INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS.....

Detection of Phishing Websites using the Hybrid Approach

Samanjeet Kaur[1], Sukhwinder Sharma[2]

[1]Student, [2]Assistant Professor
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib
[1]samanjeet023@gmail.com, [2]sukhwinder.sharma@bbsbec.ac.in

Abstract*: Phishing is a type of attack in which criminals use fake emails and bogus web sites to trap people into giving up delicate information. With the deployment of internet, keeping financial and sensitive information turn out to be more tiresome. One of the troubles faced today is rising number of phishing websites, forged webpage structured or shaped and utilized by phishers to produce the duplicate copy of the web pages of legitimate websites. Fake or bogus websites results in lack of trust in internet based services and results in financial loss to the users. So it turn out to be crucial to look for useful solution to reduce the incident of being victimized by phishing attack. The main objective of proposed approach is to provide protection and making users aware from growing phishing attack and finding better way for detection. This research employs approach that uses fuzzy logic along with classifiers like Support Vector Machine and Nearest Mean Classifier. Fuzzy based detection scheme provides useful aid in detecting phishing websites. Finally, our proposed phishing website detection approach successfully resulted in low false positive and high true positive and good accuracy results.*

Keywords: *Classifiers, SVM, phishing websites, fuzzy logic.*

1. INTRODUCTION

Phishing websites are forged WebPages created and utilized by phishers to copy the web pages of legitimate websites by which results in lack of faith in internet based services but also financial loss. Phishing attack is large scale security risk to the online community and for those who deal with the sensitive information for the reason that phishers makes identical copies of the website to direct the users to forged site that steals the information. Even if the web users are conscious of these types of attacks, then also lot of users become victimized under this attack of phishing. Only professionals or experts can recognize these types of fraudulent websites. Not all the web users are expert in recognizing them immediately; therefore web user becomes victim as a result of providing personal details to the attacker. Phishing is developing constantly as it is easy for attackers to make replica of entire website using HTML source code. By doing little changes in the source code of the website, it becomes easy to befool the victim by directing it to phishing websites. Moreover phishers make use of techniques that attract the web users, they use Greetings which attract web customers to verify their account right now without any delay or to update them otherwise their account will be terminated.

Phishers started creating fake websites to increase the successful rate of phishing. For instance, phishers list dozens of domain names that look similar to a well known brand, such as "www.citi-bank.com". Victims, who go through one of these websites, may believe that the website is the legitimate one, and their account on the website. Phishers inserts website designs into the emails, and using stolen logos and trademarks from the targeted organization, so that address look as if it comes from the legitimate. Many recent attacks consist of a link to an original banking website in the background, but a forged "login" box positioned in front of the legitimate site. Apparently it is more persuasive because the real site and the pop-up appear to be from the similar source. After giving up financial information on a fake site, the victim is directed to the real home page of the company being targeted. As a result, the victim will not believe the website of being bogus. However, there are loads of definitions of a phishing website from different viewpoint. So, we mention some of these definitions to get enhanced understanding of its features and attack strategy. According to Phish Tank "Phishing is a fraudulent attempt, usually made through email, to steal personal information".

2. LITERATURE REVIEW

Gunjan et al. [7] conducted the survey on growing problem of cyber crime in India. Different types of cyber crimes are being discussed. Survey has been conducted on cyber cases in India. Two case studies on website phishing analysis and phone phishing analysis have been done in order to implement and elaborate the process of phishing.

Madhuri et al. [14] proposed the algorithm which is end host based, which is characteristic based algorithm for detection and prevention of phishing attacks. Link guard algorithm not only detects the phishing attacks but also unknown attacks.

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY

WINGS TO YOUR THOUGHTS…..

Prasad et al. [16] highlighted the phishing attack growing effects. In the paper Client Side Protection from phishing attack, author proposed two approaches which warns user from entering bogus websites and other one to safeguarding the password. The following approach is applied on the bank websites. First approach maintains the whitelist of the indian bank websites and if user enters the bogus website. Second approach is used which prevents the user from entering the details, browser extension is applied. He et al. [8] states that Phishing attack is growing significantly each year and is considered as one of the most dangerous threats in the Internet which may cause people to lose confidence in e-commerce. In this paper, a method to determine whether a webpage is a legitimate or a phishing page is given. Support vector machine is used. Avtar et al. [3] introduces the data shield algorithm for tightening security against phishing attacks, which usually operate on hyperlinks. Phishing data is provided by anti phishing working group.

3. PROPOSED APPROACH

A hybrid approach is proposed where classifiers have been implemented and their phishing results are integrated inside fuzzy inference engine for final phishing detection, it uses classifiers like SVM, Nearest mean. Features will be collected to differentiate legitimate and phishing websites. Data set collected will be input to the method engine which contains SVM and NMC from where the rules will be generated that will be added to the fuzzy logic. The best rules are selected by fuzzy logic to generate results. In this research, by analyzing the limitations in existing methods of phishing detection, the hybrid approach will be proposed that can lessen the false positive rate and can improve true positive rate. The important advantage offered by fuzzy logic is use of linguistic variables for signifying key phishing characteristic indicator.

3.1 Classifiers used

In this study, we have used two classifiers for the detection of phishing websites. These include SVM and Nearest Mean to build better phishing detection model. In this section, we will briefly present these.

- The SVM algorithm was invented by Vladimir N. Vapnik and Alexey Ya. Chervonenkis in 1963. A Support Vector Machine (SVM) carries out classification by discovering the hyperplane that increases the margin between the two classes. The vectors (cases) that define the hyperplane are the support vectors.
- NMC, it is simple classifier presented by Fukunaga in 1990 as a classifier with lesser complexity. In machine learning, a nearest mean classifier is a classification model that assign to observations the label of class of training samples whose mean is close to the observation. The nearest centroid classifier is known as the Rocchio classifier, because of its resemblance to the Rocchio algorithm for relevance feedback.

4. EXPERIMENTAL SETUP

The proposed approach has been implemented in MATLAB.

4.1 Data Collection

Data set is collected from, ‘‘phishtank” [18] which is one of the most crucial phishing-report collector. The PhishTank database collects the URL’s of the website that are suspected as phishing which are being reported. In addition, legitimate websites were collected from yahoo directory and starting point directory. These directories contain addresses of legitimate sites for different types of services.

4.2 Performance metrics

Metrics used to calculate the performance are [12]:

- ***True positive Rate*:** It measures rate of correctly detected phishing attack in relation to all the existing phishing attacks.
  TPR = TP/(TP+FN)
- ***False positive Rate:*** It measures rate of legitimate instances that are incorrectly detected as phishing attack in relation to all existing legitimate instances.
  FPR=FP/(FP+TN)
- ***True Negative rate:*** It measures the rate of correctly detected legitimate instances in relation to all existing legitimate instances.
  TNR= TN/(TN+FP)
- ***False Negative rate:*** It measures the rate of phishing attacks that are incorrectly detected as legitimate in relation to all existing phishing attacks.
  FNR = FN/(TP+FN)
- ***Accuracy***: It measures the overall rate of correctly detected phishing and legitimate instances in relation to all instances.
  Accuracy = (TP + TN) / (TP + TN + FP +FN)

5. RESULTS AND DISCUSSION

Result is shown in Table1. The value of True positive rate is 98.2% which indicates that rate of correctly detected phishing attack in relation to all the existing phishing attacks is high, more the value of TPR, efficient is our detection. False positive rate drops to 0.95% as, lower the false positive rate better is our detection. Accuracy has reached up to 98.6% which indicates the overall rate of correctly detected phishing and legitimate instances in relation to all instances is high.

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS.....

Table1: Result of proposed approach

Proposed approach	Results
True Positive Rate	98.2%
False Positive Rate	0.95%
True Negative Rate	99%
False Negative Rate	1.7%
Accuracy	98.6%

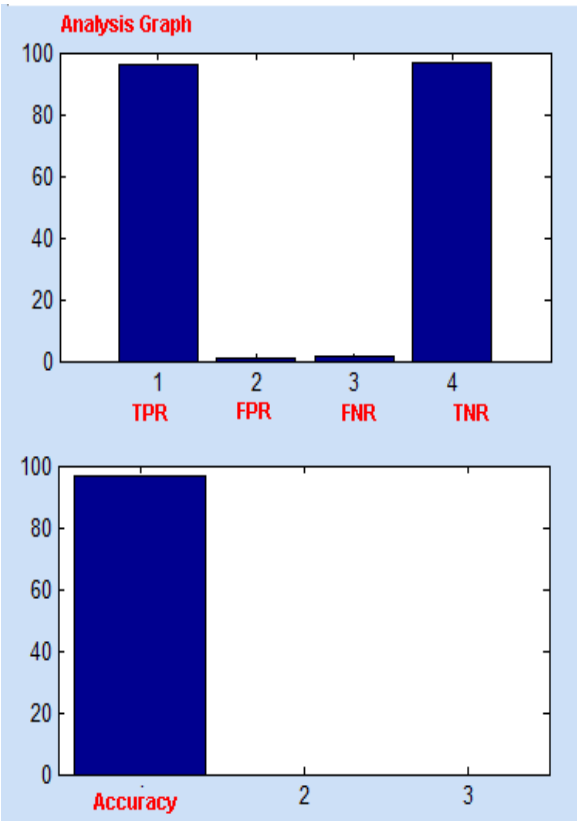

Figure 1: Analysis Graph

6. CONCLUSIONS

Phishing on the whole is organized crimes of the twenty-first century that require very little skill on part of the attacker and the challenge of securing sensitive details like bank accounts and passwords of the client secure from the hands of attackers turn out to be more crucial day by day. Phishing is a deceptive attempt to acquire personal details such credit card information, as bank information, online shopping account passwords and employment details, by thieving the trusted brands of famous e-commerce, credit card companies and banks. Results of the proposed approach indicated that the combination of classifiers with fuzzy logic helped appreciably by giving accuracy about 98.6%. Thus, proposed model put forward detection approach that shield the user about the way to deal with secretive information like bank accounts.

REFERENCES

[1] Aburrous, M., Hossain, M., Dahal, K. and Thabtah F., 2010. "Associative Classification Techniques for predicting e-Banking Phishing Websites", International Conference on Multimedia Computing and Information Technology (MCIT), Sharjah, pp.9-12.

[2] Almomani, A., Gupta, B. B., Atawneh S., Meulenberg A., and Almomani, E., 2013. "A Survey of Phishing Email Filtering Techniques", Communications Surveys & Tutorials, IEEE. Vol. 15, pp. 2070 - 2090.

[3] Avtar, R. and Verma, B. Jangra, A., 2011. "Data Shield Algorithm (DSA) for Security against Phishing Attacks", International journal of engineering sciences" Vol. 4, pp. 221-232.

[4] Banu, M.N. and Banu, S. M., 2013. "A comprehensive study of phishing attack", International journal of computer science and technology", vol.4, pp.783-786.

[5] Basnet, R.B., Sung, A.H and Liu, Q., 2014. "Learning to detect phishing urls", International Journal of Research in Engineering and Technology,Vol.3, pp.11-14.

[6] Gandhi, R. and Backiyalakshmi, R., 2014. "Intelligent phishing website detection system using fuzzy techniques", Information and Communication Technologies Discovery Publication, Vol .24, pp. 33-40.

[7] Gunjan V.K. and Gunjan, A.K., 2013. "A Survey of Cyber Crime in India", Advanced Computing Technologies 15th International Conference in Rajampet IEEE, pp. 1-6.

[8] He, M., Sutanto, A., Khan, M.K., Fan, P., Chen R. J., Ray-Shine Run and Shi-Jinn Horng, 2011. "An efficient phishing webpage detector", Journal of Expert Systems with Applications, Elsevier, Vol. 38, pp.12018–12027.

[9] Jiang, H., Zhang, D., and Yan, Z., 2013. "A Classification Model for Detection of Chinese Phishing E-Business Websites", journal of information and management, Vol.51, pp.845-853.

[10] Kandpa, V. and Singh, R. K., 2013. "Latest face of cyber crime and its prevention in India" International Journal of Basic and Applied Sciences, Vol. 2, pp. 447 – 450.

[11] Khade, A. and Shinde, S.K., 2013. "Detection of Phishing websites using Data Mining Technique", international journal of engineering and technology, Vol. 2 , pp. 3725-3729.

[12] khonji, M. and Youseff, 2013. "Phishing Detection: A Literature Survey" Cummunication survey and tutorials IEEE, Vol. 15, No. 4, pp. 2091 – 2121.

[13] Lakshmi, S. and Vijaya, 2011 "Efficient prediction of phishing website using supervised

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY

WINGS TO YOUR THOUGHTS.....

learning algorithms" International Conference on Communication Technology and System Design Elsevier, Vol. 30 , pp.798–805.

[14] Madhuri, M., Yeseswini, K. and Vidyasagar, U. ,2013. "Intelligent phishing website detection and prevention by using linkguard algorithm", International Journal of Communication Network Security Vol. 2, pp. 639-648.

[15] Mohammad, R, McCluskey, T.L. and Thabtah, Abdeljaber, Fadi 2012 "An Assessment of Features Related to Phishing Websites using an Automated Technique" International conference on Internet Technology And Secured Transactions London , pp. 492 – 497.

[16] Prasad V., Reddy and Radha, V., 2011 "Client Side Protection from phishing Attack" International journals of advanced engineering science and technology, Vol. 6, No. 1, pp. 39-45.

[17] Prevost,S., Granadillo, G. and Laurent, M. ,2011 "A dual Approach To Detect Pharming Attacks At The Client-Side", in Proceeding of the 4th IFIP International Conference on New Technologies, Mobility and Security (NTMS), Paris, France, pp.1-5.

[18] PhishTank: phishing-report collector. Available: http://www.phishtank.com/, [last accessed, 27/7/2015].

[19] Raiyn, J., 2014. " A survey of the cyber attack detection strategies" International Journal of Security and Its Applications, Vol.8, pp. 247-256.

[20] Rami M. Mohammad, Thabtah, F., 2012. "An assessment of features related to phishing websites using an automated technique", Internet Technology and Secured Transactions, pp-492-497.