, Abdulghani Ali, and Nurul Amirah Abdullah. "Real timedetection of phishing websites." Information Technology, Electronics and Mobile Communication Conference (IEMCON), 7th Annual. IEEE,2016.