Integrating Multiple Parameters to Improve Spam and Phishing Email Filtering

Shreyans Jain[1], Ekansh Dubey[2], Rachna Jain[3]

[1]Bachelor Student, Department of Computer Science, Bharati Vidyapeeth's College of Engineering, New Delhi, India

[2]Bachelor Student, Department of Computer Science, Bharati Vidyapeeth's College of Engineering, New Delhi, India

[3]Assistant Professor, Department of Computer Science, Bharati Vidyapeeth's College of Engineering, New Delhi, India

Abstract:- Phishing remains a major security threat in cyber security. In phishing, attackers steal sensitive information such as login credentials, credit card numbers from victims by providing a fake site which looks like the visual clone of a legitimate site. It can be said that single filter method is not sufficient to detect different categories of phishing emails. This paper provides a model combining various filtering methods to detect phishing. The model incorporates five layers: Block list layer, white list layer, Spam filtering layer, URL features layer, hashing layer. A prototype implementation of the proposed model is built to improve spam and phishing Email filtering over the existing techniques.

Keywords:- *Spam, Phishing, Email Filtering, Layered Filtering, Malicious Email.*

I. INTRODUCTION

Email has made the communication process become easier, faster and cheaper. It has become increasingly popular. However, phishing email is one of the major security threats. The phishing email can lead to financial loss. Attacker always sending email tends to make user believe that they are communicating with trusted entity and deceive them into providing personal credentials in order to access service, such as credit card numbers, account login credential or identity information. Phishing email causes a serious threat to information security and internet privacy. Phishing is often used to gain a foothold in corporate or governmental networks as a part of a larger attack, such as an advanced persistent threat (APT) event.

According to APWG Global Phishing Survey for 2016 there were at least 255,065 unique phishing attacks worldwide. This represents an increase of over 10% from the 230,280 attacks we identified in 2015. The attacks occurred on 195,475 unique domain names. Of the 195,475 domains used for phishing, we identified 95,424 domain names that we believe were registered maliciously by phishes.

According to Verizon's 2017 Data Breach Investigations Report, two-thirds of all malware was installed via email attachments in 2016. 60% of malware was packaged in JavaScript attachments, while 26% was packaged in malicious macros embedded in Microsoft Office documents.

Numerous techniques or methods have been proposed to detect and filter spam and phishing emails. Existing filters consume much more memory and time, and are weak in detecting zero-day attacks.

High changing rate of phishing attack techniques increase the difficulty of detecting and filtering phishing email attacks.

Phishing email messages, websites, and phone calls are designed to steal money. Cybercriminals can do this by installing malicious software on your computer or stealing personal information off of your computer.

Cybercriminals also use social engineering to convince you to install malicious software or hand over your personal information under false pretenses. They might email you, call you on the phone, or convince you to download something off of a website.

The most common type of phishing scam, deceptive phishing refers to any attack by which fraudsters impersonate a legitimate company and attempt to steal people's personal information or login credentials. Those emails frequently use threats and a sense of urgency to scare users into doing the attackers' bidding.

Spamming remains economically viable because advertisers have no operating costs beyond the management of their mailing lists, servers, infrastructures, IP ranges, and domain names, and it is difficult to hold senders accountable for their mass mailings. Because the barrier to entry is so low, spammers are numerous, and the volume of unsolicited mail has become very high. In the year 2011, the estimated figure for spam messages is around seven trillion. The costs, such as lost productivity and fraud, are borne by the public and by Internet service providers, which have been forced to add extra capacity to cope with the deluge. Spamming has been the subject of legislation in many jurisdictions.

Spam and Phishing is a big problem in terms of email communication and data security as they both adversely affect the comfort and privacy of data. System level integration to prevent these types of email can lead to improve or prevent various problems faced due to these unsolicited emails.

II. RELATED WORK

An earlier suggested model developed by A.K. Jain et al [10], used white list which updates automatically which helps to identify legitimate sites and warn the users that the URL is available on the white list or not. two components were used to identify the legitimacy of the emails, which are: 1) Domain and IP address matching module, 2) analyzing the features of the links from source code. The result from the experiment shows that the suggested model calculated 86.02% true positive rates, and 1.48% false negative rates.

One way to filter out spamming emails is to identify a sender against predesigned lists (white list or blacklist) or rules.

For example, Microsoft Outlook^{TM} interprets some rules for arrange emails. If an email meets the Junk E-mail Filter's interpretation of spam, it will be sent into the junk folder. Otherwise, if the email does meet the interpretation of a phishing rules, Outlook^{TM} keeps the email in the folder Inbox but disables the hyperlinks in the email and inhibits the user from responding to the mail, arranges it into Junk E-mail folder otherwise.

Certified e-mail (e.g., Kobe Certified by Kobe Mail, Certified Email by Good mail Systems) is an e-mail white listing approach by which an internet provider allows someone to circumvent spam filters when sending e-mail to its subscribers, in return for compensating a fee to the certifying service. A sender can then be rest assure that their messages have reached their destination without any blocking, or having links or images denuded out of them, by spam filters. The purpose of certified e-mail is to enable organizations to a reliable reach to their customers by e-mail, while giving recipients certainty that a certified message is legitimate and is not a fraudulent phishing attempt.

In recent years, numerous content-based filters have been constructed specifically for phishing emails. In [11] the authors developed a system to filter phishing emails based on the architectural properties of phishing emails. It mainly considered the dialectal properties that differentiate phishing emails from other emails. The system developed in [6] filters phishing emails based on the properties of phishing emails such as IP-based URLs and the era of domain-names. However, to identify a set of context and architectural properties that can differentiate legitimate messages from phishing messages is a problem.

A new approach that focused on keyword to identify phishing emails is overlooked in [17]. The properties represent the occurrence rate of 43 keywords that can be observed in phishing and legitimate emails. They examined on a collection of around 1700 phishing emails and 1700 legitimate emails from private mailboxes. As phishing emails appear almost similar to legitimate emails, this approach was not the most reliable anymore.

Ion Androutsopoulos et al [1] suggested a model which used naïve bayes algorithm to perform deep analysis on the data or content of the email. The system would test the content of the email and on the basis of the analysis a decision would be formed regarding the legitimacy of the email. Upon the end of the analysis the Email would be considered legitimate or not.

K. Thomas et al [2] proposed a model that would use the evaluation techniques of analyzing the URLs or Hyperlinks associated to an email to evaluate the legitimacy of an email, in this model the system performs analysis on the links that associated with the email which thereon would lead to a better understanding of the email' legitimacy.

The approach was to analyze the domain of which the link belongs to, thereon the server that is accessed upon when the link is accessed. If the link belongs to a malicious or spam domain then the chances of the link being illegitimate is very high and is considered as an illegitimate email.

M. Gleeson et al [3] derived a model that was different from all of the previous used model as it didn't use a learning mechanism to identify any illegitimate emails. In this model we used Hash sum of the email content and based on that analysis would suggest if the email is legitimate or not.

This approach was implemented over the content of the email, a hash sum is created from the content of the email and then is used to perform comparisons with previous spam emails which would lead to establish if the email is legitimate or not. If the hash sum is similar to hash sum of any previous spam email the chances of the email being illegitimate is very high.

Andre Bergholz et al [7] suggested a model which focused on phishing filter than used contest based filtering. Basically the system identifies the tags and the URLs that are associated with the email and would base the legitimacy on that analysis.

The approach suggested that the email is analyzed for tags and URL features and based on the analysis report a prior decision is taken which classifies the email as legitimate or illegitimate.

All these approaches have been prominent for a while but they all were focused on one aspect of the approach while collectively they might turn out to be a better filtering mechanism.

The earlier approaches lacked vision of fusing multiple parameters together to enable a better and more effective filter that would enable the system to have a better learning and could perform multiple analysis that would make the identification of a mail's legitimacy or illegitimacy more prominent.

III. PROPOSEDWORK

A. *ModelArchitecture*

Layer 1: This layer reduces the number of emails required for additional filtering. In this layer we have a list of domains that are considered to be legitimate in the whitelist, therefore the emails are do not require further verification of the email saving time and computation cost.

We use this layer to identify domains that are already listed as legitimate, by using this layer we can easily account an Email as legitimate on the basis of the domain its comes from, if the domain is a white list domain then it can easily be said that the Email is a legitimate one.

Layer 2: This layer deals with the block list filtering i.e. it matches the sender's domain of the email with the block list which provides the list of malicious domains. If the domain is present in the list, then the email is categorized as malicious and the email is blocked. We are using this layer as per the norm to differentiate between what could be a potential legitimate Email and an easily accountable illegitimate Email.

Layer 3: This layer enables to establish the fact that an email is spam or not, we use naïve Bayes theorem to perform analysis over the email. Deep analysis of an email is done, where the contents of the email is compared to previously established data library using which is made up of combined legitimate and spam mails based upon the data new email is categorized as spam or legitimate email.

This layer enables a deep analysis of the Email which leads the system to establish that the mail is legitimate or not.

Layer 4: In this layer, the phishing URL is compared with the popular legitimate URLs to check the similarity between the two. This process enables to filter any malicious link that could be a potential threat to the user or organization based on the system involved in place.

The basic idea behind this layer is to detect any fraudulent or malicious intentions of a person and prevent any malicious activities.

Layer 5: This layer deals with the malicious attachment, if present, in the email by computing the hash of the attachment with the list of malicious hashes.

A malicious attachment when downloaded can act in multiple ways for example, it can work as a monitoring bot which sends out sensitive data from the user end to the malicious end without being noticed. To prevent any such activities a system needs to be in place.

For the above layers to work effectively and efficiently a basic methodology or a structure needs to be in place so that the model works. For the structure to work properly, we framed an architecture which involves the following proposition.

Firstly, the Email is arrived at the system it is scanned through the list of white list domain this is our first and foremost layer. This layer helps to identify the domain easily and effectively which means we can directly identify if the Email is legitimate or illegitimate.

The Email's domain is identified and scanned in the White list domain database for identification. If the domain exists in the white list database the Email is considered as legitimate, but if the domain is not available in the white list domain the mail is moved forward to the next layer.

In the Second layer, the Email's domain is scanned through a Blacklist domain database for identification. If the domain exists in the database, we can be assured that the Email is fairly a Spam or Phishing Email. So we can directly consider the Email as Spam, Otherwise the email is moved on to the next layer.

In the third layer, deep analysis of the email is done based on Naïve Bayes algorithm. Here the Email content is analyzed with a data library which is based on the previous illegitimate mails. This layer grows stronger with time as the data library is improved as it detects more Spam mails as it adds the data to the previous data library refining it and making it even better.

This layer helps to get the best results possible for the Spam detection and allows the system to identify the spam mails and help to prevent them from any exposure to the system. But even in the case that an Email is not a Spam mail, it can possess the qualities of a phishing mail. So even if the email passes this layer, we cannot say for sure that it is legitimate so we pass this email forward onto our next layer.

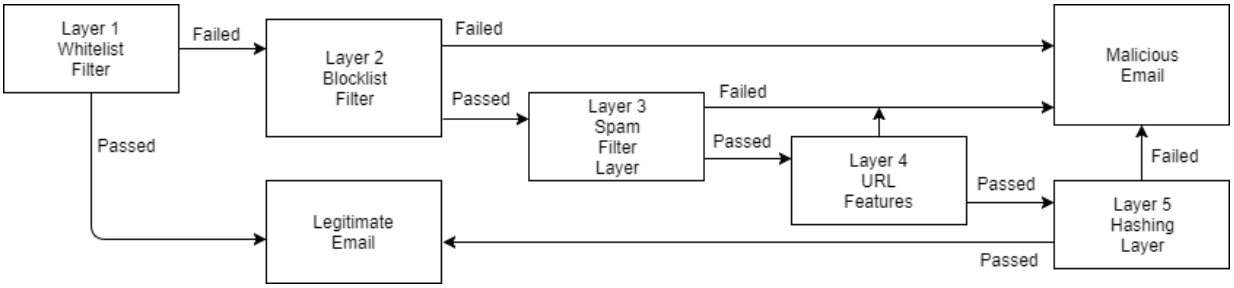

Fig 1:- Proposed Model Architecture

In the Fourth layer, we deal with all the links and URLs that are mentioned and how it can affect our system. Basically the email could possess links which are highly malicious and are illegitimate. So the legitimacy of the links is evaluated based on the credibility of the link or the domain that is accessed by the link. Analysis is done over the aspects of the domain and server links that are related to the URLs and Hyperlinks which enable to establish the credibility of the links and enable to identify illegitimate Email. These types of mails are thereon considered as Phishing Emails and are directly considered as illegitimate. If the email passes this layer, it moves to our final layer.

In the Fifth or our final layer, a Hash sum is created for the received email, if this hash sum matches the hash sum of any previous spam or phishing mail then that mail is considered as a spam or a phishing mail. This layer also identifies any malicious attachments that are involved with the spam or phishing email.

Collectively the above proposed layer architecture clearly works in a simpler manner and prevents any illegitimate spam or phishing emails to bypass the system.

This structure is highly stable and covers all the aspects of the email spectrum and helps to reduce any bypassing of illegitimate emails. The collection of various levels or layers enables to establish a strong factual analysis on why an email is considered illegitimate and also it covers a various aspects of discomfort experienced in terms of emails as there is a need of a system that can prevent Spam and Phishing mails together.

IV. RESULTS

The proposed algorithm filters both spam and phishing emails, the combined layers of the algorithm provide better email filtering and computation time. Figure 2 shows the accuracy of the spam classifier used i.e. SVM and Naïve. It can be seen in the figure 2, that as the size of the data set increases the accuracy of the spam classification decreases. Also the accuracy of SVM is significantly better than the naïve classifier for smaller data set but as the size of the data set increase the difference in the accuracy of both the algorithms reduces significantly.

Since to create an algorithm that provides high accuracy for large data set is quite difficult, but it is much easier to reduce the computation time of the algorithm without compromising the accuracy of algorithm.

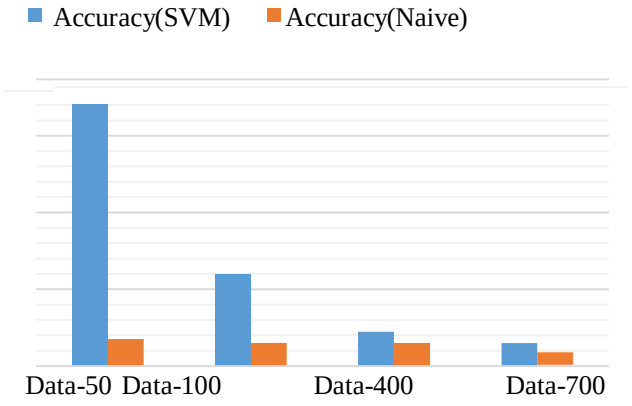

Fig 2:- Accuracy Graph

Figure 3, shows the results of the proposed method compared with the AOL email filtering. Out of 99 emails (Spam, Phishing and Legitimate Emails), AOL correctly identified 72 emails and the proposed method correctly identified 88 emails. This result gives the accuracy of 72.72% and 88.88% respectively.

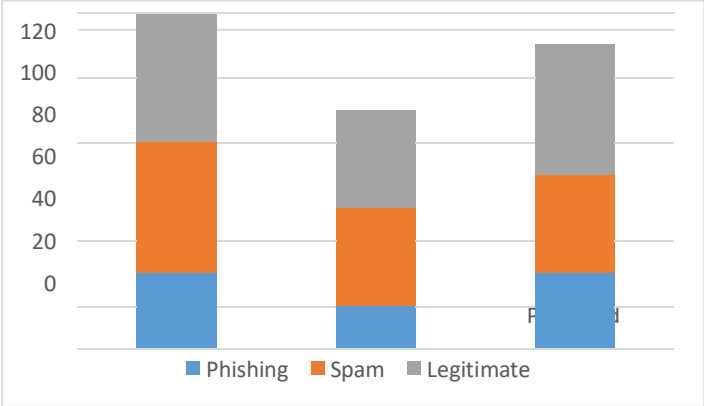

Fig 3:- Correct Email Categorization Count

V. CONCLUSION AND FUTUREWORK

The proposed method is found to be better than the AOL email client as the accuracy of the algorithm is better than the AOL filtering algorithm.

In future, the number of layers can be increased for more accuracy, to do so the number of features applied in the algorithm can be increased. Also the URL features can be enhanced by adding more parameters for checking malicious domains and URLs.

The algorithm can be optimized to reduce the computation time required for filtering of the email. Both spam and phishing algorithms can be swapped with better algorithms for higher accuracy.

REFERENCES

[1]. Androutsopoulos, I., Koutsias, J., Chandrinos, K. V., Paliouras, G., & Spyropoulos, C. D. (2000). An evaluation of naive bayesian anti-spam filtering. arXiv preprintcs/0006013.

[2]. Thomas, K., Grier, C., Ma, J., Paxson, V., & Song, D. (2011, May). Design and evaluation of a real-time url spam filtering service. In Security and Privacy (SP), 2011 IEEE Symposium on (pp. 447-462).IEEE.

[3]. Gleeson, M., Hoogstrate, D., Jensen, S., Mantel, E., Medlar, A., & Schneider, K. (2004). U.S. Patent Application No.10/846,723.

[4]. Sakkis, G., Androutsopoulos, I., Paliouras, G., Karkaletsis, V., Spyropoulos, C. D., & Stamatopoulos, P. (2003). A memory-based approach to anti-spam filtering for mailing lists. Information retrieval, 6(1),49-73.

[5]. Basnet, R. B., Mukkamala, S., & Sung, A. H. (2008). Detection of Phishing Attacks: A Machine Learning Approach. Soft Computing Applications in Industry, 226, 373-383.

[6]. Basnet, R. B., Sung, A. H., & Liu, Q. (2011). Rule-based phishing attack detection. In International Conference on Security and Management (SAM 2011), Las Vegas,NV.

[7]. Bergholz, A., De Beer, J., Glahn, S., Moens, M. F., Paaß, G., & Strobel, S. (2010). New filtering approaches for phishing email. Journal of computer security, 18(1),7-35.

[8]. Jain, A. K., & Gupta, B. B. (2016). A novel approach to protect against phishing attacks at client side using auto-updated white-list. EURASIP Journal on Information Security, 2016(1),9.

[9]. Dwyer, P., & Duan, Z. (2010, July). MDMap: Assisting users in identifying phishing emails. In Proceedings of 7th annual collaboration, ELECTRONIC messaging, Anti-ABUSE and spam conference(CEAS).

[10]. Darling, M., Heileman, G., Gressel, G., Ashok, A., & Poornachandran, P. (2015, July). A lexical approach for classifying malicious URLs. In High Performance Computing & Simulation (HPCS), 2015 International Conference on (pp. 195-202).IEEE.

[11]. Sakkis, G., Androutsopoulos, I., Paliouras, G., Karkaletsis, V., Spyropoulos, C. D., & Stamatopoulos, P. (2003). A memory-based approach to anti-spam filtering for mailing lists. Information retrieval, 6(1),49-73.

[12]. Garera, S., Provos, N., Chew, M., & Rubin, A. D. (2007, November). A framework for detection and measurement of phishing attacks. In Proceedings of the 2007 ACM workshop on Recurring malcode (pp. 1-8).ACM.

[13]. Kumar, S., Gao, X., Welch, I., & Mansoori, M. (2016, March). A machine learning based web spam filtering approach. In Advanced Information Networking and Applications (AINA), 2016 IEEE 30th International Conference on (pp. 973-980).IEEE.

[14]. Huang, H., Qian, L., & Wang, Y. (2012). A SVM-based technique to detect phishing URLs. Information Technology Journal, 11(7),921-925.

[15]. Huang, H., Tan, J., & Liu, L. (2009, June). Countermeasure techniques for deceptive phishing attack. In New Trends in Information and Service Science, 2009. NISS'09. International Conference on (pp. 636-641). IEEE.

[16]. Huh, J. H., & Kim, H. (2011). Phishing Detection with Popular Search Engines: Simple and Effective. *FPS*, *11*, 194-207.

[17]. Khonji, M., Jones, A., & Iraqi, Y. (2011, February). A novel Phishing classification based on URL features. In *GCC Conference and Exhibition (GCC), 2011 IEEE* (pp. 221-224).IEEE.

[18]. Zhou, B., Yao, Y., & Luo, J. (2014). Cost-sensitive three-way email spam filtering. Journal of Intelligent Information Systems, 42(1),19-45.

[19]. Bergholz, A., De Beer, J., Glahn, S., Moens, M. F., Paaß, G., & Strobel, S. (2010). New filtering approaches for phishing email. Journal of computer security, 18(1),7-35.

[20]. Tuteja, S. K. (2016). A Survey on Classification Algorithms for Email Spam Filtering. International Journal of Engineering Science,5937.

[21]. Tuteja, S. K., & Bogiri, N. (2016, September). Email Spam filtering using BPNN classification algorithm. In Automatic Control and Dynamic Optimization Techniques (ICACDOT), International Conference on (pp. 915-919). IEEE.

[22]. Abdelrahim, A. A. A., Elhadi, A. A. E., Ibrahim, H., & Elmisbah, N. (2013, August). Feature selection and similarity coefficient based method for email spam filtering. In Computing, Electrical and Electronics Engineering (ICCEEE), 2013 International Conference on (pp. 630-633). IEEE.

[23]. Rajendran, P., Tamilarasi, A., & Mynavathi, R. (2016). An Enhanced Approach towards Privacy Preserving Email Spam Filtering. Asian Journal of Research in Social Sciences and Humanities, 6(8),21-28.

[24]. Kotian, H., Gupta, K., & Stephy, J. J. (2015). Using Fuzzy Logic for Email Spam Filtering. International Journal, 5(10).

[25]. Stolfo, S. J., Eskin, E., Herskop, S., & Bhattacharyya, M. (2015). U.S. Patent No. 8,931,094. Washington, DC: U.S. Patent and Trademark Office.