3rd National Conference on Green Technology and Science for Sustainable Development
© 2020 IJSRST | Volume 5 | Issue 6 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
International Journal of Scientific Research in Science and Technology

Detection of Phishing Website Using Learning Techniques

Punam S. Wankhede[1], Dr. A. S. Kapse[2], Ms. Vaishnavi Hiwrale[3]

[1]Computer Science and Engineering, Sant Gadge Baba Amravati University, Chikhli, Maharashtra, India
[2]Assistant Professor (Sr. Scale) and Head, Department of Computer Science & Engineering, Anuradha Engineering College, Sant Gadge Baba Amravati University, Chikhli, Maharashtra, India
[3]M. E. CSE, Student, Anuradha Engineering College Chikhli, India

ABSTRACT

Phishing sites which expects to take the victims confidential data by diverting them to surf a fake website page that resembles a honest to goodness one is another type of criminal acts through the internet and its one of the especially concerns toward numerous areas including e-managing an account and retailing. Phishing site detection is truly an unpredictable and element issue including numerous components and criteria that are not stable. On account of the last and in addition ambiguities in arranging sites because of the intelligent procedures programmers are utilizing, some keen proactive strategies can be helpful and powerful tools can be utilized, for example, fuzzy, neural system and data mining methods can be a successful mechanism in distinguishing phishing sites. In Phishing E-mail Detection Based on Structural Properties, the proposed approach explains to find phishing through appropriate identification and usage of structural properties of email. The experiment is done by SVM and classification technique to classify phishing e-mails. The technique is used to identify phishing e- mails, which is low in efficiency and scalability. This is purely based on structural properties of e-mail and it has to extend more structural or content properties to reduce error results. Identifying phishing target based on semantic link network, the paper proposes a novel approach to discover phishing website by calculating association relation among webpages that include malicious webpages and its associated webpages to measure the combination of text relation, link relation and search relation. The semantic link network proposes a strategy based on situations to identify the suspicious webpage as phishing. The disadvantage in this approach is more kind of association has to be done, similarities between visual, layout and domain has to be related. This method is considered as a time consuming approach and also various sub-relations in the combined association relations are studied. Fuzzy Neural Network for Phishing Emails Detection deals with phishing email. It distinguishes phishing email and ham email in online mode. It is adopted on rank fetching, feature fetching and grouping similar features of email. The technique is based on binary value 0 or 1 to produce the result for all features used in this method, where 1 denotes a phishing feature and 0 for non-phishing. This technique does not have much dynamic system and thus it is inadequate in performance to produce accurate results. Intelligent Phishing

Keywords : Extreme Learning Machine, Features Classification, Information Security, Phishing.

I. INTRODUCTION

Phishing is a type of extensive fraud that happens when a malicious website act like a real one keeping in mind that the end goal to obtain touchy data, for example, passwords, account points of interest, or MasterCard numbers. In spite of the fact that there are a few contrary to phishing programming and methods for distinguishing potential phishing endeavors in messages and identifying phishing substance on sites, phishers think of new and half breed strategies to go around the accessible programming and systems. Phishing is a trickery system that uses a blend of social designing what's more, innovation to assemble delicate and individual data, for example, passwords and charge card subtle elements by taking on the appearance of a dependable individual or business in an electronic correspondence. Phishing makes utilization of spoof messages that are made to look valid and implied to be originating from honest to goodness sources like money related foundations, ecommerce destinations and so forth, to draw clients to visit fake sites through joins gave in the phishing email. The misleading sites are intended to emulate the look of a genuine organization site page. The employing so as to phishing invader's trap clients diverse social building strategies, for example, debilitating to suspend client accounts on the off chance that they don't finish the account upgrade process, give other data to approve their records or a few different motivations to get the clients to visit their satirize page. Delicate Computing strategies are progressively being utilized to address an extent of computational issues. Clustering is a kind of unsupervised learning; unsupervised learning except that there is no previous information about the class participation of the perceptions, i.e., class names of information is obscure. The reason for utilizing unsupervised learning is to specifically separate structure from a dataset without earlier preparing. On the other hand, supervised learning accommodates a vastly improved precision, unsupervised learning accommodates a quick and dependable way to deal with infer information from a dataset. That's why we used supervised learning in our work.[1]

II. IDENTIFICATION OF PHISHING WEBSITES USING VARIOUS METHODS

Phishing website is a huge effect on the financial and online commerce, detecting and preventing this attack is an important step towards protecting against website phishing attacks, there are several approaches to detect these attacks. In this section, we review existing anti phishing solutions and list of the related works. One approach is an intelligent Phishing Website Detection System using Fuzzy Techniques [5]. It is based on fuzzy logic and produces six criteria‟s of website phishing attack. There are many characteristics and factors that can distinguish the original legitimate website from the forged faked phishing website like spelling errors, long URL address and abnormal DNS record. Website phishing detection rate is performed based on six criteria and there are different numbers of components for each criterion, the criteria are:

A. URL & Domain Identity.
a) Using the IP address.
b) Abnormal request URL.
c) Abnormal URL of anchor.
d) Abnormal DNS record.
e) Abnormal URL.

B. Security & Encryption.
a) Using SSL certificate.

b) Certification authority.
c) Abnormal cookie.
d) Distinguished Names Certificate (DNC).

C. Source Code & Java script.
a) Redirect pages.
b) Straddling attack.
c) Pharming attack.
d) Using on Mouse Over to hide the Link.
e) Server Form Handler (SFH).

D. Page Style & Contents.
a) Spelling errors.
b) Copying website.
c) Using forms with "Submit" button.
d) Using Popups windows.
e) Disabling right click.

E. Web Address Bar.
a) Long URL address.
b) Replacing similar characters for URL.
c) Adding a prefix or suffix.
d) Using @ symbol to confuse.
e) Using hexadecimal character codes.

F. Social Human Factor.
a) Much emphasis on security and response.
b) Public generic salutation.
c) Buying Time to Access Accounts.
G. Common properties of Phishing attacks

The following lines represent number of properties of phishing attacks in the websites, they are [6]:

- Logos: The Phishing website uses logos found on the legitimate website to mimic its appearance. So phishers can load it from the legitimate website domain to their phishing websites (external domain).
- Suspicious URLs: Phishing websites are located on servers that have no relation with the legitimate website. The phishing website‟s URL may contain the legitimate website‟s URL as a substring (http: //www.ebaymode.com), or may be similar to the legitimate URL (http://www.paypa1.com) in which the letter „L‟ in PayPal is substituted with number „1‟. IP addresses are sometimes used to mask the host name (http://25255255255/top.htm). Others use @ marks to make host names difficult to understand (http://ebay.com:top@255255255255/top.html) or contain suspicious usernames in their URLs.
- User input: Phishing websites typically contain pages for the user to enter sensitive information, such as account number, password and so on.
- Short lived: Most phishing websites are available for only a few hours or days – just enough time for the attacker to defraud a high enough number of users.
- Copies: Attackers copy HTML from the legitimate websites and make minimal changes.
- Sloppiness or lack of familiarity with English: Many Phishing pages have misspellings, grammatical errors, and inconsistencies. [3]

III. WORKING OF PROPOSED SYSTEM

A. Submitting Information to Email

Web form allows a user to submit his personal sensitive information that is directed to some server for processing. A phisher might redirect the user's information to his personal email. To that end, a server-side script language might be used such as "mail()" function in PHP.

Figure 1 : Login Page

B. Blacklist based

A Blacklist is created in the proposed model in which the website detected as phishing is saved for the future use a to keep a track record and dataof the phishing website this can be useful in analyzing the phishing website to increase the efficiency of the system.

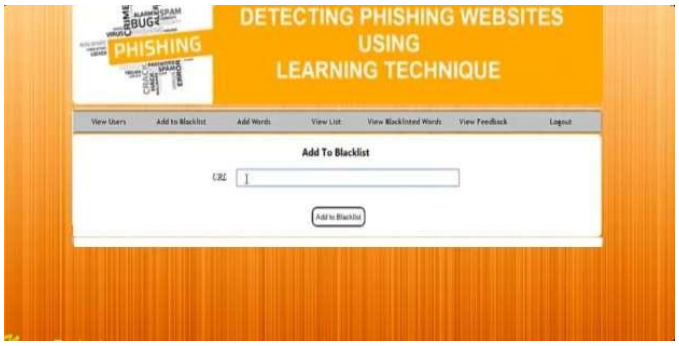

Figure 2: Blacklist Add Page

Figure 3: Adwords Add Page

C. WHOIS Database

The life of phishing site is very short, therefore; this DNS information may not be available after some time. If the DNS record is not available anywhere then the website is phishing. If the domain name of the suspicious webpage is not match with the
WHOIS database record, then webpage considers as phishing.[2]

Figure 4: Website Databas

IV. 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.

V. REFERENCES

[1]. https://arxiv.org/abs/1909.00300
[2]. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 03 | Mar 2018 www.irjet.net p-ISSN: 2395-0072
[3]. Volume 1 No. 6, October 2011 ISSN-2223-4985 International Journal of Information and Communication Technology Research

[4]. JIAN MAO1,WENQIAN TIAN1, PEI LI1, TAO WEI2, AND ZHENKAI LIANG3 Phishing-Alarm: Robust and Efficient Phishing Detection via Page Component Similarity.
[5]. Zou Futai, Gang Yuxiang, Pei Bei, Pan Li, Li Linsen Web Phishing Detection Based on Graph Mining.
[6]. Nick Williams, Shujun Li Simulating human detection of phishing websites: An investigation into the applicability of ACT-R cognitive behaviour architecture model.
[7]. XIN MEI CHOO, KANG LENG CHIEW,DAYANG HANANI ABANG IBRAHIM,NADIANATRA MUSA, SAN NAH SZE, WEI KING TIONG FEATURE-BASED PHISHING DETECTION TECHNIQUE.
[8]. Giovanni Armano, Samuel Marchal and N. Asokan Real- Time Client-Side Phishing Prevention Add-on.