Framework of Marketing or Newsletter Sender Reputation System (FMNSRS)

Akkarach Kawbunjun
and Woraphon Lilakiatsakun
Faculty of Information Science and Technology
Mahanakorn University of Technology
Bangkok, Thailand
Email: akkarach@cri.or.th, woraphon@mut.ac.th

Ubon Thongsatapornwatana
Department of Research and Development
Defence Technology Institute, Ministry of Defence
Nonthaburi, Thailand
Email: ubon.t@dti.or.th

***Abstract*—Nowadays, email advertisement is widely used in the commercials. In fact, advertising emails are often sent to undesirable recipients that might be from the outdated recipient lists. Such events can lead to boredom and annoyance to the recipients. As the consequences, the recipients might permanently deny to accept these emails or even worse the sender's businesses or product images could be damaged. Therefore, we propose Framework of Marketing or Newsletter Sender Reputation System (FMNSRS) by using sender reputation algorithm based on the centralized user feedback database to solve the problems previously mentioned. The FMNSRS can create the centralized marketing lists and newsletter senders with periodically updated. In addition, we use the sender reputation system to calculate the score of each sender by several feedback from the clients to classify senders, which the recipients can choose to receive advertising emails from the high reputation score senders. For the experiment, the results show that the FMNSRS can improve the sending capability of the marketing or newsletter email systems. Also, the FMNSRS can accurately detect the marketing or newsletter emails approximately 73.30% more than the traditional framework.**

Keywords*—*Email , Reputation System , Spam Detection , Marketing Email , Newsletter Email

I. Introduction

At present, the number of Internet users has been increasing dramatically which enables emails to be an efficient customer communication method in our daily life. In addition, the email gain lower cost and faster than the other communication methods. The recipients and senders can also send and access the information anywhere and anytime via the Internet. Because of its popularity, almost all businesses use email method to send their information to customers for various objectives as follows: advertising, purchase orders, sales, and customer care service. However, the problem of unsolicited email, commonly known as email spam, is an important issue to face.

An email spam is an email sent indiscriminately, indirectly or directly, to the unwanted recipients by the senders having no current relationship with the recipients. As a consequence, most recipients are flooded with email, some of it welcome and some not. Such event lead to affecting recipient's working efficiency. Furthermore, the recipients may be harmed by the attackers. As with an email spam, the use of the advertising emails, marketing or newsletter emails, violates the rights of the recipients that causes very important problems as follows. (1) If the advertising emails are sent randomly to unwanted recipients that might be from not up-to-date recipient list, it lead to boredom and annoyance to the recipients. (2) If the recipients receive unwanted email frequently, the recipients might permanently deny to accept these emails or even worse sender's business or products image will be damaged. (3) It consumes disk storage space, computational resources and network bandwidth of the recipient email systems.

Although the rule sending of the marketing or newsletter emails will be indicated in the CAN-SPAM that is sent only the intended recipients. The recipients can unsubscribe or use the email filtering system to identify the filter. However, this rule still can not satisfy the needs of the recipients as well as it should. To solve these problems may be too late. the recipient and the recipient email system have the negative attitude already. This result is quite serious in business. In this paper, we propose the system reliability of the marketing or newsletter email called as Framework of Marketing or Newsletter Sender Reputation System (FMNSRS). The FMNSRS uses sender reputation algorithm based on the centralized user feedback database to solve the problems previously mentioned. First, The FMNSRS creates the centralized lists of both the marketing and newsletter senders and the recipient that these lists are periodically updated. Then the sender reputation system calculate the score of senders by feedback from clients to classify both wanted senders and unwanted senders. The wanted senders have a higher reputation score that the recipients can choose to receive the advertising email from this the high reputation score senders.

In the next section, we discuss background information and previous studies of filtering spam email solutions. Section III explains our research design and process diagram. Experimental results are shown in Section IV. Section V concludes and presents future directions.

II. Literature Review

Currently, the various methods [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15] have been frequently used to manage the email, desirable and undesirable, which will be described as follows

The domain name system white list (DNSWL) [1] has been used to confirm that the customers always receive an email from the senders in white list. Besides that, the customers can refuse to receive email from the senders whose the domain name of the email system has negative history or not reliability. Hence, these the sender are stored in the domain name system blacklist (DNSBL) [1], [2], [3]. In the modern email system may be using the process to index reliability of the sender email system (Sender Reputation) of their own [4]. The recipients can choose to receive email from the high reputation score senders. Sender reputation method is used instead of DNSWL and DNSBL to simple manage the system. However, this method is suitable to be applied to the large email systems with the large volume of the receiving and sending emails. Because, reputation score senders are calculated from feedback from clients.

Besides the sender database and sender reputation method, there are methods of the receiving email analysis from the sender. For example, Sender Policy Framework (SPF) is used to identify and authenticate the senders by specifying IP address and domain name into the domain name system (DNS) [5], [6], [7]. Domain Key Identified Mail (DKIM) is a method to identify and authenticate the senders and to verify the accuracy of the letterhead and message body by Asymmetric Key [8], [9], [10]. Upasana and Chakravarty [11] and Li et al. [12] use the content filtering method to verify the email content for the final inspection. This is one method used to check the words or form sentences by using machine learning process to classify the type of received email. The recipients can configure content filtering by personal setting or specific forms from the central databases. However, It uses processing resources in relatively high and occurs errors frequently such as: the system unable to classify the type of received email if the characters of the content do not totally match with the configured words. Furthermore, some email systems use the feed back loop method to send the feedback reports back to the senders to solve the problems previously mentioned [13], [14].

Such methods are used to manage the spam email. But, they are not suitable to deal with the marketing or newsletter email. Because, they can not classify the group of marketing or newsletter email received. They can classify email only spam and not spam. However, some email systems have recently managed the marketing or newsletter email group as follows: Gmail has separate inbox of the email into groups such as Primary, Social, Promotions, Updates and etc [15]. The problems of Gmail, the system unable to classify the marketing or newsletter email group if an email with the contents in the local language is not in English. In addition to that, the business that use a single domain name to send their information to customers for various objectives and the small local business use domain name not widely known. These problems lead to affecting Gmail's classification efficiency. To solve the problems previously mentioned, we uses sender reputation algorithm based on the centralized user feedback database to classify the marketing or newsletter email group. We will explain our research design and process diagram of FMNSRS in the following section.

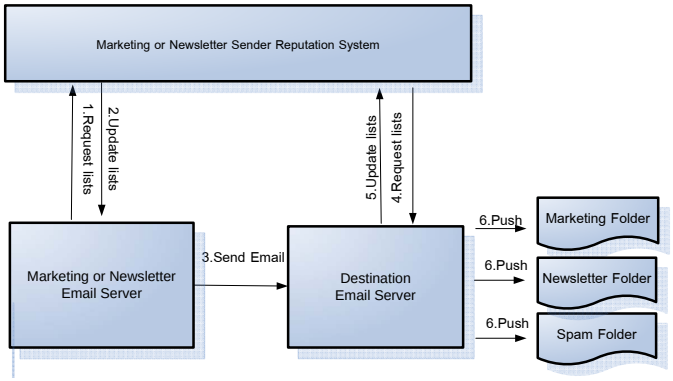

Fig. 1. Research design and process diagram.

III. Research design and process diagram

Although there are currently the various methods used to manage the wanted email and the unwanted email, it does not have a system or process that is developed to deal with the marketing or newsletter email to make effective and to meet the demand. We propose FMNSRS that the system reliability of the marketing or newsletter email. The FMNSRS acts as an intermediary to update the data sets, such as: the marketing or newsletter email requirements each of the recipient, the lists of sender and recipient systems which consist of email address, domain name, and ip address, the sender reputation scores of Marketing or Newsletter senders, to distribute these data to the email systems of the recipients and the senders. In addition, The FMNSRS uses to improve the delivery of the marketing or newsletter emails. It also does not make annoying to the recipients received the unwanted email. Fig. 1 shows the research design and the evolution processes of FMNSRS. This diagram is described in the following sub-sections.

A. The Request Method 1

This method is used to help the recipients sending their marketing or newsletter email requests to the destination email server. The request includes the email address of recipients and the marketing or newsletter email group wanted. The request method 1 is shown in Fig. 2 and described as follows: 1) The recipients submit their marketing or newsletter email groups wanted to the destination email server. 2) When the destination email server receives these requests from the recipients, it forward these requests to the Marketing or Newsletter Sender Reputation System (MNSRS) to register the wanted email with MNSRS. 3) If the MNSRS receives such requests, it stores all requests of the recipients in the database itself and distributes each of the requests to the Marketing or Newsletter Email Servers by considering the matching of these requests. For instance, the recipients want to receive movie email group that this request is sent to the Marketing or Newsletter Email Servers that provide information on the movie. The marketing or newsletter email servers store the received requests in the database itself. In this method, the marketing or newsletter email servers can send their information to the recipients efficiency that we explain in the sending marketing or newsletter email method.

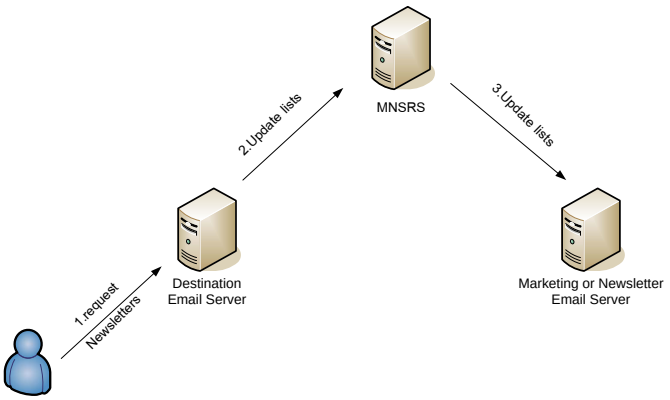

Fig. 2. The processes of the request method 1.

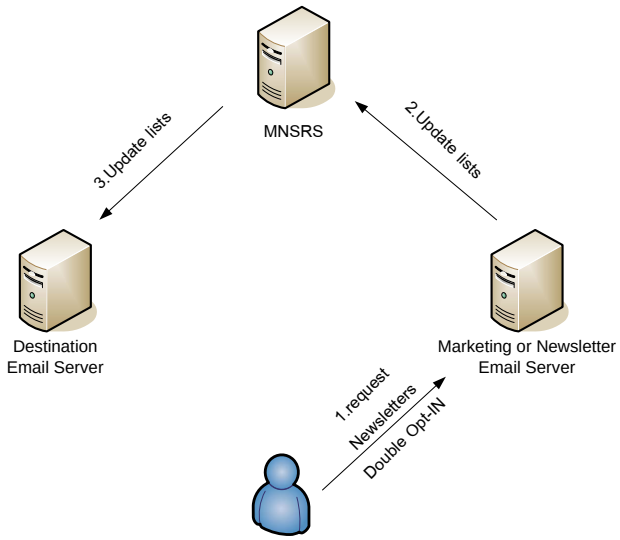

Fig. 3. The processes of the request method 2.

B. The Request Method 2

The recipients can request to receive the marketing or newsletter emails by subscription on the marketing or newsletter email servers. The request method 2 is shown in Fig. 3 and described as follows: 1) The recipients can request to receive the emails by double opt-in [16] to subscribe of the marketing or newsletter email servers wanted. 2) When the marketing or newsletter email servers receive the request and confirmation from the recipient, they send such request to MNSRS to update the list of recipients and the marketing or newsletter email group in a centralized database. 3) Then MNSRS forward this request to the destination email server of recipients registered to update such data in a database of recipients.

C. The Sending Marketing or Newsletter Email Method

The sending marketing or newsletter email method is shown in Fig. 4 and described as follows: 1) The marketing or newsletter email server sends the marketing or newsletter emails to wanted recipient systems by considering the list and marketing or newsletter email group wanted of recipients in a database of sender. To update such datas in a database,

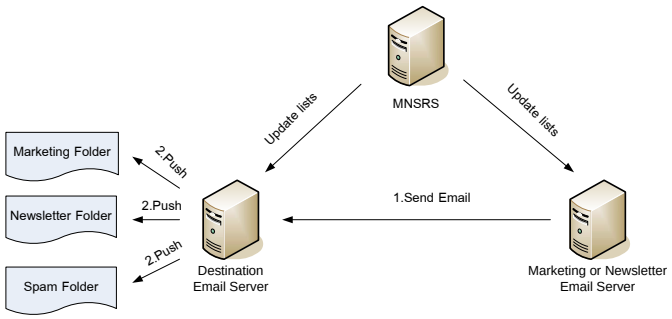

Fig. 4. The sending marketing or newsletter email method.

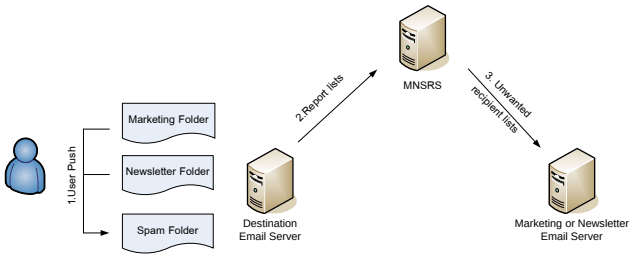

Fig. 5. The false reporting method.

there are two processes as follows: the marketing or newsletter email server can define a schedule to update automatically with MNSRS. In addition, MNSRS can also define a schedule to send these datas to the marketing or newsletter email server for automatic updating. 2) When the destination email servers receive the marketing or newsletter emails, they classify and enter these emails into the inbox categories of wanted recipients in their domain such as Marketing, Newsletter and etc. If emails received are not to be desired, they are entered into the spam inbox. In the classifying email group process, the destination email servers check the list of senders, sender reputations, and wanted email groups each of the recipients in a database itself. If such datas of the sender match datas in a database and the high reputation score sender, email is moved to the inbox categories of wanted recipients. Otherwise, email is moved to the spam inbox. A database of recipients is updated by MNSRS that both sides can define the schedule for automatic updating.

D. The False Reporting Method

The false reporting method is shown in Fig. 5 and described as follows: 1) When the recipients receive the unwanted marketing or newsletter emails and these emails are not in the spam inbox, the recipients can move these emails to spam inbox to report back to the senders. 2) Then the destination email servers of recipients reported connect to the MNSRS to unsubscribe the marketing or newsletter email servers wanted. Besides that, the MNSRS calculates new reputation score each of the senders canceled by using the statistics reports. 3) The MNSRS send the lists of unwanted recipients to the marketing or newsletter email servers that send email to unwanted recipients for updating recipient database.

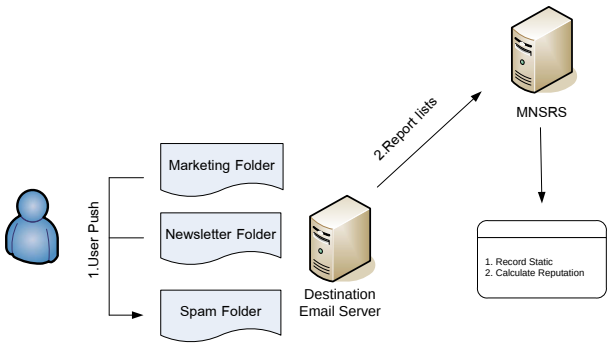

Fig. 6. The sender reputation method by using MNSRS.

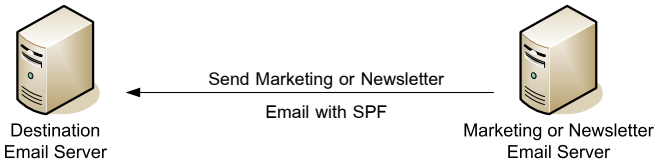

Fig. 7. The sender and recipient authentication method.

E. The Sender Reputation Method

The sender reputation method by using MNSRS is shown in Fig. 6 and described as follows: 1) While the recipients move the unwanted marketing or newsletter emails to spam inbox, the destination email servers of recipients reported will receive the these false reports 2) Such the destination email server connect to the MNSRS to unsubscribe the marketing or newsletter email servers wanted. The MNSRS updates own recipient database and calculates new reputation score each of the senders canceled based on the centralized recipient feedback by using this formula:

$$\text{reputation} = \frac{100 \times (\text{asm - aur})}{\text{asm}}$$

where
asm = Total of the sending marketing or newsletter emails each of the senders.
aur = Total of spam marketing or newsletter emails that reported by the unwanted recipients.

F. The Sender and Recipient Authentication Method

In the sending and receiving marketing or newsletter emails process between the marketing or newsletter email servers and the destination email servers, SPF [5], [6], [7] is used to confirmed both good email systems. The conditions for sending and receiving email are as follows: The registered email systems of the senders and the recipients must support SPF method to send and receive emails. The process diagram is shown in Fig. 7.

IV. Experimental Results

According to the author has designed and tested the FMNSRS system that can classify the marketing or newsletter email group matching the real needs of the recipients better than the traditional email system. It works well especially when the emails with the contents in the local language is not in English.

TABLE I. Summary of the experimental samples

The Experimental Sample	Summary
Number of the general email systems	10
Number of the marketing or newsletter email systems	10
Number of the email accounts	40
Number of the normal emails	1000
Number of the marketing or newsletter emails	1000
Number of the spam emails	1000

TABLE II. Email Data Set and Checking Record

Email Data Set	Checking Record
Email of the sender	Email records of the marketing or newsletter sender
Email of the recipient	Email records of the marketing or newsletter recipient
IP address of the sender	IP address of the sender match the DNS reliability
content (keyword)	content filter (keyword)
-	sender reputation (score1,check-push)
-	sender reputation (score2,check-push)
-	sender reputation (score3,check-push)
-	sender reputation (score4,check-push)
-	DNSBL by IP address
-	DNSWL by IP address

In addition to that, all processes of the FMNSRS system are run automation. The administrators do not need to modify the database itself. The users of FMNSRS system also feel no difference from using the general email systems. Moreover, this paper is also able to identify the senders as either quality or not quality. The experimental samples for this research are shown in Table I and Table II shows the email data set and checking record.

To analyze system performance for this paper, there are two methods for analysis described in the following subsections.

A. True positive (TP)

This method is used to analyze the accuracy of the detection email which can be calculated by using this formula:

$$\text{NEAR} = \frac{\text{NofNEDetectedNE}}{\text{NofNEDetectedNE + NofNEDetectedFault}} \times 100$$

$$\text{MEAR} = \frac{\text{NofMEDetectedME}}{\text{NofMEDetectedME + NofMEDetectedFault}} \times 100$$

$$\text{SEAR} = \frac{\text{NofSEDetectedSE}}{\text{NofSEDetectedSE + NofSEDetectedFault}} \times 100$$

Table III summarizes the notation for The TP Analysis. Fig. 8 and Table IV show the results of the TP analysis. As a results, this novel process can detect the normal emails accurately 81.90% and 45.26% more than the traditional process. Furthermore, the novel process can detect the marketing or newsletter email accurately 73.30% and the spam email accurately 62.44%. These results are an acceptable accurate rate.

TABLE III. VARIABLE DESCRIPTION FOR TP

Variables	Description
NEAR	The accurate rate of the normal email detection.
NofNEDetectedNE	The number of the normal email that are detected as the normal email.
NofNEDetectedFault	The number of the normal emails detected fault.
MEAR	The accurate rate of the marketing or newsletter email detection.
NofMEDetectedME	The number of the marketing or newsletter email that are detected as the marketing or newsletter email.
NofMEDetectedFault	The number of the marketing or newsletter email detected fault.
SEAR	The accurate rate of the spam email detection.
NofSEDetectedSE	The number of the spam email that are detected as the spam email.
NofSEDetectedFault	The number of the spam email detected fault.

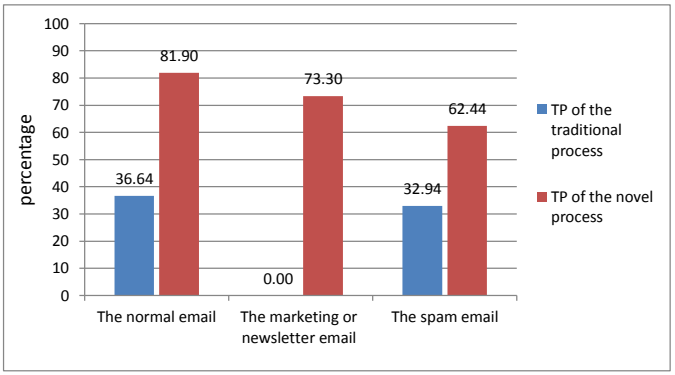

Fig. 8. The results of the true positive analysis.

TABLE IV. THE RESULTS OF THE TRUE POSITIVE ANALYSIS

Accurate Rate	Traditional process	Novel process
The normal email	36.64%	81.90%
The marketing or newsletter email	0.00%	73.30%
The spam email	32.94%	62.44%

TABLE V. VARIABLE DESCRIPTION FOR FP

Variables	Description
NEFR	The fault rate of the normal email detection.
MEFR	The fault rate of the marketing or newsletter email detection.
SEFR	The fault rate of the spam email detection.

B. False positive (FP)

This method is used to analyze email fault detection which can be calculated by using this formula:

$$\text{NEFR} = \frac{\text{NofNEDetectedFault}}{\text{NofNEDetectedNE} + \text{NofNEDetectedFault}} \times 100$$

$$\text{MEFR} = \frac{\text{NofMEDetectedFault}}{\text{NofMEDetectedME} + \text{NofMEDetectedFault}} \times 100$$

$$\text{SEFR} = \frac{\text{NofSEDetectedFault}}{\text{NofSEDetectedSE} + \text{NofSEDetectedFault}} \times 100$$

Table V summarizes the notation for The FP Analysis. Fig. 9 and Table VI show the results of the FP analysis.

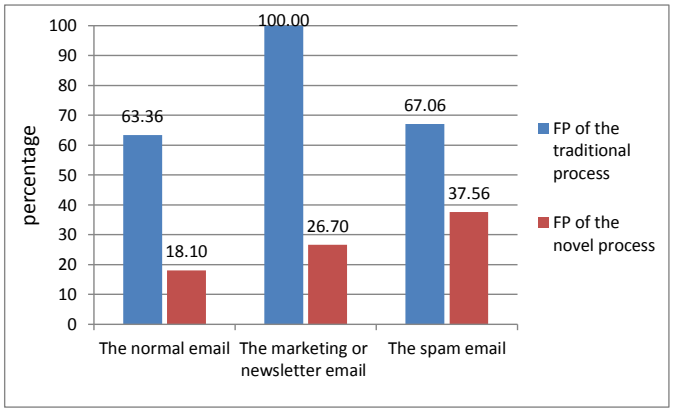

Fig. 9. The results of the false positive analysis.

TABLE VI. THE RESULTS OF THE FALSE POSITIVE ANALYSIS

Fault Rate	Traditional process	Novel process
The normal email	63.36%	18.10%
The marketing or newsletter email	100.00%	26.70%
The spam email	67.06%	37.56%

As a results, this novel process can detect the normal emails wrongly 18.10% and 45.26% more efficient than the traditional process. Moreover, the novel process can detect the marketing or newsletter email wrongly 26.70% and the spam email wrongly 37.56%.

V. CONCLUSION AND FUTURE DIRECTIONS

The testing results from the previous section show that this technique can improve the detection capability of the email detection process. The novel system can classify the type of email, marketing or newsletter, matching the real needs of the recipient better than the traditional system. In addition, it can help facilitate email management for administrators. However, there is also disadvantages as follows. 1) If the amount of the registration or the statistic are not enough, this system can detect the emails accuracy decreases. 2) This technique is necessary to use the request and response data exchange several times. In the future work, we intend to improve the exchange process less complicated. In addition to that, we intend to improve the detection accuracy although there is not enough information.

ACKNOWLEDGMENT

This work was supported by Mahanakorn University of Technology and the Defence Technology Institute (Public Organization), Thailand.

REFERENCES

[1] J. Levine, "DNS Blacklists and Whitelists," IRTF, RFC 5782, February 2010. [Online]. Available: https://tools.ietf.org/html/rfc5782

[2] T. Sochor and R. Farana, "Improving efficiency of e-mail communication via spam elimination using blacklisting," in *Telecommunications Forum (TELFOR), 2013 21st*, Nov 2013, pp. 924–927.

[3] L. Lazzari, M. Mari, and A. Poggi, "A collaborative and multi-agent approach to e-mail filtering," in *Intelligent Agent Technology, IEEE/WIC/ACM International Conference on*, Sept 2005, pp. 238–241.

[4] B. Taylor, "Sender reputation in a large webmail service," in *CEAS*, 2006.

[5] M. Wong and W. Schlitt, "Sender Policy Framework (SPF) for Authorizing Use of Domains in E-Mail, Version 1," RFC 4408, April 2006. [Online]. Available: http://tools.ietf.org/html/rfc4408

[6] N. T. Anh, T. Q. Anh, and N. X. Thang, "Spam filter based on dynamic sender policy framework," in *Knowledge and Systems Engineering (KSE), 2010 Second International Conference on*, Oct 2010, pp. 224–228.

[7] T. Seike, Y. Jin, N. Yamai, K. Okayama, K. Kawano, and M. Nakamura, "A solution for mail forwarding problem of spf by tracing recipient addresses," in *Applications and the Internet (SAINT), 2010 10th IEEE/IPSJ International Symposium on*, July 2010, pp. 129–132.

[8] D. Crocker, T. Hansen, and M. Kucherawy, "DomainKeys Identified Mail (DKIM) Signatures," IETF, RFC 6376, September 2011. [Online]. Available: http://tools.ietf.org/html/rfc6376

[9] B. Leiba, "An introduction to internet standards," *Internet Computing, IEEE*, vol. 12, no. 1, pp. 71–74, Jan 2008.

[10] Y. Higashikado, T. Izu, M. Takenaka, and T. Yoshioka, "An extension of the sender domain authentication dkim," in *Convergence and Hybrid Information Technology, 2008. ICCIT '08. Third International Conference on*, vol. 2, Nov 2008, pp. 565–568.

[11] Upasana and S. Chakravarty, "A survey on text classification techniques for e-mail filtering," in *Machine Learning and Computing (ICMLC), 2010 Second International Conference on*, Feb 2010, pp. 32–36.

[12] X. Li, J. Luo, and M. Yin, "E-mail filtering based on analysis of structural features and text classification," in *Intelligent Systems and Applications (ISA), 2010 2nd International Workshop on*, May 2010, pp. 1–4.

[13] J. Falk, "Complaint Feedback Loop Operational Recommendations," IETF, RFC 6449, November 2011. [Online]. Available: https://tools.ietf.org/html/rfc6449

[14] J. Falk and M. Kucherawy, "Battling spam: The evolution of mail feedback loops," *Internet Computing, IEEE*, vol. 14, no. 6, pp. 68–71, Nov 2010.

[15] Google, "Gmail Inbox tabs and category labels," . [Online]. Available: https://support.google.com/mail/answer/3055016?hl=en

[16] E. Allman, "Spam, spam, spam, spam, spam, the ftc, and spam," *Queue*, vol. 1, no. 6, pp. 62–69, Sep. 2003. [Online]. Available: http://doi.acm.org/10.1145/945131.945157