WEB PHISHING DETECTION

Introduction

Phishing is an online scam where criminals send alluring emails to the organization, user, and more to collect sensitive information. Mostly, this happens through a link sent by an unknown email domain. Anyone can be targeted with a phishing attack, but some types of phishing are done to very specific people. Some threat actors will send out a general email to many people, hoping a few will take the bait based on a common trait. An example would be saying something is wrong with your Facebook or Amazon account, and you need to click this link right away to log in and fix it. The link would likely lead to a spoofed webpage where you might give away your login credentials. Phishing attacks are becoming successful because lack of user awareness.

Literature Survey

1) **Topic :** *Associative classification techniques for predicting e-banking phishing websites*
   Author : (Aburrous et al., 2010)

The research proposed an intelligent, resilient aid effective model that is based on using association and classification data mining algorithms. They used a number of different listing data mining association and classification techniques.

The experimental results demonstrated me feasibility of using associative classification techniques in real applications and its better performance as compared to other traditional classifications algorithms.

2) **Topic:** *Phishing URL Detection with Lexical Features and Blacklisted Domains*
   Author: Hong J., Kim T., Liu J., Park N., Kim SW

Authors in introduced a method for phishing URLs with innovative lexical features and blacklist. They collected a list of URLs using a crawler from URL repositories and collected 18 common lexical features. They implemented advanced ML techniques consisting of under/oversamples and classification. The automated approaches outperform other existing ML apporaches. The study has focused on content features and not lexical features, which was difficult to implement in real-world environments. The experimental

results were better than the existing classification algorithms.

Limitations:

The performance evaluation was based on crawler-based dataset. Thus, there is no assurance for the effectiveness of the URL detector with real time URLs.

3) Topic: *Using case- based reasoning for phishing detection*

Author: Hassan Y.A. and Abdelfettah B.

Authors suggested a URL detector for high precision phishing attacks. They argued that the technique could be scaled to various sizes and proactively adapted. For both legitimate and malicious URLs a limited data collection of 572 cases had been employed. The characteristics were extracted and then weighed as cases to use in the prediction process. The test results were highly reliable with and without online phishing threats. For the improvement of the accuracy, Genetic algorithm (GA) has been used.

Limitations:

The performance of GA based URL detector was better; nonetheless, the predicting time was huge with complex set of URLs.

4) Topic: *PHISH-SAFE: URL Features-Based Phishing Detection System Using Machine Learning*

Author : Jain A.K., Gupta B.B.

Authors in the study proposed a URL-based anti-phishing machine learning method. They have taken 14 features of the URL to detect the website as a malicious or legitimate to test the efficiency of their method. More than 33,000 phishing and valid URLs in Support Vector Machine (SVM) and Naïve Bayes (NB) classifiers were used to train the proposed system. The phishing detection method focused on the learning process. They extracted 14 different features, which make phishing websites different from legitimate websites. The outcome of their experiment reached over 90% of precision when websites with SVM Classification are detected.

Limitation:

Both SVM and NB are slow learners and does not store the previous results in the memory. Thus, the efficiency of the URL detector may be reduced.

References:

[1] Hong J., Kim T., Liu J., Park N., Kim SW, "Phishing URL Detection with Lexical Features and Blacklisted Domains", *Autonomous Secure Cyber Systems*. Springer, 10.1007/978-3-030-33432-1_12.

[2] Hassan Y.A. and Abdelfettah B, "Using case- based reasoning for phishing detection", *Procedia Computer Science*, vol. 109, 2017, pp. 281–288.

[3] Jain A.K., Gupta B.B. "PHISH-SAFE: URL Features-Based Phishing Detection System Using Machine Learning", *Cyber Security. Advances in Intelligent Systems and Computing*, vol. 729, 2018, doi: 10.1007/978-981-10-8536-9_44.

[4] Aburrous et al." *Associative classification techniques for predicting e-banking phishing websites* ", Conference: Multimedia Computing and Information Technology (MCIT), 2010 International Conference, doi:10.1109/MCIT.2010.5444840.