LITERATURE SURVEY
WEB PHISHING DETECTION

Date	19 September 2022
Team ID	PNT2022TMID00733
Project Name	Web Phishing Detection
Maximum Marks	2 Marks

1. **Paper Name:** Detection of Phishing Websites and Secure Transactions Detection of Phishing Websites and Secure Transactions.

Author Name: Dhanalakshmi, R & Prabhu, C & Chellapan, C.

Journal Name: International Journal Communication & Network Security (IJCNS).

LS Content:

The use of a mixture of techniques of social engineering and criminals spoofing the website is an automated extortion of an online identity to trick a user to disclose sensitive data. It gathers personal identification details and financial credentials from the user. Most phishing attacks appear as spoofed e-mails that make users trust and

reveal them by clicking on the links given in the e-mail. The spoofed mails appear as legitimate ones. To describe the website, the claimed title is combined

with human experts and domain features. A variety of legal websites link to domain recognition services, while phishing generally covers domain names and suspicious domain names (fake identities). In addition to blacklists, in the state-of-the-art schemes, white lists, heuristics, and classifications used; R. Dhanalakshmi is proposing to consider the identity statements of websites. With MD5 hashing algorithms, password hashing has been done to allow secure transactions, which strengthens authentication of web passwords. Often it is, it has been shown that getting the actual password from the hashed form is not an easy task due to adding the salt meaning. Get a session key through a mobile if the user is legitimate, from which further access can be done.

2. Paper Name: Intelligent phishing detect ion system for e-banking using fuzzy data mining.

Author Name: Aburrous, Maher & Hossain, Mohammed & Dahal, Keshav & Thabtah, Fadi.

Journal Name: Expert Systems with Applications.37. 7913-7921. 10.1016/j.eswa.2010.04.044.

LS Content:

In evaluating the e-banking phishing website, Maher Aburrous introduced a new technique for fixing 'fuzziness' and proposed a smart, resilient, successful e-banking phishing website detection model. Their model is a mixture of flippant logic and data mining techniques to define the features of the phishing e- banking website, to analyze its techniques by categorizing phishing forms, and to define various parameters for attacking the structured e-banking phishing layer.

3. Paper Name: A Methodical Overview on Detection, Identification and Proactive Prevention of Phishing Websites.

Author Name: Bhagwat M. D., Dr. Patil P. H., Dr. T. S. Vishwanath.

Journal Name: Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV 2021). IEEE Xplore Part Number: CFP21ONG-ART; 978-0-7381-1183-4.

LS Content:

Bhagwat M. D., Dr. Patil P. H. and Dr. T. S. Vishwanath suggest a new approach to detect phished websites. They selected a range of features that differentiate phished websites from genuine websites and arranged these websites according to their priority by machine learning algorithms. They evaluated the features of a URL based on fuzzy rule systems. Their prototype allowed the users to enter the genuine website but if it is a phished website then this prototype sends a notification about the phished website to the corresponding host server administrator and host server administrator blocks that phished website.

4. Paper Name: Phishing website detection: an improved accuracy through feature selection and ensemble learning.

Author Name: Alyssa Anne Ubing, Syukrina Kamilia Binti Jasmi, Azween Abdullah, NZ Jhanjhi , Mahadevan Subramaniam.

Journal Name: International Journal of Advanced Computer Science and Applications.

LS Content:

A feature selection algorithm is employed and integrated with an ensemble learning methodology, which is based on majority voting, and compared with different classification models including Random Forest, Logistic Regression, Prediction model etc. It demonstrates that current phishing detection technologies have an accuracy rate between 70% and 92.52%. The experimental results prove that the accuracy rate yield up to 95%.

5.Paper Name: Improving Email Security with Fuzzy Rules.

Author Name: Chawathe, Sudarshan.

Journal Name: Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks. (ICICV 2021).
IEEE Xplore Part Number: CFP21ONG-ART; 978-0-7381-1183-4.

LS Content:

The more extreme security risks are phishing and other malicious email messages. Automated or semi-automated malicious email detection is an effective tool for combating such email threats. For such purposes, Sudarshan reviews work on using fuzzy rules to identify communications. Experimental review of the usefulness of a fuzzy rule-based classification for other classifiers like those that rely on crisp rules and decision trees, real data set and an output comparison.

6. Paper Name: Intelligent phishing detect ion system for e-banking using fuzzy data mining.

Author Name: Aburrous, Maher & Hossain, Mohammed & Dahal, Keshav & Thabtah, Fadi.

Journal Name: Expert Systems with Applications.37. 7913-7921. 10.1016/j.eswa.2010.04.044.

LS Content:

In evaluating the e-banking phishing website, Maher Aburrous introduced a new technique for fixing 'fuzziness' and proposed a smart, resilient, successful e-banking phishing website detection model. Their model is a mixture of flippant logic and data mining techniques to define the features of the phishing e- banking website, to analyze its techniques by categorizing phishing forms, and to define various parameters for attacking the structured e-banking phishing layer.

7. Paper Name: Phish net: predictive blacklisting to detect phishing attacks.

Author Name: Pawan Prakash, Manish Kumar, Ramana Rao Kompella, and Minaxi Gupta.

Journal Name: International Journal Communication & Network Security (IJCNS).

LS Content:

Phish Net is a predictive blacklisting scheme to detect phishing attacks. Traditional blacklist approaches (i.e., exact match with the blacklisted entries) are easy for attackers to evade. Instead, Phish Net uses five heuristics (i.e., top-level domains, IP address, directory structure, query string, brand name) to compute simple combinations of blacklisted sites to discover new phishing sites. Also, it proposes an approximate matching algorithm to determine whether a given URL is a phishing site or not. Phish Net consists of two major components, namely, component I: predicting malicious URLs and component II: approximate matching.

8. Paper Name: Machine learning based phishing detection from URLs.

Author Name: Ozgur Koray Sahingoz , Ebubekir Buber, Onder Demir , Banu Diri.

Journal Name: Experts Systems with Application

LS Content:

A real-time anti-phishing system, which uses seven different classification algorithms and natural language processing (NLP) based features, is proposed. The system has the following distinguishing properties from other studies in the literature: language independence, use of a huge size of phishing and legitimate data, real-time execution, detection of new websites, independence from third-party services and use of feature-rich classifiers. For measuring the performance of the system, a new dataset is constructed, and the experimental results are tested on it. According to the experimental and comparative results from the implemented classification algorithms, Random Forest algorithm with only NLP based features gives the best performance with the 97.98% accuracy rate for detection of phishing URLs%.

9. Paper Name: Intelligent Web-Phishing detection and protection scheme using integrated features of images, frames and text.

Author Name: Moruf A Adebowale, M Alamgir Hossain.

Journal Name: International Journal Communication & Network Security.

LS Content:

A phishing attack is one of the most significant problems faced by online users because of its enormous effect on the online activities performed. In recent years, phishing attacks continue to escalate in frequency, severity and impact. Several solutions, using various methodologies, have been proposed in the literature to counter the web-phishing threats. Notwithstanding, the existing technology cannot detect the new phishing attacks accurately due to the insufficient integration of features of the text, image and frame in the evaluation process. The use of related features of images, frames and text of legitimate and non-legitimate websites and associated artificial intelligence algorithms to develop an integrated method to address these together.

10.Paper Name: Large-scale automatic classification of phishing pages.

Author Name: Colin Whittaker, Brian Ryner, and Marria Nazif.

Journal Name: In NDSS, volume 10, 2010.

LS Content:

Whittaker et. al. uses a logistic regression classifier to maintain Google's phishing blacklist automatically by examining the URL and the contents of a page. The proposed scheme correctly classifies more than 90% of phishing pages several weeks after training concludes. Marchal et. al. develops a phishing detection system that requires very little training data, which is language independent, resilient to adaptive attacks and implemented entirely on client-side. The proposed target identification algorithm is faster than previous works and can help reduce false positives. The proposed scheme achieves 0.5% false positive rate and 99% true positive rate.

REFERENCES:

1. M. Korkmaz, O. K. Sahingoz and B. Diri, "Detection of Phishing Websites by Using Machine Learning-Based URL Analysis", *2020 11th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1-7, 2020.

2. A. Das, S. Baki, A. El Aassal, R. Verma and A. Dunbar, "SoK: A Comprehensive Reexamination of Phishing Research from the Security Perspective", *IEEE Communications Surveys Tutorials*, vol. 22, no. 1, pp. 671-708, First quarter 2020.

3. M. Korkmaz, O. K. Sahingoz and B. Diri, "Feature Selections for the Classification of Webpages to Detect Phishing Attacks: A Survey", *2020 International Congress on Human-Computer Interaction Optimization and Robotic Applications (HORA)*, pp. 1-9, 2020.

4. J. James, L. Sandhya and C. Thomas, "Detection of phishing URLs using machine learning techniques", *2013 International Conference on Control Communication and Computing (ICCC)*, pp. 304-309, 2013.

5. K. M. Sundaram, R. Sasikumar, A. S. Meghana, A. Anuja and C. Praneetha, "Detecting phishing websites using an efficient feature-based machine learning framework", *Revista Gestão Inovação e Tecnologias*, vol. 11, no. 2, pp. 2106-2112, Jun. 2021.

6. M. A. Adebowale, K. T. Lwin and M. A. Hossain, "Intelligent phishing detection scheme using deep learning algorithms", *J. EnterpriseInf.Manage.*,[online]Available: https://www.emerald.com/insight/content/doi/10.1108/JEIM-01-2020-0036/full/html.

7. P. Prakash, M. Kumar, R. R. Kompella and M. Gupta, "PhishNet: Predictive blacklisting to detect phishing attacks", *Proc. IEEE INFOCOM*, pp. 1-5, Mar. 2010.