PHISHING EMAIL DETECTION USING IMPROVED RCNN MODEL WITH MULTILEVEL VECTORS AND ATTENTION MECHANISM

Rathna Jyothi C.H. (Faculty Guide) Dept. of CSE, Andhra Loyola Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, India. Chrjyothi@aliet.ac.in
Vamsi Mokshagundam Computer Science and Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, India. Vamsimokshagundam@gmail.com
Sri Santosh Kumar Angadi Computer Science and Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, India. Santoshsri890@gmail.com
Gopi Chandu Battelanka Computer Science and Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, India. Gopichandu951@gmail.com

Abstract –
Phishing emails pose a significant and growing threat worldwide, leading to substantial financial losses. Despite ongoing updates to confrontation methods, the effectiveness remains unsatisfactory. This paper addresses the pressing need for more advanced phishing detection technology. This commence by analyzing email structures and propose a novel phishing email detection model, leveraging an enhanced Convolutional Neural Networks (CNN) model with multilevel vectors and an attention mechanism. This model simultaneously processes email headers, bodies, characters, and words. To assess its effectiveness, an employ an unbalanced dataset reflecting realistic ratios of phishing to legitimate emails is utilized. Experimental results demonstrate the model's superior performance in identifying phishing emails with a high probability while minimizing false positives for legitimate emails. This promising outcome surpasses existing detection methods, confirming the effectiveness of the proposed model in detecting phishing emails.

Keywords: Phishing, CNN, email, email detection, attention mechanism, Word Embedding

I. INTRODUCTION

Profitable fraud known as "phishing" takes advantage of victims' trust by tricking them into disclosing personal information. "Phishers," or dishonest actors, create emails that look like official correspondence in an attempt to trick recipients into clicking on dangerous links or divulging personal information. Cyber fraud poses a serious risk to both consumers and organizations, as the use of online services, such as banking and other technology-driven industries, becomes more and more ingrained in daily life.Our method of phishing email detection uses sophisticated data processing techniques to counter these dynamic threats. Through the use of data mining and Naive Bayes techniques, we build models that learn from available datasets and extract important features for further research. After that, stratified random sampling is used to thoroughly evaluate these models, guaranteeing a fair sample of both authentic and fraudulent emails. Emails are quickly classified according to their likelihood of being phished using binary classification methods, allowing for quick and precise identification.

Our model's ability to differentiate phishing emails from real ones is a testament to its success. Our algorithm achieves high accuracy with few false positives by using advanced analysis of email structures at several levels, such as headers, body, and word usage. These findings represent a major breakthrough in phishing detection technologies, providing improved defense against cyberattacks and enhancing trust in online security protocols.

II. LITERATURE SURVEY

Data mining involves sifting through extensive datasets to identify patterns and relationships for problem-solving. In the context of phishing email detection, the algorithm classifies emails as legitimate or phishing through a binary classification process. The goal is swift and accurate determination of the email's legitimacy by calculating its probability of being phishing, comparing it to a threshold, and classifying accordingly. The paper introduces a new phishing detection model based on naive bias, encompassing email header, body, writing, and word levels [3]. To assess performance, an unbalanced dataset reflecting real phishing and legitimate email rates is utilized. Results indicate superior effectiveness in identifying phishing emails, validating the model. The discussion extends to the application of Naive Bayes in bounding boxes prediction for unseen images. The paper underscores the significance of enhanced phishing detection technology to combat email threats effectively [3]. The narrative delves into the evolution of phishing detection, transitioning into machine learning's role, specifically the amalgamation of NLP and machine learning. Various studies are cited, showcasing diverse approaches such as decision trees, logistic regression, random forests, and SVM. The focus is on leveraging features, including basic, latent topic model, and dynamic Markov chain features, for effective email classification [8, 11].

The paper introduces a novel phishing email detection model, emphasizing simultaneous modeling at multiple levels for comprehensive analysis. Evaluation using realistic datasets reinforces the model's efficacy, outperforming existing methods [8, 9]. The discussion shifts to the importance of feature selection and the impact of information gain on classifier accuracy in machine learning [11].

The narrative broadens to the internet's transformative impact on communication and business, highlighting the necessity of web presence and the prevalence of email communication [7]. Phishing's disruptive nature, aiming to extract sensitive information through deceptive emails, prompts the need for efficient detection mechanisms. The definition and tactics of phishing are explored, emphasizing the cyber threats it poses [7].

The paper underscores the significance of anti- phishing efforts in cybersecurity, particularly in detecting fraudulent content within textual data [7]. Phishing remains a prominent threat vector, exploiting human vulnerabilities through social engineering in emails, social media, and mobile attacks [1, 2]. The evolving landscape includes a shift from individual to organizational targeting, with email and online services surpassing financial institutions as prime phishing targets [1, 2]. The discussion concludes by acknowledging the maturation of ransomware and the rising prominence of mobile malware [7].

III. METHODOLOGY

The approach for Phishing Email Detection Using Improved RCNN (Recurrent Convolutional Neural Network) Model with Multilevel Vectors and Attention Mechanism is likely to include several essential components:

Data Preprocessing:

This stage involves preparing the dataset to train the model. It comprises tasks like tokenization, which breaks down the email text into individual words or tokens, and vectorization, which converts these tokens into numerical vectors that the model can analyze.

Feature Extraction:

The model extracts features from the email content. These elements could include information from the email header, body text, sender information, and other metadata. In this context, "multilevel vectors" denotes that the model may use vectors reflecting many levels of abstraction or granularity inside the email dataset.

Attention Mechanism:

When creating predictions, attention processes are employed to assess the relative value of various incoming data points. Attention methods can assist the model in focusing on pertinent portions of the email content while disregarding background noise or extraneous information in the context of email detection.

RCNN Architecture:

Recurrent neural networks (RNNs) and convolutional neural networks (CNNs) combine their best features to create RCNNs. While CNNs are excellent at discovering spatial patterns in data, RNNs are good at capturing sequential information. These skills are probably strengthened by the upgraded RCNN model for increased phishing email detection performance.

Model Training:

The RCNN model is trained using the prepared dataset. Using gradient descent and backpropagation to update its internal parameters, the model learns to map input email data to the appropriate output labels (phishing or legitimate) during training.

Model Evaluation:

After training, a different dataset is used to evaluate the model's effectiveness in identifying phishing emails. To gauge the model's efficacy, assessment criteria including accuracy, precision, recall, and F1 score can be employed.

Fine-Tuning and Optimization:

To enhance the model's performance and capacity for generalization, additional fine-tuning may be applied using strategies like regularization or hyperparameter tuning.

All things considered, the approach probably consists of preprocessing the data, extracting features, using attention processes, and training a better RCNN model that is especially designed to identify phishing emails. The attention mechanism and multilevel vectors play a crucial role in catching intricate patterns and enhancing the model's capacity to differentiate between authentic and phishing emails.

IV. DISCUSSION

The identification of phishing emails poses a binary classification challenge. Commencing this process involves computing the probability that an email is a phishing attempt. The email corpus is subsequently categorized into two groups: legitimate and phishing emails. A critical step follows as we compare the probability value with a predefined classification threshold. If the calculated probability surpasses this limit, the email is conclusively labeled as a phishing attempt. Our objective is to swiftly and accurately discern whether the target email is legitimate or malicious. Recognizing the escalating threat of phishing emails, there is a pressing need for more potent phishing detection technology.

In this study, we embark on a comprehensive analysis of email structure as a precursor.we aim to concurrently model emails across various dimensions such as email header, email body, writing style, and word selection.

V. EXPERIMENT RESULTS

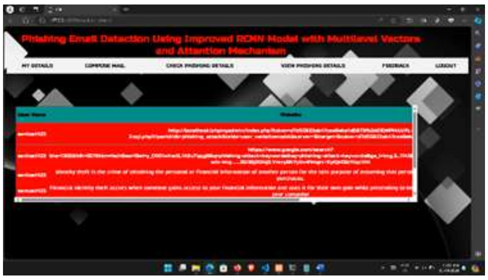
Phishing Email Detection Using Improved RCNN Model with Multilevel Vectors and Attention Mechanism
MY DETAILS
COMPOSE MAIL
CHECK PHISHING DETAILS
VIEW PHISHING DETAILS
FEEDBACK
LOGOUT

Fig.1 Attacked Emails Overview

Attacked Emails Overview provides a comprehensive snapshot of both legitimate and attacker emails encountered. The figure visually portrays the distribution of the emails, offering insights into their composition and volume. Legitimate emails, marked as non-malicious, are depicted alongside attacker emails, which may exhibit various phishing tactics and malicious intent.

By analyzing the attacked graph, one can gain insights into the progression of the attack, identify critical points of compromise, and strategize effective mitigation measures. This visual representation enhances the comprehension of complex cyber threats, facilitating a more informed and targeted response to safeguard digital assets.

Phishing Email Detection Using Improved RCNN Model with Multilevel Vectors and Attention Mechanism
USER DETAILS
ANALYSIS PAGE
CHART PAGE
PHISHING ATTACK
FEEDBACK
LOGOUT

Fig.2 Admin reply

If users encounter issues or have questions about the phishing email detection system, they might submit inquiries or support tickets. An "admin reply" would then be a response from a system administrator or support agent providing assistance or guidance on how to use the RCNN model effectively.

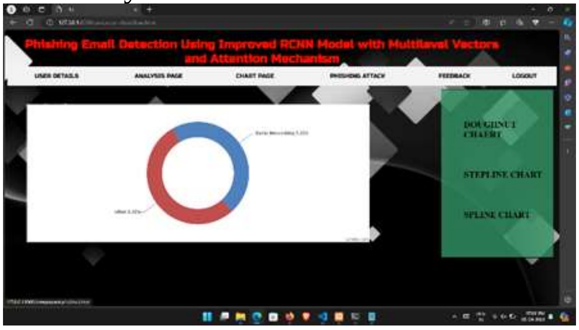

Fig.3 Attacked Graph

The attacked graph offers a visual representation of the cybersecurity landscape, illustrating the connections and interactions among various components during a security incident. This graphical depiction aids in understanding the flow of an attack, highlighting vulnerabilities, attack vectors, and compromised elements. Each node in the graph signifies a distinct component, such as devices, systems, or users, while edges represent the pathways or relationships exploited by attackers.

The analysis of the attacked graph involves identifying patterns of attack propagation, determining the source of phishing emails, understanding how they entered the system, and assessing user vulnerabilities to enhance security. The specifics of the graph depend on available data and the analysis objectives. Specialized tools can aid in the visualization and analysis of phishing attacks in email systems.

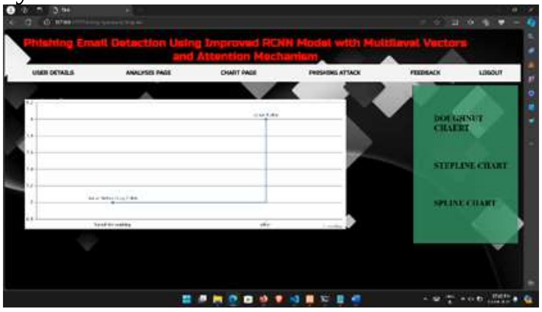

Fig.4 Non-Attacked Graph

The Non-Attacked Graph visualizes the emails in the dataset that have not been attacked or are legitimate. This graph provides insights into the distribution and features of benign emails, emphasizing key attributes or patterns that set them apart from attacker emails. The graph includes

visual components like as bar charts to depict various aspects of non-attacked emails, such as frequency, size distribution, temporal patterns, or sender demographics. By isolating non-attacked emails, this graphic helps to understand the baseline characteristics of normal email traffic, providing essential context for evaluating the effectiveness of phishing email detection algorithms and techniques.

VI. CONCLUSIONS

This model utilizes an improved Convolutional Neural Network (CNN) to effectively capture the intricacies of both the email header and body, operating at both character and word levels. By doing so, we minimize the introduction of noise into the model.

In our approach, we integrate an attention mechanism within both the email header and body, enabling the model to prioritize and focus more on the most valuable information. This attention mechanism enhances the model's ability to discern crucial patterns and features within the email content.

To validate the model's efficacy, we conduct experiments using an unbalanced dataset that closely simulates real-world scenarios. The results demonstrate the model's promising performance in detecting phishing emails. Additionally, we conduct several experiments to showcase the advantages and benefits of our proposed model.

Looking ahead, our future work will concentrate on further refining our model to address the challenge of detecting phishing emails with no email header, relying solely on the information within the email body. This ongoing improvement aims to fortify the model's capabilities and ensure its effectiveness across a broader range of phishing scenarios.

VII. FUTURE WORK

Our future efforts will concentrate on refining our model to enhance its efficacy in detecting phishing emails that lack an email header, relying solely on the email body for analysis. This involves addressing the unique challenges posed by emails without headers and devising innovative approaches to ensure accurate detection.

Furthermore, our research trajectory will delve deeper into understanding how attackers exploit the recipients' vulnerabilities. We aim to conduct an in- depth study of the psychological features manipulated by attackers to deceive users. This exploration is anticipated to yield a more comprehensive set of psychological features that can be directly employed for the detection of phishing emails.

By focusing on these future endeavours, we strive to advance the sophistication of our detection model and contribute to the ongoing development of robust and adaptive solutions to combat evolving phishing threats.

REFERENCES

[1] Anti-Phishing Working Group. (2018). Phishing Activity Trends Report 1st Quarter 2018. [Online]. Available: http://docs.apwg.org/ Preports/apwg_trends_report_q1_2018.pdf

[2] PhishLabs. (2018). 2018 Phish Trends & Intelligence Report. [Online]. Available: https://info.phishlabs.com/hubfs/2018%20PTI %20Report/PhishLabs%20Trend20Report_2018- digital.pdf

[3] M. Nguyen, T. Nguyen, and T. H. Nguyen. (2018). ‘‘A deep learning model with hierarchical LSTMs and supervised attention for anti- phishing.’’ [Online]. Available: https://arxiv.org/abs/1805.01554

[4] Anti-Phishing Working Group. (2016). Phishing Activity Trends Report 4th Quarter 2016. [Online]. Available: http://docs.apwg.org/ reports/apwg_trends_report_q4_2016.pdf

[5] Anti-Phishing Working Group. (2015). Phishing Activity Trends Report 1st-3rd Quarter 2015. [Online]. Available: http://docs.apwg.org/ Preports/apwg_trends_report_q1-q3_2015.pdf

[6] L. M. Form, K. L. Chiew, S. N. Sze, and W. K. Tiong, ‘‘Phishing email detection technique by using hybrid features,’’ in Proc. 9th Int. Conf. IT Asia (CITA), Aug. 2015, pp. 1–5.

[7] Microsoft. (2018). Microsoft Security Intelligence Report.[Online]. Available: https://clouddamcdnprodep.azureedge.net/gdc/gdc VAOQd7/original

[8] M. Hiransha, N. A. Unnithan, R. Vinayakumar, and K. Soman, ‘‘Deep learning based phishing e- mail detection,’’ in Proc. 1st AntiPhishing Shared Pilot 4th ACM Int. Workshop Secur. Privacy Anal. (IWSPA),
A. D. R. Verma, Ed. Tempe, AZ, USA, Mar. 2018.

[9] C. Coyotes, V. S. Mohan, J. Naveen, R. Vinayakumar, and K. P. Soman, ‘‘ARES: Automatic rogue email spotter,’’ in Proc. 1st AntiPhishing Shared Pilot 4th ACM Int. Workshop Secur. Privacy Anal. (IWSPA),
A. D. R. Verma, Ed. Tempe, AZ, USA, Mar. 2018.

[10] S. Sheng, B. Wardman, G. Warner, L. Cranor,
J. Hong, and C. Zhang, ‘‘An empirical analysis of phishing blacklists,’’ in Proc. 6th Conf. Email Anti-Spam (CEAS), Sacramento, CA, USA, 2009, pp. 1– 10.

[11] R. Verma and N. Hossain, ‘‘Semantic feature selection for text with appli- cation to phishing email detection,’’ in Proc. Int. Conf. Inf. Secur. Cryptol. Cham, Switzerland: Springer, 2013, pp. 455–468,

[12] G. Park and J. M. Taylor. (2015). ‘‘Using syntactic features for phishing detection.’’ [Online]. Available: https://arxiv.org/abs/1506.00037

[13] R. Verma, N. Shashidhar, and N. Hossain, ‘‘Detecting phishing emails the natural languageway,’’ in Proc. Eur. Symp. Res. Comput. Secur. Berlin, Germany: Springer, 2012, pp. 824–841.

[14] A. Vazhayil, N. B. Harikrishnan, R.Vinayakumar, and K. P. Soman, ‘‘PED-ML: Phishing email detection using classical machine learning techniques,’’ in Proc. 1st AntiPhishing Shared Pilot 4th ACM Int. Workshop Secur.Privacy Anal. (IWSPA), A. D. R. Verma, Ed. Tempe, AZ, USA, 2018, pp. 1–8.

[15] A. Bergholz, J. De Beer, S. Glahn, M.-F. Moens, G. Paaß, and S. Strobel, ‘‘New filtering approaches for phishing email,’’ J. Comput. Secur.,vol. 18, no.1,pp. 7–35,2010

[16]] J. Singh, ‘‘Detection of phishing e-mail,’’ in Proc. IJCST, vol. 2, no. 1, 2011, pp. 547–549.

[17] X. Gu and H. Wang, ‘‘Online anomaly prediction for robust cluster sys- tems,’’ in Proc. IEEE Int. Conf. Data Eng., Mar./Apr. 2009, pp.1000–1011.