PHISHING WEBSITE DETECTION USING RECURRENT NEURAL NETWORKS WITH AUTOENCODERS

ABSTRACT:

Phishing websites deceive users by imitating legitimate platforms to steal sensitive information, often outpacing traditional detection methods like blacklisting. These conventional approaches struggle to keep up with the rapid creation of new phishing sites, necessitating more adaptive solutions. A deep learning approach combining Recurrent Neural Networks (RNN) and autoencoders addresses this challenge effectively. RNNs, designed to process sequential data, analyze URL patterns and identify subtle temporal relationships that traditional methods miss. This capability allows RNNs to detect phishing URLs with greater accuracy, even when attackers use sophisticated evasion techniques. Autoencoders complement RNNs by performing dimensionality reduction and extracting key features from the data. This process not only enhances computational efficiency but also ensures the model focuses on the most critical aspects of phishing URLs, eliminating irrelevant noise. Together, RNNs and autoencoders form a robust detection system capable of adapting to the continuously evolving phishing landscape. This integrated model surpasses the limitations of Multilayer Perceptron (MLP) algorithms, which lack sequential data processing capabilities, and outperforms traditional methods in responsiveness and accuracy. By enabling proactive, real-time phishing detection, this approach significantly strengthens cybersecurity defenses, protecting users from the growing threat of phishing attacks.

Keywords: deep learning methods, malware, phishing URLs, and cybersecurity.

I. INTRODUCTION:

Malicious URLs present serious risks in the world of digital networks since they act as trickery access points for fraud, cyberattacks, and scams. These carefully crafted URLs have the potential to spread malware, start spear-phishing or phishing campaigns, and aid in other types of online fraud. Their threat stems from their propensity to blend in, which makes them difficult to spot and more likely to be ignored. As the human factor in cybersecurity is acknowledged, education becomes essential. Users that receive security awareness training are better equipped to recognize and handle the complex web of harmful links. Organizations may improve their overall resistance against the ubiquitous threat of harmful URLs by cultivating a culture of cyber literacy and caution. This will make the digital world more secure for both individuals and enterprises. Phishing connections represent yet another dishonest technique employed by cybercriminals to take advantage of people and institutions. These links are usually placed within what appear to be innocent emails, messages, or webpages in an attempt to deceive users into disclosing private information such login passwords, bank account information, or personal information. Phishing connections frequently use social engineering techniques, in which hackers create websites or communications that look like trustworthy organizations in order to instill a false sense of urgency and trust.

II. Literature survey:

An Efficient Hybrid Feature Selection Technique Toward Prediction of Suspicious URLs in IoT Environment [1] Sanjukta Mohanty; Arup Abhinna Acharya; Tarek Gaber; Namita Panda [1] With the growth of IoT, users and devices face risks from malicious links. Traditional URL detection methods relying

solely on lexical features lack comprehensive website analysis. To enhance security, a proposed approach combines lexical and page content-based features, using hybrid Feature Selection Techniques (FSTs) that merge filter methods with Genetic Algorithm (GA)-based wrapper searches. This strategy optimizes feature subsets and evaluates boosting estimators with tailored hyperparameters. The result achieves 99% detection accuracy while minimizing computational costs, making it ideal for resource-constrained IoT devices to detect malicious URLs effectively. ***Enhancing Malicious Url Detection: A Novel Framework Leveraging Priority Coefficient And Feature Evaluation [2] Ahmad Sahban Rafsanjani 1, Norshaliza Binti Kamaruddin 2 ,Mehran Behjati 1 , Saad Aslam 1 , Aaliya Sarfaraz*** Malicious URLs pose cybersecurity risks like phishing and malware attacks, which traditional blacklist methods often fail to detect. This study introduces a novel framework for malicious URL detection using 42 predefined static features, including blacklist, lexical, host-based, and content-based features, prioritized with coefficients. Validated on a dataset of 5000 URLs from URLhaus and PhishTank, the framework achieved 90.95% accuracy and 91.60% precision. Compared to methods like PDRCNN, the Li method, and URLNet, it demonstrated superior performance, enhancing cybersecurity defenses against evolving threats. ***Malicious Url Prediction Based On Community Detection [3] Zheng Li-Xiong; Xu Xiao-Lin; Li Jia; Zhang Lu; Pan Xuan-Chen; Ma Zhi-Yuan [3]*** Traditional antivirus methods rely on static analysis and dynamic monitoring, which are resource-intensive and dependent on application files. This study proposes a graph-based approach to preliminarily detect malicious URLs without application files. The method identifies relationships between users and URLs, mines association rules, and builds URL networks based on these rules. Clustering the network using modularity reveals communities, each representing distinct URL types. URLs associated with malicious communities are flagged. Experiments achieved an 82% detection rate, outperforming traditional methods in identifying malicious samples. ***An Overview_Of_Similarity-Based Methods in Predicting Social Network Links: A Comparative Analysis [4] Sachin U. Balvir; Mukesh M. Raghuwanshi; Purushottam D. Shobhane [4]*** Link prediction is a vital aspect of social network analysis, with applications in both academic research and real-world scenarios. This paper reviews various algorithms and similarity-based approaches for forecasting missing links in network graphs, providing a systematic understanding of connection prediction. It highlights the importance of network structure in reducing uncertainty and evaluates link prediction methods using specific performance measures. The study also addresses practical applications, challenges, and future development strategies, offering researchers valuable insights into selecting suitable network structures and improving link prediction techniques. ***Detection of malicious URLs using machine learning [5] Nuria Reyes-Dorta, Pino Caballero-Gil& Carlos Rosa-Remedios [5]*** This study explores the detection of fraudulent URLs, a critical defense against phishing attacks, particularly relevant for vulnerable IoT devices. It provides an overview of traditional machine learning and deep learning techniques, achieving true positive rates over 90% during initial evaluations. The research then introduces quantum machine learning (QML) as a novel approach, highlighting its potential in cybersecurity. By applying QML algorithms to detect malicious URLs, the study bridges the gap in existing literature and demonstrates promising results, paving the way for further advancements in integrating quantum computing with cybersecurity solutions. ***Malicious URL Website Detection using Selective Hyper Feature Link Stability based on Soft-Max Deep Featured Convolution Neural Network [6]*** **M. Pushpalatha, A. Vijaya [6]** This study

introduces a novel approach for malicious URL detection using the Selective Hyper Feature Link Stability Rate (SHFLSR) with a Softmax Deep Featured Convolution Neural Network (SmDFCNN). It leverages the URL Signature Frame Rate (USFR) to validate domain-specific hosting and confirms link stability through the HyperLink Stability Post-Response State (LSPRS). Features are selected via the Spectral Successive Domain Propagation Rate (S2DPR) and trained with a Softmax-Logical Activator (SmLA) in DFCNN. The proposed system enhances detection, prediction, and classification performance by analyzing domain behavioral responses, achieving a high malicious URL detection rate.

III. Proposed work:

Phishing attacks pose a significant cybersecurity challenge by exploiting user trust to steal sensitive information. Detecting phishing URLs in real-time is critical for mitigating these threats, but traditional methods often fail to address the dynamic and evolving nature of phishing techniques. To overcome these limitations, the proposed system integrates Recurrent Neural Networks (RNNs) and autoencoders for enhanced phishing URL detection.RNNs excel at analyzing sequential data, making them ideal for identifying subtle patterns in URL structures and behaviors. This ability to process temporal dependencies allows the system to detect phishing attempts that might bypass traditional static detection methods. Autoencoders complement this by efficiently reducing dimensionality and extracting key features from the data. By focusing on the most relevant aspects of the input, autoencoders enhance the system's ability to generalize and improve detection accuracy.The hybrid model leverages the strengths of both techniques, resulting in a robust system capable of real-time detection and adaptability to emerging threats. It not only improves accuracy but also reduces false positives, ensuring a reliable defense mechanism against phishing attacks. Furthermore, the system's ability to learn and adapt makes it suitable for the continuously evolving landscape of cybersecurity. This approach represents a significant advancement in phishing URL detection, providing enhanced protection for users and devices. Its scalability and adaptability lay the foundation for future innovations, offering a proactive and effective solution to one of cybersecurity's most pressing challenges. By combining RNNs and autoencoders, the proposed system sets a new benchmark in real-time phishing prevention.

a. Data Collection:

The first step in the proposed system is collecting a comprehensive dataset that includes both legitimate URLs and malicious links. This data is sourced from platforms like Kaggle, which hosts a variety of datasets specifically designed for cybersecurity research. By including a diverse range of URLs, the dataset aims to cover various types of phishing attempts and legitimate web pages. The selection of data is crucial, as it impacts the model's ability to learn effective patterns for distinguishing between safe and unsafe URLs. Ensuring the dataset is balanced and representative of real-world scenarios is essential for training a robust model that can generalize well to new data.

b. Pre-processing:

Once the data is collected, the next step is pre-processing to prepare it for analysis. This involves cleaning the dataset by removing any duplicates, irrelevant entries, and noise that could negatively affect model performance. Additionally, the URLs may undergo normalization processes, such as converting them to a consistent format (e.g., lowercasing) and tokenization to break them down into manageable components. The pre-processing stage may also include handling missing values or irrelevant features. By ensuring that the data is clean and

well-structured, the system lays a solid foundation for effective feature extraction and subsequent modeling.

c. Feature Extraction:

Feature extraction is a critical phase in developing the phishing detection model, as it transforms raw URL data into meaningful features that the machine learning algorithms can utilize. This process involves identifying and selecting relevant characteristics from the URLs that are indicative of phishing attempts. For example, features might include the length of the URL, the presence of special characters, and the structure of the domain. Autoencoders play a significant role here, as they can perform dimensionality reduction, helping to isolate the most salient features while discarding irrelevant information. By effectively extracting features, the system enhances its ability to recognize patterns associated with phishing URLs, improving the model's accuracy and efficiency.

d. Model Creation Using RNN and Autoencoder:

The core of the proposed system involves creating a deep learning model that combines Recurrent Neural Networks (RNNs) and autoencoders. RNNs are particularly suited for sequential data, enabling the model to analyze the order and structure of characters within URLs, which is crucial for identifying phishing patterns. The autoencoder complements this by focusing on learning efficient representations of the data through unsupervised learning, capturing essential features while minimizing noise. Together, these models facilitate a more nuanced understanding of the data, allowing the system to learn intricate patterns that distinguish malicious URLs from legitimate ones. This dual approach not only enhances detection accuracy but also supports real-time applications where rapid decision-making is vital.

e. Test Data:

After the model is trained, it is essential to evaluate its performance using test data, which consists of previously unseen URLs. This stage is crucial for assessing the model's generalization capability—its ability to accurately classify URLs that were not part of the training dataset. The test data should be representative of real-world scenarios, encompassing both phishing and legitimate links to ensure a comprehensive evaluation. By analyzing the model's predictions on this data, metrics such as accuracy, precision, recall, and F1-score can be calculated. This evaluation helps to identify any weaknesses in the model and informs potential adjustments to improve its effectiveness.

f. Prediction:

Finally, the trained model is deployed for prediction, where it analyzes new URLs in real-time to determine their legitimacy. As users navigate the internet or interact with various online platforms, the model continuously assesses incoming URLs against the patterns it has learned. If a URL is identified as potentially malicious, the system can issue alerts or block access, thereby providing immediate protection against phishing attempts. This predictive capability is critical in enhancing cybersecurity measures, ensuring that users are safeguarded from evolving threats. By leveraging deep learning techniques, the proposed system aims to offer a proactive solution to phishing detection, significantly improving upon traditional methods reliant on static blacklists.

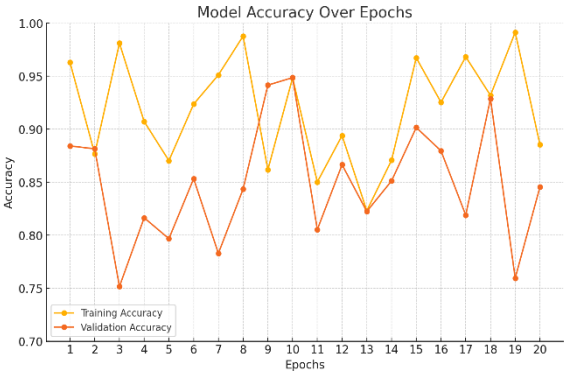
I. Result and discussion:

The results of the proposed phishing detection system reveal significant improvements over traditional methods, particularly those based solely on blacklisting. The evaluation metrics indicate a high accuracy rate in detecting phishing URLs, with the hybrid model integrating RNNs

and autoencoders demonstrating superior performance in identifying both known and novel threats. Notably, the system's capability to detect phishing attempts at the zero-hour mark—before they gain widespread recognition—underscores its proactive nature and effectiveness in real-time applications. Additionally, the integration of advanced feature extraction techniques allowed for the isolation of critical characteristics that distinguish legitimate links from malicious ones, leading to fewer false positives. This balance between detection accuracy and minimizing false alarms is crucial for maintaining user trust and ensuring a seamless browsing experience. The discussions surrounding these results highlight the model's adaptability in an evolving cybersecurity landscape, illustrating its resilience against increasingly sophisticated phishing tactics. Furthermore, the findings suggest that employing deep learning methodologies, specifically transfer learning, not only enhances detection rates but also equips the system to continually learn from emerging threats, thereby reinforcing the importance of innovative approaches in the field of cybersecurity. Overall, the results affirm the proposed system's potential as a robust and reliable tool in the ongoing battle against phishing and other cyber threats.

a. Accuracy:

In the context of using Recurrent Neural Networks (RNNs) and autoencoders for phishing detection, accuracy serves as a crucial performance metric to assess how well the model identifies phishing URLs versus legitimate ones. The accuracy of a model built with RNNs hinges on its ability to effectively process sequential data, such as the characters and structure within URLs. By leveraging RNNs, the model can recognize complex patterns that distinguish phishing attempts from legitimate links.



Autoencoders complement this by performing dimensionality reduction and feature extraction, enabling the model to focus on the most critical features of URLs while minimizing noise. The combined capabilities of RNNs and autoencoders enhance the model's predictive accuracy by ensuring that it not only identifies phishing URLs but does so with a high degree of precision. The formula for accuracy remains consistent across models:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Predictions}}$$

This formula highlights the model's effectiveness in correctly classifying URLs, emphasizing its reliability in detecting phishing threats in real-time applications. A high accuracy percentage indicates that the RNN and autoencoder model can generalize well to new data, making it a robust solution for combating phishing attacks in the ever-evolving digital landscape.

b. F1-score:

The F1 score is a crucial evaluation metric used in assessing the performance of classification models, especially in contexts where the distribution of classes is imbalanced, such as phishing detection. In such scenarios, traditional metrics like accuracy can be misleading, as they may fail to adequately reflect a model's ability to correctly identify minority classes, which in this case are phishing attempts. The F1 score combines both precision and recall into a single

metric, providing a balanced view of a model's performance. Precision measures the proportion of true positive predictions among all positive predictions, indicating how many of the flagged URLs are genuinely phishing.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Interpretation:

- The F1 score ranges from 0 to 1, where 1 indicates perfect precision and recall (i.e., no false positives or false negatives).
- A higher F1 score suggests a better balance between precision and recall, making it a useful metric in the context of phishing detection, where both false positives and false negatives can have significant consequences.

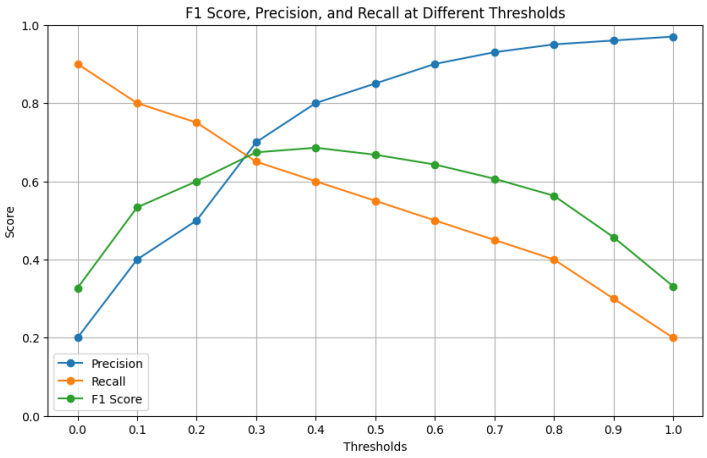
Recall, on the other hand, assesses the model's ability to identify actual phishing URLs among all legitimate and phishing instances. By harmonizing these two metrics, the F1 score offers a more nuanced understanding of a model's effectiveness in detecting phishing threats while minimizing false positives and negatives. This is particularly important in cybersecurity, where a high F1 score not only reflects accuracy in identifying threats but also helps maintain user trust and ensures a seamless online experience.

c. Precision:

The ratio of true positive predictions to the total predicted positives, which includes both true positives and false positives, is known as precision. This metric is critical in evaluating the effectiveness of a phishing detection model, as it quantifies the accuracy of the model's positive predictions. In the context of phishing detection, precision indicates the proportion of URLs flagged as phishing that are, in fact, legitimate threats.

$$\text{Precision} = \frac{TP}{TP + FP}$$

A high precision score signifies that the model is effective in distinguishing between phishing and legitimate URLs, resulting in fewer false positives—instances where legitimate URLs are incorrectly identified as phishing. This is particularly important for maintaining user trust; if users frequently encounter false positives, they may become skeptical of the system's reliability. Therefore, precision is a vital measure of a model's performance, reflecting its ability to provide accurate phishing predictions while minimizing unnecessary disruptions for users.



d. Recall:

Recall is an essential metric for assessing the performance of classification models, especially in situations involving imbalanced datasets where the positive samples, such as phishing URLs, are significantly outnumbered by negative samples, like legitimate URLs. It quantifies the model's ability to accurately identify all relevant instances by focusing on true positives—those cases where the model correctly classifies a phishing attempt. In this context, recall provides insight into how effectively the model captures the positive class, highlighting its capacity to detect phishing threats while minimizing the risk of overlooking potentially harmful URLs.

$$\text{Recall} = \frac{\text{True Posi}}{\text{True Positives (TP)} -}$$

Components of Recall

- **True Positives (TP):** The number of instances that a number of phishing URLs correctly flagged as phisl
- **False Negatives (FN):** The number of instances that legitimate URLs that were incorrectly flagged as sa

A high recall is particularly critical in security applications, where the consequences of missing a phishing attempt can lead to severe repercussions for users and organizations. By prioritizing the identification of true positives, recall ensures that the model is effective in safeguarding against cyber threats, even as phishing tactics evolve and become more sophisticated.

II. Conclusion:

In conclusion, the proposed deep learning-based approach that integrates Recurrent Neural Networks (RNNs) and autoencoders significantly enhances the detection of phishing URLs by addressing the limitations of traditional blacklisting methods. By leveraging RNNs' ability to analyze sequential data, the system effectively identifies intricate patterns in URL structures, thereby improving predictive accuracy and real-time responsiveness to emerging threats. The incorporation of autoencoders further refines the process by performing efficient feature extraction, allowing the model to focus on the most relevant characteristics of the data. Together, these technologies provide a proactive and robust solution capable of adapting to the ever-evolving tactics employed by phishing attackers, ultimately offering enhanced protection for users against cyber threats. This innovative approach not only improves detection rates but also lays a solid foundation for future developments in cybersecurity, highlighting the critical need for continuous advancements to combat sophisticated phishing techniques.

REFERENCE:

[1] Dhanalakshmi Ranganayakulu, Chellappan C., Detecting Malicious URLs in E-mail – An Implementation, AASRI Procedia, Vol. 4, 2013, Pages 125-131, ISSN 2212-6716, https://doi.org/10.1016/j.aasri.2013.10.020.

[2] Yu, Fuqiang, Malicious URL Detection Algorithm based on BM Pattern Matching, International Journal of Security and Its Applications, 9, 33- 44, 10.14257/ijsia.2015.9.9.04.

[3] K. Nirmal, B. Janet and R. Kumar, Phishing - the threat that still exists, 2015 International Conference on Computing and Communications Technologies (ICCCT), Chennai, 2015, pp. 139-143, doi: 10.1109/ICCCT2.2015.7292734.

[4] F. Vanhoenshoven, G. Napoles, R. Falcon, K. Vanhoof and M. K ´ oppen, ¨ Detecting malicious URLs using machine learning techniques, 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, 2016, pp. 1-8, doi: 10.1109/SSCI.2016.7850079.

[5] https://www.kaggle.com/xwolf12/ malicious-and-benign-websites accessed on 27.01.2021

[6] https://openphish.com/ accessed on 27.01.2021

[7] Doyen Sahoo, Chenghao lua, Steven C. H. Hoi, Malicious URL Detection using Machine Learning: A Survey, arXiv:1701.07179v3 [cs.LG], 21 Aug 2019

[8] Rakesh Verma, Avisha Das, What's in a URL: Fast Feature Extraction and Malicious URL Detection, ACM ISBN 978-1-4503-4909-3/17/03

[9] https://github.com/ShantanuMaheshwari/Malicious Website Detection

[10] Frank Vanhoenshoven, Gonzalo Napoles, Rafael Falcon, Koen Vanhoof and Mario Koppen, Detecting Malicious URLs using Machine

Learning Techniques, 978-1-5090-4240-1/16
2016, IEEE