Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, No. 4, April : 2024

DETECTING PHISHING WEBSITE USING DEEP CONVOLUTION NEURAL NETWORK

Shyamsundar Yadav Dept. Of Computer Science, Trinity college of Engineering and Research Pune University
Lalit Kesare Dept. Of Computer Science, Trinity college of Engineering and Research Pune University
Atharva Shinde Dept. Of Computer Science, Trinity college of Engineering and Research Pune University
Mrs Barkha Shahaji Dept. Of Computer Science, Trinity college of Engineering and Research Pune University

Abstract
In this joint research initiative, we investigate diverse methods targeting the detection and prevention of phishing websites. The first study employs fuzzy logic strategies in real-time detection, offering flexibility through a smart model that integrates smooth logic and machine learning with 30 characteristic features. The second paper provides an organized overview of phishing detection, refining URLbased features and conducting a comparative analysis of anti-phishing tools. The third introduces a deep convolutional neural network (DCNN) model, achieving 99% accuracy in identifying phishing sites. The fourth conducts a comprehensive survey, employing machine learning to enhance performance and comparing various techniques. The fifth segment of our research delves deeply into the exploration of various classifiers and sophisticated techniques for feature selection, shedding light on the remarkable efficacy achieved by the Address bar-based feature group, boasting an impressive accuracy rate of 96.25%. Additionally, the sixth installment of our investigation introduces the innovative concept of "Phishdentity," a novel approach that harnesses the power of website favicons to effectively combat the ever-evolving threat posed by polymorphic phishing websites. The seventh proposes a phishing detection system notifying users about blacklisted URLs through email and alerts. In the eighth installment of our research, we present WC-PAD, a sophisticated system designed specifically for the detection of phishing attacks through web crawling methodologies. This innovative approach incorporates advanced techniques to scrutinize web content, effectively identifying potential threats with remarkable accuracy. Notably, WC-PAD demonstrates an impressive accuracy rate of approximately 98.9%, underscoring its efficacy in safeguarding against malicious cyber activities. Through meticulous analysis and comprehensive monitoring, WC-PAD stands as a formidable defense mechanism, offering proactive measures to mitigate the risks associated with phishing attacks in today's increasingly interconnected digital landscape. The ninth presents a novel approach for real-time detection of phishing websites through URL characteristics analysis. Finally, the tenth addresses web spoofing with an innovative detection approach, achieving low false alarm rates and preemptive detection of various phishing attacks. This collective effort contributes insights into phishing detection, encompassing fuzzy logic, machine learning, deep neural networks, and innovative detection techniques.

Keywords:
phishing detection, fuzzy logic, machine learning, convolutional neural network , deep neural network , web crawling, URL characteristics.

I. Introduction

Phishing involves pretending to be a legitimate website to trick users into sharing personal information like usernames, passwords, and account numbers. It's a common cybercrime, impacting various sectors such as online payments, webmail, financial institutions, file hosting, and more. Webmail and online payments are particularly targeted by phishing.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, No. 4, April : 2024

Phishing can take different forms, such as email phishing scams and spear phishing. Users need to be cautious and not rely solely on standard security applications. Machine Learning is an effective method for detecting phishing because it addresses the limitations of existing approaches.

A URL phishing attack seeks to fraudulently obtain sensitive data, such as usernames and passwords, by tricking users into clicking on deceptive links. To combat this threat, we're rolling out a fresh system. This system bolsters our cybersecurity defenses, employing advanced technology to identify and thwart phishing attempts proactively.

Phishing is a sneaky online scam where scammers create a fake version of a real website. They often send deceptive emails or messages to trick people into sharing personal, financial, or password information, making them believe it's a legitimate website.

To stay safe, it's important to use tools that can detect and block these fraudulent websites.

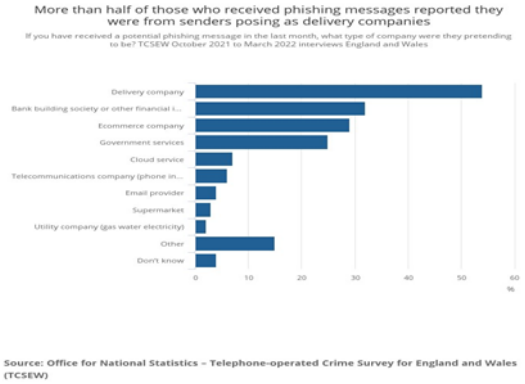

Figure 1 Survey on Phishing websites

Our project provides advanced phishing detection using state-of-the-art machine learning techniques, seamlessly integrated into Python for unparalleled accessibility and efficiency. By harnessing Python's versatile interface, we offer users a streamlined and intuitive platform for robust phishing detection. Our innovative approach leverages the power of machine learning algorithms to deliver superior accuracy and reliability in identifying fraudulent activities online. With direct integration with Python, we ensure a user-friendly experience, enabling seamless implementation and customization according to individual needs. Stay ahead of evolving threats with our solution, confidently safeguarding your digital assets and privacy. Experience the next generation of phishing detection with our Python-powered solution, setting new standards in cybersecurity excellence.

II. Literature

Phishing, a pervasive cyber threat, has prompted extensive research to develop sophisticated detection techniques. Dr. T. S Vishwanath and Dr. Patil P. H propose a real-time approach leveraging fuzzy logic, acknowledging the nuanced and dynamic nature of phishing attempts. Their method utilizes machine learning and smooth logic, employing 30 characteristics specific to phishing websites for accurate identification.

Inspired by a surge in phishing attacks on banking services, Sudhir Dhage and Srushti Patil conduct a methodical overview of phishing detection. Their work emphasizes the refinement of URL-based features, contributing practical insights to enhance existing antiphishing tools. This research provides a valuable foundation for proactive measures against phishing threats.

K. M. Zubair Hasan, Md Zahid Hasan, and Nusrat Zahan tackle the challenge of accurate phishing identification on the World Wide Web. They introduce a deep convolutional neural network (DCNN) model, demonstrating a remarkable 99% accuracy. This marks a significant improvement over traditional methods, showcasing the potential of deep learning in cybersecurity.

Mohammed Hazim Alkawaz and Stephanie Joanne Steven focus on machine learning methods for phishing detection. Their comprehensive survey introduces trained machine learning models and

analyzes various techniques, contributing insights into the identification and analysis of phishing websites.

Shihabuz Zaman, Shekh Minhaz Uddin Deep, Zul Kawsar, Md. Ashaduzzaman, and Ahmed Iqbal Prito investigate phishing through effective classifiers and feature selection techniques. Their emphasis on the Address barbased feature group achieves a high accuracy of 96.25%, demonstrating the efficacy of integrated multi-classification using top algorithms.

Soon Fatt Choo, Jeffrey, Kang Leng Chiew, and San Nah Sze propose "Phishdentity," a method leveraging website favicons for polymorphic phishing detection. Their approach achieves a true positive accuracy of 97.2%, introducing a novel dimension to phishing website identification.

Shihabuz Zaman et al. contribute a phishing detection system with user notifications, enhancing user awareness and preventive measures. Nathezhtha, Sangeetha, and Vaidehi propose WC-PAD, a web crawling-based phishing attack detection system, achieving around 98.9% accuracy, particularly effective against zero-day attacks.

Abdullah introduce a real-time detection approach, scrutinizing url characteristics for diverse phishing attacks. Muhammet baykara and zahit ziya gu¨rel focus on web spoofing, presenting an innovative solution capable of low false alarm rates and preemptive reporting against various phishing attacks.

III. Proposed methodology

3.1 Phishing Detection

1. Module: Administrator Access and Management

In this comprehensive section, the administrator is required to log in using valid credentials to access the system. Once successfully logged in, a plethora of tasks become accessible, empowering the administrator to oversee user management, grant authorizations, evaluate e-commerce websites, review products and associated feedback, access early product reviews, scrutinize keyword search details, analyze product search ratios, and assess product review ranks, fostering robust control and insights within the system.

2.User Viewing and Authorization Module

Within this pivotal module, the administrator gains visibility into the roster of registered users, wielding authority to review and manage user details such as usernames, emails, and addresses, and pivotal decision-making capabilities to grant or revoke user authorizations. Additionally, the module offers detailed insights through chart results viewing, including comprehensive analyses of product search ratios, keyword search results, and product review rankings, augmenting the administrator's ability to make informed decisions.

3. User Interaction Module: Enhancing User Experience

Within this dynamic segment, a diverse user base engages with the system, where registration is a mandatory precursor for participation. Post-registration, user details are securely stored in the database, ensuring data integrity and confidentiality. Upon successful registration, users gain access to a myriad of functionalities, including account management, seamless product searches via keywords, streamlined purchasing processes, and the ability to review their transactional history, fostering a user-centric environment focused on convenience and satisfaction.

4. Phishing Detection System Goals and Objectives: Fortifying Cybersecurity

The overarching goals of the phishing detection system revolve around fortifying cybersecurity by mitigating the prevalence of successful phishing attacks, safeguarding user credentials, and protecting sensitive information from malicious exploitation. To achieve these objectives, a diverse dataset comprising both phishing and legitimate websites is meticulously curated, encompassing various types of phishing attacks and incorporating multifaceted features such as URLs, website content, and metadata, ensuring comprehensive coverage and efficacy in threat detection.

5. Harnessing Machine Learning for Proactive Defense

In conclusion, the integration of machine learning algorithms into phishing detection systems emerges as a promising and effective strategy to combat the persistent threat of fraudulent activities

online. The inherent adaptability and analytical prowess of machine learning algorithms empower the system to analyze evolving patterns in phishing attacks, thereby enhancing detection accuracy and efficiency over time. By incorporating advanced features such as URL analysis, content inspection, and user behavior modeling, our phishing detection system epitomizes a proactive defense strategy against a spectrum of phishing techniques, continually evolving and improving to ensure resilience against emerging threats in cyberspace.
Collectively, these studies contribute a multifaceted understanding of phishing detection, incorporating fuzzy logic, deep learning, machine learning, and innovative methodologies. These advancements lay a foundation for more adaptive and robust systems in the ongoing battle against phishing threats.

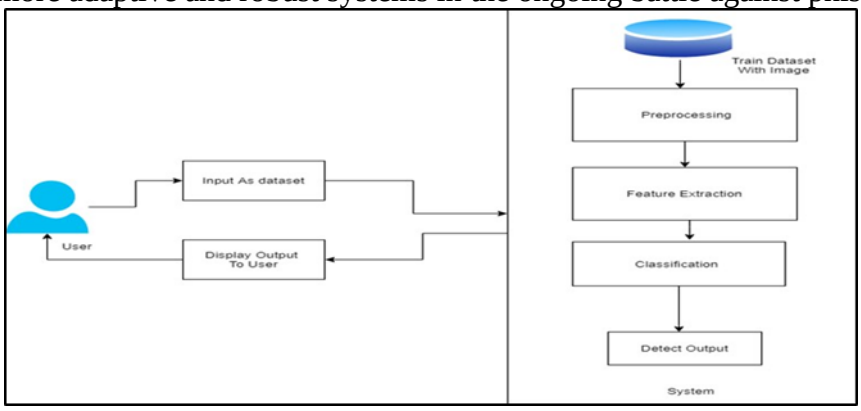

Figure 2: Proposed Architecture Of Phishing Detection

Figure 2 illustrates a sophisticated architecture for phishing detection, encompassing a multi-layered approach to combat fraudulent activities online. Beginning with data acquisition from diverse sources, including website content and user behavior, the system proceeds with preprocessing to refine and format the data for analysis. Feature extraction isolates crucial attributes and patterns, aiding in the differentiation between legitimate and phishing websites. Employing machine learning algorithms, the architecture builds predictive models capable of discerning fraudulent behavior, continuously updated with new data to adapt to evolving threats. Incorporating a feedback loop mechanism for user input and manual validation enhances detection accuracy, while real-time monitoring and alerting promptly flag suspicious activity for mitigation. Emphasizing scalability and efficiency, the architecture integrates seamlessly with existing security frameworks, ensuring comprehensive protection against online threats .

IV. Conclusion

In the application of machine learning in phishing detection for websites proves to be a promising and effective approach in mitigating the persistent threat of fraudulent activities. The ability of machine learning algorithms to analyze and adapt to evolving patterns in phishing attacks enhances the accuracy and efficiency of detection systems . The integration of machine learning techniques into the realm of phishing detection for websites presents a robust and promising strategy for combating the ever-present menace of fraudulent activities. By harnessing the sophisticated capabilities of machine learning algorithms, these detection systems can adeptly scrutinize and adapt to the evolving intricacies of phishing attacks, thereby significantly enhancing the precision and efficacy of their detection mechanisms. This is facilitated by the incorporation of diverse features such as URL analysis, content inspection, and the modeling of user behavior, which collectively fortify the defenses of our phishing detection system against a wide spectrum of phishing techniques. Furthermore, the inherent dynamism of machine learning enables our system to perpetually evolve and refine its methodologies, ensuring a proactive stance in the face of emerging threats and bolstering its resilience over time.

References

[1] Bhagwat M. D, Dr. Patil P. H. Dr. T. S. Vishawanath “A Methodical Overview on Detection, Identification and Proactive Prevention of Phishing Websites” (ICICV 2021)

[2] S. Saudagar, M. Kulkarni, I. Raghvani, H. Hirkani, I. Bassan and P. Hole, "ML-based Java UI for Residence Predictor," 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2023, pp. 838-843, doi: 10.1109/IDCIoT56793.2023.10053480.

[3] Srushti Patil , Sudhir Dhage “A Methodical Overview on Phishing Detection along with an Organized Way to Construct an Anti-Phishing Framework” 2019 (ICACCS)

[4] K. M. Zubair Hasan , Md Zahid Hasan , Nusrat Zahan “Automated Prediction of Phishing Websites Using Deep Convolutional Neural Network” Srushti Patil, Sudhir Dhage g (IC4ME2), 11-12 July, 2019.

[5] Mohammed Hazim Alkawaz , Stephanie Joanne Steven , Asif IqbalHajamydeen , Rusyaizila Ramli “A C o m p r e h e n s i v e S u r v e y o n Identification and Analysis of Phishing Website based on Machine Learning Methods” SCAIE51753.2021.9431794.

[6] Saudagar, S. and Ranawat, R. 2023. Attack Classification and Detection for Misbehaving Vehicles using ML/DL. International Journal on Recent and Innovation Trends in Computing and Communication. 11, 8s (Aug. 2023), 491–496. DOI:https://doi.org/10.17762/ijritcc.v11i8s.7230.

[7] G. A. Jagnade, S. I. Saudagar and S. A. Chorey, "Secure VANET from vampire attack using LEACH protocol," 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Paralakhemundi, India, 2016, pp. 2001-2005, doi: 10.1109/SCOPES.2016.7955799.

[8] Shihabuz Zaman , Shekh Minhaz Uddin Deep , Zul Kawsar , Md. Ashaduzzaman and Ahmed Iqbal Pritom “Phishing Website Detection Using Effective Classifiers and Feature Selection Techniques” (ICIET) 23-24 December, 2019 Soon Fatt Choo, Jeffrey , Kang IEEE DOI 10.1109/ARES.2014.21

[9] S. Saudagar and R. Ranawat, "Detecting Vehicular Networking Node Misbehaviour Using Machine Learning," 2023 International Conference for Advancement in Technology (ICONAT), Goa, India, 2023, pp. 1-3, doi: 10.1109/ICONAT57137.2023.10080114.

[10] Saudagar, S.I., Chorey, S.A., Jagnade, G.A. (2019). Review on Intrigue Used for Caching of Information in View of Information Density in Wireless Ad Hoc Network. Emerging Technologies in Data Mining and Information Security. Advances in Intelligent Systems and Computing, vol 814. Springer, Singapore. https://doi.org/10.1007/978-981-13-1501-5_59.

[11] Mohammed Hazim Alkawaz , Mohammed Hazim Alkawaz , Asif Iqbal Hajamydeen “Detecting Phishing Website Using Machine Learning” (CSPA 2020), 28-29 Feb. 2020.

[12] Nathezhtha , Sangeetha ,Vaidehi , “WC-PAD: Web Crawling based Phishing Attack Detection” 2019 IEEE

[13] Denso Wave's article titled "The Evolution of QR Code Development," published in 2019, was accessed on March 28, 2019. Additionally,

[14] A. Dabrowski, K. Krombholz, J. Ullrich, and E. R. Weippl authored the paper "QR Inception: Exploiting Barcodes within Barcodes," presented at the 4th ACM Workshop on Security and Privacy in Smartphones and Mobile Devices in 2014, with their work spanning pages 3 to 10