Anti-phishing Email Detection Framework for New-Age Phishing Attacks

Muningandu Tjingaete
School of Computing
Asia Pacific University of Technology and Innovation (APU)
Kuala Lumpur, Malaysia
tp066859@apu.edu.my

Julia Juremi
School of Computing
Asia Pacific University of Technology and Innovation (APU)
Kuala Lumpur, Malaysia
julia.juremi@apu.edu.my

***Abstract*— Email as a form of communication has always been the major target as the co-operate world including the rest of the internet users tend to use emails as a form of communication. The rise of covid-19 has prompt the use of emails much more as employees are working from home meaning that the day-to-day work is now confirmed and communicated via email. This has prompt the rise of phishing attacks as cyber criminals are aware that users are frequently using emails as a mode of communication now more than ever before. The protection of crucial data has always been the main aim of cyber security but, however cybercriminals continue finding new ways and techniques to steal data, as data is always the main target, and the confidentiality of data is always bridged meaning users tend to reveal different types of crucial information without intending to do so. Phishing email remains to be one of the most forms of attack which prevails regardless of the different types of techniques used. This paper focuses on the different anti-phishing methods which have been introduced, an exploration of the traditional existing solutions will be reviewed including the limitations it has when it comes to comparting new age-phishing attacks. Existing machine language techniques will further be analyzed and reviewed compared in terms of their effectiveness. A review on the use of deep learning techniques that have been used including blacklisting methods as well. Several awareness techniques to prevent users from falling victims to phishing attacks will further be explored. Observations and analyses made from previous research work will however show that there is still a need for anti-phishing improvement.**

Keywords—Anti-phishing methods, Machine learning, blacklisting, Deep learning, Phishing email

I. INTRODUCTION

Phishing is one of the most common types of security attack which occurs online with the aim stealing user's personal data. This is achieved by using various phishing methods and techniques to unlawfully attain confidential information be it username, passwords, credit card numbers etc. This is an important issue as its one of the most troublesome types of attack which is used by cybercriminals to bypass security. This remains to be one of the most common techniques used amongst users, which is constantly rising at the same pace as technology. Due to the fact that most of the daily transactions are carried out online hackers have a way of sending fake emails asking clients to reset their banking password or send fake discounts which requires users to enter their username and passwords thus, giving the hackers full access control of their accounts (Patil & Dhage, 2019). Various type of anti-phishing techniques can be introduced to mitigate such attacks but however human knowledge still plays an important role as the security system must be accurate all the time while the hackers just must be correct one time to be able to steal confidential data. The user knowledge plays an important role in securing critical infrastructure as the protection of against phishing attacks is not only depended upon numerous technologies but also on the basic human knowledge (Nmachi & Win, 2021). Thus, it's important to use various types of campaigns, programs, or training methods to educate users on how to distinguish between legitimate and fake emails.

The main purpose of this paper is to gather various types of articles and research paper to demonstrate sufficient knowledge on phishing with the aim of reviewing all the different types of techniques which were used to mitigate phishing attacks. This will be done by evaluating several types of anti-phishing methods both technological and human techniques that were previously created to mitigate phishing attacks. However, a review will be done in comparison with the aim of identifying and justifying any form of research gap. Lastly, this paper will analyze whether there are any phishing tools which can detect emails which has made it inside the email box already and if there is any method or feature's which are preinstalled to help users identify phishing attacks.

II. ANTI-PHISING TECHNIQUES

There are various types of anti-phishing techniques which are used both technical and non-technical with the aim of mitigating phishing attacks. There are certain features which are used to detected whether it's a phishing email or not. The body character is often analyzed by certain models with the aim of discovering if the email is legitimate or not. URL based characters is also used by certain modules and users are further educated by about the significant of URL in detecting phishing sites. Various types of email features such as subject based, the variety of sender-based characters and script features are all the different types of methods which are used with the aim of identifying whether an email is malicious or not. However, there is diversity in the different techniques used to detect phishing emails (Yi & Kamsin, 2022). The use of machine language and artificial intelligence is used quite more often to detect phishing sites as there are different methods and algorithms which can be used with ML. An exploration of the

different type of technique used including their effectiveness and limitations will be discussed and explored.

A. Anti-spam

Software tool can identify phishing emails from legitimate emails thus, block phishing emails from entering the email Inbox. All the phishing emails are sent to spam but, however, this phishing techniques block genuine email and send them to junk meaning this technique deals more with false positive. Thus, with the use computerized method of compacting phishing email the aim here is to increase the rate of true positive without giving the users additional work of having to constantly check the spam folder and retrieve some legitimate email which was sent there during false positive. In research from Gangavarapu et al. (2021) there are different types of methods which are used to countermeasure the anti-spam but, however the unsolicited bulk emails (UBEs) need machine learning models for content and behavior-based features. The help of machine language to prevent UBEs comes with a high number of true positive but, however this method lacks security measures to handle UBE filter attacks.

B. Databases

There are predefined malicious domains/URL which are recognized as harmful, and they are added into a blacklisting database. This means that every time a user clicks on a certain URL which is attached within the email, its checked within the blacklisting database and will either allow the user to open the web browser if the link is not recognized within the database but, however, if the link is matched within the database a warning is given to the user when the link is clicked. Blacklisting falls under the detection methods for anti-phishing methods that have been introduced to solve the problem of email phishing attacks. In research from Sahingoz et al. (2019) the use of backlisting alone is not sufficient as its not able to detect any first-time attack and works on predefined set of URL's and IP addresses. Other limitations is due to the fact that the URL are updated periodically at a specific time and days which leaves a loophole and efficient time for attackers to create newly URL and attack the system as they will not be recognized when matched against the database as its not updated (Issa , Thabtah, & Chiclana, , 2018). Thus, the using of blacklisting alone is not efficient enough as it can't constantly recognize new phishing URL which are not added on the database. Blacklisting is known to be more time consuming as it requires users' constant identification and reporting of this malicious URL (Fang, Zhang, Huang, Lui, & Yang, 2019).

A suggestion to use whitelisting was further introduced which wasn't sufficient as users browse multiple websites. This came with certain limitations such as users are always assuming that they are dealing with trusted websites which is not always the case (Issa , Thabtah, & Chiclana, , 2018).

As the use of blacklisting is not efficient enough, the use of visual similarity is further used with the aim of using different type of features such as images, website logo and source code to compare legitimate websites for clones. However, this technique is also not efficient enough as it cannot detect new phishing websites and this further produces high false positive (Jain & Gupta, 2018).

C. Machine Language

The use of artificial intelligence (AL) through Machine learning (ML) is further adopted to reduce phishing attacks this method has proven to work to a certain extend only but however, this comes with certain limitations as this method requires the manual work of feature engineers. The engineers need to manually find all the illustrative features which are not conducive to the relocation of application developments (Fang, Zhang, Huang, Lui, & Yang, 2019). On the other hand, deep learning is a subset of ML which comes with several benefits of catching new phishing URLs unlike ML which needs manual work to constantly update features DL can catch and doesn't require constant manual work. However, the use of DL comes with certain limitations as it requires large sets of data for accurate performance. (Nmachi & Win, 2021). Further the use of deep learning together with natural language processing detection to mitigate phishing attacks. This came with a high right of true positive but however this method has its own limitation problems as technique is limited to word embedding and fails to focus more on the specificness of phishing email detection. The THEMIS model was further introduced which worked on an improved RCNN but comes with certain limitations as it cannot detect emails with no email header (Fang, Zhang, Huang, Lui, & Yang, 2019)

A combination of ML and Natural language processing (NLP) has further played an important role in mitigating phishing attacks, this didn't include the previous features such as semantics, syntax, and context. However, ML was used with Support Vector Machine (SVM), Random Forest(RF), Decision Trees(DT), Logistic Regression(LR) but this came with certain limitations. The major drawback of this method was the lack of deep semantics as it dependent more on surface text, this meant that when a structure, synonyms or the use of different words are used it was hard for this technique to notice. The major problem of NLP built on ML is the fact that it cannot detect new phishing emails (Salloum, Gaber, Vadera, & Shaalan, 2021). According to new research using NLP on ML has proven to be a great detection algorithm to a certain extended as its user's semantics to catch and verify whether an email is phishing or not (Peng et al., 2018). However, this method comes with certain limitations and not fully accurate as it only relies on email text analysis, it needs the constant use of blacklist to be generated by using machine language with a Naïve bayes classifier.

The use ML on client-side detection is used which comes with several benefits such as real time phishing detection, this method also has a high detection accuracy meaning it produces truer positive but however, this requires the downloading of the whole page to be able to detect the phishing website and it also involves a limited amount of dataset which is only applicable for HTML source code only (Jain & Gupta, 2018).

In research from Bhardwaj et al. (2021) a new privacy detection framework is introduced which aims at mitigating new age phishing techniques this follows zero trust policy with the assumption that the user has opened or clicked on the link which was sent by the hacker. In this instant a privacy security framework mainly on end users including a local DNS which comes with certain features such as blocking trackers and ads. However, this method has certain limitations as its developed using Linux OS and Python meaning its only applicable to certain end users only. An implementation of machine learning with the use of seven different types of ML algorithms such as Decision Trees, Random Forest, SMO etc. is used to increase the rate of phishing detection. This method came with several benefits such as independency of third

parties, new phishing websites could however, be detected using this method and all of this can be done in real-time execution. There are certain drawbacks which comes with this technique as new subsystem still needs to be conducted for shorter URL's (Sahingoz, Buber, Demir, & Diri, 2019). The use of ML modeling cycle with a combination of decision trees and Naïve bayes has been used to try and mitigate phishing emails, the main aim of this model is to deal with emails that has already entered the email box but however, this model comes with certain limitations such as the non-accuracy (Espinoza, et al., 2019).

D. Awareness

In research from Jain and Gupta (2018) user education remains to be the key to reduce the likelihood of users falling victims to phishing attacks several games have been proposed and are used with the aim teaching users to identify between legitime webpages compared to fake websites. However, other methods are further used in which users can be educated about phishing attacks and how to protect themselves from these techniques. Through the use of a game users are taught different types of identification techniques such as how to identify a phishing URL, the users are further being educated about dangers of short URL including how and when they can search the internet for the legitimate URLs instead of falling victims to fake URL (CJ, et al., 2018).

E. Stylometric Analysis

This method is introduced with the aim of capturing writing behavior of a legitimate email in comparison with the phishing email, it focuses more on users writing habits, the vocabulary used by the users including the complexity of the email text, the Source Code Author Profiles (SCAP) method came with certain limitations as it produced a high false positive. This method was improved with a novel automated approach which can link the email with the sender's profile. This meant that if the sender's profile cannot be matched with a specific profile it's recognized as a phishing email but, however, a small email size will affect the accuracy of the data (Nmachi & Win, 2021).

Conclusion

In conclusion, Phishing attack remains to be one of the most growing attacks thus, there is a constant need for improvements to mitigate users falling victim to phishing attacks via email. The proposed framework and modules which are analyzed and viewed showed a lack in helping users detect email which was passed through the mailbox. The focus is more on building algorithms with additional features but this, however, shows that there is still a lack of accuracy and users still end up falling for this phishing attacks. The use of blacklisting and whitelisting was further reviewed and an observation on how they can't detect new phishing attacks and how this database needs more of human effort which is time consuming. Spam filters are further used in different type of mail server but however, they have a high false positive rate. The use of Machine language to detect and prevent phishing attacks is highly used with different type of algorithms but however this doesn't provide a full accuracy on detecting phishing attacks. Different type of awareness training is given to users which prove to help users to a certain extend only but however they still do fall victim to phishing email. Thus, there is still a gap in the solutions provided and more work still needs to be done to help mitigate phishing attacks and to further help users identify this phishing emails.

REFERENCES

Bhardwaj, A., Al-Turjman, F., Sapra, V., Kumar, M., & Stephan, T. (2021). Privacy-aware detection framework to mitigate new-age phishing attacks. *Computers & Electrical Engineering*, 1-13.

CJ, G., Pandit, S., Vaddepalli, S., Tupsamudre, H., Banahatti, V., & Lodha, S. (2018). PHISHY - A Serious Game to Train Enterprise Users on Phishing Awareness. *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, 1-13.

Espinoza, B., Simba, J., Fuertes, W., Benavides, E., Andrade, R., & Toulkeridis, T. (2019). Phishing Attack Detection: A Solution Based on the Typical Machine Learning Modeling Cycle. *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, 1-6.

FANG, Y., ZHANG, C., HUANG, C., LIU, L., & YANG, Y. (2019). Phishing Email Detection Using Improved RCNN Model With Multilevel Vectors and Attention Mechanism. *IEEE Access*, 1-12.

Gangavarapu, T., Jaidhar, C., & Chanduka, B. (2020). Applicability of machine learning in spam and phishing email filtering: review and approaches. *Artificial Intelligence Review*, 5019-5081.

Issa, Q., Thabtah, F., & Chiclana, , F. (2018). A recent review of conventional vs. automated cybersecurity anti-phishing techniques. *Computer Science Review*, 44-55.

Jain, A., & Gupta, B. (2018). Towards detection of phishing websites on client-side using machine learning based approach. *Telecommunication Systems*, 687-700.

Nmachi, W. P., & Win, T. (2021). Phishing Mitigation Techniques: A Literature Survey. *SSRN Electronic Journal*, 1-10.

Patil, S., & Dhage, S. (2019). A Methodical Overview on Phishing Detection along with an Organized Way to Construct an Anti-Phishing Framework. *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, 1-6.

Peng, T., Harris, I., & Sawa, Y. (2018). Detecting Phishing Attacks Using Natural Language Processing and Machine Learning. *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*, 1-2.

Sahingoz, O., Buber, E., Demir, O., & Diri, B. (2019). Machine learning based phishing detection from URLs. *Expert Systems with Applications*, 1-13.

Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2021). Phishing Email Detection Using Natural Language Processing Techniques: A Literature Survey. *Procedia Computer Science*, 19-28.

Yi, C. X., & Kamsin, I. F. B. (2022). Research of Detection and Prevention of Phishing and Proposal of Phishing Detector Solution. Journal of Applied Technology and Innovation, 6(3), 1–5.