LITERATURE SURVEY

WEB PHISHING DETECTION

[1] Paper Name: 'Phishing Scams Cost American Businesses Half A Billion Dollars A Year'.

Author Name : Dr. Gunikhan Sonowal

Content: Phishing remains a basic security issue in cyberspace. In phishing, assailants steal sensitive information from victims by providing a fake site which looks like the visual clone of a legitimate site. Phishing shall be handled using various approaches. It is established that single filter methods would be insufficient to detect different categories of phishing attempts.

[2] Paper Name: Phish net: predictive blacklisting to detect phishing attacks.

Author Name: Pawan Prakash, Manish Kumar

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.

[3] Paper Name: A machine learning based approach for phishing detection using hyperlinks information

Author Name: Ramana Rao Kompella, and Minaxi Gupta.

Content: This paper presents a novel approach that can detect phishing attack by analysing the hyperlinks found in the HTML source code of the website. The proposed approach incorporates various new outstanding hyperlink specific features to detect phishing attack. The proposed approach has divided the hyperlink specific features into 12 different categories and used these features to train the machine learning algorithms. We have evaluated the performance of our proposed phishing detection approach on various classification algorithms using the phishing and non-phishing websites dataset. The proposed approach is an entirely client-side solution, and does not require any services from the third party. Moreover, the proposed approach is language independent and it can detect the website written in any textual language. Compared to other methods, the proposed approach has relatively high accuracy in detection of phishing websites as it achieved more than 98.4% accuracy on logistic regression classifier.

[4] Paper Name: Phishing websites detection using a novel multipurpose dataset and web technologies features.

Author Name: Sahingoz et al

Content: Phishing attacks are one of the most challenging social engineering cyberattacks due to the large amount of entities involved in online transactions and services. Phishing is a fraudulent technique that is used over the Internet to deceive users with the goal of extracting their personal information such as username, passwords, credit card, and bank account information.

[5] Paper Name: Phishing-Alarm: Robust and Efficient Phishing Detection via Page Component Similarity

Author Name: Jian Mao, Wenqian Tian, Pei Li, Tao Wei and Zhenkai Liang

Content: The paper proposes a robust solution to identify phishing pages according to the visual similarity of web page components, which are difficult to be evaded by attackers. The authors developed techniques to select the effective features on a web page, and propose an efficient method for page similarity detection according to these features. Their approach was prototyped and evaluated using a large set of phishing pages. The results illustrate that the approach is efficient and effective.

[6] Paper Name: PDGAN: Phishing Detection With Generative Adversarial Networks

Author Name: Saad Al-Ahmadi, Afrah Alotaibi and Omar Alsaleh

Content: A website is classified as phishing in a machine learning approach if the tested website results match the predefined feature set. The performance of this approach depends on the feature set, training data, and classification algorithm. Using machine learning algorithms can enable unseen URLs to be easily detected. A phishing website detection approach PDGAN, which does not depend on webpage content but rather only on a webpage's URL is designed. PDGAN uses a deep learning model, namely a GAN, whose adversarial process allows the model to learn different variations in phishing features and produce a final model that provides better detection results.

[7] Paper Name: Detection of Phishing Websites and Secure Transactions Detection of Phishing Websites and Secure Transactions.

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

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.

References

[1] Higashino, M., et al. An Anti-phishing Training System for Security Awareness and Education Considering Prevention of Information Leakage. in 2019 5th International Conference on Information Management (ICIM). 2019.

[2]Dr. Gunikhan Sonowal: ‘Phishing Scams Cost American Businesses Half A Billion Dollars

A Year'. Forbes, 5 May 2017. Accessed Jan 2018.
[1] Pawan Prakash, Manish Kumar 'Phish net: predictive blacklisting to detect phishing attacks. SANS Institute, 2007. Accessed Jan 2018.
[2] Ramana Rao Kompella, and Minaxi Gupta. 'A machine learning based approach for phishing detection using hyperlinks information' vol.12, no.2, pp.1–27, 2007.
[3] Sahingoz et al 'Phishing websites detection using a novel multipurpose dataset and web technologies features'. vol.55, no.1, pp.74– 81, 2012.
[4] Jian Mao, Wenqian Tian, Pei Li, Tao Wei and Zhenkai Liang 'Phishing-Alarm: Robust and Efficient Phishing Detection via Page Component Similarity'. IEEE Access (Volume: 5) 23 August 2017
[5] Saad Al-Ahmadi, Afrah Alotaibi and Omar Alsaleh 'PDGAN: Phishing Detection With Generative Adversarial Networks' IEEE Access (Volume: 10) 18 April 2022
[6] Dhanalakshmi, R & Prabhu, C & Chellapan, C ' Detection of Phishing Websites and Secure Transactions Detection of Phishing Websites and Secure Transactions'. International Journal Communication & Network Security (IJCNS).