WEB PHISHING DETECTION

TEAM ID : PNT2022TMID44694

TEAM MEMBERS : DHIVYA S (732519104004)

ABITHA J (732519104002)

VASUNTHARA R (732519104030)

PRIYATHARSHINI G (732519104703)

DEPARTMENT : CSE

COLLEGE NAME : SHREE VENKATESHWARA HI-TECH ENGINEERING COLLEGE, GOBI.

LITERATURE SURVEY

BASE PAPER

JAIN, A.K., GUPTA, B.B.: 'A MACHINE LEARNING BASED APPROACH FOR PHISHINGDETECTION USING HYPERLINKS INFORMATION', J. AMBIENT INTELL. HUMANIZ. COMPUT,2019. Jain and Gupta proposed a novel web phishing detection approach by extracting hyperlinks of the web pages. The proposed approach has extracted 12 specific hyperlink features such as total hyperlink feature, no hyperlink feature, internal hyperlinks, external hyperlinks, null hyperlink, internal CSS, external CSS, internal redirection, external redirection, internal error, external error, login form link, internal favicon, and external favicon. The extracted features are then fed into ML algorithms such as naïve Bayes, random forest, SVM, Ada boost, neural network, C4.5, and logistic regression. The performance of all the ML algorithms was measured and reported

REFERENCE PAPER

AKANBI, O.A., AMIRI, I.S., FAZELDEHKORDI,E: 'A MACHINE LEARNIN APPROACH TO PHISHING DETECTION AND DEFENSE' (SYNGRESS, 2014, 1ST EDITION). Phishing is the most widespread and pernicious cyber attack , which mostly targets the human rather than the computer by exploiting their vulnerabilities. According to Anti-Phishing Working Group (APWG), phishing is a criminal mechanism employing both social engineering and technical tricks to steal user's identity data and financial account credentials by disguising as a trusted one.

GOEL, D., JAIN, A.K.: 'MOBILE PHISHING ATTACKS AND DEFENCE MECHANISMS: STATEOF ART AND OPEN RESEARCH CHALLENGES', COMPUT. SEC., 2018. To lure the end-users, attackers use malicious websites and e-mails by posing themselves as a trusted one. The supreme goal of phishing is to abduct confidential data such as username, password, bank details, credit card details etc. Attackers perform phishing for many reasons – to gain benefits financially, to steal personal information, to ruin the reputation of the organisations and sometimes just to get fame.

SAHINGOZ, O.K., BUBER,E., DEMIR O.,ET AL.: 'MACHINE LEARNING BASEDPHISHING DETECTION FROM URLS', EXPERT SYST. APPL., 2019. Applied heuristics to extract natural language processing (NLP) features from the URL to detect the URL-based web phishing attacks. The heuristics are derived base don parameters such as raw word count, short word length, Alexa ranking, similar brand name count etc. Yukun Li et al.

PROBLEM STATEMENT

Phishing is one of the techniques which are used by the intruders to get access to the user credentials or to gain access to the sensitive data. This type of accessing the is done by creating the replica of the websites which looks same as the original websites which we use on our daily basis but when a user click on the link he will see the website and think its original and try to provide his credentials . To overcome this problem we are using some of the machine learning algorithms in which it will help us to identify the phishing websites based on the features present in the algorithm. By using these algorithm we can be able to keep the user personal credentials or the sensitive data safe from the intruders.