JCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-6, Issue-5, May 2018 E-ISSN: 2347-2693

Analysis of Filtering Techniques for Spam Email Detection

A. Ahuja

Department of Computer Science, Guru Nanak Dev University, Amritsar, India

Corresponding Author: annieahuja89@gmail.com, Tel.: +91-8054915241

Available online at: www.ijcseonline.org

Accepted: 21/May/2018, Published: 31/May/2018

Abstract— Email is considered to be one of the most effective ways of communication source. It has gained attention because of the fastest and cost-effective means of source of communication. But with the enormous increase in its usage leads to its exploitation as it has become fascinated approach for the today's businesses. Email spam is the sending of unsolicited email in bulk to the randomly selected recipients for the purpose of advertising has become a serious concern. These unwanted emails not only occupy network bandwidth and memory space for communicating but can be used by the attackers in order to steal the user's identity. By looking at the prevailing scenarios there is a need for a solution that can manage the spam issue quite efficiently. The goal of this paper is to provide insight into an issue of spam email, and the highlight of this paper is the key findings of filtering techniques used for spam detection based on analysis of the content and non-content part of email.

Keywords- Spamming, Whitelist, Blacklist, Greylist , ham, CR systems, Heuristics, Signatures

I. Introduction

The ever-increasing growth in Internet provides an efficient way of communication in the form of email. Its wide use is seen in the last couple of years as adopted by both individuals and organizations and considered to be one of the reliable sources of communication. Internet revolutionized the way of how a person communicates with his/her family, friends and with the outside world. Today millions and billions of people are able to communicate with each other because of this powerful tool and the estimated count is 294 billion mails sending per day [1]. The alluring thing about this tool is that it is trustworthy source of information exchange, cost-effective and within no time the data/information gets communicated to the other person by sitting at your own place. The term 'spam' refers to the sending of unwanted emails in bulk [1].

Email has become a lucrative approach for advertisers which are the main reason behind the generation of spam emails. The email is considered to be spam when the intended user does not wish to receive. In spamming, sending email messages in bulk not only creates a problem for the email users but also for the Internet Service Providers (ISPs) as they are responsible for providing Internet services to the users as well as have a drastic impact on the usage of resources. Resources are more consumed as these mails are always sent in bulk and the affected resources include network bandwidth, memory usage and computational power, increase in investment by the companies leads to financial loss etc. It has now become a serious concern for the recipients to identify an email as spam or ham or legitimate email. It is a time-consuming task, firstly it involves the identification of email messages and then deleting all those spam messages which is annoying for the users as well as for the organizations. Sometimes it may happen that even legitimate emails comes to your spam box and spam emails like winning prize money, etc comes to your inbox. With the advancement in technology, spamming has many forms that cannot be easily detected by using only a single approach. Multiple approaches need to be inculcated in this scenario in order to get rid of this ever-increasing threat. The possible solution for this tremendously growing problem is to have spam filters that are capable enough in detecting spam emails and thus helps the users such that they do not receive any unwanted email. In the existing scenario, though it is not possible to eradicate this issue as a whole but can be reduced to a great extent by adopting the suitable filter classification techniques.

Anti-spammers play a key role in handling this issue of spamming by putting their efforts in this direction. With the evolving distinct types of spam filters in classifying the text as spam are available now. Open source spam filters are also available for analysing the frequency and classification of spam emails [2]. Email consists of subject that represents purpose of the email and body of the email represents its description. In spam filtering technique, the address of email, subject and description or content of the email message and it is generally assumed that distinction of spam and ham lies in the content of email message [3]. There are various approaches that can be used in detecting spam emails that includes network based approach, machine learning and non-machine learning approaches [3, 4]. In network based approach, some rules are used that needs to be updated timely with IP address and network address for classifying email as spam or legitimate. The results are reliable but the approach consumes a lot of time. The machine learning approach extracts the knowledge from email messages then use that extracted information further in the classification of newly received email messages. The most commonly used machine learning classifier for filtering emails is Naive Bayes Classifier [5]. In non-machine learning approach, several techniques are used which includes signature based schemes, heuristic approach, white-list, blacklist, grey-list, sandboxing, mail header scanning, etc [6].

It is observed that machine learning approach is more efficient than network based approach and non-machine learning approach [6, 7].

This paper reviews the basic terminology and discussed the need of the hour is to develop more and more intensive filtering techniques for detection of spam emails. The rest of the paper is structured as follows: Section II represents the filtering of spam emails at user level and enterprise level. Section III represents the working principle for selecting the best features from the data available in terms of header and content of email message. Section IV discusses the existing spamming techniques. Section V discusses the filtering spam techniques based on content and non-content analysis of email and how they are useful in reducing the spamming issue in existing and upcoming scenarios. Section VI contains the key findings in this research direction.

II. PROCESS OF SPAM FILTERING AT USER AND ENTERPRISE LEVEL

The process of filtering spam emails can be attained at user level or an enterprise level as represented in Figure 1.

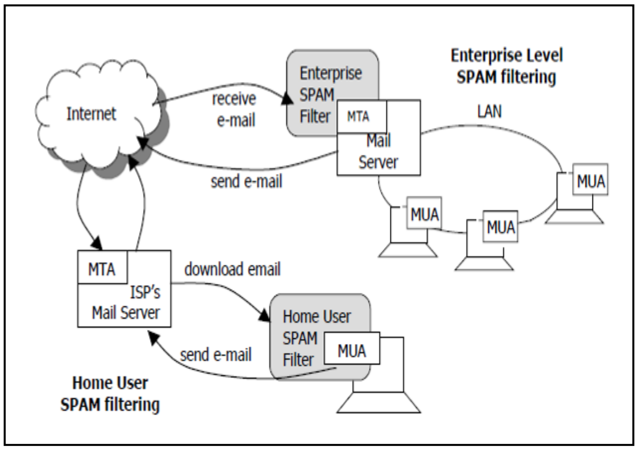


Figure 1. Spam filtering at User and Enterprise level [adapted from [5]]

At user level, if a person wants to filter the spam messages then he/she can install a spam filtering system on their PC. Sending and receiving of emails is possible for an individual by sitting at their own place with Internet Service Providers (ISPs) such as Bharat Sanchar Nigam Limited (BSNL), etc. The filtering system has the capability to either interface directly with Mail User Agent (MUA) or it can itself act as MUA for sending and receiving messages and managing emails. At user level, the other way for filtering spam mails is by using network i.e. Local Area Network (LAN). When email gets downloaded on the LAN, using the required filter system can help you to classify emails locally.

At enterprise level, filtering of mails is done when they enter internal operating network of an enterprise. The emails get classified with the interaction between spam filter which is installed on server side and Mail Transfer Agent (MTA). If the mail is identified as spam, then it is for all the users on that specific enterprise network.

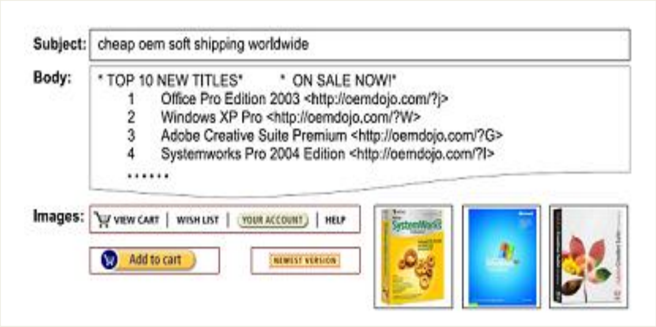
A. *Criterion for classifying emails used by current spam filtering systems*

The criterion used by most of the current filtering systems for classifying emails is rule-based scoring. A set of rules are matched against an incoming email, if score exceeds the desired limit i.e. threshold limit then it is considered to be spam email. There are hundreds of rules to be maintained and must be regularly updated as spammers are always trying to evade the rules by manipulating their earlier techniques of spamming.

III. WORKING PRINCIPLE

The working principle of these techniques lies in selecting the best features from the available data and then classifies email as spam or ham. There are two ways in order to carry out the selection of best feature process:

- *Header Based Selection* contains the email address of person who sent the email, Blind Carbon Copy (BCC) field that is used to send the copy of email message to other person but it is not known to receiver that the same copy has been sent to other person as well, Carbon Copy (CC) field used to send same message to other person but it is known to all the receivers that the same copy of message has been sent and to whom, To field contains receiver’s email id, From field contains sender’s email id and Subject states the purpose of sending email [7]. This is how best feature is selected from the header part of an email. Figure 2 representing header based selection of features for classifying emails as spam or legitimate.



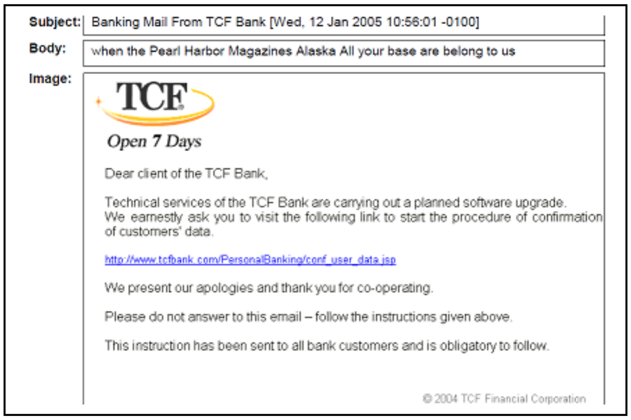
Subject: cheap oem soft shipping worldwide

Body:
* TOP 10 NEW TITLES* * ON SALE NOW!*
1 Office Pro Edition 2003 <http://oemdojo.com/?j>
2 Windows XP Pro <http://oemdojo.com/?W>
3 Adobe Creative Suite Premium <http://oemdojo.com/?G>
4 Systemworks Pro 2004 Edition <http://oemdojo.com/?I>
******

Images:

Figure 2. Text written in subject and body portion of an email message can be easily identified as spam by users simply by look up at the keywords used [adapted from [7]]

- *Content Based Selection* considers the content part of an email in order to classify an email. Figure 3 representing content based selection of features for classifying emails as spam or legitimate.



Subject: Banking Mail From TCF Bank [Wed, 12 Jan 2005 10:56:01 -0100]

Body: when the Pearl Harbor Magazines Alaska All your base are belong to us

Image:

Dear client of the TCF Bank,

Technical services of the TCF Bank are carrying out a planned software upgrade. We earnestly ask you to visit the following link to start the procedure of confirmation of customers' data.

http://www.tcfbank.com/PersonalBanking/conf_user_data.jsp

We present our apologies and thank you for co-operating.

Please do not answer to this email – follow the instructions given above.

This instruction has been sent to all bank customers and is obligatory to follow.

© 2004 TCF Financial Corporation

Figure 3. Spam email contains whole message text embedded in attached image and body field has bogus text [adapted from [7]]

The best feature is selected from the content part which is also body part of an email message. The content of message can be in different forms like text-based, audio, video, attachments, etc. Content Based Selection is proven to be more successful approach in terms of authentication in comparison to Header Based Selection approach which can easily by intruded by spammers or attackers [7]. The focus of this review article is in content based analysis of emails using different filtering techniques.

IV. SPAMMING TECHNIQUES

The users are able to receive unintended emails as the advertisers who can be marketers, pay certain amount for the matching of data that they have elicited with the external database that contains email addresses. Therefore, by using these means the marketers or other advertisers are able to send their advertisement in bulk to the random set of recipients [5].

The different spamming techniques are presented in Figure 4.

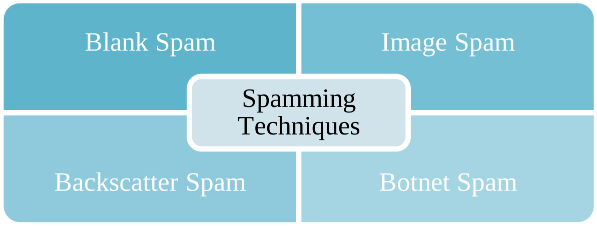
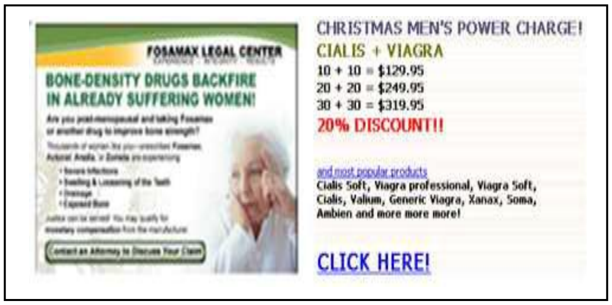


Figure 4. Different spamming techniques

A. *Image Spam*

In image spam technique, spammers used to incorporate text as an image which is then displayed in the email or can be received as an attachment, then it is not easily detected by text-based spam filters. The extension of stored image is generally GIF (Graphics Interchange Format) or JPEG (Joint Photographic Experts Group). With the advancing technology, new techniques are evolving for reading images in order to find embedded text but still results are not yet satisfied. Spammers have started using animated GIF images and obscuring content so as to evade detection of spam emails by OCR (Optical Character Recognition) tools. OCR plugins like Bayes OCR Plug-in, etc are provided by Spam Assassin which is used open source spam filter for detecting image spam [8]. The popular example of image spam used by spammers in mid 2000s was for the advertisement of "pump and dump" stocks [5].



FOSAMAX LEGAL CENTER
BONE-DENSITY DRUGS BACKFIRE IN ALREADY SUFFERING WOMEN!
Contact an Attorney to Discuss Your Claim
CHRISTMAS MEN'S POWER CHARGE!
CIALIS + VIAGRA
10 + 10 = $129.95
20 + 20 = $249.95
30 + 30 = $319.95
20% DISCOUNT!!
and most popular products
Cialis Soft, Viagra professional, Viagra Soft, Cialis, Valium, Generic Viagra, Xanax, Soma, Ambien and more more more!
CLICK HERE!

Figure 5. Spam images [adapted from [1]]

B. *Blank Spam*

In blank spamming, subject and body of email is generally missing but still regarded as spam mail as it is sent in bulk.
The purpose of generating blank spam is to gather information regarding valid email addresses from the email service providers in the form of dictionary based attack [5]. Spam email is blank when spammer is unable to fill the subject and body message of an email while spam set up runs. Sometimes, you might have seen that the received email in your inbox has truncated headers, this is also a blank spam may be due to poorly-written software. Sometimes, you have an illusion of spam email as blank while it is not the case. For example, VBS.Davinia.B email worm that spreads

through email messages with no subject line and uses HTML code for downloading other files [5].

C. Backscatter Spam

In backscatter spam that is generated as a result of email spam, viruses and worms, in which bounce messages are send to an innocent one by the email servers. In other words, when email is sent by sender, he/she gets informed regarding the status of sent email by email servers whether it is delivered successfully or not in terms of messages called Delivery Status Notification (DSN) such as email could not be delivered, etc. These messages are quite helpful to senders in a way that they communicate the senders mistake in terms of incorrect filling of receivers email address, etc. But the problem occurs when backscatter spam occurs. This happens as the sender's message envelope is copied fradulently so as to with held email address of the victim. Backscatter are Delivery Status Notifications (DSN) from another server rather than the intended one [2]. It is sent in bulk with a fake From: header, that matches with original sender's envelope. Backscatter spam emails generated by the systems are listed on various Domain Name System Blacklists (DNSBLs) and also in violation terms and services of Internet Service Providers (ISPs). It comes out in variety of forms like you get the request from mailing list- "please confirm your subscription", challenge requests, etc. Servers that are responsible for generating backscatter spams get included in DNSBLs. Email servers must be properly configured in order to reduce this spamming to a great extent. The results concluded using open source filters states that 1% of test set of 49,086 messages are in the category of backscatter spam [2].

D. Botnet Spam

With the evolving anti-spamming techniques, spammers are forced to think of developing new strategies so as to overcome with their reduced profits. The solution opted by spammers who are working at large scale collaborated with virus and exploit coders so that they can get control of bots or zombies on the Internet. 'Botnets' refers to set of machines that are controlled by 'botmaster' [8]. A bot when maliciously attacked can be made to send spam emails, malware, and used to harvest password which is critically vital part of defense mechanism that keeps your account safe, login details and also users are re-routed to spoofed websites or leads to generation of new bots and so on. Botnets pose a major threat to Internet infrastructure as they are capable enough to make Denial of Service (DoS) attacks on servers, send out unwanted emails in bulk; can break weak passwords, etc. Shadow server, a public tracker conducted a survey that the count of active zombies or bots is at least one million and still constantly increasing [8].

CISCO reported that one of the biggest security concerns on the Internet today is Botnet. The mechanism used by botnets is to do spamming but not at such a large-scale so that they evade detection [8]. For example, Grum botnet had only 600,000 members before taking it down [8]. The challenge is to detect botnets as they have devastating impact on the cyber-security front due to the provision of providing distributed platforms for illegal and exploiting activities such as phishing, click fraud, etc. Due to its great potential for being a security threat, single technology is not enough to deal with it.

V. FILTERING TECHNIQUES

The various filtering techniques that are used for detecting spam emails by analyzing the content and non-content part of an email are represented in Figure 6.

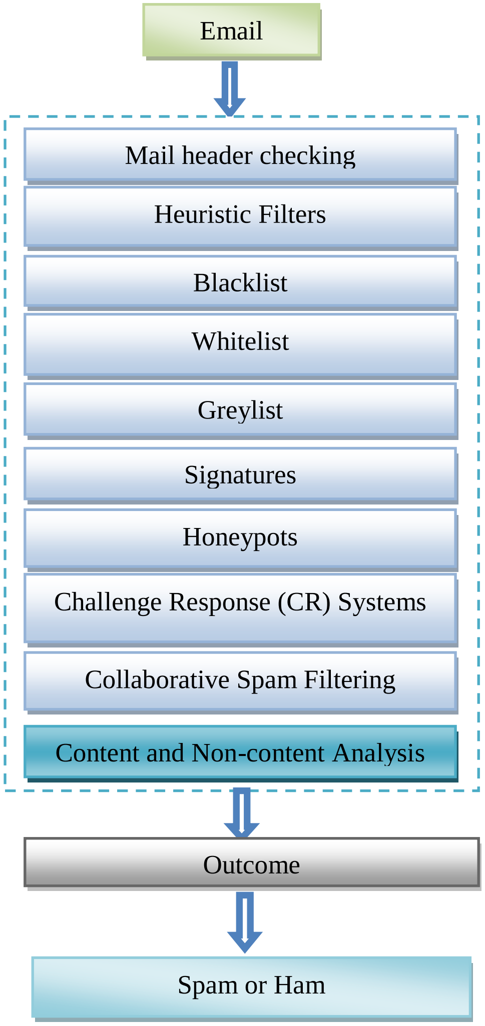


Figure 6. On applying different filtering techniques to incoming email, outcome is either spam or ham

A. Mail header checking:

In mail header analysis, some rules are laid down for matching with mail headers. If the outcome is positive i.e. mail header matches with the set of framed rules then server gets invoked which return mails with empty From field, conflicts in subject field, etc [1]. We can rely on email headers as it is one of the powerful sources that possess discriminative features to be used in filtering of spam emails other than subject field and body content. Basically, it determines the receiver of a message and maintains the routing path which the email message takes as it passes through each mail server.

The statistical analysis presented that 92.5% of emails were filtered out, out of 10,024 junk emails using mail header analysis of features include message ID, Mail User Agent (MUA), sender's address, recipient's address, etc [9]. The evaluation of performance of several header based spam filtering techniques is performed by Hayat, Basiri, Seyedhossein, Shakery in 2010 and by Al-jarrah, Khater, Al-duwairi in 2012 [10, 11].

B. Heuristic Filters:

Heuristic filters also use set of coded rules termed as heuristics for filtering the emails as spam or ham [12]. Heuristic approach that is based on content observes the content of email and then put it in the category of spam or legitimate based on the occurrence of words like 'lottery', 'prize winning', etc. In order to prevent these commonly used lucrative words from detection by spam filters, spammers made the obscuring content like 'l*o*T*T*e*E*R*Y instead of writing 'lottery', can be done in other ways too. Commonly used words by spammers are represented in Figure 7. A solution is provided by Sanz in 2008 that uses rule-based filter approach and even it is capable of tracking IP addresses that are behind spamming [8].

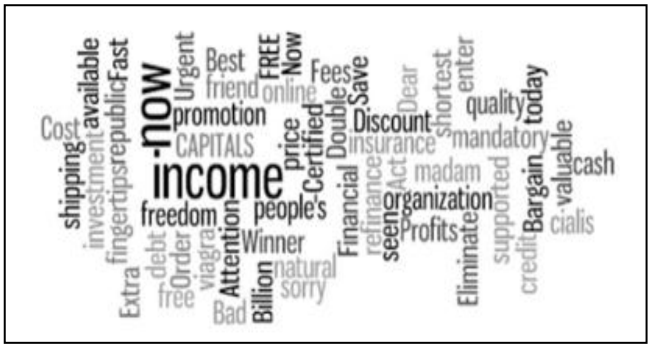


Figure 7. Commonly used words in spam emails [adapted from [8]]

C. Blacklist:

Blacklist is a list that is maintained at user level or server level and keeps the email addresses or IP addresses that are responsible for sending spam emails and prevent that email to come at client's inbox [1, 6]. Whenever email is received by user from these blacklisted addresses, it gets blocked at SMTP connection phase. Real-time Blackhole Lists (RBL) and Domain Name System Black Lists (DNSBLs) are part of blacklist [8]. Commonly, the database of black list contains network addresses, proxies, individual addresses responsible for sending spam. Examples include Google blacklists and SpamHaus [8].

D. Whitelist:

Whitelist is a list of email addresses or IP addresses that are authenticated and approved contacts or domains from which the user can receive an email [1, 6]. The addresses which are not part of whitelist get blocked whether it is spam or ham message using this filtering technique.

E. Greylist:

In greylist technique, whenever email is sent by sender, the receiver's server checks whether the address resides in the blacklist or white-list depending upon that it classifies email as spam or ham [8]. The message is rejected temporarily if the address does not identified in white-list or blacklist and then Mail Transfer Agent (MTA) of receiver side gives a response with an SMTP temporary error message but MTA then starts recording and analysing recent attempts from sender side and database gets updated with that client's information. In the next attempt, if the sender is legitimate it might be possible that it is accepted for sending mail. In this method, there is an assumption that whenever spammers try to resend their messages they get blocked in public blacklist i.e. DNSBLs while attempting two times the same scenario.

F. Signatures:

In signature technique of filtering spam emails, the hash values of previously identified spam email are maintained in database at MTA level. In order to classify an incoming email message, previous stored hash values are matched against the new incoming email as signatures possess unique identity and able to provide exact pattern matching.

G. Honey pots:

The purpose of honey pot server or system lies in collecting spam emails and elicit information regarding intruders or attackers [13, 14]. It is for content based spam filters which are based on fingerprints.

H. Challenge Response (CR) systems:

In case of white-list the authentication of addresses is done on receiver side but in CR systems, addresses authentication takes place on sender side. Whenever sender sends an email to receiver, it has to deal with challenge like identifying and then marking cars in a view received from Mail Transfer Agent (MTA). The sender's response determines the receiving of an email to recipient. If the response is satisfactory then only mail gets received by recipient otherwise get deleted or goes to spam folder.

I. Collaborative Spam Filtering:

It is a collaborative spam email filtering technique that gathers the spam email related information like subject, sender, mathematical function over message body, etc. rather than content of emails. The previous researchers or receivers share the digital footprints of spam messages which form the basis for classifying messages as spam or ham. Examples of Collaborative Spam filtering techniques are Vipuls Razor, Pyzor, and Distributed Checksum Clearinghouse (DCC) on the web [8].

VI. RELATED WORK

Table 1. Represents Filtering Techniques for spam detection with their descriptions in terms of Classification Principle, Findings and Disadvantages [1, 6, 8, 15]

Filtering Techniques for Spam Detection
A. Mail header checking • *Classification Principle*- Matching headers with the set of framed rules • *Findings*- Low false positive rates with the careful framing of rules • *Disadvantages*- Sometimes legitimate emails headers are matched with the set of rules and then it is considered as spam and not ham.
B. Heuristic Filters • *Classification Principle*- Based on set of coded rules or heuristics • *Findings*- Not as satisfactory as difficult to interpret obscuring content prior to rule based solution in 2008, afterwards performed competitively to Naïve-Bayes anti-spamming solution. • *Disadvantages*- The rules need to be regularly updated as spammers keep on employing new techniques. Therefore, it is time-consuming approach.
C. Blacklist • *Classification Principle*- Maintain a blacklist of email addresses and IP addresses • *Findings*- Computational cost is less as it needs only to look into blacklist whenever email comes. • *Disadvantages*- Blacklist need to be always up-to-date otherwise it can lead to vulnerabilities. Sometimes, valid addresses get blocked by filter by mistake or arbitrarily. Victims are users like us and domains like Hotmail might get blocked when used by spammers without even asking permission from their owners [8].
D. Whitelist • *Classification Principle*- Maintain a white-list of email addresses and IP addresses • *Findings*- High false positive rate • *Disadvantages*- Hinders the establishing of new contacts as the legitimate email messages also get blocked if they are not part of white-list. It is easy for spammers to avoid the filtering mechanism by imitating the addresses or can also use existing well-known mail white-list. As a consequence, provides average filtering rate and needs to be carefully maintained.
E. Greylist • *Classification Principle*- If sender's address is not present in blacklist or whitelist then it is temporarily rejected then again decided whether to accept or reject the sender's message by analysing sender's attempt. • *Findings*- Effective technique • *Disadvantages*- Zombies can be used by spammers for retrying for spamming.
F. Signatures • *Classification Principle*- Signatures i.e. hash values • *Findings*- It provides detection of known spam emails quite efficiently. • *Disadvantages*- New spam emails can easily pass through this filter technique. Moreover, the database keeping signatures need to be updated on hourly, once a day, or weekly basis. For generating different hash values, random string is introduced into spam messages by the spammers.
G. Honey pots • *Classification Principle*- Fingerprint based technique for content based spam filtering • *Findings*- Provides you insight details that helps security professionals in learning the techniques used by spammers or attackers. • *Disadvantages*- Not used in non-content based spam emails.
H. Challenge Response (CR) systems • *Classification Principle*-Authentication is performed at sender's side • *Findings*- Effective technique in identifying spam emails from automated systems or botnets. • *Disadvantages*- The encountered delay in the delivery process in terms of communication overhead leads to inconvenience. Moreover, ham mails can also be blocked if they fail in challenge given by MTA. The reason behind backscatter email spam can be these systems.

I. Collaborative Spam Filtering

- *Classification Principle*- Spam Fingerprints
- *Findings*- Collaborative techniques give more promising results.
- *Disadvantages*- Scalability issues, working as a collaborative team itself is also a challenge

In the Table 1, the filtering techniques which are mainly focussed on email header and its content for spam email detection are evaluated with the parameters- Classification Principle, Findings, and Disadvantages. Each and every filtering technique has its own classification mechanism for classifying incoming emails into spam or ham. The evaluation is theoretical key findings by the author by studying the literature review. All the filtering techniques have both sides like the two sides of a coin. One side represents advantages and the other one disadvantages. In order to utilize their full potential in detecting spam emails, need of the hour is to carefully analyze their disadvantages and firstly try to overcome those limitations up to certain extent for better yielding of results.

VII. Conclusion and Future Scope

This paper provides an insightful study of the process of filtering spam emails at user and an enterprise level. The different spamming techniques and how they are exploiting Internet infrastructure is highlighted. Then the different filtering spam detection approaches are discussed in detail with respect to header and content based analysis of emails.

It is observed that among all the spamming techniques- image spam, blank spam, backscatter spam and botnet spam, the backscatter spam emails count is negligible as compared to image spam. Botnet spam has the potential to exploit the security. Image spam has been widely used technique for spamming. It is observed by studying literature that most of the researchers have utilized widely used commercial and open source- 'Spam Assassin'[8] for filtering spam emails but the experimental results most of the times are on machine learning filtering spam detection approaches and are based on content analysis of email. Non-machine learning approaches, never gained attention as they are less efficient than machine learning approaches. Therefore, author has tried to contribute towards non-machine learning approaches by analyzing content and non-content part of email for detecting spam emails. Spamming has become a serious concern as it poses threat on security of data, so the need of the hour is to do further research in this direction that will surely give promising results.

Acknowledgment

I would like to thank my colleagues for their invaluable guidance.

References

[1] Rekha, S. Negi, " *A Review on Different Spam Detection Approaches",* International Journal of Emerging Trends and Technology , Vol. 11, No.6, pp. 315-318, 2014.

[2] G. V. Cormack, T. R. Lynam, "*On-line Supervised Spam Filter Evaluation*" , ACM Transactions on Information Systems (TOIS), Vol.25, No.3, 2007.

[3] P. Sharma, U. Bhardwaj, " *Machine Learning Based Spam E-Mail Detection*" , International Journal of Intelligent Engineering & Systems , Vol.11, No.3, pp. 1-10, 2018.

[4] R. Bansod, R. S. Mangrulkar, V.G.Bhujade, "*Text and Image based Spam Email Classification using an ANN Model- an Approach",* International Journal on Recent and Innovation Trends in Computing and Communication , Vol. 3, No. 5, pp.115-118, 2015.

[5] O. Saad, A. Darwish, R. Faraj, "*A Survey of machine learning techniques for Spam Filtering*" , International Journal of Computer Science and Network Security, Vol.12, No.2, pp.66-73, 2012.

[6] H. Kaur, P. Verma, "*SURVEY OF E-MAIL SPAM DETECTION USING SUPERVISED APPROACH WITH FEATURE SELECTION",* INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY (IJESRT) , Vol.6, No. 4, pp. 1-10, 2017.

[7] G. Fumera, I. Pillai, F. Roli, " *Spam Filtering Based On The Analysis Of Text Information Embedded Into Images*" , Journal of Machine Learning Research, Vol.7, pp.2699-2720, 2006.

[8] A. Bhowmick, S. M. Hazarika, " *Advances in Electronics,Communication and Computing*", Springer Publication, Singapore, pp. 573-582, 2016

[9] C.C. Wang, S.Y. Chen, '*Using Header Session Messages to Anti-Spamming'*, Computers & Security, Vol.26, No.5, pp. 381-390, 2007.

[10] M. Z. Hayat, J. Basiri, L. Seyedhossein, A. Shakery, "Content-Based Concept Drift Detection for Email Spam Filtering", In the Proceedings of 2010 5th International Symposium on TeleCommunications (IST'2010), pp. 531-536, 2010.

[11] O. Al-jarrah, I. Khater, B. Al-duwairi, "Identifying Potentially Useful Email Header Features for Email Spam Filtering", In the Proceedings of The Sixth International Conference on Digital Society, pp. 140-145, 2012.

[12] E. P. Sanz, "*E-mail Spam Filtering*", Advances in Computers, Vol. 74, pp. 45-114, 2008.

[13] M. Andreolini, A. Bulgarelli, M. Colajanni, F. Mazzoni, "*Honeyspam : Honeypots Fighting Spam at the Source*", In the Proceedings of 2005 the Steps to Reducing Unwanted Traffic on the internet Workshop, Cambridge, MA, pp. 77-83, 2005.

[14] M. Dagar, R. Popli, "*Honeypots: Virtual Network Intrusion Monitoring System*", International Journal of Scientific Research in Network Security and Communication, Vol. 6, No. 2, pp.45-49, 2018.

[15] D. Mallampati, "*An Efficient Spam Filtering using Supervised Machine Learning Techniques*", International Journal of Scientific Research in Computer Sciences and Engineering", Vol. 6, No. 2, pp.33-37, 2018.

Authors Profile

Ms. A Ahuja pursued Bachelor of Computer Appli-Cations from Guru Nanak Dev University, Amritsar In 2011 and Master of Computer Applications from Punjab Technical University in 2014. She is currently Working as Assistant Professor in the Department of Computer Science, Guru Nanak Dev University, Amritsar. She is IBM DB2 Academic Associate. She has 4 years of teaching experience.