# Spam Mail Detection and Blockingfor E-Mail Security by Cascade Hybridization and Collaborative Recommender

Adapa Chandrakala
*Asst. Professor, Baba Institute of Technology and Sciences, VSKP*
*chandrakalaadapa28@gmail.com,*

Gangu DharmaRaju
*Asst. Professor, Baba Institute of Technology and Sciences,VSKP.*
*dharmaa.surya@gmail.com,*

A V S Pavan Kumar
*Asst. Professor, Baba Institute of Technology and Sciences,VSKP.*
*avspavankumarmca@gmail.com*

**Abstract-There are copious ways of communication methods in this digitally advanced world but Electronic mail which is also known as e-mail or email is the utmost competent method to communicate or transfer our data from one to another. There is the likelihood of going astraywhen transferring or communicating during e-mail. One of those misbehaviours is receiving huge number of undesirable e-mails from a set of unfamiliarsenders. A huge number of these mails frequently consist of commercial content. In the current actual system to avoid the undesirable Email receiving Spam method is used. The other terms used for Email spam are unsolicited bulk email (UBE), unsolicited commercial email (UCE), direct mail, third-class mail or junk mail.Sending huge quantity of messages to haphazard set of recipients constitutes spam. This method can differentiatejunk messages from other messages in many times but not always. None of the ways we have can becounteracting these undesirable messagesor Email receiving in spam method.To clear up this enigma of receiving undesirable messages or Emails we contemplate the abstraction of Spam mail blocking system. In the contemplated method we can permanently counteract the incoming of undesirable messages or Email through Spam mail blocking system.**

**Keywords— Email, Digital, UCE, Spam.**

## 1. INTRODUCTION

Emails are the utmost competent way of sending messages or sharing data in the digital media. But the enigma arising from the junk messages is decreasing the efficiency of this Email system. Junk email is frequently adopted for advertisement or marketing. These advertisers send huge quantity of undesirable and unrelated messages to haphazard set of recipients. The different enigmas arising from the junk are

- Gigantic in bulk- According from the data from esecurityplanet.com approximately 97.4 billion junk messages are sent in a day. Sometimes 90% of the emails are junks that fill his/her inbox.
- Behemoth in magnitude of contents
- Occupy giant chunk of user space
- May consist ofapocryphal information- The United State Federal Trade Commission noticed that 66% of junks have apocryphal information.
- Not akin to the recipient's interests- 18% of junks consist of "Adult" material.
- User time gets exhausted on these mails- Based on another report 12% of users give half an hour or more in a day in negotiating with junk emails.
- Dissipate the user network resources-they cost money for ISPs because the bandwidth and the memory of system are exhausted.
- Causes a lot of security enigmas- because most of them include Trojan, viruses and Malwares.

Considering all these drawbacks many techniques are developed to filter and check the current of junk emails. But even with these filtering techniques the flow could not be controlled and the number of junk messages is growing. Till now there is no specific solution for these junk messages as the spammers are unidentified users and they have adequate knowledge to bypass filters. They are well versed with the filtering mechanisms. So the spammer can bypass the filtering mechanisms with different techniques and they can easily send the spam. Such spammers can make junk emails undetectable from the normal mails and these junk emails are misjudged as legitimate ones.

Plenty of studies have been performed on spam email filtering. The main issue with these filtering systems are they are due to their situation. They are mostly kept at the receiver's end. So the flow from the sender cannot be stopped which results in no decrease in the network load and network resources are exhausted.

This paper by experimental results shows that this new approach to stopjunk mails is more competent and works well when compared with the previously contemplated approaches. Our abstraction and the attempt results are discussed under.

As we know many filtering techniques are there to negotiate with the junk mails. They check the current of these junk mails and stop them reaching the users mail boxes as much as possible. Figure 1 shows the classification of spam e-mail filtering techniques. It accommodatesstatic algorithm, list-based filtering, and IP-based filtering. The list-based filtering is further divided into three types; Whitelist, Blacklist, and Grey list. Static algorithm is divided into rule based and the content-based filtering. At last, IP-based filtering comprise of revers-lookup.

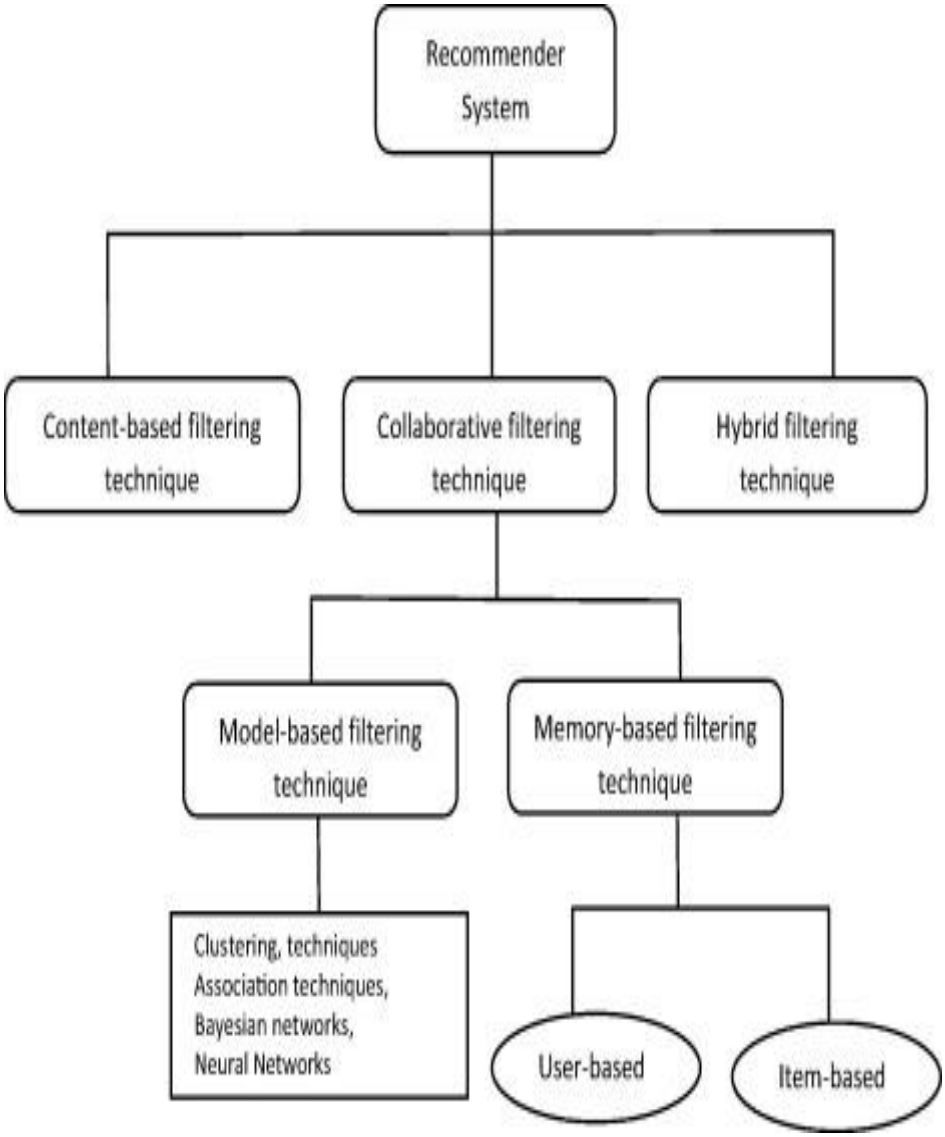


Fig.1: Different types of filtering techniques

In this method we use two different types of filtering techniques simultaneously. One is collaborative filtering which is not dependant on doomain prediction. It is well used for content that cannot easily and adequately be described by metadata. Collaborative filtering technique works by building a database (user-item matrix) of preferences for items by users. It then matches users with relevant interest and preferences by calculating similarities between their profiles to make recommendations. On the other hand the cascade hybridization technique uses an iterative refinement process in constructing an order of preference among different items. The results of one recommendation technique are refined by another recommendation technique. The first recommendation technique outputs a coarse list of recommendations which is in turn refined by the next recommendation technique. The hybridization technique is very competent and tolerant to noise due to the coarse-to-finer nature of the iteration.

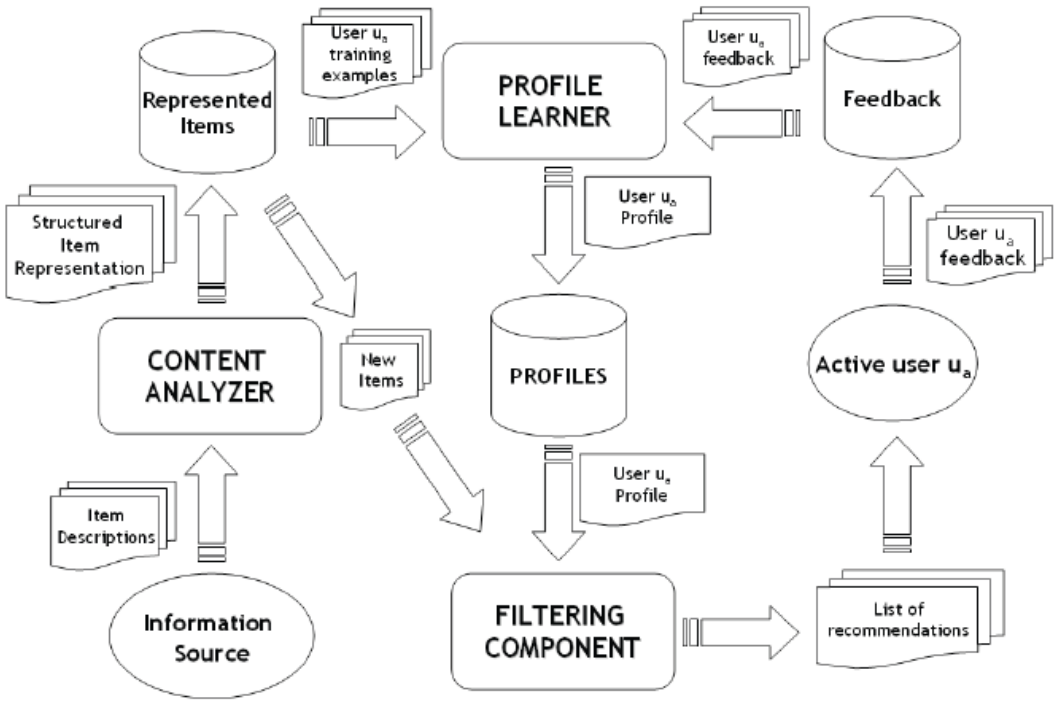


Fig.2: Internal working model of filtering system

## 2. PROBLEM DEFINITION

If a gigantic number of e-mails are sent to diverse users, those are called as spam e-mails. These kinds of mails are usually sent as a marketing or advertisement purposes. These mails are undesirable by users and dissipate user network resources. Many people have addressed the drawbacks of junk e-mails. Sometimes 90% of the emails are junks that fill his/her inbox. High in volume- According from the data from esecurityplanet.com approximately 97.4 billion junk messages are sent in a day. The contents of these mails are also huge in quantity and so these mails occupy a lot of memory in the available user space. And the information provided in these mails is not necessarily true. 2 out of 3 junk mails have apocryphal information. Many of them consist ofapocryphal information that causes huge amount of confusion and frustrating to the users. Approximately every $7^{th}$ user exhausts a minimum of half an hour in a day for negotiating these junk mails. These mails may consist of adult material as well. Every $5^{th}$ mail of junk messages may consist of adult material. So these junks create a huge amount of issues by consuming network resources and bandwidth and memory of the system. Another potential and dangerous conundrum is that they may consist ofTrojan, viruses or Malwares.These security akin issues associated with junk mails are used by criminals in accessing user accounts in getting sensitive information from the restricted areas. Another irritating conundrum to the users is that these junk mails are not even akin to the users interests. Many junk mails are having uninteresting contents which causes burden over users.

Nowadays spam is sent from personal computers, offices and other gadgets around the world. These mails are sent via “ZOMBIE NETWORKS”. Spam detection is mainly dependant on 2 aspects. One is content

based which is based on keywords or content of the email and the other is non-content based which is based on the statistical means.Detecting keywords or Content based statistical means can give very good filtering results if they are correctly fed with the information. Other enigmas with junk messages are

- We can't counteractjunk mails from getting received but those can be moved only into spam folder.
- Due to the reception of large number of huge junk mails, it occupies memory which gets exhausted on undesirable things.
- To release this memory the user needs to delete those junk mails manually which needs confirmation of authorization to the system. It is a time exhausting process to delete all these junk mails.
- When we are going to negotiate with these junk mails, Trojans, viruses or malwares can get access into our system which can cause security issues.

To counteract all these drawbacks which are due to reception of spam mails, we introduce an email blocking system. The uniqueness of this system is that, it blocks the third-class mails from getting received. That means the third-class mails will not come into the users account either in mailbox or in the spam. It almost completely counteracts undesirable mails and spams to enter. Since the undesirable messages are not even getting received the user memory is not occupied by those unsolicited messages which save user memory. Anyway we usually not see the unsolicited messages, so there is no necessity to enter into spam folder section which saves a lot of time and other issues. If we want to communicate with a spammer we can unblock the spammer which enables him to communicate with us as a normal user. The spammer can always be blocked again if there are frequent mails or any other issues so that the spammer cannot communicate with the recipient later.

*Advantages*

- We can filter both domain dependent and domain independent data which gives more efficient way of blocking.
- The advantages of both the filtering systems will get add up and the disadvantages of each system can be overcome by the other system increasing the combined system's efficacy (By detecting more true positive spams and identifying false negatives as well).
- Both these filtering techniques are used at the sender mail server so that the mails send by sender are immediately analysed in a short span of time.
- As the spam is not received there is no question of receiving Trojans, viruses or Malwares.
- Lot of time and memory can be saved in negotiating with unsolicited messages as they cannot even reach the receiver mail.
- Receiver's bandwidth memory and network resources are saved as the unsolicited mails are not getting flown in the network stream.

## 3. CONCLIUSION

A combination of collaborative recommendation and cascade hybridization filtering of junk mails at the sender end is an excellent solution for unsolicited mails. There are copious of ways to negotiate with

unsolicited mails like classifiers and filters. We have analysed the previous akin works on junk mails. Our contemplated work used cascade hybridization filtering technique which filters junk mails by different recommendation techniques.

Collaborative filtering which is not dependant on doomain prediction. It is well used for content that cannot easily and adequately be described by metadata. Collaborative filtering technique works by building a database (user-item matrix) of preferences for items by users. It then matches users with relevant interest and preferences by calculating similarities between their profiles to make recommendations. The cascade hybridization technique uses an iterative refinement process in constructing an order of preference among different items. The results of one recommendation technique are refined by another recommendation technique. The first recommendation technique outputs a coarse list of recommendations which is in turn refined by the next recommendation technique. The hybridization technique is very proficient and tolerant to noise due to the coarse-to-finer nature of the iteration. By using these to techniques simultaneously we can block both domain dependant and domain independent data.Both these filtering techniques are used at the sender mail server so that the mails send by sender are immediately analysed in a short span of time. If they are not legitimate they can be blocked immediately even before the entry into network stream. So the network workload can be reduced and network resources can be saved. As they cannot reach the receiver mailbox or spam, users are counteracted to lose their space and lot of time is saved which is spent on negotiating with these spams. As the receiver's bandwidth and memory is preserved there is a better performance by the users.

BIBLIOGRAPHY

1. Herlocker JL, Konstan JA, Terveen LG, Riedl JT. Evaluatingcollaborative filtering recommender systems. ACM Trans InformSyst 2004;22(1):5–53.

2. Burke R. Hybrid recommender systems: survey and experiments.User Model User-adapted Interact 2002;12(4):331–70.

3. Rashid AM, Albert I, Cosley D, Lam SK, McNee SM, KonstanJA et al. Getting to know you: learning new user preferences inrecommender systems. In: Proceedings of the internationalconference on intelligent user interfaces; 2002. p. 127–34.

4. 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.

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

6. 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

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

8. 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.

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

10. Acilar AM, Arslan A. A collaborative filtering method based onArtificial Immune Network. ExpSystAppl 2009;36(4):8324–32.

11. Murat G, SuleGO. Combination of web page recommendersystems.ExpSystApplicat 2010;37(4):2911–22.

12. Park DH, Kim HK, Choi IY, Kim JK. A literature review andclassification of recommender systems research. Expert SystAppl2012;39(11):10059–72.

13. Bojnordi E, Moradi P. A novel collaborative filtering model basedon combination of correlation method with matrix completiontechnique. In: 16th CSI international symposium on artificialintelligence and signal processing (AISP), IEEE; 2012.

14. https://www.esecurityplanet.com/network-security/almost-100-billion-spam-e-mails-sent-daily-in-q1-2013.html