A Study on Email Spam Filtering Techniques

Rahul Chandra
17BCS1811
CSE-6(GROUP B
CHANDIGARH UNIVERSITY

Arindam Bagchi
17BCS1780
CSE-6(GROUP A)
CHANDIGARH UNIVERSITY

Suraj Sachan
17BCS1769
CSE-6(GROUP A)
CHANDIGARH UNIVERSITY

ABSTRACT
Electronic mail is used daily by millions of people to communicate around the globe and is a mission-critical application for many businesses. Over the last decade, unsolicited bulk email has become a major problem for email users. An overwhelming amount of spam is flowing into users‟ mailboxes daily. Not only is spam frustrating for most email users, it strains the IT infrastructure of organizations and costs businesses billions of dollars in lost productivity. The necessity of effective spam filters increases. In this paper, we presented our study on various problems associated with spam and spam filtering methods, techniques.

General Terms
Spam, spam filtering

Keywords
Email, spam, spam filtering

1. INTRODUCTION
The internet has become an integral part of everyday life and e-mail has become a powerful tool for information exchange. Along with the growth of the Internet and e-mail, there has been a dramatic growth in spam in recent years. Spam can originate from any location across the globe where Internet access is available. Despite the development of anti-spam services and technologies, the number of spam messages continues to increase rapidly. In order to address the growing problem, each organization must analyze the tools available to determine how best to counter spam in its environment. Tools, such as the corporate e-mail system, e-mail filtering gateways, contracted anti-spam services, and end-user training, provide an important arsenal for any organization. However, users cannot avoid the very serious problem of attempting to deal with large amounts of spam on a regular basis. If there are no anti spam activities, spam will inundate network systems, kill employee productivity, steal bandwidth, and still be there tomorrow.

2. SPAM – UNSOLICITED BULK EMAIL
E-mail spam, known as unsolicited bulk Email (UBE), junk mail, or unsolicited commercial email (UCE), is the practice of sending unwanted e-mail messages, frequently with commercial content, in large quantities to an indiscriminate set of recipients. The technical definition of spam is „An electronic message is "spam" if (A) the recipient‟s personal identity and context are irrelevant because the message is equally applicable to many other potential recipients; and (B) the recipient has not verifiably granted deliberate, explicit, and still-revocable permission for it to be sent‟. The risks in filtering spam are sometimes legitimate mails may be rejected or denied and legitimate mails may be marked as spam. The risks of not filtering spam are the constant flood of spam clogs networks and adversely impacts user inboxes, but also drain valuable resources such as bandwidth and storage capacity, productivity loss and interfere with the expedient delivery of legitimate emails. General Advice to avoid spam is, Avoid giving your “real” email address to all but close associates, Setup web mail accounts (Google, hotmail etc.) for registering with web sites or for communicating with people you do not know, Educate your contacts to exercise caution with email address, Do not open junk email, just delete it (note that auto preview is the same as opening), Never click to unsubscribe to a mailing unless you are sure it is a reputable entity.

3. SPAM FILTER ARCHITECTURE
Spam filters can be implemented at all layers, firewalls exist in front of email server or at MTA(Mail Transfer Agent), Email Server to provide an integrated Anti-Spam and Anti-Virus solution offering complete email protection at the network perimeter level, before unwanted or potentially dangerous email reaches the network. At MDA (Mail Delivery Agent) level also spam filters can be installed as a service to all of their customers. At Email client user can have personalized spam filters that then automatically filter mail according to the chosen criteria. Figure 1. shows the typical architecture of spam filter.

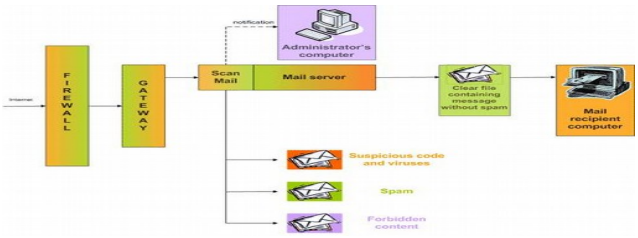

Figure 1. Spam Filter Architecture

4. SPAM IDENTIFICATION METHODS
The several different methods to identify incoming messages as spam are, Whitelist/Blacklist, Bayesian analysis, Mail header analysis, Keyword checking.

- ***Whitelists/Blacklists***

The functionality of these filters is simple: a whitelist is a list, which includes all addresses from which we always wish to receive mail. we can add email addresses or entire domains, or

functional domains. An interesting option is an automatic whitelist management tool that eliminates the need for administrators to manually input approved addresses on the whitelist and ensures that mail from particular senders or domains are never flagged as spam. The number of records can be configured. When an overflow occurs, obsolete records are overwritten. A blacklist works similarly to competitive alternatives: this is a list of addresses from which we never want to receive mail.

- **Mail header checking**

This is a fairly known method. Mail header checking consists of a set of rules that, if a mail header matches, triggers the mail server to return messages that have blank "From" field, that lists a lot of addresses in the "To" from the same source, that have too many digits in email addresses (a fairly popular method of generating false addresses). It also enables to return messages by matching the language code declared in the header.

- **Bayesian analysis**

The word probabilities (also known as likelihood functions) are used to compute the probability that an email with a particular set of words in it belongs to either category. This contribution is called the posterior probability and is computed using Bayes' theorem. Then, the email's spam probability is computed over all words in the email, and if the total exceeds a certain threshold (say 95%), the filter will mark the email as a spam.

- **Keyword checking**

Another method widely used in filtering spam. It works by scanning both email subject and body. Using "conditions" i.e. combinations of keywords is a good solution to enhance filtering efficiency. We can specify combinations of words and update the list that must appear in the spam email. All messages that include these words will be blocked.

5. SPAM FILTERING TECHNIQUES

The various spam filtering techniques adopted to get rid of the problem of spam are discussed.

Distributed adaptive blacklists: This technique can be used at the mailserver. When a message is received by a MTA, a distributed blacklist filter is called to determine whether the message is a known spam. These tools use clever statistical techniques for creating digests. Tools like Razor and Pyzor operate around servers that store digests of known spams.

Rule based filtering: Evaluate a large number of patterns--mostly regular expressions--against a candidate message. Some matched patterns add to a message's score, while others subtract from it. If a message's score exceeds a certain threshold, it is filtered as spam; otherwise it is considered as legitimate. Some ranking rules are fairly constant over time. Other rules need to be updated as the products and scam advanced by spammers evolves. SpamAssassin is one of the popular rule based spam filtering tool.

Bayesian classifier: Particular words have particular probabilities of occurring in spam email and in legitimate email. The filter doesn't know these probabilities in advance, and must first be trained so it can build them up. After training, the word probabilities (also known as likelihood functions) are used to compute the probability that an email with a particular set of words in it belongs to either category. Each word in the email contributes to the email's spam probability, or only the most interesting words. This contribution is called the posterior probability and is computed using Bayes' theorem. Then, the email's spam probability is computed over all words in the email, and if the total exceeds a certain threshold (say 95%), the filter will mark the email as a spam. Some spam filters combine the results of both Bayesian spam filtering and other heuristics (pre-defined rules about the contents, looking at the message's envelope, etc.), resulting in even higher filtering accuracy, sometimes at the cost of adaptiveness. Server-side email filters, such as DSPAM, SpamAssassin, SpamBayes, Bogofilter and ASSP, make use of Bayesian spam filtering techniques.

K **nearest neighbors:** If at least *t* messages in *k* neighbors of the message *m* are unsolicited, then *m* is unsolicited email, otherwise, it is legitimate. The tool TiMBL uses k nearest neighbour technique.

Support vector machine (SVM): It can be used to classify spam emails. It is assume that we are in a hyperspace of *n* dimensions, and that the training sample is a set of points in the hyper-space. In the case of spam problem it is of just two classes. The classification using Support vector machine look for the hyper plane able to separate the points of the first class from those of the second one such that the distance between the hyper plane and points of each class is maximum.

Content based Spam Filtering Techniques - Neural Networks: The neural networks are quite famous to be well adapted for problems of classification. Without being spread out over the model, we will retain in what follows the characteristics which contribute to the design of an antispam filter. Spams filtering and if one makes a point of applying the technique of the perceptron, it is enough to choose a characteristic vector larger than that of the training sample to ensure the convergence. However such practice will heavily weigh down the computation.

The multi-layer networks: As its name indicates, the multi-layer neural net is a network of connected perceptrons which form a network with successive layers. The outputs of each perceptron are inputs of perceptrons of the following layer. The inputs of the neurons of the first layer are the components of the characteristic vector, while the outputs of the last layer are the results of the classification.

Technique of search engines: When it acts on text e-mails, classification techniques of text seem to be efficient. However, spammers do not cease to invent tricks to circumvent filters. One

of these tricks is to include in the body of the message only the hyperlink to a Web page which contains the advertising text. The problem become then a web content classification. A proposed technique to overcome this kind of spams is to use the public search engines which offer a mean to classify the websites. The principle of this technique is to analyze automatically the contents of the pages referred by the links sent in the messages likely to be spams.

Technique of genetic engineering: In the design of a bayesian filter, the characteristic vector may include the frequencies of some words generally selected by human experts. In fact, this construction is sometimes decisive in the performances of the filter. In Hooman proposes a method to build automatically the bayesian filter. This method is based on the genetic programming. Thus, the frequencies of a word in E-mail can argument the classification of the message as unsolicited. As genetic programming, the filter is represented by a syntactic tree where nodes are numbers that represent the frequencies, operations on numbers, words and operations on words. A syntactic tree of a filter should be built according to a precise syntax. Syntactic rules then can be used to check the correctness of the tree by checking whether we are able to reduce the tree to some number.

Technique of artificial immune system: Anti-spams filter based on the generation of artificial lymphocytes using gene database. Genes are regular expressions which represent mini-languages likely to contain keywords that are usually checked in spam. The use of the regular expressions aims according to the author at increasing the accuracy as well as the general information hold in the detecting lymphocytes. The generation of lymphocytes is based on a training sample. The lifespan of these lymphocytes can be tuned in order to ensure the system dynamicity.

Above all, in spam filtering, False negatives just mean that some spam mails are classified as legitimate and moved to inbox. False positive mean that legitimate emails that get mistakenly identified as spam and moved to spam folder or discarded. For most users, missing legitimate email is an order of magnitude worse than receiving spam, so spam filters that yields less % of false positives are called as effective spam filters.

6. CONCLUSION

Spam or unsolicited e-mail has become a major problem for companies and private users. This paper explored the various problems associated with spam and different methods and techniques attempting to deal with it. From the study we identified that, many of the filtering techniques are based on text categorization methods and there is no technique can claim to provide an ideal solution with 0% false positive and 0% false negative. There is lot of scope for research in classifying text messages as well as multimedia messages.

7. REFERENCES

[1] Ahmed Khorsi, “An Overview of Content-Based Spam Filtering Techniques”, Informatics 31 (2007) 269-277 269

[2] David Mertz, “Comparing a Half-Dozen Approaches to Eliminating Unwanted Email”, August 2002

[3] Aditya Bakshi,”Github Link for the Project”,May 2016

[4] Analytic Mind,”Buid a Spam Filter Blog from Machine Learning Expert Yuva”,February 2019-04-10

[5] Medium,”Building a Spam Filter from Scratch Using Machine Learning—Machine Learning Easy and Fun”, September 2019

[6] MIT,”CS430 Class Spam Filter”,June 2019

[7] Spam Filter Blog from the KD Nuggets,August 2019