Enhanced Deep Learning Model for Phishing Email Detection Utilizing Multilevel Vectors and Attention Mechanism

P Manjulatha [1] , Dr P Rama Koteswara Rao[2]

[1]pg scholar , department of computer science and engineering , SreeDattha institute of engineering and science sheeriguda, Ibrahimpatnam Hyderabad , Telangana,India.

[2]professor, department of computer science and engineering , Sree Dattha Institute of Engineering and Science, Sheriguda ibrahimpatnam Hyderabad Telangana, India.

Abstract:The phishing email is one of the significant threats in the world today and has caused tremendous financial losses. Although the methods of confrontation are continually being updated, the results of those methods are not very satisfactory at present. Moreover, phishing emails are growing at an alarming rate in recent years. Therefore, more effective phishing detection technology is needed to curb the threat of phishing emails. In this paper, we first analyzed the email structure. Then based on an improved Recurrent Convolutional Neural Networks (RCNN) model with multilevel vectors and attention mechanism, we proposed a new phishing email detection model named, which is used to model emails at the email header, the email body, the character level, and the word level simultaneously. To evaluate the effectiveness of, we use an unbalanced dataset that has realistic ratios of phishing and legitimate emails. Experimental results show that the. Meanwhile, the ensure that the filter can identify phishing emails with high probability and filter out legitimate emails as little as possible. This promising result is superior to the existing detection methods and verifies the effectiveness of in detecting phishing emails.

Keywords : Email Phishing, RCNN, Deep Leaning, Classification

INTRODUCTION

Humans are often perceived as the weakest link in cybersecurity defense, particularly due to social engineering attacks, such as phishing emails. Phishing emails often involve impersonating a trustworthy entity and utilizing urgency or emotional manipulation tactics to make the message appear authentic. The goal is to trick victims into providing sensitive information or clicking on attachments or links that can lead to further attacks. Phishing emails tend to target specific groups of people or exploit critical moments in time in reality, which only account for a small percentage. For example, since the outbreak of the novel coronavirus in late 2019, attackers have been exploiting people's fears by sending phishing emails that are closely related to COVID-19. As the COVID-19 pandemic has spread, people have become less sensitive to it, and phishing emails related to COVID-19 are no longer as common [1]. Before holidays, such as the traditional Mid-Autumn Festival in China, attackers take advantage of people's greed by posing as organizations and sending phishing emails claiming to offer free mooncakes, which are notoriously rare throughout the year [2].

Therefore, the proportion of benign emails to phishing emails is unbalanced in reality.

Phishing emails are also a significant method used by advanced persistent threats (APTs). Currently, over 80% of reported APT attacks involve phishing emails. According to the 2022 China Corporate Email Security Study, corporate email users in China received 42.59 billion phishing emails in 2022, an increase of 24.5% from 34.22 billion phishing emails in 2021 [3]. According to the Global Email Threat Report for 2022, the average number of phishing email attacks per 1000 email addresses worldwide was 299.27 per month, which represents a 12.36% increase from the previous year [4]. Recently, the cybersecurity firm Cofense released the 2023 Email Security Report, which revealed a 569 percent surge in malicious email attacks in 2022 [5]. Therefore, there is an urgent need for an effective method to detect phishing emails.

With the emergence of machine learning (ML) and deep learning (DL) in recent years, numerous researchers have utilized them for detecting phishing emails. The steps can be summarized as follows: (1) selecting and extracting features; (2) choosing a machine learning classifier model; (3) training and testing the model. Email features generally fall into one of two categories: the email's contents and the email's body text. The former contains the structural properties of the email, while the latter contains the semantic features of the email body. The common ML and DL algorithms used to detect phishing emails include support vector machine (SVM), naïve Bayes (NB) decision tree (DT) logistic regression (LR) etc.Although these algorithms have achieved good results in the field of detecting phishing emails, there are many limitations in the experimental design process as follows: (1) Relying only on text-based features or content-based features may not provide comprehensive information; (2) The dataset used is outdated; (3) Using balanced datasets that do not match real-world scenarios; (4) Evaluation metrics are not comprehensive enough. Bountakas et al. propose HELPHED, which addresses the aforementioned issues and obtains satisfactory results [23]. Although they used imbalanced datasets and the experimental design was closer to real scenarios, they did not take any measures to address the imbalance between benign and phishing emails, which may have hindered the model performance and generalization ability. To investigate the impact of the imbalance between benign and phishing emails on the classifier performance and improve the detection rate of phishing emails, this paper presents the following contributions:

To mitigate the impact of unbalanced datasets on classifier performance, two novel algorithms based on undersampling are proposed: FMPED and FMMPED. These algorithms differ in their approach to undersampling benign emails;

To improve classification performance, we use an ensemble learning approach for handling the hybrid features of emails, which combines decision tree (DT) and support vector machine (SVM) as basic classifiers to train content-based features and text-based features respectively;

To simulate real-life scenarios, the ratio of benign emails to phishing emails of the

dataset used in this article is almost 10:1. In order to conduct a more comprehensive evaluation of the classifier performance, we utilize a variety of evaluation metrics that apply to an unbalanced dataset, including the F1-score, accuracy, AUC, G-mean, and Matthews correlation coefficient (MCC).

RELATED WORK

There are various methods for detecting phishing emails, including blacklist-based, machine learning, deep learning, natural language processing, and combinations of these techniques. We compile a list of recent studies on machine learning-based phishing email detection methods that have used both balanced and unbalanced datasets. Due to the fact that our subsequent work is based on [23], we also provide a detailed introduction to it.

2.1. Works Based on Balanced Datasets

Dutta et al., focused on designing a model for phishing email detection and classification using biogeography-based optimization with deep learning [24]. The dataset used in this article is sourced from the CLAIR dataset [25], which comprises 3685 phishing emails and 4894 legitimate emails. The ratio of legitimate emails to phishing emails is 1.3:1. However, it is important to note that the dataset is quite old, dating back to 2008, and therefore may have limited relevance to current trends in email phishing. Magdy et al. proposed a three-classifier system based on deep learning to classify emails into legitimate, spam, and phishing categories based on content characteristics in the dataset [26]. The dataset used in the experiment consisted of 2758 legitimate emails and 2432 phishing emails, which is a ratio of almost 1:1.

Alhogail et al., proposed a model for classifying phishing emails that combines graph convolutional network (GCN) and natural language processing techniques [27]. The dataset used in the experiment is the fraud dataset [25], which includes 3685 phishing emails and 4894 legitimate emails. Two-thirds of the dataset was used for training. The remaining data are for testing purposes. The accuracy of detecting phishing emails in this dataset is 98.2%, with a false-positive rate of only 0.015.

Somesha et al., proposed a method for classifying phishing emails using a combination of word embedding and machine learning algorithms [28]. They used three datasets for the experiment, one of which was a balanced dataset containing 6295 benign emails and 9135 phishing emails, resulting in an accuracy of 99.5%. Valecha et al. [29] proposed a model for detecting phishing emails that utilizes Word2vec [30] and four machine learning classifiers, which takes into account gain and loss clues present in the emails. The dataset included 19,153 legitimate emails and 17,902 phishing emails. It achieved the highest accuracy for gain (96.52%), loss (96.16%), and a combined accuracy of 95.97%. It is evident that the datasets used in the mentioned work are close to balance or outdated, which is far from the actual scenario.

PROBLEM STATEMENT :

Various techniques for detecting phishing emails are mentioned in the literature. In the

entire technology development process, there are mainly three types of technical methods including blacklist mechanisms, classification algorithms based on machine learning and based on deep learning. From previous work, the existing detection methods based on the blacklist mechanism mainly rely on people's identification and reporting of phishing links requiring a large amount of manpower and time. However, applying artificial intelligence to the detection method based on a machine learning classification algorithm requires feature engineering to manually find representative features that are not conducive to the migration of application scenarios. Moreover, the current detection method based on deep learning is limited to word embedding in the content representation of the email. These methods directly transferred natural language processing (NLP) and deep learning technology, ignoring the specificity of phishing email detection so that the results were not ideal Given the methods mentioned above and the corresponding problems, we set to study phishing email detection systematically
based on deep learning. Specifically, this paper makes the following contributions:

Disadvantages

1. With respect to the particularity of the email text, we analyze the email structure, and mine the text features from four more detailed parts: the email header, the email body, the word-level, and the char-level.
2. The RCNN model is improved by using the Then, the email is modelled from multiple levels using an improved RCNN model. Noise is introduced as little as possible, and the context information of the email can be better captured.

PROPOSED MODEL :

With the emergence of email, the convenience of communication has led to the problem of massive spam, especially phishing attacks through email. Various anti phishing technologies have been proposed to solve the problem of phishing attacks. studied the effectiveness of phishing blacklists. Blacklists mainly include sender blacklists and link blacklists. This detection method extracts the sender's address and link address in the message and checks whether it is in the blacklist to distinguish whether the email is a phishing email. The update of a blacklist is usually reported by users, and whether it is a phishing website or not is manually identified. At present, the two well-known phishing websites are PhishTank and OpenPhish. To some extent, the perfection of the blacklist determines the effectiveness of this method based on the blacklist mechanism for phishing email
detection. The current situation is that new threats may not only cause severe damage to customers' computers but also aim to steal their money and identity. Among these threats, phishing is a noteworthy one and is a criminal activity that uses social engineering and technology to steal a victim's identity data and account
information. According to a report from the Anti-Phishing Working compared with the fourth quarter of According to the striking data, it is clear that phishing has shown an apparent upward trend in recent years.

Similarly, the harm caused by phishing can be imagined as well.

Advantages

1. Phishing email refers to an attacker using a fake email to trick the recipient into returning information such as an account passwordto a designated recipient.
2. Additionally, it may be used to trick recipients into entering special web pages, which are usually disguised as real web pages, such as a bank's web page, to convince users to enter sensitive information such as a credit card or bank card number and password. Although the attack of phishing email seems simple, its harm is immense.

ALGORITHM

R-CNN Algorithms

Let's quickly summarize the different algorithms in the R-CNN family (R-CNN, Fast R-CNN, and Faster R-CNN) that we saw in the first article. This will help lay the ground for our implementation part later when we will predict the bounding boxes present in previously unseen images (new data). R-CNN extracts a bunch of regions from the given image using selective search, and then checks if any of these boxes contains an object. We first extract these regions, and for each region, CNN is used to extract specific features. Finally, these features are then used to detect objects. Unfortunately, R-CNN becomes rather slow due to these multiple steps involved in the process. Fast R-CNN, on the other hand, passes the entire image to ConvNet which generates regions of interest (instead of passing the extracted regions from the image). Also, instead of using three different models (as we saw in R-CNN), it uses a single model which extracts features from the regions, classifies them into different classes, and returns the bounding boxes. All these steps are done simultaneously, thus making it execute faster as compared to R-CNN. Fast R-CNN is, however, not fast enough when applied on a large dataset as it also uses selective search for extracting the regions.

RESULTS :

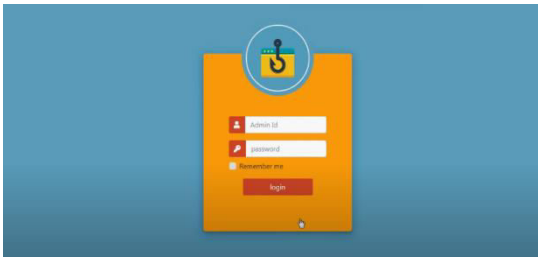

Fig: 1 Admin page

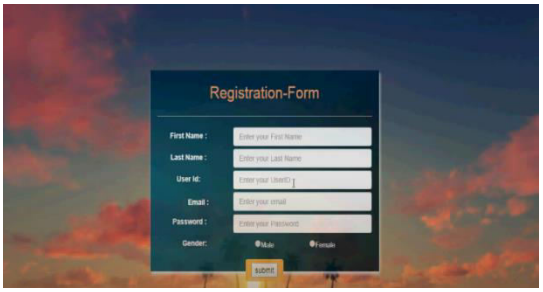

Fig: 2 Registration page

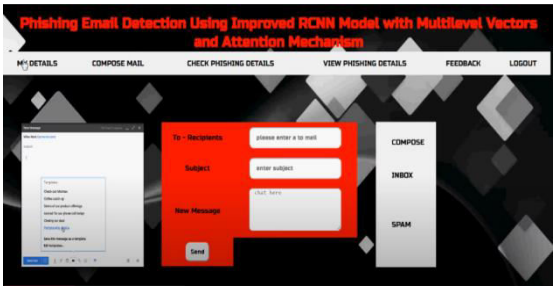

Fig:4 Details page

Fig:5Spam page

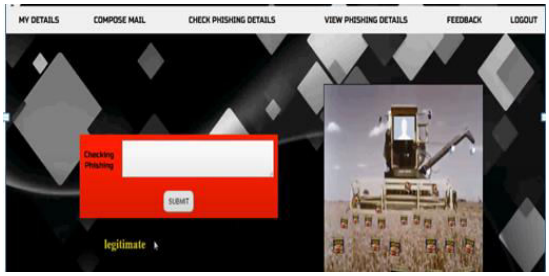

Fig:6 Checking page

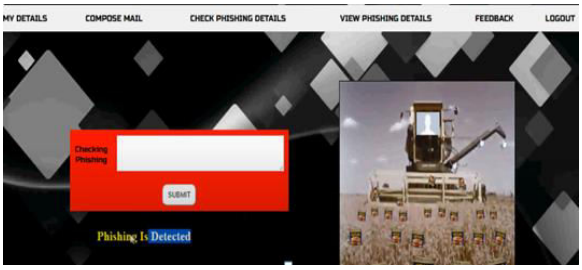

Fig:7 Phising Detected page

CONCLUSION:

we use a new deep learning model named to detect phishing emails. The model employs an improved RCNN to model the email header and the email body at both the character level and the word level. Therefore, the noise is introduced into the model minimally. In the model, we use the attention mechanism in the header and the body, making the model pay more attention to the more valuable information between them. We use the unbalanced dataset closer to the real-world situation to conduct experiments and evaluate the model. The model obtains a promising result. Several experiments are performed to demonstrate the benefits of the proposed model. For future work, we will focus on how to improve our model for detecting phishing emails with no email header and only an email body.

REFERENCES

[1]INTERPOL Report Shows Alarming Rate of Cyberattacks during COVID-19. Available online:

https://www.interpol.int/News-and-Events/News/2020/INTERPOL-report-shows-alarming-rate-of-cyberattacks-during-COVID-19 (accessed on 4 September 2020).

[2]The University of Science and Technology of China Sent 40,000 “Free Mooncake Giveaway” Phishing Emails. Available online: https://www.thepaper.cn/newsDetail_forward_19819224 (accessed on 8 August 2022).

[3]2022 China Corporate Email Security Study. Available online: https://www.qianxin.com/threat/reportdetail?report_id=294 (accessed on 27 March 2023).

[4]Global Email Threat Report for 2022. Available online: http://mailsec.cn/news/html/?539.html (accessed on 31 January 2023).

[5]2023 Email Security Report. Available online: https://cofense.com/blog/phishing-emails-increased-in-2022-according-to-annual-report-from-cofense/ (accessed on 29 March 2023).

[6]Verma, P.; Goyal, A.; Gigras, Y. Email phishing: Text classification using natural language processing. Comput. Sci. Inf. Technol. 2020, 1, 1–12. [Google Scholar] [CrossRef]

Vinayakumar, R.; Soman, K.P.; Poornachandran, P.; Mohan, V.S.; Kumar, A.D. ScaleNet: Scalable and hybrid framework for cyber threat situational awareness based on DNS, URL, and email data analysis. J. Cyber Secur. Mobil. 2019, 8, 189–240. [Google Scholar] [CrossRef]

[7]Kumar, A.; Chatterjee, J.M.; Díaz, V.G. A novel hybrid approach of SVM combined with NLP and probabilistic neural network for email phishing. Int. J. Electr. Comput. Eng. 2020, 10, 486. [Google Scholar] [CrossRef]

[8]Niu, W.; Zhang, X.; Yang, G.; Ma, Z.; Zhuo, Z. Phishing emails detection using CS-SVM. In Proceedings of the IEEE International Symposium on Parallel and Distributed Processing with Applications and IEEE International Conference on Ubiquitous Computing and Communications, Guangzhou, China, 15 December 2017. [Google Scholar]

[9]Hamisu, M.; Mansour, A. Detecting advance fee fraud using nlp bag of word model. In Proceedings of the IEEE 2nd International Conference on Cyberspac,

Nagoya, Japan, 26–29 June 2020. [Google Scholar]

[10]Junnarkar, A.; Adhikari, S.; Fagania, J.; Chimurkar, P.; Karia, D. E-mail spam classification via machine learning and natural language processing. In Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, Tirunelveli, India, 4–6 February 2021. [Google Scholar]

[11]Castillo, E.; Dhaduvai, S.; Liu, P.; Thakur, K.S.; Dalton, A.; Strzalkowski, T. Email threat detection using distinct neural network approaches. In Proceedings of the 1st International Workshop on Social Threats in Online Conversations: Understanding and Management, Marseille, France, 10 May 2020. [Google Scholar]

[12]Peng, T.; Harris, I.; Sawa, Y. Detecting phishing attacks using natural language processing and machine learning. In Proceedings of the IEEE 12th International Conference on Semantic Computing, Laguna Hills, CA, USA, 31 January–2 February 2018. [Google Scholar]

[13]Unnithan, N.A.; Harikrishnan, N.B.; Vinayakumar, R.; Soman, K.P.; Sundarakrishna, S. Detecting phishing E-mail using machine learning techniques. In Proceedings of the 1st Anti-Phishing Shared Task Pilot 4th ACM IWSPA Co-Located 8th ACM Conference on Data and Application Security Privacy, Tempe, AZ, USA, 21 March 2018. [Google Scholar]

[14]Swetha, M.S.; Sarraf, G. Spam email and malware elimination employing various classification techniques. In Proceedings of the 2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology, Bangalore, India, 17–18 May 2019. [Google Scholar]

[15]Chowdhury, M.U.; Abawajy, J.H.; Kelarev, A.V.; Hochin, T. Multilayer hybrid strategy for phishing email zero-day filtering. Concurr. Comput. Pract. Exper. 2017, 29, e3929. [Google Scholar] [CrossRef]

[16]Harikrishnan, N.B.; Vinayakumar, R.; Soman, K.P. A machine learning approach towards phishing email detection. In Proceedings of the Anti-Phishing Pilot at ACM International Workshop on Security and Privacy Analytics, Tempe, AZ, USA, 21 March 2018. [Google Scholar]

[17]Rastenis, J.; Ramanauskaitė, S.; Suzdalev, I.; Tunaitytė, K.; Janulevičius, J.; Čenys, A. Multi-Language spam/Phishing classification by Email Body text: Toward automated security Incident investigation. Electronics 2021, 10, 668. [Google Scholar] [CrossRef]

[18]Sharma, V.D.; Yadav, S.K.; Yadav, S.K.; Singh, K.N.; Sharma, S. WITHDRAWN: An effective approach to protect social media account from spam mail—A machine learning approach. Mater. Today Proc. 2020, 12, 377. [Google Scholar] [CrossRef]

[19]Das, A.; Baki, S.; El Aassal, A.; Verma, R.; Dunbar, A. SoK: A comprehensive reexamination of phishing research from the security perspective. IEEE Commun. Surv. Tut. 2019, 22, 671–708. [Google Scholar] [CrossRef] [Green Version]

[21]El Aassal, A.; Baki, S.; Das, A.; Verma, R.M. An in-depth benchmarking and evaluation of phishing detection research for security needs. IEEE Access 2020, 8, 22170–22192. [Google Scholar] [CrossRef]

[22]Gangavarapu, T.; Jaidhar, C.D.; Chanduka, B. Applicability of machine learning in spam and phishing email filtering: Review and approaches. Artif. Intell. Rev. 2020, 53, 5019–5081. [Google Scholar] [CrossRef]

[23]Bountakas, P.; Xenakis, C. Helphed: Hybrid Ensemble Learning Phishing Email Detection. J. Netw. Comput. Appl. 2023, 210, 103545. [Google Scholar] [CrossRef]

[24]Dutta, A.K.; Meyyappan, T.; Qureshi, B.; Alsanea, M.; Abulfaraj, A.W.; Al Faraj, M.M.; Sait, A.R.W. Optimal Deep Belief Network Enabled Cybersecurity Phishing Email Classification. Comput. Syst. Sci. Eng. 2023, 44, 2701–2713. [Google Scholar] [CrossRef]

[25]Clair Collection of Fraud Email, ACL Data and Code Repository. Available online: http://aclweb.org/aclwiki (accessed on 8 June 2008).