E-Mail Security Using Spam Mail Detection and Filtering System

[1]K. Ravikumar, [2]P. Gandhimathi
[1]Asst.professor, Dept.of.Computer Science, Tamil University, Thanjavur-613010
[2]Research Scholar, Dept.of.Computer Science, Tamil University, Thanjavur-613010

Abstract: Electronic mail, also known as email or e-mail, is a method of exchanging digital messages from an author to one or more recipients. Email is the most efficient way to communicate or transfer our data from one to another. While transferring or communicating through email there is the possibility of misbehave. In the existing system Spam method is used to avoid the unwanted Email receiving. Email spam, also known as unsolicited bulk Email (UBE), junk mail, or unsolicited commercial email (UCE), is the practice of sending unwanted email messages, frequently with commercial content, in large quantities to an indiscriminate set of recipients. But in Spam method there is no way to prevent the unwanted messages or Email receiving. To solve these unwanted messages or Email receiving we propose the concept Email misbehave blocking system. In the proposed method we permanently prevent the incoming unwanted messages or Email through blocking system.

Keywords: *Email, Digital, UCE, Spam*

*****

I. INTRODUCTION

An e-mail is considered “spam” when a massive number of them are sent to multiple recipients. Spam email is usually used for advertisement or marketing. These unwanted emails cause drawbacks to the recipient, and consume the users’ network resources. The disadvantages of spam emails have been addressed in many occasions. In some cases for a single user 9 out of 10 emails are spams that fill his/her inbox. The United States Federal Trade Commission described that 66% of spams have false information somewhere in the message and 18% of spams advertise “Adult” material. According to another report 12% of users spend half hour or more per day dealing with spam emails. There are several major problems with spam mails. First of all, they are high in volume and fill in mailbox of users. Secondly, there is no correlation between receivers’ area of interests and the contents of spam mails. Thirdly, they cost money for ISPs because the bandwidth and the memory of system are wasted. Finally, Spam e-mails cause a lot of security problems because most of them include Trojan, Malwares, and viruses. Many filtering techniques have been developed to control the flow of spam emails. Unfortunately, even with these available techniques, the number of spam emails is growing and the flow has not been controlled completely. The setback is that there is no actual solution because a spammer; an unidentified user with enough knowledge is able to be familiar with the logic of the filtering mechanisms. As a result, bypassing the filter and sending the spam seems not to be a difficult task for such spammers. In such cases, the spam emails are not detected and are considered as legitimate ones.

There are studies regarding spam email filtering. The common issue with the usage of all of these techniques is that the filtering systems are set up in the receiver mail server, consequently, causing network load and wasting network resources. To preserve network resources such as bandwidth and memory, and to reduce network load, this paper proposes to locate spam email filtering in the sender mail server rather than the receiver mail server.

Moreover, this paper by experimental results shows that this novel approach works more efficiently compared with the previously proposed approaches. This paper is organized as follows. In section 2, the related work to the subject will be highlighted. The Overview of email system and its operation are described in. Our proposal and the experiment results are presented in respectively.

As stated before, there are many filtering techniques to stop the flow of spam emails to mail boxes. Figure 1 simply illustrates the classification of spam email filtering techniques. The classification includes list-based filtering, static algorithm, and IP-based filtering. The list-based filtering is classified into three categories; Blacklist, Whitelist, and Greylist. Static algorithm is classified into content-based, and the rulebased filtering. Finally, IP-based filtering consists of revers-lookup. In the Blacklist filtering, the IP address and the domain name of the sender server is stored in a list called Blacklist and the emails from that IP address and domain are blocked. Then, based on the policy of the receiver side, the emails from the Blacklisted IP addresses are deleted or sent to spam folder. Conversely, there are some limitations for the Blacklist filtering. First, since the spammer uses several IP addresses with a variety of domain names, updating these lists is a difficult task for the client. Consequently, updating the Blacklist regularly is costly.

Second, Blacklist filtering may result in identification of an email as false negative because of minimal control in this methodology. On the other side of the Blacklist, is the Whitelist filtering. In this technique, any user stores his/her email contacts in a list called the

Whitelist. Therefore, any received email with the correspondent address from this list is accepted, and all other addresses out of this list are considered uncertain. In this technique, also there are certain obstacles. The obvious one is that, since the sender is unidentified and unpredictable, it is difficult to insert all possible sender addresses in this list. Similar to the Blacklist, the Whitelist filtering needs to be updated regularly; which is a costly task for the user. Another major issue is that if the email address of a spammer is added in the Whitelist of an email client once, this will provide access to all of the addresses in the Whitelist of that specific client without any boundaries or limits. As a result, this will ensure the spammer more reachable email addresses.

II. PROBLEM DEFINITION

An e-mail is considered "spam" when a massive number of them are sent to multiple recipients. Spam email is usually used for advertisement or marketing. These unwanted emails cause drawbacks to the recipient, and consume the users' network resources. The disadvantages of spam emails have been addressed in many occasions. In some cases for a single user 9 out of 10 emails are spams that fill his/her inbox. The United States Federal Trade Commission described that 66% of spams have false information somewhere in the message and 18% of spams advertise "Adult" material. According to another report 12% of users spend half hour or more per day dealing with spam emails. There are several major problems with spam mails. First of all, they are high in volume and fill in mailbox of users. Secondly, there is no correlation between receivers' area of interests and the contents of spam mails. Thirdly, they cost money for ISPs because the bandwidth and the memory of system are wasted. Finally, Spam e-mails cause a lot of security problems because most of them include Trojan, Malwares, and viruses. Many filtering techniques have been developed to control the flow of spam emails.

In recent years Email spam is sent via "zombie networks", from personal computers in homes and offices around the globe. Detecting spam based on the content of the email, either by detecting keywords or by statistical means i.e., content or non-content based, is widely used technique to find spam messages. Content based statistical means or detecting keywords can be very accurate when they are correctly tuned to the types of legitimate email that an individual gets. The content also doesn't determine whether the email was either unsolicited or bulk, the two key features of spam. Non-content base statistical means can help lower false positives because it looks at statistical means vs. blocking based on content/keywords.

- We can only move the incoming spam messages to the spam folder but we can't prevent receiving spam mails.
- It occupies the more mail memory wastage by receiving this kind of spammed mails.
- We need to delete the unwanted spam messages manually; it is time consuming process to the users and it takes the request or response service from the email server.
- By this method there is the possibility of receiving virus or warm emails from the spammer.

We propose an Email spam blocking system, which blocks the incoming spam or unwanted messages. In this method users can able to prevent the spam messages entering into their inbox or spam box. Since the unwanted emails are blocked from the spammer. So we can save the mail memories, because of mail memories are limited we need to save our memory. We are not in need to view our Spam box. In case of emergency communication the blocked person email can be unblocked by the recipient user who blocked the spammer. This emergency communication is possible only once for a blocked account. If they are unblocked that particular account they can communicate frequently like normal user, if they are misusing the emergency communication then the same spammer account can be blocked again by then the spammer cannot communicate with the recipient in future.

ADVANTAGES:

- By this method we can permanently block the receiving messages.
- There is no mail memory wastage for the unwanted emails, because the unwanted emails are blocked.
- There is no need of opening the spam folder so we can save our time.
- By this method there is no way of receiving virus or warm emails from the spammer.

III. CONCLUSION

A content based classification of spam mails with fuzzy word ranking. There are many classifiers and filters available for classifying and filtering spam mails. This study analyzed the previous related works. The proposed work used two sets of linguistic terms for ranking and classifying spam mails. This method has extracted only the features from the content of an email instead of extracting all the features from the mail. The actual words are extracted from the inbox of an email are compared with a list of spam words in the database and the words are categorized according to its rank value. This input value is passed to the fuzzy inference system. FIS classifies the spam and

produces the output. This work obtains a better result from ranking and classifying of spam words. An efficient approach for spam email detection. To shift the location of spam email filtering system from receiver mail server to sender mail server. The purpose of this novel idea is to detect spam emails in the shortest time and consequently to prevent wasting the network resources from misusage of spammers. In addition, by experimental results we proved that our idea is efficient because just the resources in the sender side are accessed. This implies that if an email is identified as spam one, the receiver's bandwidth and memory is preserved which will assure a better performance. Finally, by locating the filtering system in the sender mail server; the processed time becomes n times less than the time when the filtering system is in the receiver mail server when n indicates the number of processed emails.

REFERENCES

[1] C. MacFarlane, (2003), "FTC Measures False Claims Inherent in Random Spam," *Federal Trade Commission*, http://www.ftc.gov/opa/2003/04/spamrpt.shtm, Accessed Jul. 20, 2011.

[2] L. Nosrati & A. Nemaney Pour, "Dynamic Concept Drift Detection for Spam Email Filtering," *Proceedings of ACEEE 2nd International Conference on Advances Information and Communication Technologies (ICT 2011)*, Amsterdam, Netherlands, pp. 124-126, Dec. 2011.

[3] A. Ramachandran, D. Dagon & N. Feamster, "Can DNS-Based Blacklists Keep Up with Bots?," *The Third Conference on Email and Anti-Spam (CEAS 2006)*, California, USA, pp.1-2, Jul. 2006.

[4] J. Goodman, "Spam: Technologies and Policies," *White Paper, Microsoft research*, pp.1-19, Feb. 2004. International Journal of Network Security & Its Applications (IJNSA), Vol.4, No.2, March 2012 62

[5] A.Ramachandran & N. Feamster, "Understanding the Network-Level Behavior of Spammers," *Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM 2006)*, Pisa, Italy, pp. 291-302, Sep. 2006.

[6] E. Harris, (2003), "Greylisting: The Next Step in the Spam Control War," *White Paper*, http://projects.puremagic.com/greylisting/whitepaper.html, Accessed Dec. 20, 2011.

[7] J.R. Levine, "Experience with Greylisting," *Proceedings of Second Conference on Email and Anti-Spam (CEAS 2005)*, CA, USA, pp. 1-2, Jul. 2005.

[8] P. Graham, "Better Bayesian filtering," *MIT Spam Conference*, Jun. 2003..

[9] H. Yin & Z. Chaoyang, "An Improved Bayesian Algorithm for Filtering Spam E-Mail," *IEEE 2nd International Symposium on Intelligence Information Processing and Trusted Computing(IPTC 2011)*, Huangzhou, China, pp. 87-90, Oct. 2011.

[10] A. Ciltik & T. Gungor, (2008), "Time-efficient spam e-mail filtering using n-gram models," *Elsevier,Pattern Recognition Letters*, Vol. 29, No. 1, pp. 19-33.