International Journal of Scientific Research and Engineering Development-– Volume 4 Issue 3, May -June 2021

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Email Spam Filtering Techniques – A Survey

Karthikeya H S*, Ganashree K C**

*(Department of Computer Science, R.V. College of Engineering, Bangalore
Email: karthikeyahs@gmail.com)

** (Department of Computer Science, R.V. College of Engineering, Bangalore
Email:ganashree@rvce.edu.in)

----------**********----------

Abstract:

Email spam is one of the serious issues of the present-day Internet, carrying monetary harm to organizations. It also causes irritation in individual clients. Among the techniques created to stop spam, filtering is a significant and famous one. In this paper we discuss filtering techniques other than machine learning methods. Some of the techniques that can be mentioned are IP reputation filtering, mail policies, content filters and sender domain reputation filtering. Machine learning methods are the most reliable techniques for filtering spam. But, in many scenarios some simple methods such as the above mentioned can be used when the spam mail passes the machine learning filters or to reduce the load on machine learning filters.

Keywords **—Email spam filtering, IP reputation filtering, mail policies, sender domain reputation filtering, content filters**

----------**********----------

I. INTRODUCTION

In most parts of the world emails is the official mode of communication. Emails which originate at the sender travels through many components of the network mainly the Internet to reach the receiver. The components in the email flow are sender, mail user agent, mail submission agent, mail transfer agent, mail delivery agent and receiver. The sender is a person who writes the email or is an automatic email generator. Emails can be typed using a mail user agent like outlook, Gmail, Mozilla Thunderbird etc. Some of the mail user agents like outlook and Gmail also does the work of authentication like sender authentication, email content validation and email encryption. The next component in the email flow is mail submission agent. The mail submission agent is responsible for passing on the email from the private network to the sender's outgoing mail server. Every email user has 2 servers, An incoming mail server and outgoing mail server. The outgoing mail server forwards the email to the Internet. The Internet is responsible for routing the email to the receivers incoming mail server. The next component in the mail flow is mail delivery agent. The incoming mails are stored in the incoming mail server of the recipient. It is the responsibility of the mail delivery agent to pull the emails from the incoming mail server. On the receiver side also, there exists a mail user agent. The emails reside in the inbox of the receiver's mail user agent. the receiver is a person whom the mails are intended for. This email flow is depicted in figure 1. One of the main devices which lies at the border of the Internet and the private network is a security appliance or a firewall which filters unsafe mails. The functions of this security appliance are:

- Receipt — As the appliance connects to a remote host to receive incoming email, it adheres to configured limits and other receipt policies. For example, verifying that the host can send your users mail, enforcing incoming

connection and message limits, and validating the message's recipient.

- Work Queue — The appliance processes incoming and outgoing mail, performing tasks such as filtering, safelist/blocklist scanning, anti-spam and anti-virus scanning, Outbreak Filters, and quarantining.
- Delivery — As the appliance connects to send outgoing email, it adheres to configured delivery limits and policies. For example, enforcing outbound connection limits and processing undeliverable messages as specified.

Comparing the email flow to the traditional post mechanism, emails are like letters. The mail user agent can be compared to the post box. The mail submission agent is the postman and the post office. The Internet is like the transmitter of post. The mail delivery agent can be compared to the receiver's post office and postman. The receivers post box is again the mail user agent.

Emails being widely used across the world is also the major threat space which is explored by spammers and hackers. Some of the attacks that can happen through emails are email spams, virus emails, email phishing attacks, email bomb attacks, Gray mails and outbreak mails.

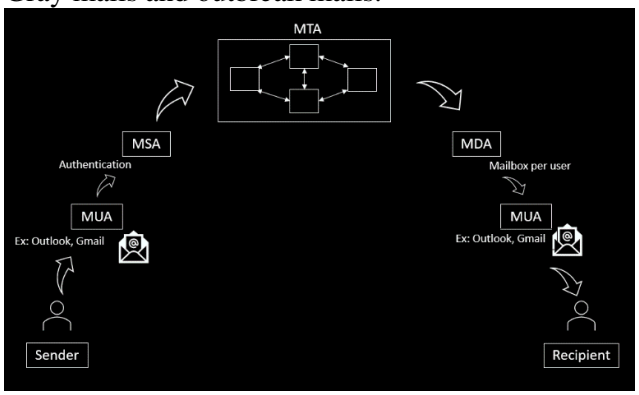

Fig. 1Email end-to-end flow

A brief introduction about these email attacks is provided in section 2.

II. EMAIL ATTACKS

A. Spam Attack

Email spams can be defined as unsolicited bulk mails. Here unsolicited means uninvited or unsubscribed emails. The receiver and the sender do not know each other in this case. The receiver wouldn't have subscribed to the sender's emails. As email ID's can be looked upon on the web and on the social media platforms, nowadays it is easy for spammers to send spam emails. The second characteristic of a spam mail is that it is sent in bulk. To counterattack spam mails spam filters which uses machine learning techniques are in place. Any machine learning algorithm is not perfect. Hence some of the spam mails might pass the machine learning filter test. To mitigate this problem, we explain some of the most important techniques other than machine learning methods which can be used to filter spam. These methods can be used on top of the spam filters which use machine learning techniques. Some of these methods are discussed in Section 4.

B. Phishing Attack

Phishing attack can be defined as an email attack where the attacker impersonates himself as a legitimate party. The attacker makes the receiver believe that the email has been sent from a trusted sender. The objective is to deceive the email beneficiary into accepting that the message is something they need - a solicitation from their bank or a note from somebody in their organization. Phishing emails usually tend to ask the email beneficiary their personal information or sensitive data like bank account details, credit card details etc.

Phishing attacks started late back in 1990s. 2021 statistics say that 96% of phishing attacks happens in email space. The rest 4% attacks happen through phones and malicious websites. Federal Bureau of Investigation has reported that 75% of the companies around the globe have experienced one or the other forms of phishing attacks in 2020. 74% of the companies based in United States have experienced successful phishing attacks which is 14% higher than 2019.

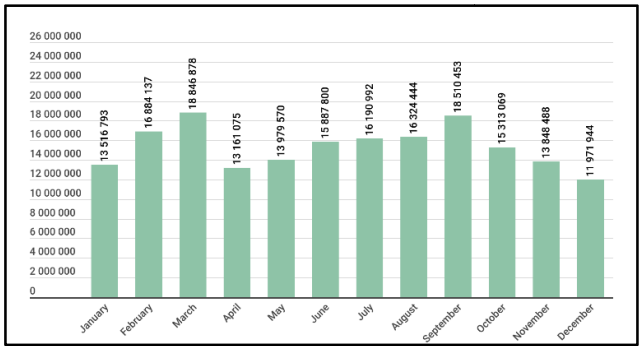

Fig. 2 Combined data of Spam and Phishing attacks. Credits: Kaspersky

The above bar graph shows the total number of email spam and email phishing attacks that have happened in each month of 2020. The trend can be explained as neither increasing nor decreasing throughout the year. An average of 15 million such attacks have happened every month. This is the data captured by Kaspersky, an anti-virus product company.

C. Virus Attack

A computer virus can be defined as a malicious code or malware which when enters the victim's computer can spread by modifying the computer programs and inserting its own code. Email virus constitutes the major portion of computer viruses. The malicious code is spread in the email message. This code gets activated when the innocent receivers click on the email in the inbox, or when they open a web link or when they download an email attachment. Hence, malicious code enters the victim's computer. Though viruses have various means to enter a computer, emails are the easiest ways to enter as emails are still the official means of communication.

Phishing attacks when combined with email virus attack is even more dangerous. The attacker usually performs 2 techniques to get sensitive information from the email recipient. The recipient is either made to believe that the sender is a trusted entity and hence he can share sensitive information or email virus is made to enter the recipient's computer after making him believe that the email sender is a legitimate sender.

Statistics of 2021 tell that 94% of malware is delivered by email. Among those, 66% of malware on a computer gets installed from email attachments. It has been recently discovered that among all email attachments which contain malware, the major file types are .doc and .dot which comprise about 37% and .exe files which comprise about 19.5%. Moreover, 47% of the malware email attachments are office files. The Covid-19 pandemic also gave a boost to email attacks. There have been email malware and phishing attacks in the name of corona virus. Google alone has blocked 18 million malware and phishing emails on a daily basis.

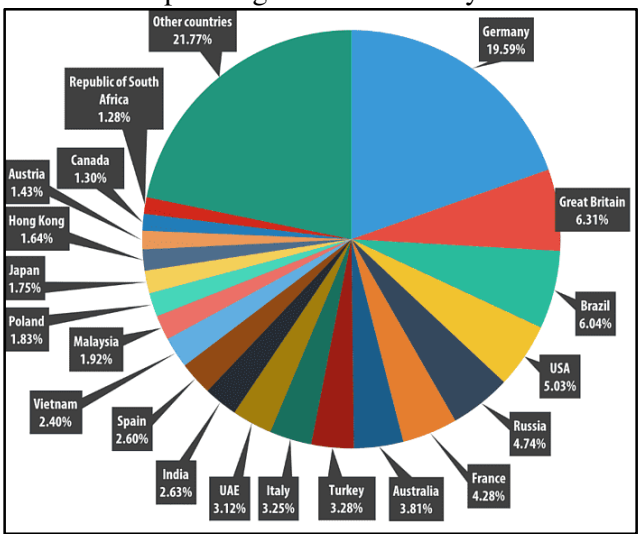

Fig. 3 Country wise statistics of email spam, virus and phishing attacks combined. Credits: SecureList

The above pie chart gives country wise email spam, virus and phishing attack statistics. From the pie chart, we can conclude that Germany tops the list of countries which experience the above-mentioned email attacks. Germany is followed by Great Britain, Brazil, USA and Russia. India experiences around 2.63% of total email attacks of the world.

D. Email Bomb Attack

Email Bomb attacks started in late 1990s. Email bomb attack can be defined as a scenario when someone receives thousands of emails in a short span of time. An official definition of email bomb is "It is a denial-of-service attack (DoS) against an

email server, designed to make email accounts unusable or cause network downtime.". When thousands of emails are sent to one recipient, the incoming mail server of that recipient will face huge incoming traffic. This causes network down time of mail servers which can disrupt one's business for a short span of time. It also disrupts one's ability to communicate. The main intention behind email bomb attacks is to hide certain important emails regarding the email account or the business. Usually, this happens because the recipient tends to miss out an important mail among thousands of unwanted mails.

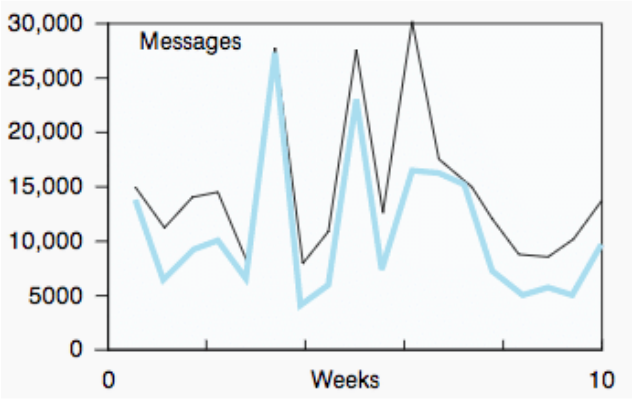

Fig. 4 Daily number of email bombs (shaded) vs total emails. Picture Credits: E-Mail Bombs and Countermeasures: Cyber Attacks on Availability and Brand Integrity [1]

E. Graymail

Graymails are similar to email spams except that they are solicited mails. Graymails is not classified as an email attack by many experts. Advertisement and announcements are a part of graymails. These mails can be annoying for some and useful for some others.

III. LITERATURE SURVEY

Bing Zhou, Yiyu Yao and Jigang Luo [2] in their paper titled "A Three-Way Decision Approach to Email Spam Filtering" have proposed a email spam filter based on Bayesian Theory. Email spam filtering is often dealt as a binary classification problem. hence there is a possibility of getting false positives and false negatives if the classification model isn't accurate. To reduce the number of false positives and number of false negatives a three-way decision approach is proposed in this paper. The model thus developed flags the emails as spams, non-spams or cannot be decided. The third category of spams reduces the number of false positives and false negatives. False positive is a case where a legitimate email is flagged as a spam mail. False negative is a case where a spam mail is flagged as a legitimate mail. The false positive case is quite dangerous as the business might miss out some important emails.

J Attenberg, K Weinberger, A Dasgupta and A Smola [3] in their paper titled "Collaborative email-spam filtering with the hashing trick" have proposed a two-pronged approach for email spam filtering. A local email spam filtering algorithm which works for an individual client also takes the input from the client who flags the mails as spams. Some malicious clients intently flag legitimate mails as spams. Hence such a behaviour should not affect a global email spam filtering algorithm which is applied across all the clients. Hence, they propose a local email spam filtering algorithm for such individual clients and a global spam filtering algorithm for general clients.

V Christina, S Karpagavalli and G Suganya [4] in their paper titled "Email spam filtering using supervised machine learning techniques" have proposed and email spam filter based on supervised learning algorithms such as decision tree classifier, multilayer perceptron and Naïve Bayes classifier. These supervised learning algorithms learn the features of spam mails. The learned features are then used to classify emails as spams or non-spams.

CT Wu, KT Cheng, Q Zhu and YL Wu [5] in their paper titled "Using visual features for anti-spam filtering" have proposed an email spam filtering algorithm based on visual clues. As email spam filtering techniques become robust, spam attackers also become smart. The content of emails and the type of content has changed in the last decade or so. Nowadays it is easier to send images and videos in emails. Hence spam filtering methods based on text analysis is not enough. In this paper the authors have developed an email spam filter

which analyses images in spam mails (as flagged by the client). Later, the spam filtering algorithm applies these image features on new emails to classify them as spam or legitimate mails.

W Liu and T Wang [6] in their paper titled "Online active multi-field learning for efficient email spam filtering" have proposed a spam filtering algorithm which is an online learning algorithm based on multiple fields in emails. The algorithm is based on three characteristics of an email. We know that an email is an online application. Hence the algorithm to filter spam mails should also be an online learning algorithm. An email contains multiple fields like subject, body, to address and file attachments. Hence to classify an email as spam or legitimate mail it is not enough to analyse only the body of the email. Therefore, this model takes into consideration all the fields of an email to classify an email as spam or not. Also, it is not cost efficient to get a label for a real time spam filtering algorithm. Hence the algorithm should be an active learning algorithm. Hence this algorithm is named online active multi field learning algorithm for email spam filtering. The algorithm meets the state-of-the-art performance and requires smaller number of features which in turn reduces time and memory issues.

O Al-Jarrah, I Khater and B Al-Duwairi [7] in their paper titled "Identifying potentially useful email header features for email spam filtering" have proposed an email spam filtering algorithm which uses email headers. We know that each transmission protocol over the Internet has its own headers. In this model email headers are used to classify email as spam or not. Publicly available data sets are used for this purpose. The classification algorithms used are decision tree classifier, naïve bayes classifier, support vector machines, multilayer perceptron, bayes network and random forest.

IV. METHODOLOGY

Most of the email spam filtering algorithms are based on one or the other machine learning techniques. The literature on machine learning based email spam filtering techniques is in abundance. Hence, this paper will not discuss any machine learning algorithms for email spam filtering. Though, the prime algorithms are based on machine learning, some simple techniques can be used to avoid false positives. Some off these techniques are message filters, content filters, IP reputation filtering, mail policies, sender domain reputation filtering. These techniques are discussed here:

A. IP Reputation Filtering

Every networking device on the globe have an IP address assigned to it. Therefore, IP address can be used as a factor based on which emails are flagged as spam or not. Usually, emails from highly reputable senders like customers and partners can be accepted. Mails from less reputed senders can be passed for further content scanning and spam filters. This reduces the load on spam filters. Here, the sender's trustworthiness is decided based on IP address. An IP Reputation Score (IPRS) can be assigned for each sender based on the volume of email data sent by the sender, complaint rates, data from public blacklists and open proxy lists. Highly reputed senders can be assigned a positive IPRS and IP addresses which are open proxies or that send high volumes of spam or viruses to users or spam traps can be assigned a negative IPRS.

B. Mail Policies

Mail policies can be defined as rules that govern what type of data should enter the private network and what should not. It also decides what data can be sent out from private network of the organization to the Internet to prevent sensitive data loss. The content may include spams, legitimate marketing messages, graymail, phishing attacks, virus mails, organization's private data and might also contain personal data. Hence, mail policies can be designed to not allow plausible spam mails. For example, if mail policies suspect a mail to be spam, then the spam mails can be quarantined in the Email Service Provider's (ESP) database. The ESP's mail managing admin is responsible for manually

checking if the mail is spam or not. Later, the admin can decide on either deleting the mail if it is spam or releasing the mail if it is not spam. The other alternative is to send the suspected spam mail to the recipient with a warning which tells that the mail is a spam. The mail subject can be manipulated to let the email recipient know that the email is spam.

C. Message Filters

Message filters can be defined as rules for handling the mail messages. An email body and attachments are passed through filters like text filter, image filters, document filters. Filter actions can be defined as the action to be performed when a filter filters a part of the email. The filter actions can be 'drop the mail', 'bounce the mail to the bounce address', 'archive the mail', 'quarantine the mail', 'blind carbon copy the mail or 'alter the mail'. For example, a text filter which tells that a mail with subject or body which contains the word "terrorism" should be deleted. Hence, this text filter will delete any email which contains "terrorism" in its subject or body.

D. Content Filters

Content filters can be used for customizing the factors that decide whether an email is spam or not. It is similar to message filters except that it is applied later in the email pipeline after splintering the email message into multiple parts.

E. Sender Domain Reputation Filtering

Every email address consists of 3 parts – username, '@' symbol and 'domain name'. For example, 'abc@rvce.edu.in'. In this email address, 'abc' is the username and 'rvce.edu.in' is the hostname. Sender domain reputation filtering is a simple filter which assigns each domain a Sender Domain Reputation Score (SDRS). Any email received from a sender whose domain has SDRS less than a threshold value is dropped off. Other emails are sent for further processing like IP reputation filtering, mail policies, message filters and content filters.

V. CONCLUSIONS

This survey paper has covered the email flow pipeline, why email is the biggest threat space for spammers and hackers, what are the different email attacks, statistics related to those email attacks, literature survey section which mainly includes machine learning based email spam filtering techniques and methodology section which covers 5 important non-machine learning based spam filtering techniques.

ACKNOWLEDGMENT

I would like to thank my college, R.V. College of Engineering and Computer Science and Engineering Department for giving me an opportunity to conduct this extensive research survey. I also thank my professor Prof. Ganashree K C for providing valuable guidance during the research.

REFERENCES

[1] Bass, Tim &Freyre, Alfredo & Gruber, David & Watt, Glenn. (1998). E-Mail Bombs and Countermeasures: Cyber Attacks on Availability and Brand Integrity. Network, IEEE. 12. 10 - 17. 10.1109/65.681925.

[2] Zhou, Bing & Yao, Yiyu& Luo, Jigang. (2010). A Three-Way Decision Approach to Email Spam Filtering. 28-39. 10.1007/978-3-642-13059-5_6.

[3] Attenberg, J. & Weinberger, Kilian & Dasgupta, Anirban &Smola, A. &Zinkevich, Martin. (2009). Collaborative Email-Spam Filtering with the Hashing Trick.

[4] V.Christina, &S.Karpagavalli, &G.Suganya,. (2010). Email Spam Filtering using Supervised Machine Learning Techniques. International Journal on Computer Science and Engineering. 2.

[5] Ching-Tung Wu, Kwang-Ting Cheng, Qiang Zhu and Yi-Leh Wu, "Using visual features for anti-spam filtering," IEEE International Conference on Image Processing 2005, 2005, pp. III-509, doi: 10.1109/ICIP.2005.1530440.

[6] Liu, Wuying& Wang, Ting. (2011). Online Active Multi-Field Learning for Efficient Email Spam Filtering. Knowledge and Information Systems. 33. 10.1007/s10115-011-0461-x.

[7] Washha, Mahdi &Khater, Ismail &Qaroush, Aziz. (2012). Identifying Spam E-mail Based-on Statistical Header Features and Sender Behavior. ACM International Conference Proceeding Series. 10.1145/2381716.2381863.

[8] Blanzieri, E., Bryl, A. A survey of learning-based techniques of email spam filtering. ArtifIntell Rev 29, 63–92 (2008). https://doi.org/10.1007/s10462-009-9109-6

[9] Nurul FitriahRusland et al Analysis of Naïve Bayes Algorithm for Email Spam Filtering across Multiple Datasets; 2017 IOP Conf. Ser.: Mater. Sci. Eng. 226 012091

[10] Christina, V &Karpagavalli, S &Suganya, G. (2010). A Study on Email Spam Filtering Techniques. International Journal of Computer Applications. 12. 10.5120/1645-2213.

[11] Gordon V. Cormack, José María Gómez Hidalgo, and Enrique PuertasSánz. 2007. Spam filtering for short messages. In Proceedings

of the sixteenth ACM conference on Conference on information and knowledge management (CIKM '07). Association for Computing Machinery, New York, NY, USA, 313–320. DOI:https://doi.org/10.1145/1321440.1321486

[12] Bhowmick, Alexy& Hazarika, Shyamanta. (2018). E-Mail Spam Filtering: A Review of Techniques and Trends. 10.1007/978-981-10-4765-7_61.

[13] Bhuiyan, Hanif &Ashiquzzaman, Akm&Juthi, Tamanna & Biswas, Suzit& Ara, Jinat. (2018). A Survey of Existing E-Mail Spam Filtering Methods Considering Machine Learning Techniques.

[14] Delany S.J., Cunningham P., Tsymbal A., Coyle L. (2005) A Case-Based Technique for Tracking Concept Drift in Spam Filtering. In: Macintosh A., Ellis R., Allen T. (eds) Applications and Innovations in Intelligent Systems XII. SGAI 2004. Springer, London. https://doi.org/10.1007/1-84628-103-2_1

[15] S. K. Tuteja and N. Bogiri, "Email Spam filtering using BPNN classification algorithm," 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), 2016, pp. 915-919, doi: 10.1109/ICACDOT.2016.7877720.

[16] Song J., Lee S., Kim J. (2011) Spam Filtering in Twitter Using Sender-Receiver Relationship. In: Sommer R., Balzarotti D., Maier G. (eds) Recent Advances in Intrusion Detection. RAID 2011. Lecture Notes in Computer Science, vol 6961. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-23644-0_16

[17] Gray, Alan & Haahr, Mads. (2004). Personalised, Collaborative Spam Filtering.

[18] Bhuiyan, AkmAshiquzzaman, Tamanna Islam Juthi, Suzit Biswas, Jinat Ara, Hanif. " A Survey of Existing E-mail Spam Filtering Methods Considering Machine Learning Techniques." Global Journal of Computer Science and Technology [Online], (2018): n. pag. Web. 12 Jun. 2021

[19] [19] T. A. Almeida and A. Yamakami, "Content-based spam filtering," The 2010 International Joint Conference on Neural Networks (IJCNN), 2010, pp. 1-7, doi: 10.1109/IJCNN.2010.5596569.

[20] Anirban Dasgupta, Maxim Gurevich, and Kunal Punera. 2011. Enhanced email spam filtering through combining similarity graphs. In Proceedings of the fourth ACM international conference on Web search and data mining (WSDM '11). Association for Computing Machinery, New York, NY, USA, 785–794. DOI:https://doi.org/10.1145/1935826.1935929

[21] Gaurav, D., Tiwari, S.M., Goyal, A. et al. Machine intelligence-based algorithms for spam filtering on document labeling. Soft Comput 24, 9625–9638 (2020). https://doi.org/10.1007/s00500-019-04473-7

