International Journal of Modern Computation, Information, and Communication Technology
2023;6(2):38-43.
ISSN: 2581-5954

Email Phishing Filtering Techniques

R. Monisha[1*], Veto Dey[2], N. Jayapriya[3], I. Shanmugapriya[4]

[1]*Department of Management Studies, St. Joseph's College of Engineering, Chennai, Tamil Nadu, India*
[2]*Department of Computing Technologies, College of Engineering and Technology, SRMIST, Kattankulathur, Chennai, Tamil Nadu, India.*
[3]*Sri Venkateshwara College of Education, Telangana, India*
[4]*Postgraduate of Management (PGDM), Rajalakshmi School of Business, Chennai, Tamil Nadu, India*

[*]*Corresponding author: monisharamaraj89@gmail.com*

Abstract. One of the biggest issues with the Internet today is phishing emails, which can be both frustrating and costly for businesses and customers. There are several methods for identifying and blocking phishing emails, but none of them are entirely effective. This article provides a comprehensive overview of the present level of knowledge on these types of assaults. This study is the first large-scale examination of how to prevent phishing emails. Provide a survey of the current methods used to identify phishing emails at various points in the attack lifecycle, with an emphasis on machine learning approaches. These filtering strategies are analyzed and compared with one another. This sheds light on the nature of the issue, the range of possible solutions, and the likely future of research.

Keywords: Phishing email, Classifiers, Authentication, Network-level protection, Filtering

INTRODUCTION

Millions of dollars are lost every year due to Spam and other forms of unwanted email communication. Although many models and techniques have been developed to automatically find spam mail, nothing has shown 100% predictive accuracy. Machine learning and deep learning algorithms performed better than any of the other suggested models. The models' precision was boosted by using NLP (natural language processing). Word embedding is presented here for its usefulness in the classification of spam mail. To perform the job of distinguishing spam mails from non-spam mails (HAM), the BERT (Bidirectional Encoders Representation from Transformer) transformers models have been pre-trained. By using attention layers, BERT can consider the broader context of the text [1]. For comparisons, use standard deep neural networks (DNN) models with two stacked dense layers and BiLSTM (bidirectional Long short-term Terms Memories) layers. The findings are also compared to those obtained by using the traditional classifier k-NNs and NBs. One open-sourced data set is utilized for model training, while another is used to evaluate the model's durability and resilience on novel information. The suggested technique achieved best-in-class F1 scores of 98.66% and accuracies of 98.67%.

Data mining's feature selection (FS) is a difficult but crucial part of pattern recognition. This article introduces HWOAFPA, a novel model that combines the benefits of the whale optimization algorithms (WOA) and the flower pollination algorithms (FPA) to address the FS issue considering the ideas of opposition-based learning (OBL). The OBL is executed concurrently with the WOA run, and the WOA population is updated at the same time. It serves as FPA's seed population, improving the accuracy and speed of convergences [2]. The effectiveness of the suggested strategy was determined via a series of two-stage studies. Ten datasets from the University of California, Irvine data repositories and Email spam detection datasets were used for the tests. The initial findings demonstrated that the suggested technique outperformed other basic metaheuristic algorithms in terms of the mean sizes of the selections and the accuracy of the classifications. Furthermore, the second step's finding demonstrated that the suggested technique, when applied to the Email spam dataset, outperformed comparable algorithms with respect to the accuracy with which they detected Email spam.

The need for improved anti-spam tactics or filters, which are used to prevent Spam from entering inboxes, has arisen in response to the exponential growth in the volume of spam emails. Spam email classification techniques based on machine learning have become widespread and very effective. This study gives comprehensive

Received: 19.05.2023 Revised: 08.07.2023 Accepted: 23.07.2023

overviews of previous machine learning-based approaches to spam filtering in email. Researchers will benefit from this analysis since it evaluates the pros and cons of existing machine-learning methods for combating spam [3]. Email has replaced postal mail as the preferred method of communication in the corporate world. Unsolicited commercial email, sometimes known as Spam, has been on the rise with the meteoric rise in email use. There are many motivations for sending such emails: Information gathering, adult content promotion, and product/service advertising and marketing fall under this category. The development of a complete Spam categorization system based on semantics-based text categorization with NLP and URL-based filtering is, thus, of critical relevance. The goal is to develop a model with high performance and efficiency, and therefore, a review of Machine Learning algorithms has been conducted.

Intelligent decision-making and self-driven analytics are within the capabilities of the developing trend of Artificial Intelligence (AI) combined with the Internet of Things (IoT) in industrial applications. Attackers may use the large amounts of data generated by IoT devices to wreak havoc on regular business and service operations. As a result, protecting networks and people from criminal activity requires proactive data analysis. Since the semantics of communications aid in identifying the sources of possible evidence, it is necessary to analyze both the header and the email content while investigating crimes using Electronic Mail (email) [4]. The inquiry process is slowed down because of the increasing difficulty in extracting the necessary semantics information from the massive amounts of emails that are being sent. This offers the perpetrator an advantage in covering up their misdeeds. Current keyword-based search and filtering methods sometimes miss important information in favor of irrelevant, short-sequence communications. To address the shortcoming, offer a new efficient method for multiclass email classification called SeFACED, which makes use of Long Short-Terms Memory (LSTM) based Gated Recurrent Neural Networks (GRU) [5]. SeFACED can process a dependency of 1000 characters or more. Therefore, it is not limited to working with short sequences. To achieve optimal performances, SeFACED evaluates LSTM-based GRU in comparison to more conventional machine learning techniques, deep learning models, and state-of-the-art research. Experimental findings on the self-extended benchmark dataset show that SeFACED successfully outperforms state-of-the-art approaches while maintaining the robustness and reliability of the classification processes [6].

LITERATURE SURVEY

The diffusion of knowledge is speeding up because of technological development. Businesses may now reach customers anywhere in the world because of the Internet's ability to link millions of devices and their users. Every day, people all around the globe use email to communicate with one another. Communication by email is efficient, easy, quick, and low-cost. There are two main categories of email, and they are Spam and ham. Spam makes up more than half of the user's inbox. Building a spam filtering system is essential for secure and productive email use [7]. The goal of this effort is to use a classifier to identify Spam and minimize its prevalence. Machine learning techniques allow for the most precise spam categorization. To identify Spam, the email's text was analyzed using a natural language processing technique. Machine learning methods were chosen as follows for evaluation purposes: Some common methods are Naive Bayes, K-Nearest Neighbor, SVMs, Logistic regressions, Decision trees, and Random forests. The training was done using preexisting data collection. Accuracy of up to 99% may be achieved with the help of NB and logistic regression. The findings may be used to improve existing spam detection algorithms or develop new filtering techniques [8]. There has been a significant evolution in assaults over the last several years, from the very simple and random to the highly complex and intentional. Spam, or unsolicited email, is one source of the sophisticated cybercrime strategies used to deceive individual users. In the recent decade, spam detection has emerged as one of the most prominent machine learning-oriented applications. This article provides a novel approach to spam email detection using deep learning architecture in the context of NLP. Various feature engineering techniques have been used in previous publications on spam email detection using conventional machine learning [9].

Finding a reliable engineering approach is challenging and risky in a hostile setting. The suggested approach utilizes the natural language processing text representation to focus on the identification of Spam in electronic mail. As a prerequisite for machine learning techniques, emails are converted into email word vectors using a variety of representation approaches. In addition, the hyper-parameter tuning method is used to determine the best settings for a wide variety of DL architectures and a variety of email representations. Using three publicly accessible email datasets, the effectiveness of many traditional machine learning classifiers and deep learning architectures with different text representations is assessed [10]. In terms of accuracy, precision, recall, and F1 scores, the experimental findings favor deep learning architectures over traditional machine learning classifiers.

This is mostly because deep learning architecture allows for the acquisition of sequential, abstract, and hierarchical feature representations of email messages. In addition, when compared to the other conventional email representation approaches, word embedding with deep learning has fared quite well. Learning the syntactic, semantic, and contextual similarities across emails is simplified by word embedding. This equips word embedding with deep-learning approaches for use in actual spam email screening [11].

Spam filtering, or the elimination of unsolicited, undesired, and virus-infested emails, is a major issue since spam emails consume Internet bandwidth, waste users' time, and may even result in financial loss. When it comes to spam filtering, the state-of-the-art method is the support vector machine (SVM). Because of the large dimensionality of emails, SVM incurs substantial temporal complexity. This research advocates for a manifold learning-based strategy for spam filtering [12]. Experiments showing that most characteristics are not decisive lead to the conclusion that only a small percentage of spam emails can be recognized using these criteria. In light of this realization, we suggest using the Laplace features map approach to extract the crucial geometrical characteristics from the email text dataset. The collected characteristics are then sent into a support vector machine (SVM) to filter out unwanted messages. Do comprehensive trials across three datasets, and the evaluation findings show that the proposed technique is both accurate and efficient in terms of time [13].

Constant difficulties with spam and phishing emails are being spread more and further with the aid of the botnet. Current methods of protecting against Spam and phishing through email rely mainly on some kind of spam filtering technology. However, spam filtering requires many defensive resources, most of which are scarce. Therefore, this work addresses the following issues to have the most possible impact with constrained defensive means: When securing a network, how many individual user nodes are most important? When and how do these variables play a role? To address these issues, we offer an evolutionary game model in which numerous attackers and multiple defenders engage in a strategic arms race [14]. First, an efficient defensive plan is established by determining the appropriate number of users to shield from attacks. This helps cut down on energy usage and the frequency of assaults. Then, learn that the attack cost, the attack loss, and the percentage of users that read their emails all play a role in this defensive approach. Managers of email networks may use findings to implement spam filtering based on the chance of an email being opened, an area in which the defense has asymmetric information advantages. In addition, investigates how factors like assault cost and loss affect defensive tactics. It turns out that the defensive approach is significantly influenced by the assault cost rather than the attack loss. To conclude, conduct experiments on a huge real-world data set consisting of emails exchanged over a period of two years. The defense plan has been shown to be successful in making the most of available defensive assets [15].

PROPOSED SYSTEM

Malware-based phishing and misleading phishing are the two main categories of phishing assaults. Malware-based phishing occurs when a piece of harmful software is sent to a user through spam email or downloaded and installed on the user's computer via exploiting a security flaw. If the user enters sensitive information while using the virus, it might be stolen and transferred to the phisher. This study focuses on a specific kind of phishing called "deceptive phishing," in which the phishers send out emails that seem to have originated from a trusted source like a bank. Typically, a phisher would send a victim an email with a link to a malicious website, where the user will be requested to enter sensitive information (such as a password). The phisher then uses this information for their own gain, such as making unauthorized withdrawals from the victim's bank account. Several solutions have been presented to combat phishing. Provide an overview of these safeguards in the next section. Differentiate between blacklisting and white-listing, as well as content-based and network-based methods. The goal of any efforts used to improve communication is to make it more secure. Virus prevention software and timely software upgrades are the initial lines of defense against malicious software (malware) phishing assaults. Email authentication is an extra safety step.

While there are a number of technical suggestions, at present, the great majority of email messages do not make use of any of them. Password hashing (where the password is hashed and sent with the domain name) and two-factor authentications (where multiple credentials, such as smartcards and passwords, are used) are two further methods. Some financial institutions (such as the Post-bank in the Netherlands and the Bank of Austria) have begun employing transaction authentication numbers (TANs) that are supplied to the customer on demand over the mobile phone network's Short Message Service (SMS). A man-in-the-middle attack cannot occur since the TAN encodes both the amounts being transferred and the recipient account numbers. The potential for phishing has just been drastically reduced. Keep in mind that authentication methods, such as the mobile TAN, need complex

infrastructure, take time, and cost money. This is not possible now for many forms of communication that are vulnerable to phishing. Search engines for products and social media interactions are two such examples. A phisher may post low-priced offers to a shopping search engine like Froogle, therefore attracting many potential victims. In most cases, the search engine has no say over the monetary dealings and will not need a cumbersome authorization procedure. Therefore, further anti-phishing measures are required. Email and online content screening is where their attention is focused. When a page is displayed in a browser, one method of preventing phishing involves validating the URL. An overview of the proposed system's design is shown in Figure 1.

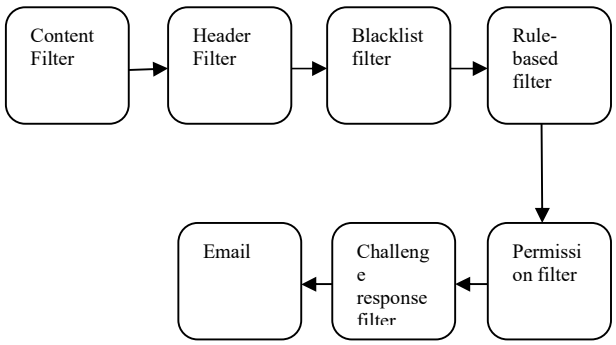

FIGURE 1. System architecture of the proposed system

Each website a user requests via the Mozilla Firefox browser, for example, is verified against blacklists of known phishing sites. This list is periodically updated and is downloaded to the local PC automatically. It is common knowledge, however, that new phishing sites pop up often; for instance, in December 2007, over 35 new phishing sites were discovered per hour. Phishing sites often only last three days, with many being down within hours. A new phishing site won't be added to the blacklists until it has been reported and investigated for some time. As a result, blacklisting can only do so much. Another issue is distributed phishing assaults, in which several separate servers operated on a botnet are redirected from links in a single phishing email. Gupta draws the conclusion that spam and phishing emails that are well-written often get through the blacklist filters. Since there is a limited set of organizations that are often the targets of phishing attempts, white-list strategies have been offered. In this case, incoming email links are matched against a whitelist of "good" URLs. While this strategy seems to have more potential, it may be difficult and time-consuming to compile a reliable source list. Another problem with whitelists is that they may filter out legitimate emails (false positives), whereas blacklists can only overlook Spam (false negatives). The filters use statistical methods to determine if a message is phishing or not by weighing the evidence from a wide variety of attributes. Many characteristics, to be detailed below, may serve as warning signs of phishing. In the first filters, each characteristic was given a "weight," and an email or website was marked as phishing if its score was high enough.

RESULTS AND DISCUSSIONS

The Offer a wide range of assessment metrics using a 10-fold cross-validation procedure. True positives (TP) mean that the email was accurately categorized as either phishing or non-phishing (ham), whereas false positives (FP) mean that the email was incorrectly classified as phishing when it was non-phishing, and false negatives (FN) mean that the email was incorrectly classed as ham when it was phishing. The percentage of properly categorized emails (accuracy) is not reported since it is not very useful until the distribution of classes is highly skewed. Not only do they give true positive and false negative rates, but also the more common accuracy, recall, and F-measure. Keep in mind that mistakes are not created equal when it comes to email categorization. It is far more expensive to get a false positive than a false negative. A classifier with a low rate of false positives is therefore preferable. Adjusting a classifier's decision threshold might decrease the false positive rates but increase the false negative rates.

Get first-rate outcomes for the specified real-world data. Classification outcomes for various feature sets, in particular, model-based feature sets like semantic topics features and dynamics Markov chains features. Utilizing a paired t-test determined that there is a statistically significant improvement when utilizing all features rather

than just the minimum required set of characteristics; this improvement has t-values of 8.70 with nine degrees of freedom or a probability of greater than 0.999. As can be seen, the model-based features on their own are already rather potent. More than 99% F-measure performance may be attained using only the DMC Text features. However, when additional characteristics are considered during feature selection, an additional relative error reduction of almost 85% is achieved. This fits along with what was seen before. Salting, picture distortion, and logo identification are all brand-new additions to this work, as are the DMC Link capabilities. Keep in mind that the emails in the corpus were gathered between April 2007 and November 2007. The findings suggest that maintaining the same classifier for a whole month leads to a little performance drop. To prevent a decline in performance, it may be possible to retrain the classifier using newly annotated "suspect" emails. Expanded corpus to include 40,000 emails by including 20,000 spam emails sent over the same time frame. The phishing class was combined with the spam email category to create the "unwanted" category. Used the features from the feature selection to run the program unaltered on this expanded corpus and saw a little drop in performance (F-measure: 98.48%). The email spam volume is shown in Figure 2, and the effectiveness of the infection vector is displayed in Figure 3.

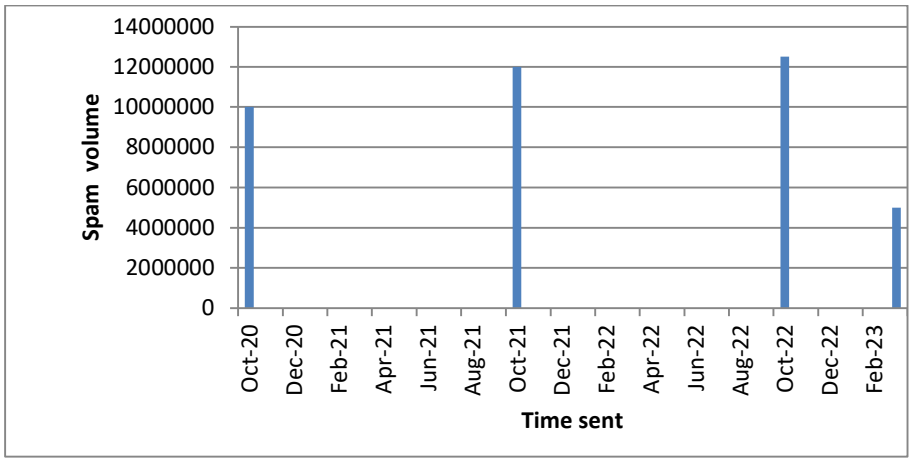

FIGURE 2. Spam volumes of the email with time sent

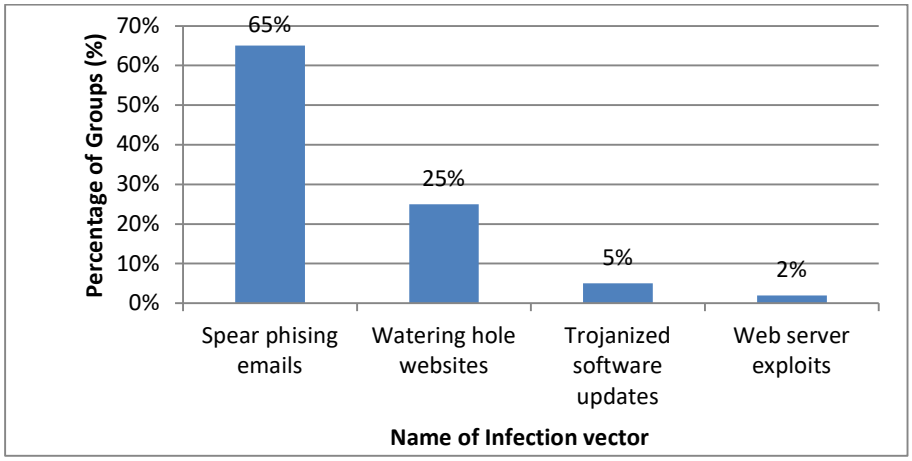

FIGURE 3. Performance analysis of the infection vector

It's important to keep in mind that spammers use a variety of salting techniques; thus, identifying Spam may need a different set of attributes than identifying phishing attempts. Simply a raw Spam Assassin implementation in the classifier is included. Even though the output was still good. The efficiency might be greatly improved with the installation of a specialized spam filter.

CONCLUSIONS

Email phishing is one of the most recent and pervasive network security issues. It's the practice of tricking people into giving up private information by sending fake but official-looking emails. A review of methods forwarding against phishing emails are included. The issue of phishing emails, the existing solutions, and the potential for future advancements in phishing email filtering are all better understood thanks to this study. Even with zero-day phishing email attacks, the approaches presented in the literature have a long way to go in terms of accuracy and performance. Although several algorithms have been implemented, there is currently no universal method that can prevent phishing assaults or zero-day phishing emails. Not only is that but the bulk of the effort is put in while no one is online. First, I need to finish the data-gathering, data-analyzing, and profile-making steps. A common characteristic of offline methods is that they are reactive. The whole process must be redone if the phishing email's characteristics are altered. The delay caused by this adaptation leaves users vulnerable to fraudulent emails. Considering these difficulties, novel technologies are still needed to identify and forecast zero-day phishing emails in real-time.

REFERENCES

[1]. Q. Yaseen, "Spam email detection using deep learning techniques," *Procedia Computer Science*, **184**, pp. 853-858, 2021.

[2]. H. Mohammadzadeh, and F.S. Gharehchopogh, 2021, "A novel hybrid whale optimization algorithm with flower pollination algorithm for feature selection: Case study Email spam detection," Computational Intelligence. **37(1)**, pp. 176-209.

[3]. N. Nisar, N. Rakesh, and M. Chhabra, 2021, "Review on Email Spam Filtering Techniques," International Journal of Performability Engineering. **17(2)**, pp. 1-6.

[4]. A. Junnarkar, S. Adhikari, J. Fagania, P. Chimurkar, and D. Karia, 2021, "E-mail spam classification via machine learning and natural language processing," *Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks*, pp. 693-699.

[5]. M. Hina, M. Ali, AR. Javed, F. Ghabban, LA. Khan, and Z. Jalil, 2021, "Sefaced: Semantic-based forensic analysis and classification of e-mail data using deep learning," IEEE Access. **9**, pp. 98398-98411.

[6]. Y. Kontsewaya, E. Antonov, and A. Artamonov, 2021, "Evaluating the effectiveness of machine learning methods for spam detection," *Procedia Computer Science*. **190**, pp. 479-486.

[7]. S. Srinivasan, V. Ravi, M. Alazab, S. Ketha, AM. Al-Zoubi, and S. Kotti Padannayil, 2021, "Spam emails detection based on distributed word embedding with deep learning," *Machine intelligence and big data analytics for cybersecurity applications*, pp. 161-189.

[8]. C. Wang, Q. Li, TY. Ren, X.H. Wang, and G.X. Guo, 2021, "High efficiency spam filtering: a manifold learning-based approach," *Mathematical problems in engineering*, pp. 1-7.

[9]. M. Wang, and L. Song, 2021, "Efficient defense strategy against spam and phishing email: An evolutionary game model," *Journal of Information Security and Applications*, **61**, pp. 1-5.

[10]. A.A. Ojugo, and D.A. Oyemade, 2021, "Boyer moore string-match framework for a hybrid short message service spam filtering technique," *IAES International Journal of Artificial Intelligence*, **10(3)**, pp. 1-8.

[11]. A. Velu, and P. Whig, 2021, "Protect personal privacy and wasting time using Nlp: a comparative approach using Ai," *Vivekananda Journal of Research*, **10**, pp. 42-52.

[12]. T. Mehrotra, G.K. Rajput, M. Verma, B. Lakhani, and N. Singh, 2021, "Email spam filtering technique from various perspectives using machine learning algorithms," *In Data Driven Approach Towards Disruptive Technologies: Proceedings of MIDAS*, pp. 423-432, 2021.

[13]. J. Batra, R. Jain, V. A. Tikkiwal, and A. Chakraborty, 2021, "A comprehensive study of spam detection in e-mails using bio-inspired optimization techniques," *International Journal of Information Management Data Insights*, **1(1)**, pp. 1-7.

[14]. P. Garg, and N. Girdhar, 2021, "A systematic review on spam filtering techniques based on natural language processing framework," *11th International Conference on Cloud Computing, Data Science and Engineering*, pp. 30-35,

[15]. M. Jazzar, R.F. Yousef, and D. Eleyan, 2021, "Evaluation of machine learning techniques for email spam classification," *International Journal of Education and Management Engineering*, **11(4)**, pp. 35-42.