International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 09 | Sep 2019 www.irjet.net p-ISSN: 2395-0072

Phishing Website Detection based on Machine Learning

Nandini Patil[1], Priyanka Kathare[2]

[1]Professor, CSE, Sharanbasva University, Kalaburagi, India
[2]Student, CSE, Sharanbasva University, Kalaburagi, India

Abstract - *Classification or association Data Mining algorithms based Phishing Detection website is an intelligent and effective model. To identify and characterize all rules and factors in order to classify the phishing website and relationship that correlates them with each other purpose these algorithms used. Phishing websites can be detect them by their performance, accuracy, number of rules generated and speed. This proposed system optimizes the system which is more efficient and faster than existing system. With the help of these two algorithms with WHOIS protocol the error rate of the existing system decreases by 30% so by using this method proposed system create an efficient way to detect the phishing website. This creates a most efficient way to detect the phishing website.*

Keywords: Phishing Detection, Data Mining.

1. INTRODUCTION

A common security threat attack is a Social engineering which is used to reveal private and confidential information. It is implemented by simply tricking the users without being detected. Using this attack we can gain the sensitive information such as username, password and account numbers. Messaging, SMS, VOIP and fraudster emails are the different types communication forms of Phishing attack.

We people have user accounts on different websites including social network, email and also accounts for banking. Many innocent web users are the most vulnerable targets towards this attack. So most people are unaware of their valuable information, which helps to make this attack successful. Such phishing attack exploits the social engineering to tempt to the victim through sending a spoofed link by redirecting the victim to a fake web page. The spoofed link is placed on the popular web pages or sent via email to the victim. Similar to the legitimate webpage a fake webpage is created. Thus, rather than directing the victim request to the real web server, it will be directed to the attacker server.

Antivirus, firewall and designated software do not fully prevent the web spoofing attack by The present solutions. The implementation of Secure Socket Layer (SSL) and digital certificate (CA) also does not protect the web user against such attack. In web spoofing attack, the attacker diverts the request to fake web server. In fact, a certain type of SSL and CA can be forged while everything appears to be legitimate. According to, secure browsing connection does virtually nothing to protect the users especially from the attackers that have knowledge on how the "secure" connections actually work. An anti-web spoofing solution is developing in this system based on inspecting the URLs of fake web pages. To check characteristics of websites Uniform Resources Locators (URLs), this solution developed series of steps. A phishing website's URL have some unique characteristics which make it different from the URLs of the legitimate web page. To determine the location of the resource in computer networks, URL is used here.

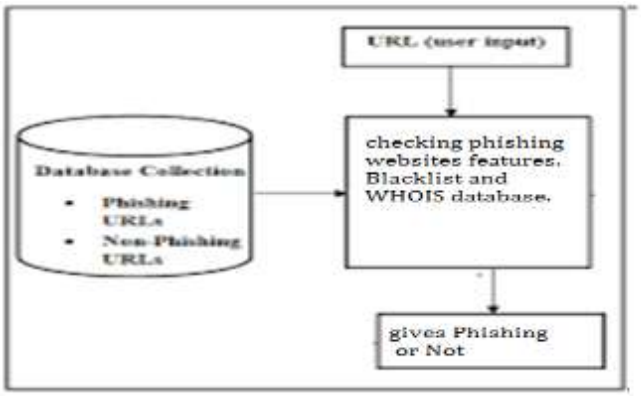

Fig. 1. Block Diagram Detection of Phishing Website

1.1 RELATED WORK

- **JAIN MAO, WENQIAN TIAN and ZHENKAI LIANG** has proposed a system which detect the phishing using page component similarity which analyzes URL tokens to increase prediction accuracy phishing pages typically keep its CSS style similar to their target pages. Based on the observation, a straightforward approach to detect phishing pages is to compare all CSS rules of two web pages, It prototyped Phishing-Alarm as an extension to the Google Chrome browser and demonstrated its effectiveness in evaluation using real-world phishing samples.
- **ZOU FUTAI, PEI BEI and PAN LI** Uses Graph Mining technique for web Phishing Detection. It can detect some potential phishing which can't be detected by URL analysis. It utilize the visiting relation between user and website. To get dataset from the real traffic of a Large ISP. After anonymizing these data, they have cleansing dataset and each record includes eight fields: User node number (AD), User SRC IP(SRC-IP) access time (TS), Visiting URL (URL), Reference URL(REF), User Agent(UA), access server IP (DSTIP), User cookie (cookie). For a client user, he is assigned a unique AD but a variable IP selected from ISP own IP pool. Therefore, we build the visiting relation graph with AD and URL, called AD-URL Graph and the Phishing website is detected through the Mutual behavior of the graph
- **NICK WILLIAMS and SHUJUN LI** proposed a system which analysis ACT-R cognitive behavior architecture model. Simulate the cognitive processes involved in judging the validity of a representative webpage based primarily around the characteristics of the HTTPS padlock security indicator. ACT-R possesses strong capabilities which map well onto the phishing use case and that further work to more fully represent the range of human security knowledge and behaviors in an ACT-R model could lead to improved insights into how best to combine technical and human defenses to reduce the risk to users from phishing attacks
- **XIN MEI CHOO, KANG LENG CHIEW and NADIANATRA MUSA** this system is based on utilizing support vector machine to perform the classification. This method will extract and form the feature set for a webpage. It uses a SVM machine as a classifier which has two phase training phase and testing phase during training phase it extracts feature set and while testing it predict the website is legitimate or a phishing.

1.2 PROPOSED SYSTEM

- Data mining algorithm used in this system provides better performance as compared to other traditional classifications algorithms.
- With the help of this system user can also purchase products online without any hesitation.
- Admin can add phishing website url or fake website url into system where system could access and scan the phishing website and by using algorithm, it will add new suspicious keywords to database.
- System uses machine learning technique to add new keywords into database.
- This section describes the proposed model of phishing attack detection. The proposed model focuses on identifying the phishing attack based on checking phishing websites features, Blacklist and WHOIS database.
- According to few selected features can be used to differentiate between legitimate and spoofed web pages. These selected features are many such as URLs, domain identity, security & encryption, source code, page style and contents, web address bar and social human factor.
- This study focuses only on URLs and domain name features. Features of URLs and domain names are checked using several criteria such as IP Address, long URL address, adding a prefix or suffix, redirecting using the symbol "//", and URLs having the symbol "@".These features are inspected using a set of rules in order to distinguish URLs of phishing webpages from the URLs of legitimate websites.

ADVANTAGES OF PROPOSED SYSTEM

This system can be used by many E-commerce or other websites in order to have good customer relationship.

- User can make online payment securely.
- Data mining algorithm used in this system provides better performance as compared to other traditional classifications algorithms.
- With the help of this system user can also purchase products online without any hesitation.

MODULES

- **Admin**
- **ViewUsers**
- **Addtoblacklist**
- **Viewfeedback**
- **Userregister**
- **Checkurl**
- **Feedback**

2. RESULTS

- **Admin**
  - **View Users** :This is used to view the registered users. The admin can view the details of the registers who registered.
  - **AddToBlacklist** : This module enters the URL, and validates whether it is blacklisted or not using url validation function. If it is blacklisted then it adds it is Yes else No.
  - **View feedback** : This module is used to view the feedback.
- **User**
  - **Register** : This is used to register the users. Using this the user register himself so that they can login to check whether the url is blacklisted or not.
  - **CheckURL** : This is used to check that whether the url is blacklisted or not.
  - **Feedback** This is used to give the feedback.

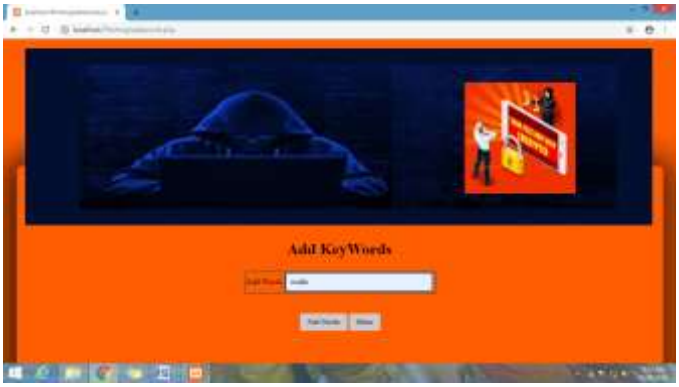

Fig 2. Add Keywords

Fig 3. Add to blacklist

Fig 4.Viewlist

Fig.5 Checkwebsite

3. CONCLUSION

The most important way to protect the user from phishing attack is the education awareness. Internet users must be aware of all security tips which are given by experts. Every user should also be trained not to blindly follow the links to websites where they have to enter their sensitive information. It is essential to check the URL before entering the website. In Future System can upgrade to automatic Detect the web page and the compatibility of the Application with the web browser. Additional work also can be done by adding some other characteristics to distinguishing the fake web pages from the legitimate web pages. PhishChecker application also can be upgraded into the web phone application in detecting phishing on the mobile platform.

REFERENCES

1] JIAN MAO1,WENQIAN TIAN1, PEI LI1, TAO WEI2, AND ZHENKAI LIANG3 Phishing-Alarm: Robust and Efficient Phishing Detection via Page Component Similarity.

[2] Zou Futai, Gang Yuxiang, Pei Bei, Pan Li, Li Linsen Web Phishing Detection Based on Graph Mining.

[3] Nick Williams, Shujun Li Simulating human detection of phishing websites: An investigation into the applicability of ACT-R cognitive behaviour architecture model.

[4] XIN MEI CHOO, KANG LENG CHIEW,DAYANG HANANI ABANG IBRAHIM,NADIANATRA MUSA, SAN NAH SZE, WEI KING TIONG FEATURE-BASED PHISHING DETECTION TECHNIQUE.

[5] Giovanni Armano, Samuel Marchal and N. Asokan Real-Time Client-Side Phishing Prevention Add-on.

[6] Trupti A. Kumbhare and Prof. Santosh V. Chobe an Overview of Association Rule Mining Algorithms.

[7] S.Neelamegam, Dr.E.Ramaraj Classification algorithm in Data mining: An Overview

[8] Varsharani Ramdas Hawanna, V. Y. Kulkarni and R. A. Rane A Novel Algorithm to Detect Phishing URLs.

[9] Jun Hu, Xiangzhu Zhang,Yuchun Ji, Hanbing Yan, Li Ding, Jia Li and Huiming Meng Detecting Phishing Websites Based on the Study of the Financial Industry Webserver Logs.

[10] Samuel Marchal, Giovanni Armano and Nidhi Singh Off-the-Hook: An Efficient and Usable.

[11] U.Naresh1 U.Vidya Sagar2 C.V. Madhusudan Reddy3 IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 14, Issue 3 (Sep. - Oct. 2013), PP 28-36 www.iosrjournals.org

[12] W. D. Yu, S. Nargundkar, N. Tiruthani, "Phishcatch – a phishing detection tool", 33rd Annual IEEE International on Computer Software and Applications Conference 2009. COMPSAC '09, pp. 451-456, 2009.

[13] How to recognize phishing email messages or links", March 2011, [online] Available: http://www.microsoft.com/security/online-privacy/phishing-symptoms.aspx.

[14] P. Likarish, D. Dunbar, T. E. Hansen, "B-apt: Bayesian anti-phishing toolbar", IEEE International Conference on Communications 2008. ICC '08, pp. 1745-1749, 2008.

[15] A. Y. Fu, L. Wenyin, X. Deng, "Detecting phishing web pages with visual similarity assessment based on earth mover's distance (emd)", IEEE Trans. Dependable Secur. Comput., vol. 3, no. 4, pp. 301-311, Oct. 2006.

[12] H. Shen, Y. Zhu, X. Zhou, H. Guo, and C. Chang, ``Bacterial foraging optimization algorithm with particle swarm optimization strategy for global numerical optimization,'' in *Proc. 1st ACM/SIGEVO Summit Genet. Evol. Comput.*, Shanghai, China, 2009, pp. 497_504.

[13] P.-G. Kou, J.-Z. Zhou, Y.-Y. He, X.-Q. Xiang, and C.-S. Li, ``Optimal PID governor tuning of hydraulic turbine generators with bacterial foraging particle swarm optimization algorithm,'' *Proc. CSEE*, vol. 29, no. 26, pp. 101_106, 2009.

[14] K. M. Passino, ``Biomimicry of bacterial foraging for distributed optimization and control,'' *IEEE Control Syst.*, vol. 22, no. 3, pp. 52_67, Jun. 2002.

[15] W. B. Heinzelman, ``Application-speci_c protocol architectures for wireless networks,'' Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, Jun. 2000.

[16] Q. Jiang and D. Manivannan, ``Routing protocols for sensor networks,'' in *Proc. Consum. Commun. Netw. Conf.*, Las Vegas, NV, USA, Jan. 2004, pp. 93_98.