International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 4 | ISSN : 2456-3307

Finding Fraud Websites by Using Data Mining Techniques

Maheswari G, Madhura P

Department of Computer Applications, Rayalaseema Institute of Information and Management Sciences, Tirupathi, Andra Pradesh, India

ABSTRACT

Globally the internet is been accessed by huge amount people inside their restricted domains. once the client and server exchange messages among one another, there's an activity that may be observed in log files. Log files provides a elaborated description of the activities that occur in an exceedingly network that shows the IP address, login and logout durations, the user's behavior etc. There are many varieties of attacks occurring from the net. Our focus of analysis is on Denial of Service (DOS) attacks with the help of pattern recognition techniques in data processing. Through that the Denial of Service attack is known. Denial of service is a terribly dangerous attack that jeopardizes the IT resources of a corporation by overloading with imitation messages or multiple requests from unauthorized users. But we cannot detect the fake website in this criteria. In order to detect and predict e-banking phishing website, we proposed an intelligent, flexible and effective system that is based on using classification Data mining algorithm. We implemented classification algorithm and techniques to extract the phishing data sets criteria to classify their legitimacy.

Keywords: Phishing websites, DOS attacks, Data mining, Association rules, cluster analysis, Log File, Cyber Crimes.

I. INTRODUCTION

Cyber refers to one thing that may be done on internet. Crime refers to one thing that's done lawlessly or without authorization. All those crimes that area unit done on the internet so as to achieve access to secured info or authorization rights is termed as "Cyber Crime". Globally the cyber-crime hindrance is unfold across profusely. In existing paper they applied the data mining techniques for identifying the Denial of Service attack. As this attack is extremely dangerous because it threatens the IT resources. It makes the server busy by imitation messages and recurrent queries. The server is engorged by traffic packets, so as to mitigate the server performance. If the amount of comparable requests square measure received at the server, that is larger than the edge price, we tend to assume this as associate attack and therefore the administrator is been wise to.

Social engineering attacks targeting users not computers or systems are designed to obtain sensitive or confidential information from users. Most social engineering attacks are classified as phishing attacks. And there are different techniques for phishing such as phishing by email, instant messages, SMS and website. These techniques help the phisher to lure unsuspecting online users into divulging personal information such as bank account information, website login information, and other sensitive information that can be used by a third party for illegal profit, blackmailing etc.

Phishing is a form of internet scam in which an attacker makes use of an email or website to illegally obtain private information.As explained in the complexity of understanding and analyzing phishing website is as a result of its involvement with technical and social problems. Simply, the aim is to lure users to phishing websites that mimics a legitimate websites to ruse users in order to get their sensitive information such as passwords, credits card, e-bank account, etc. As a result, the attacker can abuse the user's information in various ways from using it to gain illegal profit, blackmail, or even impersonate the user.

Although, phishing is a relatively new type of cyber security threat - the increasing sophistication of phishers in recent years have led to great harm in e-commerce services and information security . According to the Anti-Phishing Working Group (2013), 49,480 unique phishing websites were detected in the first quarter of 2013 and stayed at the higher rate through the third quarter. Hence, the need to efficiently resolve the outbreak of phishing in our online environment cannot be over exaggerated considering the danger of phishing websites to unsuspecting online victims. Due to the ever increasing phishing websites springing up by the day, it has become increasingly difficult to track and block them as attackers are coming up with innovative methods every day to entice unsuspecting users into divulging their personal information Cyber Security is that branch of pc Technology that deals with security in computer network. Cyberspace refers to the outline of policies relating to the networks and pc systems. The policies arranged go in the Cyber security area unit for the rationale of avoiding the malicious activity or unauthorized access to secured information. Since the emergence of high structured networks there arises a priority concerning however showing intelligence these networks area unit secured. These problems area unit major considerations in the web era.

But by this they are not going to find the which website is fraud and which website is good. In this paper we are going to deal the above drawback by taking e-banking website as an example.

Phishing is a fraudulent attempt, usually made through email, to steal your personal information. The best way to protect yourself from phishing is to learn how to recognize a phish.

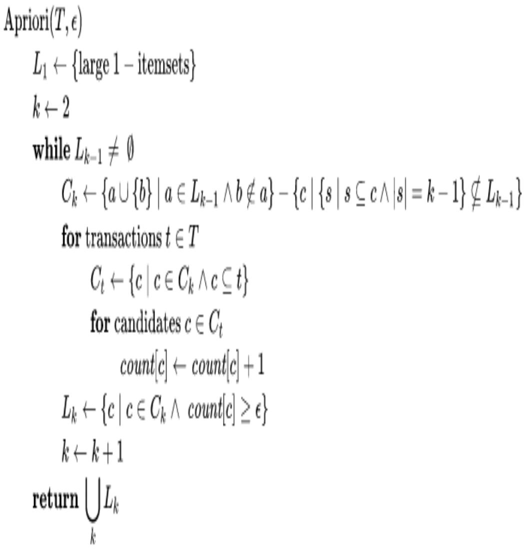
There are number of users who purchase products online and make payment through e- banking. There are e- banking websites who ask user to provide sensitive data such as username, password or credit card details etc. often for malicious reasons. This type of e-banking websites is known as phishing website. In order to detect and predict e-banking phishing website, we proposed an intelligent, flexible and effective system that is based on using classification Data mining algorithm. We implemented classification algorithm and techniques to extract the phishing data sets criteria to classify their legitimacy. The e-banking phishing website can be detected based on some important characteristics like URL and Domain Identity, and security and encryption criteria in the final phishing detection rate. Once user makes transaction through online when he makes payment through e-banking website our system will use data mining algorithm to detect whether the e-banking website is phishing website or not. This application can be used by many E-commerce enterprises in order to make the whole transaction process secure. 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. This system can be used by many E-commerce 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.

II. ALGORITHM

Apriori algorithm:

Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database.

This algorithm will detect if similar patterns of requests exist in the normal records prior to consider it as attack. If the algorithm finds out the pattern and or finds the number of request for the same transaction more than the threshold value it is considered as an attack and it sends signal or message to the administrator about the suspected attack.



```
Apriori$(T, \epsilon)$
    $L_1 \leftarrow \{\text{large } 1 - \text{itemsets}\}$
    $k \leftarrow 2$
    while $L_{k-1} \neq \emptyset$
        $C_k \leftarrow \{a \cup \{b\} \mid a \in L_{k-1} \wedge b \notin a\} - \{c \mid \{s \mid s \subseteq c \wedge |s| = k-1\} \nsubseteq L_{k-1}\}$
        for transactions $t \in T$
            $C_t \leftarrow \{c \mid c \in C_k \wedge c \subseteq t\}$
            for candidates $c \in C_t$
                $count[c] \leftarrow count[c] + 1$
        $L_k \leftarrow \{c \mid c \in C_k \wedge \; count[c] \geq \epsilon\}$
        $k \leftarrow k + 1$
    return $\bigcup_k L_k$
```

III. CONCLUSION

Generally cyber crimes are going on in the cyberspace continuously as many people are showing interest to use the technology. In the existing systems they are going to find one of the attack like Denial Of Service (DOS) attack which automatically changes the system configurations. This attack can be found by the administrator by setting one threshold value. The attacker can attack easily by using logfiles. But nobody can detect which website is fraudulent. Some e-banking websites may use the sensitive information like username and password of our account and they can do some malicious attacks on our account. These are called Phishing websites. Nowin this paper we are going to fingd the fake or fraud website by taking reviews from many people. The e-banking phishing website can be detected based on some important characteristics like URL and Domain Identity, and security and encryption criteria in the final phishing detection rate. Once user makes transaction through online when he makes payment through e-banking website our system will use data mining algorithm to detect whether the e-banking website is phishing website or not.

IV. REFERENCES

[1]. Design and Implementation of Small and Medium Sports Events Management Platform for Colleges,Wang wei,Xuan Lingqiang.

[2]. D. Zhang, Z. Yan, H. Jiang, and T. Kim, "A domain-feature enhanced classification model for detection of Chinese phishing e-business websites," Information & Management, 2014.

[3]. G. Liu, B. Qiu, and L. Wenyin. "Automatic detection of phishing target from phishing webpage." in Pattern Recognition (ICPR), 2010 20th International Conference on. 2010. IEEE.

[4]. H. Zhang, G. Liu, T. W. Chow, and W. Liu, "Textual and visual content-based anti-phishing: a Bayesian approach," Neural Networks, IEEE Transactions on, 2011. 22(10): p. 1532-1546.

[5]. G. Ramesh, I. Krishnamurthi, and K. Kumar, "An efficacious method for detecting phishing webpages through target domain identification," Decision Support Systems, 2014. 61: p. 12-22.

[6]. P. Garrard, V. Rentoumi, B. Gesierich, B. Miller, and M. L. Gorno-Tempini, "Machine learning approaches to diagnosis and laterality effects in semantic dementia discourse," Cortex, 2014. 55: p. 122-129.

[7]. A. Abunadi, O. Akanbi and A. Zainal "Feature extraction process: A phishing detection approach." in Intelligent Systems Design and Applications 2013. ISDA 2013. 13th International Conference. ISDA.

[8]. L. A. T. Nguyen, B. L. To, H. K. Nguyen, and M. H. Nguyen. "Detecting phishing web sites: A heuristic URL-based approach." in Advanced Technologies for Communications (ATC), 2013 International Conference on. 2013. IEEE.

[9]. G. Xiang, J. Hong, C. P. Rose, and L. Cranor, "Cantina+: A feature-rich machine learning framework for detecting phishing web sites," ACM Transactions on Information and System Security (TISSEC), 2011. 14(2): p. 21.

[10]. Y. Li, R. Xiao, J. Feng, and L. Zhao, "A semi-supervised learning approach for detection of phishing webpages," Optik-International Journal for Light and Electron Optics, 2013. 124(23): p. 6027-6033.

[11]. C.-R. Huang, C.-S. Chen, and P.-C. Chung, "Contrast context histogram—An efficient discriminating local descriptor for object recognition and image matching," Pattern Recognit., vol. 41, no. 10, pp. 3071–3077, Oct. 2008.

[12]. M. Dunlop, S. Groat, and D. Shelly. "GoldPhish: using images for content-based phishing analysis." in Internet Monitoring and Protection (ICIMP), 2010 Fifth International Conference on. 2010. IEEE.

[13]. S. Afroz and R. Greenstadt. "Phishzoo: Detecting phishing websites by looking at them." in Semantic Computing (ICSC), 2011 Fifth IEEE International Conference on. 2011. IEEE.

[14]. W. Zhuang, Q. Jiang, and T. Xiong. "An intelligent Anti-phishing strategy Model for Phishing website Detection." in Distributed Computing Systems Workshops (ICDCSW), 2012 32nd International Conference on. 2012. IEEE.

[15]. H. Kazemian and S. Ahmed, "Comparisons of machine learning techniques for detection.