Ideation Phase

Literature Survey

Team id	PNT2022TMID42127
Project Name	Web Phishing Detection

INTRODUCTION

Phishing is that the fraudulent plan to to obtain sensitive information like username, password, and creditcard details, often malicious purposes, by disguising as a trustworthy entity in an electronic communication. 'Phishing' recorded on 2nd January, 1996 according to Internet records.[6] Social media phishing is when attackers use social networking sites like Facebook, Twitter, and instagram rather than email to obatain your sensitive personal information

LITERATURE SURVEY

Rathore, S., Sharma, P.K., Loia, V., Jeong, Y.S. and Park, J.H. [11] presents a comprehensive survey of different security and privacy threats that target every user of social networking sites. A Social Network Service (SNS) is a type of web service for establishing a virtual connection between people with similar interests, backgrounds, and activities. In recent years, SNSs become a well-liked medium of communication. The number of SNS users worldwide is continuously increasing every year. This paper separately focuses on various threats that arise due to the sharing of multimedia content within a social networking site. In this, describing three classes of threats – Multimedia Content Threats, Traditional Threats and Social Threats [11].

Pujara, P. and Chaudhari, M.B., [10] Phishing frauds might be the most popular cybercrime used today. This paper detailed literature survey and proposed new approach to detect phishing website by features extraction and machine learning algorithm. In this paper author describe different methodologies such as Blacklist method, Heuristic based method, Visual similarity and Machine learning for phishing detection. Blacklist method is used in which list of phishing URL is stored in database and then if URL is found in

database, it is known as phishing URL and gives warning otherwise it is called legitimate. Heuristic based method is extension of blacklist and able to detect new attack as use features extracted from phishing site to detect phishing attack. Visual similarity approach deceive user by extracting image of legitimate site. Machine Learning approach works efficiently in large dataset [10].

Jain, A.K. and Gupta, B.B., [8] Attackers steal sensitive information like personal identification number (PIN), credit card details, login, password, etc., from internet users. In this paper, author proposed a machine learning based anti-phishing system based on Uniform Resource Locator (URL) features. To evaluate the performance of proposed system, author taken 14 features from URL to detect a website as a phishing or non- phishing. The proposed system is trained using quite 33,000 phishing and legitimate URLs with SVM and Naïve Bayes classifiers. Experiment results show quite 90% accuracy in detecting phishing websites using SVM classifier [8].

Sahingoz, O.K., Buber, E., Demir, O. and Diri, B., [12] tries to detect phishing site using url for preventing User's sensitive information. Computer users fall for phishing due to the five main reasons:

- Users don't have detailed knowledge about URLs,
- Users don't know, which web pages can be trusted,
- Users don't see the whole address of the web page, due to the redirection or hidden URLs,
- Users don't have much time for consulting the URL, or accidentally enter some web pages,
- Users cannot distinguish phishing web pages from the legitimate ones.

In proposed system, author used NLP based features and Word features for classification of phishing and non-phishing sites. For classification used Decision Tree, Adaboost, K-star, kNN(n=3), Random Forest, SMO(Squential Minimal Optimization) and Naïve Bayes [12].

Machado, L. and Gadge, J. [9] Phishing sites are the fake websites created by phishers with intent of stealing user's personal information to carry out fraudulent activities. This paper proposes an efficient way for detection of the phishing website using C4.5 decision tree approach. The method proposed in this paper uses various URL features and also uses C4.5 decision tree approach for better results [9].

Jain, A.K. and Gupta, B.B., [7] presents a novel approach that can detect phishing attack by analyzing the hyperlinks found in the HTML source code of the website. A phishing attack is performed by taking advantage of the visual resemblance between the fake and the authentic web-pages. The proposed approach has divided the hyperlink specific features into 12 different categories and used these features to train the machine learning algorithms. Author evaluated the performance of proposed phishing detection approach on various classification algorithms using phishing and non-phishingwebsite dataset

REFERENCES

[1] Godbole, N. and Belapure, S., 2011. Cyber Security, Understanding Computer Forensics and Legal Perspectives.

[2] https://apwg.org/trendsreports/.

[3] https://computer.howstuffworks.com/phishing.htm.

[4] https://inspiredelearning.com/blog/social-phishing/

[5] https://timesofindia.indiatimes.com/city/vadodra/multi-state-job-racket-busted-by-cybercrime-cell/articleshow/66421586.cms

[6] https://www.webopedia.com/DidYouKnow/Internet/phishing.asp.

[7] Jain, A.K. and Gupta, B.B., 2019. A machine learning based approach for phishing detection using hyperlinks information. *Journal of Ambient Intelligence and Humanized Computing*, *10*(5), pp.2015-2028.

[8] Jain, A.K. and Gupta, B.B., 2018. PHISH-SAFE: URL features- based phishing detection system using machine learning. In *Cyber Security* (pp. 467-474). Springer, Singapore.

[9] Machado, L. and Gadge, J., 2017, August. Phishing Sites Detection Based on C4. 5 Decision Tree Algorithm. In *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)* (pp. 1-5). IEEE.

[10] Pujara, P. and Chaudhari, M.B., 2018. Phishing Website Detection using Machine Learning: A Review.

[11] Rathore, S., Sharma, P.K., Loia, V., Jeong, Y.S. and Park, J.H., 2017. Social network security: Issues, challenges, threats, and solutions. *Information sciences*, *421*, pp.43-69.

[12] Sahingoz, O.K., Buber, E., Demir, O. and Diri, B., 2019. Machine learning based phishing detection from URLs. *Expert Systems with Applications*, *117*, pp.345-357.