Phishing Detection Method using Naïve base Algorithm

A. Gayathri[1] J. Helenprincy[2] P. Leelavathi[3] M. Subasini[4]
[1,2,3,4]Department of Computer Science & Engineering
[1,2,3,4]Tejaa Shakthi Institute of Technology for Women, Coimbatore

Abstract— In computing, phishing is a criminal activity using social engineering techniques. Phishers attempts to abuse sensitive information such as passwords and credit card details by masquerading as a trustworthy person or business in an electronic communication. Phishing is typically carried out using email or an urgent message, although phone contact has been used as well. While the Off-the-Hook technique have language independence, speed of decision and user privacy. They suffer from several drawbacks including time delay, inaccuracy, information retrieval, storage media, reports and enquiry. These drawbacks are slowdown the speed of the browser. To address these limitations, we present a new approach for detecting and identifying phishing URL using Naive Bayes algorithm. It is an efficient and effective technique to identify the phishing websites. In addition, Naive Bayes identifies the target websites that a phishing webpage is attempting to mimic and include this target in its warning.

Key words: Naivebase Algorithm, Phishing Detection Method, Off-the-Hook technique

I. INTRODUCTION

Phishing is the attempt to obtain impassioned information by disguising as a reliable entity in an electronic communication. The word is a neologism created as a homophone of fishing due to the similarity of using a bait in an endeavor to catch an immolation. According to the 2013 Microsoft Computing Safety Index released in February 2014, the annual worldwide consequence of phishing could be high as US$5 billion. Phishing is typically carried out by email spoofing or instant messaging, and it often directs users to enter personal information at a

Fake website, the look and feel of which are identical to the legal one and the only dissention is the URL of the website in diligence. Communication intension to be from social web sites, action sites, banks, online transmission processors or IT administrators are often used to lure hierophant. Fake emails may contain links to websites that apportion malware.

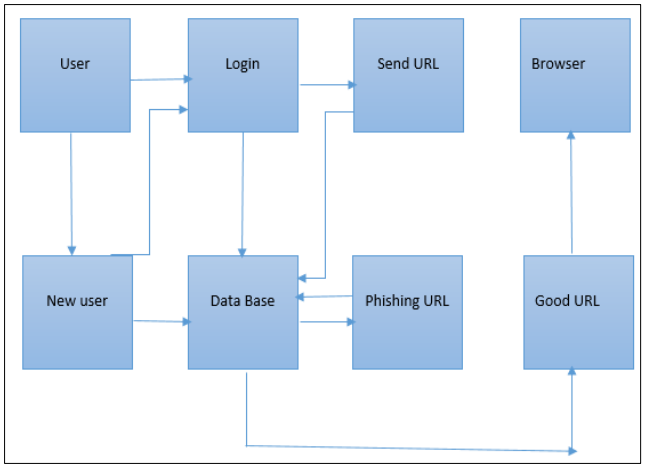

Fig. 1: Architecture diagram

II. PRELIMINARIES

A. *User interface:*

1) Login Module:

Login module will help in authentication of user accounts. Users should have valid login id and password to login into the system

2) Registration Module:

These modules help users to get registration. Using this system user can get registered by filling up the registration form online. After registration student can also edit the personal information.

3) Message Communication:

Texting is one alternative in a larger universe of messaging modalities, including email, instant messaging and messaging within social media platforms (including, for example, Facebook and Twitter). However, texting is differentiated by its broad carrier, platform and device support; simplicity and ease of use; global availability; and cultural pervasiveness. Best known and most popular on wireless (cellular) wide area networks, text messaging (sometimes called texting or wireless messaging) has numerous applications, from casual, consumer-to-consumer communications, to information services and alerts, notifications, premium (paid) services, e-commerce, mobile marketing, healthcare, security and more.

4) Phishing Detection:

To detect phishing from a number of different sources. Since companies are now affected by the recent surge in phishing scams, many companies are trying to protect Consumers using software. Anti-phishing software is available that may identify phishing contents on websites, act as a toolbar that displays the real domain name for the visited website, or spot phishing attempts in email. Spam filters also help protect users from phishers, because they reduce the number of phishing-related emails that users receive. Many organizations have introduced a feature called challenge questions, which ask the user for information that should be known only to the user and the bank. Sites have also added verification tools that allow users to see a secret image that the user selected in advance; if the image does not appear, then the site is not legitimate. However, most of the phishing detection is going to have to be done within the user's browser – and many browser manufacturers are taking steps to help consumers.

III. EXISTING SYSTEM

In the existing system Samuel Marchal make use of Off-the-Hook is as a fully-client-side browser add-on which preserves user privacy. Phishing is a major problem on the web. The attacker's uses phishing emails to distribute malicious links. They propose a new phish detection tool, Off-the-Hook. It is implemented as a browser add-on that can decide in real time if a visited webpage is a phish. The decision process relies solely on information extracted from the web browser while

loading a webpage. In addition, Off-the-Hook identifies the target website that a phishing webpage is attempting to mimic and includes this target in its warning. We evaluated Off-The-Hook in two different user studies.

A. Disadvantage of Existing System

The following are the drawbacks of the existing manual system..,

1) Time Delay:

In the existing system, data related to all avocation is stored in different registers. Since, all the transactions are stored in different schedules, it takes lot of time to prepare variety of reports.

2) Thesaurus:

As the information passes through different registers, each register consolidated and sent to next register. So, the same information is being tabulated at each register which, involves lot of entanglement and gemination in work thus, it causes thesaurus.

3) Correctness:

Since, the same data is compiled at different sections, the possibility of tabulating data wrongly increases. Also, if the data is more validations become difficult. This may result in loss of accuracy of data.

4) Information Retrieval:

As the information is stored in the particular format, it can only be retrieved in the same format. But if it is to be retrieve in different format, it is not possible.

5) Storage Media:

In the existing system, data transaction being stored on too long registers it is very difficult to refer after some time.

6) Reports:

At the various reports are tabulated manually. They are not such Attractive and require more time. They do not provide adequate help in maintaining the accounts.

7) Enquiry:

Enquiry for different level of information is much more difficult. On Line enquiry of data is not possible.

IV. PROPOSED SYSTEM

In the system, we propose that initially observed and studied a 2000 records database including 1000 fake website records built from the Phish Tank database. In this paper, the targeted websites of the phishing attack are primary. Therefore, all the retained 1000 fake website records must contain their appropriate target. Moreover, the studied database consist also of 1000 legal websites which we collected ourselves by combining Alexa's 500 top global website with 500 websites resulted from queries to Google search engine, as for the queries we used to feed our database are (*.bank.*, *.commerce.*, *.trade.*) in notion of the phishing attack and the websites more likely to be ta rgeted. Our analysis shows that the URL portion of interest is composed of several parts.

A. Advantage

- The project uses Naive Bayes algorithm is one of the most common efficient and effective, supervised learning algorithm.
- Naive Bayes test classification is that it is a low cost method for classification.
- It is fast and cheap because it only base on conditional probability calculation.

V. CONCLUSION

The project has met the standards required to work at Web Site. If the business logic remains same the project can be ported to any Website with minor changes in the working procedure of the project. The project can be used as an availability to develop a project for a different company with different business logic wherein the commonalties in certain areas remain the same at any business level. By using the common features in future development the development time as well as the cost of development can be decreased considerably.

To modify the project to Dot net 3.0 and extending this functionality to mobile internet platform using mobile ASP.NET platform by which the restrictions of the software & hardware requirements can be scaled down, which is not possible using ASP.NET 2.0.

REFERENCES

[1] Abdelhamid N., Ayesh A., Thabtah F. (2014) Phishing detection based associative classification data mining. Expert Systems with Applications 41 (13) Pages 5948–5959, Oct 2014.

[2] Abdelhamid N, Ayesh A., Thabtah F. (2013) Phishing Detection using Associative Classification Data Mining. ICAI'13 - The 2013 International Conference on Artificial Intelligence, pp. (491-499). USA.

[3] R. Basnet, A. Sung, and Q. Liu, "Feature selection for improved phishing detection," Advanced Research in Applied Artificial Intelligence, pp. 252–261, 2012.

[4] Cohen W. (1995) Fast effective rule induction. In machine learning: Proceedings of the 12th International conference, pp. 115-123. Lake Tahoe, California. Morgan Kaufmann.

[5] Mohammad R. Thabtah F. McCluskey L., (2015) Phishing websites dataset. Available: https://archive.ics.uci.edu/ml/datasets/Phishing+Websites Accessed January 2016.

[6] Muhammad R., Thabtah F., McCluskey L., (2014) Predicting Phishing Websites based on Self-Structuring Neural Network. Journal of Neural Computing and Applications, (3)1-16. Springer.

[7] Mohammad R., Thabtah F, McCluskey L (2012) An Assessment of Features Related to Phishing Websites using an Automated Technique. In The 7th International Conference for Internet Technology and Secured Transactions (ICITST-2012); 2012; London: ICITST.

[8] Peng, H.C., Long, F., and Ding, C. (2005). Feature selection based on mutual information: criteria of maxdependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (8): 1226–1238. doi:10.1109/tpami.2005.159. PMID 16119262.

[9] Qabajeh I, Thabtah F. (2014) An Experimental Study for Assessing Email Classification Attributes Using Feature Selection Methods. Proceedings of the 3rd IEEE conference on Advanced Computer Science

Applications and Technologies (ACSAT), pp. 125-132, 2014.
[10] Quinlan, J. (1993) C4.5: Programs for machine learning. San Mateo, CA: Morgan Kaufmann

