LITERATURE SURVEY
ON
WEB PHISHING DETECTION

TEAM ID: PNT2022TMID39819

TEAM MEMBERS

M. KOKILA VANI

M. THAMARAISELVI

R. PRIYANKA

M. PRIYADHARSHINI

DATE: 21-09-2022

Survey of review spam detection using machine learning techniques

AUTHOR	YEAR OF PUBLICATION
Michael Crawford ,Taghi M. Khoshogoftaar, joseph D, Prusa, Aaron N. Richter &Hamzah Al Najada	05 October 2015

ABSTRACT

This article illustrates online reviews are often the primary factor in a customer’s decision to purchase a product or service, and are a valuable source of information that can be used to determine public opinion on these products or services. Because of their impact, manufactures and retailers are highly concerned with customer feedback and reviews. Reliance on online reviews gives to the potential concern that wrongdoers may create false reviews to artificially promote or devalue products and services. This practice is known as opinion (Review) spam, where spammers manipulate and poison reviews (i.e., making fake, untruthful, or deceptive reviews) for profit or gain. Since not all online reviews are truthful and trustworthy, it is important to develop techniques for detecting review spam.

TECHNIQUES USED

- Machine Learning
- Review centric review spam detection

PROS

- In recent years, review spam detection has received significant attention in both business and academia due to the potential impact fake reviews can have on customer behaviour and purchasing detection.
- The paper discusses various types of phishing attacks such as spoofing emails, hacking and how to prevent them.

CONS

- The studies discussed in this paper have primarily focused in the area of feature engineering. But which combination of features is best remain unclear.
- Although there are a large number of machine learning algorithms(learners) available, current research using supervised learning methods has been, for the most part, limited to three learners: Logistic Regression (LR), Navie Bayes (NB) and support Vector Machine (SVM).

A Survey and classification of web Phishing detection schemes

AUTHOR	**YEAR OF PUBLICATION**
Manoj Misra, Pradeep K. Atrey	26, October 2016

ABSTRACT

Phishing is a fraudulent techniques that is used over the internet to device users with the goal of extracting their personal information such as username, passwords, credit card, and bank account information. The key to phishing is deception. Phishing uses email spoofing as its initial medium for deceptive communication followed by spoofed websites to obtain the needed information from the victims. This paper studies ,analyses, and classifies the most significant and novel strategies proposed in the area of phished website detection, and outlines their advantages and drawbacks. Furthermore, a detailed analyses of the latest schemes proposed by researchers in various subcategories is provided.

TECHNIQUES USED

- Search engine-based technique
- Heuristics and machine learning based technique
- Phishing blacklist and whitelist-based technique
- Visual similarity-based techniques
- DNS-based techniques

PROS

- The paper focuses on the fact that phishing detection schemes perform better than phishing prevention and user training solutions because they do not require changes in authentication platforms and do not rely on the user's ability to detect phishing.
- Furthermore, phishing detection solutions are cheaper than the phishing prevention solutions in terms of the extra hardware required and password management.

CONS

- If a webpage is carefully designed by Phisher the extracted features might not give enough information to detect Phishing
- It is difficult to detect phishing websites from their visual appearance or via security indicators on mobile phones due to their small screen size.

Phishing Detection: Analysis of Visual Similarity Based Approaches

AUTHOR	YEAR OF PUBLICATIONS
Ankit Kumar Jain and B.B. Gupta	10 January 2017

ABSTRACT

Phishing is one of the major problem faced by cyber-world and leads to financial losses for both industries and individuals. Detection of phishing attack with high accuracy has always been a challenging issue. At present, visual similarities based techniques are very useful for detecting phishing websites efficiently. Phishing websites looks very similar in appearance to its corresponding legitimate websites to deceive users into believing that they are browsing the correct website. Visual similarity based phishing detection techniques utilize the feature set like text content, text format, HTML tags, Cascading Style Sheet(CSS), image, and so forth ,to make the decision. This paper presents a comprehensive analysis of phishing attacks, their exploitation , some of the recent visual similarity based approaches for phishing detection and its comparative study.

TECHNIQUES USED

- Antiphishing Technique: Modus operandi
- Phishing Mechanism
- Pixel based Techniques

PROS

- Visual similarity based approaches compare the visual appearance of the suspicious website to its corresponding legitimate website by using parameters.
- This technique has high accuracy of identifying phishing website.

CONS

- The problem is the zero-hour phishing attack. Most of the antiphishing technique compare the suspicious website from the pool of legitimate sites.
- If attack the designs the new webpage and its target(corresponding legitimate page) is not available in the data set then this technique fails to detect new fake webpage(zero-hour attack).

Intelligent web-phishing detection and protection scheme using integrated features of Images, frames and text

AUTHOR	YEAR OF PUBLICATIONS
M.A.Adebowale, K.T.Lwin, E.Sanchez, M.A.Hossain.	January 2019

ABSTRACT

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.

TECHNIQUES USED

This paper presents an Adaptive Neuro-Fuzzy Inference System (ANFIS) based robust scheme using the integrated features of the text, images and frames for web-phishing detection and protection.

PROS

- ❖ Adaptive Neuro-Fuzzy Inference System based robust scheme provide more accuracy.
- ❖ Combine features text, images & frames for phishing detection proof more detection.
- ❖ This is the first work that reflects the best unified text, image and frame feature.
- ❖ Using SVM for phishing web classification and relate the use with the current result.
- ❖ The proposed solution achieves 98.3% accuracies.

CONS

- ❖ Rules are generated by Neuro-Fuzzy logic are completely in agreement with the findings based on statistical analysis.
- ❖ The structure is not total interpretable.

Web phishing detection techniques: a survey on the state-of-the-art, taxonomy and future directions

AUTHOR	**YEAR OF PUBLICATIONS**
M. Vijayalakshmi, S. Mercy Shalinie, Ming Hour Yang, Raja Meenakshi U.	23 September 2020

ABSTRACT

Internet dragged more than half of the world's population into the cyber world. Unfortunately, with the increase in internet transactions, cybercrimes also increase rapidly. With the anonymous structure of the internet, attackers attempt to deceive the end-users through different forms namely phishing, malware, SQL injection, man-in-the-middle, domain name system tunneling, ransomware, web trojan, and so on. Amongst them, phishing is the most deceiving attack, which exploits the vulnerabilities in the end-users. Phishing is often done through emails and malicious websites to lure the user by posing themselves as a trusted entity. Security experts have been proposing many anti-phishing techniques. Till today there is no single solution that is capable of mitigating all the vulnerabilities. A systematic review of current trends in web phishing detection techniques is carried out and a taxonomy of automated web phishing detection is presented.

TECHNIQUES USED

- List-based detection techniques
- Heuristic rule-based detection techniques
- Learning based detection techniques
- Zero-hour attack

PROS

- A systematic review of current trends in web phishing detection is carried out and a taxonomy of web phishing detection is proposed based on the input parameters chosen.
- High accuracy of finding detection.

CONS

- The limitations of the state-of-the-art web phishing detection approaches are explored by means of detection time, detection rate, and storage complexity to verify the level of robustness against the phishing attack.
- Most of the recent web phishing detection approaches lag in feature selection mechanism as they use handcrafted features to detect the attack.

Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions

AUTHOR	YEAR OF PUBLICATIONS
Nguyet Quang Do, Ali Selamat, Ondrej Krejcar, Enrique Herrera-Viedma	17 February 2022

ABSTRACT

Phishing has become an increasing concern and captured the attention of end-users as well as security experts. Existing phishing detection techniques still suffer from the deficiency in performance accuracy and inability to detect unknown attacks despite decades of development and improvement. Motivated to solve these problems, many researchers in the cybersecurity domain have shifted their attention to phishing detection that capitalizes on machine learning techniques. Deep learning has emerged as a branch of machine learning that becomes a promising solution for phishing detection in recent years. As a result, this study proposes a taxonomy of deep learning algorithm for phishing detection by examining 81 selected papers using a systematic literature review approach.

TECHNIQUES USED

- Deep Learning
- Machine Learning
- Intrusion and Malware detection

PROS

- Recent research by MIT suggested that the DL model’s computational requirements have been growing significantly, which exceeds the ability that specialized hardware can handle.
- There is no standard guideline for an optimal set of parameters that can produce the best performance accuracy.

CONS

- Classical ML techniques still suffer from the lack of efficiency in detecting zero-day phishing attacks .
- Furthermore, complex DL models using GPUs and TPUs in their implementation have certain effects on the environment and energy consumption.
- The amount of carbon dioxide emitted from such models is approximately five times an average car’s lifetime emission.