Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 11, No. 2, 2025

Email Security Detection Solution Based on Multi-Scenario Classification

Xin Liu

College of Computer Science and Technology, Qingdao University, Qingdao, Shandong, China

Abstract: In the backdrop of the rapidly advancing global network informatization, cybercriminals frequently utilize a variety of phishing emails to launch attacks against numerous crucial organizations and information systems. Once these attacks succeed, they can inflict substantial damage on physical and digital systems as well as assets, resulting in the leakage of sensitive information, reputation impairment, and economic losses. The issue of how to detect phishing emails from a vast number of emails has long drawn significant attention. However, with the increasingly complex deception techniques of phishing emails, the existing solutions for detecting phishing attacks are no longer sufficient to tackle these problems. In this paper, a multi - scenario classification - based email security detection scheme is proposed. By analyzing the different stylistic features and deception differences of emails in various application scenarios, the collected email datasets are classified into multiple scenarios. Subsequently, the Long Short - Term Memory (LSTM) is employed to train the data under different classifications, and the trained model is used to classify phishing emails. The results demonstrate that the detection scheme proposed in this study exhibits relatively high accuracy.

Keywords: Phishing Emails; Email Security Detection; Multi-scenario Classification.

1. Introduction

In recent years, cyber security incidents have occurred frequently. In most of these incidents, attackers have used phishing emails as a means to gain initial access, successfully infiltrating government systems (such as the U.S. Department of State and the White House [1]), well-known companies (such as Google and RSA [2]), as well as the websites of politicians and social organizations in many countries (such as John Podesta and the Democratic National Committee [DNC] [3]). This series of highly publicized incidents has highlighted the growing prevalence and potency of phishing attacks.

On the one hand, phishing emails often cause economic damage to enterprises. On the other hand, phishing emails lead to the leakage of private information, causing harm to industries and even countries. Unlike attacks that exploit specific technical vulnerabilities in software and protocols, phishing attacks are based on social engineering [4-8]. By sending fraudulent emails, attackers trick recipients into taking some dangerous actions unknowingly (such as clicking on links, entering passwords, etc.).

In the field of email security detection, traditional detection devices are mostly ineffective in the face of phishing email attacks.

Phishing emails are a prominent and targeted attack medium in today's Internet. More than 100,000 Internet users around the world fall victim to phishing attacks every day. The Gartner Group estimates that the losses caused by phishing attacks amount to billions of dollars each year [9]. Therefore, phishing detection has received even more attention in recent years [10]. The problem is further exacerbated when phishers manage to influence the victims' responses while maintaining an appearance of authenticity and legitimacy, which makes it extremely difficult for filters to classify phishing content as fraudulent [11,12], and users remain vulnerable to phishing messages. Thus, flagging emails as phishing content is a crucial task in cyber security.

2. Related Work

At present, there are numerous network attack methods. Among them, phishing is a major threat to the Internet economy, causing losses of billions of dollars annually. Nowadays, phishing attacks are most commonly carried out through emails, which has led to extensive research on phishing email attack detection both at home and abroad.

Phishing is a form of social engineering, aiming to obtain information from unsuspecting victims. Attackers usually disguise themselves as legitimate institutions to deceive users into disclosing sensitive information, which can be later used for fraudulent activities. Phishing emails are different from ordinary spam. Attackers can utilize the sender information obtained in advance to impersonate trustworthy email senders, forge more believable emails, and deceive victims into downloading malicious attachments or clicking on malicious links.

Moore and Clayton [13] analyzed the empirical data regarding the removal time of phishing websites and the number of visitors attracted by these websites. It concluded that for attackers to launch a phishing attack, they must create a fake website and send emails to attract visitors. Although removing phishing websites is part of the counter - phishing efforts, the speed at which this is done is insufficient to completely alleviate the problem.

Abbasi and Mariam [14] have discovered that the existing anti-phishing tools lack the accuracy and universality required to protect Internet users and organizations from the diverse attacks encountered daily. As a result, users often overlook the warnings issued by these tools. In this study, we adopted a design-science approach and proposed a novel method for detecting phishing websites. By adopting the perspective of genre theory, the proposed genre - tree - kernel method makes use of the fraud clues associated with the differences in purpose between legitimate and phishing websites. These clues are manifested through the genre composition and design structure, thereby enhancing the anti

-phishing capabilities.

Park and Taylor [15] conducted a comparative study on the use of subjects and objects of verbs in phishing emails and legitimate emails. The purpose of this study was to explore whether syntactic structures, as well as the subjects and objects of verbs, could serve as features for differentiating phishing emails from legitimate ones. A comparison of the syntactic similarity of sentences and the subjects and objects of verbs was carried out. The experimental results show that these two features can be applied to certain verbs, but further research is needed for other verbs.

Valecha, Mandaokar and Rao [16] analyzed the role of gain and loss related persuasion cues in phishing emails. Phishing attackers often employ persuasion techniques to obtain positive responses from recipients. Three machine - learning models were created, using gain - related persuasion cues, loss - related persuasion cues, and a combination of gain and loss persuasion cues respectively. The estimation results were compared with a baseline model that does not consider persuasion cues. The results show that the three phishing - detection models with relevant persuasion cues significantly outperform the baseline model in terms of F-score, indicating that a deep understanding of persuasion cues can provide information for the design of effective countermeasures to detect and block phishing emails.

Chandrasekaran, Narayanan and Upadhyaya [17], developed a technique for detecting phishing emails. They introduced 25 phishing email features, including 23 stylistic marker features and two structural features. These features were used to train an SVM classifier, which achieved a classification accuracy of 88%.

Gascon, Ullrich and Stritter [18] demonstrated that senders leave content - agnostic features in the structure of emails. Based on these features, a method was developed that can learn the profiles of a large number of senders and identify forged emails as deviations from them. More than 700,000 emails from 16,000 senders were evaluated, and it was proven that this method can distinguish among thousands of senders, with a detection rate of 90% and less than 1 false positive in 10,000 emails for forged emails. Additionally, it was also shown that individual features are difficult to guess, and deception can only succeed when the attacker has access to and can simulate the entire email of the sender.

3. The Details of Phishing Email Detection

3.1. Data Collection

3.2. Overview Of Detection Framework

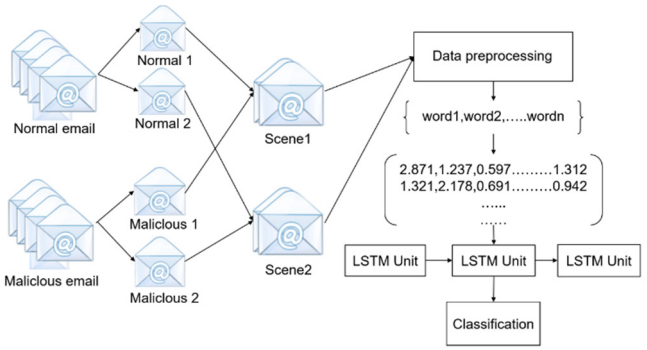

Figure 1. Phishing Email Detection Framework

In this study, the data was collected from publicly available datasets on the Internet, such as the publicly available Enron email dataset and other phishing email datasets. The total number of emails is approximately 10,000.

In this section, we introduce the main research ideas and methods within the proposed phishing email detection framework. In existing research works, a large number of detection methods based on phishing blacklists are too rigid and can only detect phishing websites that have already been recorded. However, a large number of newly - built phishing websites cannot be identified at any given moment. Some other studies use conventional machine - learning methods such as support vector machines and decision trees to extract features from the email body and classify emails. In these methods, the detection party needs a great deal of relevant professional knowledge to select features and analyze whether these features are useful for the detection work, and the corresponding detection work is also relatively complex.

Therefore, we choose the deep - learning method to detect phishing emails. Compared with the existing research schemes that simply and directly conduct deep - learning training on a large number of emails, through research and comparison, we find that there are significant differences in the writing characteristics, sentence structures, and word usages of emails in different application scenarios. For example, the contents discussed in emails in daily - life scenarios and work scenarios are quite different. Accordingly, the deception methods of phishing emails also vary. Thus, simply mixing emails from these different application scenarios for detection may affect the detection results. As shown in Figure 1, the main idea of our research is to classify a large number of emails according to different application scenarios, mainly into daily - life, work, advertising, and other categories, and then perform corresponding data processing and model training.

3.3. Classification of Different Scenarios for Emails

In this study, we classify a large number of mixed emails into daily - life, work, advertising, and other categories based on different email subject contents and writing characteristics. There are significant differences in the characteristics among emails in different scenarios.

Emails in daily - life scenarios often contain information such as congratulating friends on graduation, marriage, or the birth of a child; expressing gratitude for others' help, gifts, or support; inviting friends or family to parties and celebrations; and sending condolences and words of encouragement when the recipient encounters difficulties or misfortunes.

Emails in work scenarios often have clear subjects, such as "Project Progress Update", "Meeting Arrangement Notice", etc.; formal salutations, like "Dear General Manager Zhang", "Dear Manager Li", etc.; detailed and specific requests, information, suggestions, or feedback, often accompanied by documents; as well as names, positions, company names, and contact information.

Emails in advertising scenarios often have attractive subjects, using words like "Limited- time Offer", "Exclusive Discount", and verbs such as "Get", "Snap up"; personalized salutations to make the emails seem closer to the audience; clear content themes, such as exclusive news, special offers, or professional advice; and explicit calls - to - action, such as "Click", "Buy", "Register", etc., to guide the recipients to take the next step, often accompanied by electronic links.

Emails that do not fall into the daily-life, work, and advertising categories are classified as other - type emails.

3.4. Model Training

Recurrent Neural Networks (RNNs), due to their special network models, take into account the previous time - series information when learning the current time - series information. Therefore, RNNs have unique advantages in handling time - series and text - sequence problems. However, for RNNs, it is difficult to learn long - distance information. The Long Short - Term Memory (LSTM) model, a special form of the RNN model, overcomes the problem that classical RNN models have difficulty in learning long - distance information.

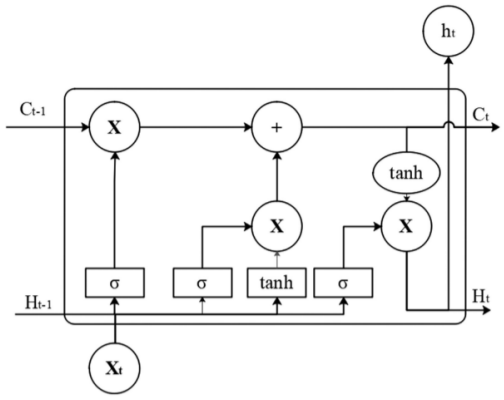

Figure 2. LSTM Neurons

In the LSTM model, the state of a neuron is similar to a data line. Data can be transmitted along the entire line with only a small number of linear interactions. It can easily maintain the flow of information. An LSTM neuron is mainly composed of three gate structures, which can select the information to be transmitted. The gate structures are mainly implemented by the sigmoid neural layer and point - wise multiplication operations.

Forget Gate: The left - hand side of the figure shows the forget gate of an LSTM neuron. The inputs of the forget gate are $H_{t-1}$and X_t. By processing the inputs, the forget gate can output a number between 0 and 1, representing the degree of forgetting. If the output is 1, it means that all the information has been memorized; if it is 0, it means that all the information has been forgotten.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

In the formula (1), $W_f$is the weight of the forget gate, $[h_{t-1}, x_t]$combines the two vectors together, b_f is the bias of the forget gate, and σ represents the sigmoid function.

Input Gate: In the figure, the middle part represents the input gate. The input gate layer determines which values need to be updated. The tanh layer generates a new vector, and the input gate layer and the tanh layer work together to update the state.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

In the formula (2), W_i is the weight of the input gate, and b_i is the bias of the input gate.

Output Gate: In the figure, the right - hand part is the output gate, which determines the output. First, a sigmoid layer is run to determine which part of the state we need to output. Then, the state is input into the tanh function to limit the value of the output function between - 1 and 1. Finally, we can obtain the output by multiplying the output obtained in the previous step by the sigmoid gate.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3)$$

In the formula (3), h_{t-1}represents the output value of the last neuron. The LSTM model fully preserves and controls information through these three gate structures. In the research, for faster computational speed, the softsign function is used as the activation function, and Adam is utilized as the optimizer to adjust the learning rate, enabling the model to converge rapidly. Orthogonal initialization is also employed to address the problems of gradient vanishing and gradient explosion caused by overly long message bodies in deep networks.

4. Research Evaluation

4.1. Evaluation Criteria

In this study, the experimental results will be evaluated through four different parameters: Acc, P, R, and F1-score. The definitions of the four parameters are as follows:

$$\text{Acc} = \left(1 - \frac{errorsum}{sum}\right) \times 100\% \quad (4)$$

$$\text{P} = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

$$\text{R} = \frac{TP}{TP+FN} \times 100\% \quad (6)$$

$$\text{F1} - \text{score} = \frac{2 \times P \times R}{P+R} \times 100\% \quad (7)$$

In the formula (4-7), errorsumrepresents the number of misclassified samples, and sum represents the total number of samples. TP stands for True Positive, which represents the number of phishing emails. FN and FP represent the numbers of False Negative and False Positive respectively. R represents the recall rate, indicating the proportion of phishing emails correctly classified by the model. The $\text{F1} - \text{score}$ is based on the harmonic mean of the accuracy rate and the recall rate, and comprehensively evaluates the performance of the model.

4.2. Evaluation Results

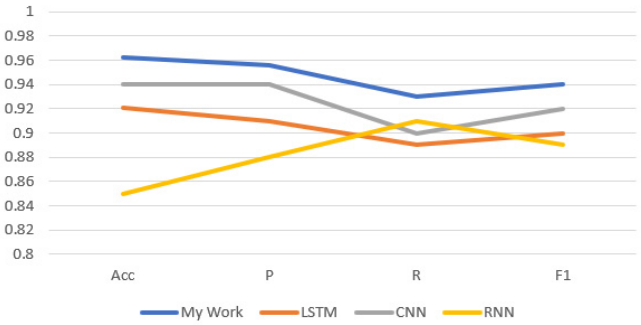

Figure 3. Comparison Of Experimental Results

First, we compare the results of our research with those of other different neural network models. These selected neurons are mainly used for processing sequential data, including the standard RNN neural network model, the CNN neural network model, and the conventional LSTM neural network model. The results are shown in the figure 3. Our research has achieved relatively good results. Due to its simple model, the RNN neural network has the worst performance among the four neural networks. The results of the other several neural networks are superior to those of the

RNN neural network. The features extracted by the CNN model are similar to n - grams that ignore word order and cannot achieve satisfactory results in sentiment analysis tasks. Relatively speaking, LSTM can better capture the induced statements in the messages and thus can be better used for phishing detection.

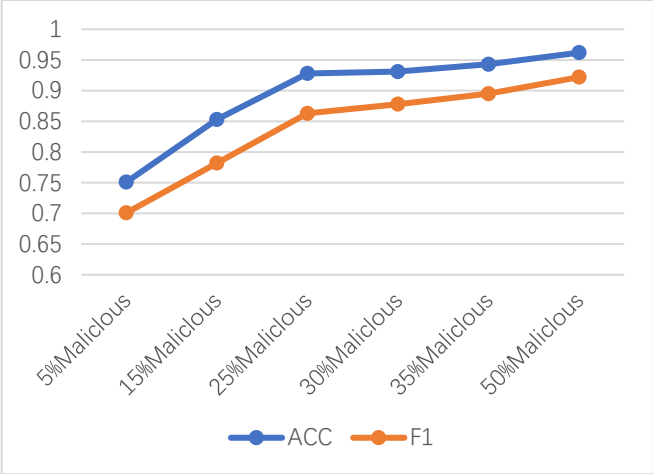

Figure 4. Comparison Of Samples with Different Proportions

Secondly, in our research, the model is also trained by adjusting the different proportions of malicious phishing emails in the dataset. The results are shown in the figure 4. When the proportion of positive and negative samples in the dataset is close, the performance of our research model is more ideal. During the training process, when the proportion of a certain part of the samples is too large, the results will tend towards the side with the excessive proportion, which may lead to a large number of false - positive detection results.

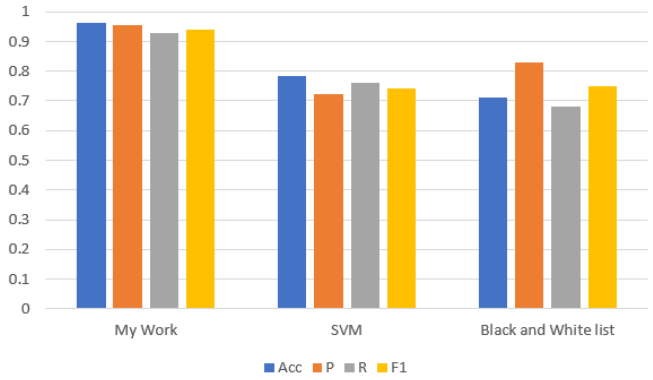

Figure 5. Compared With Other Schemes

Finally, since the focus of a large number of existing research works lies in the Support Vector Machine (SVM) and the black - and - white malicious list mechanism, we have also made a simple comparison between our research and these works. The results are shown in the figure. Our research has achieved relatively good experimental results. The black - and - white list mechanism has the worst detection effect. It can only detect phishing information that has already been recorded, while a large number of newly - built phishing websites and mailboxes cannot be identified at any time. Some other studies use conventional machine - learning methods such as support vector machines to extract features from the email body and classify emails. In these methods, the detection party needs a great deal of relevant professional knowledge to select features and analyze whether these features are useful for the detection work. It is also difficult to regulate the importance of different features in the detection model.

5. Conclusion

In this paper, we analyzed the existing phishing email detection methods and found that traditional detection methods, such as the black and white list mechanism and traditional machine-learning based detection schemes mainly relying on SVM, have difficulty accurately detecting phishing emails. Therefore, we proposed and designed an email security detection scheme based on multi - scenario classification. By analyzing the differences in the subject contents and writing characteristics among different emails in various application scenarios, we classified the collected large volume mixed email data and then trained it using the LSTM model. Finally, through comparison with other detection models and traditional detection mechanisms, the detection scheme in this paper has a relatively high accuracy.

Acknowledgments

I am deeply grateful to my supervisor; your expertise and patience have been invaluable throughout my research. Your guidance has been the compass that steered me through the complexities of this study. I also want to thank my family.

References

[1] CHEN P, DESMET L, HUYGENS C. A study on advanced persistent threats[A]. Communications and Multimedia Security-15th International Conference[C]. 2014. 63-72.

[2] NIKOS V, DIMITRI G. The big four—what we did wrong in advanced persistent threat detection[A]. International Conference on Availability, Reliability and Security[C]. 2013. 248-254.

[3] YANG G M Z, TIAN Z H, DUAN W L. The prevent of advanced persistent threat[J]. Journal of Chemical and Pharmaceutical Research, 2015, 6(1):572-576.

[4] FRIEDBERG I, SKOPIK F, SETTANNI G, et al. Combating advanced persistent threats: from network event correlation to incident detection[J]. Computers & Security, 2015, 48(2):35-57.

[5] BUTT M I A. BIOS integrity: an advanced persistent threat[A]. Conference Proceedings - 2014 Conference on Information Assurance and Cyber Security[C]. 2014. 47-50.

[6] CHRISTOS X, CHRISTOFOROS N. Advanced persistent threat in 3G networks: attacking the home network from roaming networks[J]. Computers & Security, 2015, 40(2): 84-94.

[7] ZHAO W T, ZHANG P F, ZHANG F. Extended Petri net-based advanced persistent threat analysis model[J]. Lecture Notes in Electrical Engineering LNEE, 2014, 277: 1297-1305.

[8] GIURA P, WANG W. Using large scale distributed computing to unveil advanced persistent threats[J]. Science, 2013, 1(3): 93-105.

[9] A. Litan, “Phishing victims likely will suffer identity theft fraud,” Gartner Res., ID Number: FT-22-8873, 2004.

[10] A. Vishwanath, T. Herath, R. Chen, J. Wang, and H. R. Rao, “Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model,” Decis. Support Syst., vol. 51, no. 3, pp. 576–586, Jun. 2011, doi: 10.1016/j.dss.2011.03.002.

[11] A. Abbasi, F. Mariam Zahedi, D. Zeng, Y. Chen, H. Chen, and J. F. Nunamaker Jr, “Enhancing predictive analytics for anti-phishing by exploiting website genre information,” J. Manage. Inf. Syst., vol. 31, no. 4, pp. 109–157, 2015.

[12] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer, “Social phishing,” Commun. ACM, vol. 50, no. 10, pp. 94–100, 2007.

[13] T. Moore and R. Clayton, “Examining the impact of website take-down on phishing,” in Proc. Anti-Phishing Work. Groups 2nd Annu. eCrime Res. Summit, 2007, pp. 1–13.

[14] A. Abbasi, F. Mariam Zahedi, D. Zeng, Y. Chen, H. Chen, and J. F. Nunamaker Jr, “Enhancing predictive analytics for anti-phishing by exploiting website genre information,” J. Manage. Inf. Syst., vol. 31, no. 4, pp. 109–157, 2015.

[15] G. Park and J. M. Taylor, “Using syntactic features for phishing detection,” 2015, arXiv150600037.

[16] R. Valecha, P. Mandaokar and H. R. Rao, "Phishing Email Detection Using Persuasion Cues," in IEEE Transactions on Dependable and Secure Computing, vol. 19, no. 2, pp. 747-756, 1 March-April 2022, doi: 10.1109/TDSC.2021.3118931.

[17] M. Chandrasekaran, K. Narayanan and S. Upadhyaya, Phishing email detection based on structural properties, in Proc. of the NYS Cyber Security Conference (2006), pp. 1–7.

[18] Gascon H , Ullrich S , Stritter B ,et al.Reading Between the Lines: Content-Agnostic Detection of Spear-Phishing Emails [C]// 2018.DOI:10.1007/978-3-030-00470-5_4.