Phishing E-mail Detection Using CNN

Ch. Satyanarayana Reddy [1], K. Pavani [2], Ch. Geeth Adarsh[3]
#1 Assistant Professor in the Department of MCA, SRK Institute of Technology, Vijayawada..

#2 Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada

#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada.

Abstract: One of the biggest problems facing the globe today is phishing emails, which have resulted in enormous financial losses. Even while confrontation techniques are always being improved, their current outcomes are not very good. Furthermore, the number of phishing emails has increased alarmingly in recent years. Therefore, to reduce the threat of phishing emails, more efficient phishing detection technology is required. The email structure was the first thing we examined in this article.

We then suggested a novel phishing email detection model based on an enhanced Convolutional Neural Networks (CNN) model with multilevel vectors and an attention mechanism. This model is used to simultaneously model emails at the header, body, character, and word levels. We utilize an unbalanced dataset with actual ratios of phishing and legitimate emails to assess the efficacy. In the interim, make sure the filter can detect phishing emails with a high likelihood and remove as few legitimate emails as feasible. This encouraging outcome outperforms current detection techniques and confirms the efficacy of phishing email detection...

Index terms *- Phishing Email Detection, Deep Learning, Convolutional Neural Network (CNN), Attention Mechanism, Cybersecurity, Email Classification, Machine Learning, Phishing Attack Prevention, Natural Language Processing (NLP), Email Header Analysis, Feature Extraction, Multilevel Vector Representation, Spam Detection, Fraud Detection, Intelligent Email Filtering..*

1. INTRODUCTION

Phishing is a profitable type of fraud in which dishonest recipients deceive and get private information under false pretenses. Fisher emails will direct users to click on the link or link on the website where you must enter private information, such credit card numbers and passwords. The majority of commercial and technology businesses, including banks, have been drawn to offer their services online by recent advancements in web technology. People's security and privacy can be seriously threatened by cyber fraud since they depend on Internet services for their transactions. Give up any personal information the fraudster may have used to access the user's account without authorization. For instance, malware in the form of malware plug-ins or email attachments may be included in a fraudulent email sent to a user. The malware will install itself on the desktop and transfer funds to the fraudster's bank account each

time the user attempts to complete an online transaction if the user downloads a link on the computer.

Phishing attacks employ websites and emails that mimic those of reputable businesses in order to trick consumers into divulging personal or financial information. Sensitive user data could be used maliciously by an attacker. Users may be duped into replying to bogus emails, publishing sensitive information via a web form, or downloading and installing Trojan horses that search users' systems or monitor their online activity.

The number of fraudulent operations is growing every day, and victims of prior attacks are now looking for measures to defend themselves. They must strengthen their security mechanism in order to accomplish this, which means that the current security system must be significantly upgraded. Fraudulent activity can be detected by the system and stopped.

Phishing is a dishonest attempt to pose as a reliable business in order to obtain private information. These dispute resolution techniques are continuously being modified, and their current outcomes are not very favorable. We solve this issue by analyzing emails using the Naive Bayes theorem. The concept of online consumers has significantly changed due to the quick development of Internet innovation. As of right now, the new attacks are intended to steal as well as seriously harm consumers' systems..

2. LITERATURE SURVEY

Data mining is the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. In the process of discovering patterns in large volumes of datasets involving methods at the intersection of machine learning, classifiers are used for classification of phishing emails. Researchers have aimed to maximize the accuracy and minimize the number of features required for classification. A Naïve Bayes phishing detection approach with high accuracy is presented.

Detection for phishing emails is a binary classification problem. To begin this process, we calculate the probability that the email is a phishing email. Emails are divided into two categories: legitimate emails and phishing emails. Then, the probability value is compared to the classification threshold. If it exceeds the classification limit, it is determined as a phishing email. Our goal is to determine whether the target email is legitimate or phishing quickly and accurately. More effective phishing detection technology is needed to control the threat of phishing emails [1], [2].

In this paper, we first analyze the email structure. A new phishing email detection model based on Naïve Bayes is proposed, which can simultaneously model emails at the email header, email body, writing level, and word level. To evaluate performance, we use an unbalanced dataset with realistic rates of phishing and legitimate emails. Experimental results show that the filter can identify phishing emails with high probability while filtering out legitimate emails as little as possible. This promising result is superior to existing detection methods and verifies the effectiveness of phishing email detection [3], [6], [8].

With the emergence of email communication, the convenience of communication has also led to the problem of massive spam, especially phishing attacks through email. Various anti-phishing technologies have been proposed to solve phishing attacks. Sheng et al. [9] 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, two well-known phishing websites are PhishTank and OpenPhish. To some extent, the effectiveness of blacklist-based phishing email detection depends on the quality and completeness of the blacklist [4], [5].

With the development of Artificial Intelligence, phishing email detection has entered the era of machine learning. In particular, the combination of Natural Language Processing (NLP) and machine learning has played a significant role in phishing email detection. Machine learning techniques such as decision trees, logistic regression, random forests, and Support Vector Machines (SVM) have been widely applied for phishing email classification [6], [8].

Deep learning approaches have also shown promising results in phishing detection. Nguyen et al. [3] proposed a deep learning model with hierarchical LSTMs and supervised attention for anti-phishing. Hiransha et al. [8] introduced a deep learning-based phishing email detection system, while Coyotes et al. [9] proposed ARES, an automatic rogue email spotter system. These approaches demonstrate the effectiveness of deep learning and NLP techniques in identifying phishing emails.

The growth of the Internet has revolutionized the digital era. This revolution has completely changed the way we communicate, conduct business, and advertise. In today's world, establishing a successful business requires a strong web presence, and most important communications take place through email. At the same time, phishing emails are sent to users with the objective of stealing sensitive information. These emails often appear to originate from trusted sources and contain links or attachments designed to obtain confidential information from users.

Phishing can be defined as an attempt to steal valuable information such as usernames, passwords, debit/credit card details, and other personal information for malicious purposes while pretending to be a legitimate organization. Phishing attacks rely on fooling users into sharing their sensitive details through emails, SMS, or phone calls. Such attacks persuade users to enter their details into fraudulent websites that act as intermediaries between the victim and the attacker. Most phishing attacks rely on emails and websites that are designed to closely resemble genuine organizations in order to trick users into revealing financial or personal information [1], [2], [7].

Anti-phishing aims to detect phishing content and documents within large pools of textual data. This is an important problem in cybersecurity that helps protect users from fraudulent information. Phishing continues to be one of the top threat vectors for

cyberattacks. Threat actors increasingly leverage social engineering through email, social media, and mobile attacks to deceive users. Industry trends show that attackers are shifting from targeting individuals to targeting organizations. Email and online services have overtaken financial institutions as the top phishing targets, while attacks on social media platforms continue to increase significantly [1], [2], [7]...

3. METHODOLOGY

i) Proposed Work:

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.

ii) System Architecture:

The system architecture presented is designed to detect phishing emails using an advanced machine learning technique called Recurrent Convolutional Neural Network (RCNN). The process begins when a user logs into the system and composes an email. Once the email is composed, it is automatically passed through a detection phase where the RCNN algorithm analyzes the contents of the email to identify whether it is a phishing email or a legitimate one. This detection process includes scanning the text of the email and verifying any URLs present in the message to check for phishing links.

Once the analysis is complete, the email is classified accordingly—either as a phishing email or a normal, legitimate mail. This classification data, along with the email content, is stored in a central database. The stored data can later be accessed and prepared by an admin for further analysis and refinement of the model.The RCNN model plays a vital role in learning and understanding the text features of emails, which helps it distinguish between malicious and safe content. On the administrative side, the admin logs into the system to review the dataset, analyze results, and ensure the accuracy of the phishing detection model. The admin can also address personal concerns or security threats posed by certain emails.

Based on the detection results, the admin can determine the effectiveness of the RCNN model and make improvements to enhance its accuracy. This feedback loop helps in creating a more efficient phishing detection mechanism that adapts over time to new phishing strategies. The final output of the system is displayed in the form of results and graphs, which visually represent the detection performance and help in understanding the success rate of the model.

Overall, the system emphasizes the need for more effective phishing detection technology to reduce the growing threat of phishing attacks. By leveraging the capabilities of RCNN, the system aims to provide a more secure and intelligent email environment for users and administrators alike.

.

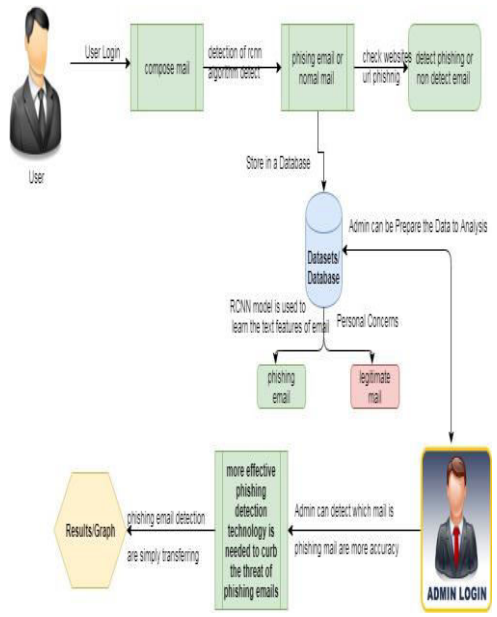


Fig1 proposed architecture

iii) Modules:

1. USER

The client is designed to get the data from the platform. Here the client sends user name and password for getting authentication. The authentication for client access is given, if and only if both the user name and password matches to the details in database. Else access is denied. After authentication, client get session via device using which the client can perform further operations to send and receive mail .Every user should register in that and then we have to send and receive the emails because with that only we can find the phishing mails send by the malicious user or hacker to steal our data . Everyone can prevent their data by using this type of the software.

2. ADMIN:

The admin is designed to get the data from the platform. Here the client sends user name and password for getting authentication. The authentication for client access is given, if and only if both the user name and password matches to the details in database. Else access is denied. After authentication, client get session via device using which the client can perform further operations to maintain everything.

Admin has a default user name and password i.e. admin only, with that credentials only we can log into that. There we can the user who have registered and the mails which have attacked and mails which are not attacked and main thing we can also find the resuts in the form of graph.

3. User Queries:

Users can have queries about the process. This part of project is dedicated to make and get response for queries that are needed to answerable. The major part of the modules is making project as interactive one, queries have been very normally arise to users regarding different details about the process.

4. Graph Analysis:

Graph analysis is the part where admin can knows the statistics about process of details. The data are taken from the project flow and it shows until updated value. The data are gives clear solution to admin that part of improvement and user satisfaction and other factors.

5. Analysis:

Analysis of email structure. a circle represents a character, and a rectangle represents a word. A rectangle is filled with an indefinite number of circles, indicating that the word consists of an indefinite number of characters.

iv) Algorithms:

1. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a deep learning architecture specifically designed for image processing and classification. CNN automatically extracts hierarchical features from images through convolution, pooling, and fully connected layers, enabling accurate detection and classification of fake and genuine logos.

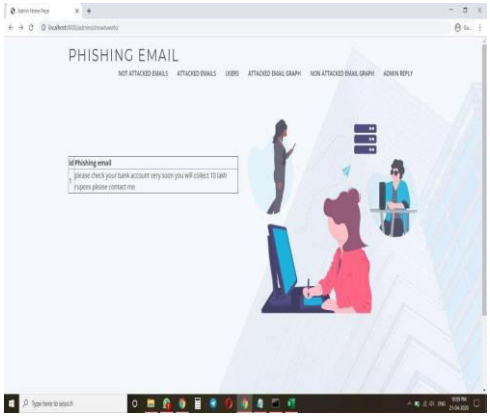
.4.EXPERIMENTAL RESULTS

The experimental evaluation of the proposed system was carried out using IoT network traffic datasets containing both normal and malicious traffic records. The system was implemented using Python, Flask framework, and machine learning libraries such as Scikit-Learn, NumPy, and Pandas. Machine learning algorithms including Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) were trained using preprocessed IoT traffic data to perform anomaly detection and attack classification. The experimental results demonstrate that the proposed system successfully identifies abnormal traffic behavior and classifies various cyber-attacks such as DoS, DDoS, malware attacks, reconnaissance attacks, botnet activities, and data theft attacks with improved accuracy and reduced false positive rates.

The developed web-based application provides real-time anomaly detection, live input analysis, traffic monitoring, attack confidence evaluation, and risk reference analysis through an interactive user interface. Experimental outputs show that the system effectively analyzes network traffic samples, detects suspicious activities, and generates attack alerts along with confidence scores and traffic feature analysis. The system also supports manual live input analysis and reference-based risk prediction for identifying high-risk and low-risk network traffic conditions. The obtained results confirm that the proposed machine learning framework provides efficient, scalable, and intelligent security protection for IoT environments against modern cyber threats.

Accuracy: A test's accuracy is its capacity to distinguish healthy from ill cases. Find the percentage of instances with genuine positives and negatives to assess test accuracy.

Accuracy = TP + TN /(TP + TN + FP + FN)



$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: Classification accuracy or positive cases constitute precision. The formula for accuracy is:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\ cision = \frac{TP}{(TP + FP)}$$

Recall: A model's recall measures its ability to recognize all appropriate machine learning class instances. The ratio of accurately predicted positive observations to total positives indicates a model's class instance detection skill.

$$Recall = \frac{TP}{(FN + TP)}$$

mAP: Mean Average Precision ranks quality. It considers the number and order of relevant ideas. Calculating MAP at K uses the arithmetic mean of each user or query's Average Precision (AP).

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{the AP of class } k$$

$$n = \text{the number of classes}$$

F1-Score: A high F1 score suggests an accurate machine learning model. Integrating recall and precision improves model correctness. Accuracy measures how often a model predicts a dataset correctly.

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\ cision)}{(Recall + \Pr e\ cision)}$$

Fig2 Admin reply

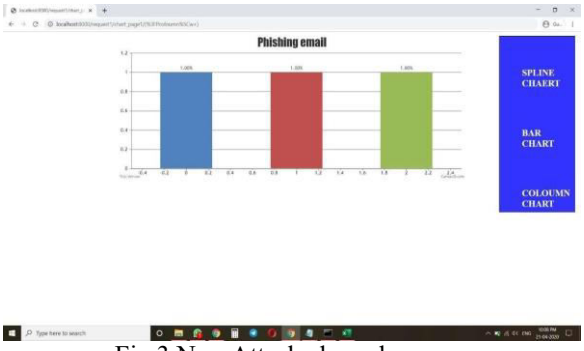
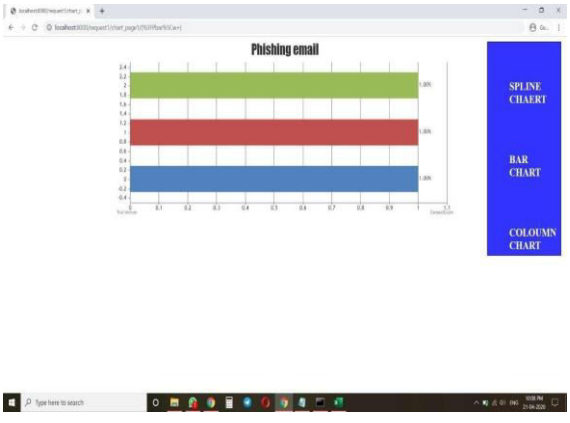


Fig 3 Non-Attacked graph



.

Fig 4 Attacked graph

4. CONCLUSION

To identify phishing emails, we employ a novel deep learning model. The email header and text are modeled at both the character and word levels using an enhanced CNN. As a result, very little noise is added to the model. In order to have the model focus more on the most important information between the header and the body, we employ the attention mechanism in both. We test and assess the model using the unbalanced dataset that is more representative of the real world. The model produces an encouraging outcome. To illustrate the advantages of the suggested model, a number of experiments are conducted. Future research will concentrate on refining our model for identifying phishing emails that just include an email body and no email header..

5. FUTURE SCOPE

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. In the future, we will further study the means by which an attacker users the recipient's weakness. We expect to find more effectively psychological features which can be used directly to detect the phishing emails.

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%20Trend %20Report_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]. Avail- able: https://clouddamcdnprodep.azureedge.net/gdc/gdcVAOQd7/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)*,

Author Profiles

Mr. Ch. Satyanarayana Reddy Completed his MCA, BCA. He has System Administrator. He also a web developer and python developer, currently working has an Assistant Professor in the department of MCA at SRK Institute of Technology, Enikepadu, NTR District. His area of interest includes Artificial Intelligence and Machine Learning.

Ms. K. Pavani Working as Assistant professor & Head of Department of MCA, in SRK Institute of technology in Vijayawada. She done with MCA, M. Tech in Computer Science. Her area of interest includes Artificial intelligence, Machine Learning with Python and DBMS.

Mr. Ch. Geeth adarsh is MCA Student in the Department of Computer Applications at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. He has Completed Degree in B.Sc. (computers) from Sarada degree College Vijayawada. His area of interest are DBMS and Machine Learning with Python.