Webpage: www.ijaret.org
UGC Approved Journal-43847

Volume 6, Issue III, Mar 2018
eISSN 2320-6802

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY

WINGS TO YOUR THOUGHTS…..

Review on Design and analysis of various data mining clustering techniques for segmentation of mobile Customer

Monika Rawat[1], **Sandeep Garg** [2]

[1,2]RPIIT College of Engineering, Kurukshetra University
Bastara, Karnal, Haryana
[1]monikarawat79.mr@gmail.com
[2]sandeep.1091@gmail.com

***Abstract**: Phishing Detection and prevention of the attacks are the very big challenges as the phisher perform attacks to bypass the existing anti-phishing techniques. In this paper, we focus on detecting login phishing pages, pages that contain forms with email and password fields to allow for authorization to personal/restricted content. This research works with any authentication technologies which are based on exchange of credentials. One of the effective solutions to prevent a phishing attacks is to integrate security features with the web browser to raise alerts when-ever a phishing site is accessed by an internet user. Generally, web browsers provide security against phishing attacks with the help of white list-based solutions The methodology introduced in our paper identifies duplicate websites by submitting incorrect credentials and analyzing the response. We have also proposed a mechanism for analyzing the responses from server against the submissions of all those credentials to determine the legitimacy of a given website*

Keywords: *Web Clustering, Web site, Domain Name, Heuristics, URL analysis, Login pages, White-list, Web security*

1. INTRODUCTION

There are many techniques exist to detect phishing attack but no one bullet proof solution yet to present which detect all type of phishing attack. In order to detect whether the website is phishing or fake website, the first question to ask is: how to discriminate phishing website and the legitimate website as the reason is that the phishing website is look alike to the legitimate website .Where if the we have the portrayed identity of the query website then we can find out that if it a legitimate or a phishing website. (if the doubt website is a phishing website, the portrayed identity will be the identity of the besieged legitimate website)[2].. The phisher sends an email including the link of the phishing website to it to his victims. In the case of spike phishing, a mail is sent to individual targeted victims. When the victim opens the email, and visits the phishing website, the phishing website prompts the victim to insert private data, for example, if the phisher copycats the phishing website of a famous organization, then the users of organization are expected to willingly reveal their private credentials to the phishing website.

Phishing is the act of mimicking a trusted website to gain sensitive information from online users like detail of credit card, personal identification number etc[3].

2. LITERATURE SURVEY

Bian et.al [1]: Proposed a method to assess the effectiveness of three popular online resources in identifying phishing sites-viz,Yahoo! Inlink data and Yahoo! directory service, Google Page Rank system. Their results point towards that these online resources can be used to boost the accuracy of phishing site detection when used in combination with existing phishing countermeasures. The proposed loom involves investigate the following three attributes of a goal site (site being check up): (1) the reliability of the target site hosting domain, (2) the reliability of in-neighbor sites that link to the hosting domain, and (3) the connection between the aim site web category and its hosting domain web kind. The abovementioned online resources by themselves are insufficient to concentrate on the phishing attack problem. This approach provide convention on how each of those resources may be included with existing phishing detection techniques to offer a more efficient solution.

DeBarr et.al [2]: Proposed a approach as a first step the exercise of Spectral Clustering to analyze messages based on traffic behavior. specifically, Spectral Clustering analyzes the association between URL substrings for web sites originate in the message contents. Cluster membership is then employ to assemble a Random Forest classifier for phishing. Data from the Phishing Email quantity and the Spam killer Email quantity are used to evaluate this approach. Performance assessment metrics include the region Under the receiver operating characteristic Curve (AUC), as well as accurateness, exactness, evoke, and the (harmonic mean) F measure. Presentation of the incorporated Spectral Clustering and Random Forest loom is found to provide important developments in all the metrics listed, contrasted to a satisfied filtering technique such as LDA joined with text message deletion done arbitrarily or in an adaptive fashion using adversarial learning. The Spectral Clustering approach is strong against the lack of content. website, assert, and figure the documentary significance between this claimed identity and other description in the website. Their phishing detection system then employs this textual significance as one of the sort for classification.

Tan et.al in [3]: Proposed an anti-phishing method to protect users against phishing attacks in the internet. The scope of this approach study focuses mainly on the detection of phishing websites with English content. In order to encourage users on whom the website claims to be, phishers usually place brand names in different parts of the URL. They oppressed this phishing pattern by conveying weights to words take out from the HTML content, based on their co-appearance at path, hostname and file names of URLs. These weights are then supplementary to their equivalent TF-IDF weights. The most likely words are particular and submitted to Yahoo Search to recover the highest frequency domain name amongst the top 30 search results. A WHOIS lookup is carry out to disclose the vendor behind the selected domain name. A phishing website can be easily illustrious if the vendor of query domain name be different from the owner of domain name returned by the search engine.

Nguyen et.al in [4]: Proposed an efficient approach for identifying phishing websites foundation on the single-layer neural network. Particularly, the proposed technique calculates the value of heuristics impartially. Then, the weights of heuristic are produced by a single-layer neural network. The proposed technique is assessed with a dataset of 11,660 phishing sites and 10,000 legitimate sites.

Deshmukh et al.[5]: Proposed a approach as cyber crime is technology based fault committed by technocrats. This paper deals with modification of cyber crime like Packet Sniffing, Salami Attack, Bot Networks and Tempest Attacks. It also contains real world cyber crime suitcases their situation and modus operandi. The worldwide malware, rate spam rate and phishing rate is rising speedily. And there is a latent shock of cyber crime on consumer trust, economics and production time. The contradict ways similar to Intrusion Detection, GPRS Security architecture and Agent Based Distributed Intrusion Detection System and prevention System are utilized for safety reason.

Ali et.al in [6]: Proposed a approach of confidentiality in Instant Messengers (IM) by means of Association Rule Mining (ARM) method a Data Mining approach included with Speech Recognition system. verbal skills are acknowledged from words with the help of FFT spectrum analysis and LPC coefficients methodologies. Online criminal's at the present time modified voice chatting technique along with text messages collaboratively or either of them in IM's and squashing out personal information direct to intimidation and barrier for privacy. To facilitate centre of attention on privacy preserving this approach residential and try out Anti Phishing Detection system (APD) in IM's to detect unreliable phishing for text and audio collaboratively.

3. PROPOSED METHODOLOGY

The proposed insignificance aims to study the phishing detection by using website approach. The methodology comprises of following two processes: First process will capture the record and credentials of the login user and screenshot. This approach has a few advantages. As the research work will focus on as a replacement for finding the record of the user in white list if the information in white list so extract the valid detail and after authicate the user successfully. This approach has a few advantages. As, the captured the login detail is actual offer the web content, which means there is no other secret task. If two or more host or web link are joined in network then it is consider as a cluster.

Phishing login pages are designed to lure victims into willingly giving their credentials. However, Phishing websites has no information regarding the victim's real credentials. Hence, it is expected to have one of the following scenarios [5]:

1. Phishing website shows a success message.
2. Phishing website redirects to another website.
3. Phishing website shows a failure message.
4. Phishing website shows the same login page again.

Uniform Resource Locator and Domain Analysis Module

The purpose of the URL and domain analysis module is to minimize the rate of false positives and reduce analysis time. The URL and domain analysis module act as a filter for URLs that are clearly not legitimate before testing with the phishing identification module [4].

There are many algorithms available to determine the best weights for the heuristics like domain creation and expiry date or @ in URL is >=1,Number of dots in URL is >=5 However, for simplicity forward linear model will be used or store the credentials of the user in cookies to create the session.

4. CONCLUSIONS AND FUTURE WORK

In this technique we have developed a new method to detect phishing websites based on the URL of the website using the white list and create the session and all the relevant information are used in authentication process are stored in session or database to hack the records. The system has shown a 97% detection rate of phishing website because the white list used by phishing website returns the search of original websites or other websites that have back linked the original website but the test website's URL never appears in the search result. Thus making a 97% accurate detection because some of the records are stored in white list.

In the future work, we can add more parameters like Google Page Rank, number of back links etc in order to increase the overall confidence towards phishing as well as non-phishing website.

REFERENCES

[1] Bian, K., Park, J.M., Hsiao, M.S., Belanger, F. and Hiller, J., 2009, July. Evaluation of online resources in assisting phishing detection. In Applications and the Internet, 2009. SAINT'09. Ninth Annual International Symposium on (pp. 30-36). IEEE.

[2] DeBarr, D., Raman than, V. and Wechsler, H., 2013, June. Phishing detection using traffic behavior, spectral

clustering, and random forests. InIntelligence and Security Informatics (ISI), 2013 IEEE International Conference on (pp. 67-72). IEEE.

[3] Tout, H. and Hafner, W., 2009, August. Phishpin: An identity-based anti-phishing approach. In Computational Science and Engineering, 2009. CSE'09. International Conference on (Vol. 3, pp. 347-352). IEEE

[4] Nguyen, L.A.T., To, B.L., Nguyen, H.K. and Nguyen, M.H., 2014, October. An efficient approach for phishing detection using single-layer neural network. In Advanced Technologies for Communications (ATC), 2014 International Conference on (pp. 435-440). IEEE.

[5] Deshmukh, J.J. and Chaudhari, S.R., 2014. Cyber crime in indian scenario–a literature snapshot. International Journal of Conceptions on Computing and Information Technology, 2(2).

[6] Ali, M.M. and Rajamani, L., 2012, March. Deceptive phishing detection system: from audio and text messages in instant messengers using data mining approach. In Pattern Recognition, Informatics and Medical Engineering (PRIME), 2012 International Conference on (pp. 458-465). IEEE