# **POST: Prevention of Spam Transmission**

<sup>1</sup>Parimal Tank <sup>2</sup>Mehul Gundaliya <sup>3</sup>Chetan Ajudiya

<sup>1, 2</sup>PG Student at Noble Group of Institution, Gujarat, India<sup>3</sup> ME, Assistant Professor, NGI, Junagadh

Abstract—this paper introduced the newer approach POST, based on some new technique to stop the irrelevent communication over network or called spam communication. Online communication media such as social networking, emails, messaging allows user to reach number of peoples at neglible cost. This attribute enables any information roaming freely in network. So as a result unwanted comunication happened with great degree. As number of email users increses there is also increase tremendous growth in irrelavent messages called spam or ham. So develop the system that filters the spams with great efficiency to prevent the uwanted communication and thus to reduce cost of the network also. Threre are many of techniques developed for filtering spams as content based, classification based and others. POST is the system that prevents to route unwanted content over the network. POST implements the algorithm at client side to stop the spam. POST provides the best identification environment to check legitimity of sender to prevent spam transmission.

**Keywords:** -POST (Prevention of Spam Transmission), Spam transmission, Content based filtering, Classification based filtering, Legitimity

### I. INTRODUCTION

Internet users are increase in numbers day-by-day as numbers of facility available on internet. We cannot find a single person from any field, which not knows about the internet. Online social networks, emails, content-sharing sites and mails are internet based communication systems. These facilities are for any one at negligible cost.

Information sharing on these network is the most acceptable and important feature of the internet. Internet is any time any place service. So it is also very much useful also.

With an inexpensive Internet connection, any user has the potential to reach millions of users by posting messages to an email list or by uploading content to a sharing site. Actually, this attribute is good to reach beyond limit but it has democrat content publication, as anyone can publish and anyone interested in the content can obtain it.

Because of this better facility, the same attribute can be used with negative purpose of disruption of legitimate communication, unsolicited marketing, to down the network by hackers. These unwanted and irrelevant communications are known as Spam Communications.

In 2012, around 75% [17] of the emails were spams and spam incurs a cost of \$235 [17] billion. Therefore, to prevent them is very much essential.

Based on different facilities different types of spam are there. To judge that spams and deal with spam is necessary for every service providers.

There are many of techniques available to stop the spams as content-based filtering, using classification tree

analysis, using header based techniques. All filters the spam at certain levels.

Some of the manufacturer of spam filter and their product name.

SPAM FILTERS WITH PROVIDER	
Manufacturer	Product name
SonicWall	Email Security Appliances
Symantec	Brightmail Anti-Spam
Symantec	Norton AntiSpam
Google	Mail
spamcop.net	SpamCop
Apple	Mac Mail
mozilla.com	Thuderbird built-in spam filter
Microsoft	Exchange Server spam filter
McAfee	SpamKiller 6

All techniques use different methods, different models to detect whether the content transmitted is spam or not.

POST is the system that cop with these irrelevant. The POST creates the virtual group type of system that helpful to stop the spam transmission. To stop spams is very much difficult and thus try to stop the transmission of the spam. Thus it not travels through the network and do not cost to network.

The focus of this paper is to provide the system POST that prevents spam transmission. POST is very much powerful that it can handle this transmission.

The structure of the rest of the paper is as follows:

Second part includes study about related work done so far in the area of spam transmission. Next part inlcudes proposed algorithm and it also describes how whole system works. It includes how this new system will help to prevent spam transmission. Final part includes conclusion and future work.

#### II. WORK DONE SO FAR

Anti-spam approaches are as follows comprised one or several of the following basic approaches [9].

#### A. Spam Filtering By Content Rating Approaches:

Content rating [10] used by many content-sharing sites (e.g., YouTube [18]). Users can rate the level of interest, relevance, and appropriateness of a content item they have viewed. The content is then tagged with the aggregated user ratings. Data mining offers value across a broad spectrum of industries.

This techniques can help users to identify relevant content and avoid unwanted. These ratings can also help administrators to identify potentially inappropriate content, which they can then inspect and possibly remove.

Content rating is applicable only to one-to-many communication. Moreover content-rating systems can be

**45** 

manipulated, particularly in a system with weak user identities.

## *B.* Filter Based On Social Network:

To detect spam now a days get information from social networks. They construct a graph, whose vertices represent email addresses. A directed edge is added between two nodes A and B, if A has sent an email to B.

Boykin and Roychowdhury [2] initially classify email addresses based on the clustering coefficient of the graph subcomponent: For spammers, this coefficient is very low because they typically do not exchange emails with each other. While in contrast, the clustering coefficient of the subgraph representing the actual social network of a non-spammer (colleagues, friends, etc.) is rather high.

Golbeck and Hendler propose another scheme to rank email addresses, based on exchange of reputation values [11]. The main problem of this approach is that its attack resilience has not been verified.

# C. Filter by Content based Approaches:

This approach analyzes the subject or body of an email for certain keywords (may dynamically learn using a Bayesian filter or statically provided) or patterns that are typical for spam emails (e.g., URLs with numeric IP addresses in the email body).

The great thing about content-based schemes is their ability to filter quite a high number of spam messages. But also main drawback is that they (e.g., the set of static keywords) have to be adapted continuously since otherwise the high spam recognition rate will decrease. [12]

D. Filter by Header based Approaches:

To detect spam this approach examines the headers of email messages. Blacklist schemes store the IP addresses (email addresses can be forged easily) of all known spammers and refuse to accept emails from them. While, Whitelist schemes, to decrease the number of false positives from content-based schemes, collect all email addresses of known non-spammers in a whitelist.

For higher accuracy user can manually create such blacklists but it is quite burdening for user to maintain it regularly.An automatic creation can be realized, for instance based on previous results of a content-based filter as is done with so-called *autowhit elists* in SpamAssassin [16].

Both blacklists and whitelists are rather difficult to maintain, especially when faced with attacks from spammers who want to get their email addresses on the list (whitelist) or off the list (blacklist).

## E. Protocol based Approaches:

This approach proposes changes to the underlying email protocol. Challenge-response schemes [9] require a manual effort to send the first email to a particular recipient. For example, the sender has to go to a certain web page and activate the email manually, which might involve answering a simple question (such as solving a simple mathematical equation).

Afterwards, the sender will be added to the recipient's whitelist such that further emails can be sent without the activation procedure. The activation task is considered too complex for spammers, who usually try to send millions of spam emails at once.

An automatic scheme is used in the *greylisting* approach [19], where the receiving email server requires each unknown sending email server to resend the email again later.

# III. LITERATURE REVIEW

Till date many attempts has been made to deal with spam. Manyresearchers have tried a lot to prevent spam transmission. Many of techniques they used to filters non related spam data. They tried many a time with different strategy and compare their system with previous one. Most of them provers better in perticular area, while not as good in some area.

For ex. Blacklist and whitelist techniques can better filters the spam if mentioned in lists and not as good if not included in list.

There is a need of the system (filter) that filters non regular or can say irrelevent or ham content effectively with great effort. Spam prevention techniques must be strong enough to detect spam data and spammers also.

Many of literature available about spam transmission. There are many of techniques [2] to [9] to prevent.

A. An MCL Based Approach [13] For Spam Profile Detection inOSN.

MCL is applied on the weighted graph to generate different clusters containing different categories of profiles. Majority voting is applied to handle the cases in which a cluster contains both spam and normal profiles.

Experimental results of this paper show that majority voting not only reduces the number of clusters to a minimum, but also increases the performance.

*B.* SOAP – A social network aided personalized and effective spam filters to clean your e-mail box. [6]

Current many spam filters uses social networks itself to moniter spam detection. To develop the perfect spam filter this paper unlightens the way. They proposed a new filter called SOAP: That is network aided spam filter.

As seen in techniques many of filters (Bayesia) emphasis on static keywords or lists (Black or White). Unlike many of filters, SOAP not depends on a single method to filter spams rather it uses more than one technique to filter spams. The system integrates trust management, social relations and basic one that is bayesia filter.

This system also checked with real dataset of Facebook profiles, which includes both regular and spam profiles. The system prooves better to scan the spams.

C. Preventing Unwanted Social Inferences with Classification Tree Analysis [14]

Here in this paper uses decision tree method to differentiate normal situations and high risk situation. To evaluate this methodology, test and training datasets were collected during a large mobile-phone field study for a location-aware application.

For the current and past situations, the classification tree employs two inference functions. Results show

That the achieved true classification rates are significantly better than approaches that employ other available features for the internal nodes of the trees. The results also suggest that common classification tools cannot accurately capture the helpful information for social applications. This is mostly due to the lack of enough training data for high-risk, low-entropy situations and outliers.

Thus, paper concludes that estimating the information entropy and the relevant inference risk using a pre-processor can yield a simpler and more accurate classification tree.

## D. Detecting Spammer on Twitter [15]

This paper discuss about to deal with spammers on Twitter. To cope with spams on Twitter manually classified the legitimate users and spammers. For that real dataset of Twitter about 54 million users is collected, along with 1.9 billion links, and almost 1.8 billion tweets. [15]

To detect spammers they identify number of attributes or behaviours related to content andbehaviour. This is very much useful to detect spammers. To detect spammers or non-spammers uses this attributes to MLP (Machine Learning Process) for classification.

This strategy succeeds to detect irrelevent data or spam data (content) with great percentage approx. 70% of dummies and 96% of regular one.

## E. Detecting Spammer with SNARE [5]

*Summary:* SNARE is the system that works on sender reputation engine that automatically and accurately classify email senders based on previous history.

SNARE is the type of reputation engine that uses more than one mehtod to classify spammers and nonspammers. The regular spam filtering technique like listing is not easy to maintain and error prone also if attacker attacks on lists.

SNARE examines features rather than contents that are why it is very much lightweight. They encorporate this feature in classification algoritham and tests wheher it can classify as spammer or legitimate one.SNARE is build using this feature kept in mind. This engine can be used as first pass in the blacklists.

# F. Personal email network: An effective anti-spam tool. [2]

To find trusted networks of friends in cyberspace personal email network provide automated graph theoretic methood. Network keeps history of users. Mail user can use their mail network to differentiate irrelavent or can say unsolicited mail, named spam. Now this mail network is generally constructed from historical information available in the header of email.

Paper focus to construct a trusted like of network in which network must know about all the users resides in the network. This personallized network thus helps to identifies legitimate data and spam data. With 100% accuracy, algorithm of this tool cans classifiy approx. 53% of all emails as spam or non-spam.

## G. Mail Rank: using ranking for spam detection. [3]

This technique uses ranking system to rate the emails which are arrived. As a result from that rank sender can be identified as spam or non-spam.

There are two possibilities for Mail-Rank system.

First one is Basic Mail-Rank, which calculates anoverall (global) rank for every mail address. Second one

is Personalized Mail-Rank, in which for every mail address score is different.

The system, Mail-Rank is very much reliable and highly resistant against spam attack. In sparse network, the network of a small set of peers, Mail-Rank can also performs well.

From this survey we can say that there are many of practise done so far to prevent irrelevent data transmission called spam transmission.

But still due to weak sets in every method not a single method can say that it exactly classifies the spammers. Not any of tools or system can say that their technology can fight against spam in all environments or in any condition. So try to develop such method or technology that can handle any such situation and become 100% reliable for filtering of spam.

## IV. Post

Till this point we now sure about that spam filtering faces many of the problems. Develop such system or software or tool that deal with spam data. "POST: Prevention of Spam Transmission" is the system that cop this problem.

POST system faces spam transmission with higher resistance with much more relaibility and also much more security. It must filters spam and not misjudge a single one.

**POST** implemented at Mail Server. POST applies quite different logic then any others.



POST architecture is divided into three parts.

i. Group creation,

ii. Main User Node

iii. Authenticator

- A. How it Works?
- 1) Group creation:

The group creation is done as that types of friends are decided. Here for this group we have to manage the user

friends are known as Direct Friend (DF) and Friends of that DF are Neighbour Friend (NF).

The combination of both will create a group. The NF will act as a guard of the group from outliers. The group creation should do carefully for proper management.

## 2) Main User Node (MUN):

The Main User Node is the main part of the POST architecture as it is the part which concern about head of group, that why it is Main User Node (MUN).

MUN is selected carefully as it gives the referrence of the all nodes resides in the group.

There are four situations for MUN selection.

*a)* Select MUN within the group.

Node with highest nodes of DF select as a MUN.

*b)* Select MUN for adjacent the group,

Node which is outside to group, to deside MUN takes help of NF. From many any one must be connected with any DF or NF. Thus with that it will suggest the MUN in adjacent community.

Select MUN for diffent MS group.

Here after process B completes the MS of both sender and reciever communicate with each other to form a new group.

Select MUN for new node.

Here for this problem first it starts communication with some authentication process. The authentication process may be any to check legitimity of user. May ask questions, or to do calculation or to identify numbers etc.

The POST is very much dependent on this MUN management. To successful impementation of POST MUN have to work properly

3) Authenticator:

c)

d)

This part of the system is responsible for legitimity selection of the user amonst the group. This part classifies the user type: Regular User, Spammer and New user.

B. Proposed Algorithm

The POST will implement the SAA (Sender Authenticaion Algorithm). This algorithm copes with spam sender. This algorithm checks the legitimity of the user based upon the group of the sender group.

This algorithm, with help of authenticator checks the rank or status of the sender, if the sender ranked worst by MUN (Main User Node), Sender is blocked.

Sender is first checked from group which are already created. Then after on stages checks from next adjacent group and then after it will direct to find MUN.

The MUN finds based on sender's group. Next MUN gives suggestion about sender nodes. If it is find *legitimate*, *new* or *illegitimate* (spammer)

#### V. CONCLUSION

Spam prevention is the biggest problem in today's social world, and to cope with them POST implements a newer approach. The paper, introduce POST, a spam protection system based on the social networking paradigm.

With the help of Mail User Node (MUN) it counters spams that are not from individuals' social circle. This approach help user to be free from spam attacks and free from attacks of data which are totally irrelavents. It is also desirable that POST never stops the legitimate senders and never catches legitimate mail as a spam mail.

We hope that POST remains lightweight and give better performance than any other spam filtering tool or techniques. Also much needed that POST will not decrease performace of the system as numbers of users increase.

## VI. ACKNOWLEDGEMENTS:

I am heartily thankful to Mr Rahul Mandaliya, InfoString TechnoLabs, and Gujarat, India. He supports a lot to develop the system. Also thankful to internal guide Prof. Chetan Ajudiya, who encourages me and provides necessary guidance I also want to thank my family and friends for helping and supporting me

### REFERENCES

- [1] Enrico Blanzieri and Anton Bryl, "A survey of learning based techniques of Email spam filtering", A Technical Report, University of Trento, 2008..
- [2] P. O. Boykin and V. Roychowdhury. Personal email networks: An effective anti-spam tool. IEEE COMPUTER, 2004.
- [3] Paul Alexandru Chirita, J<sup>\*</sup>org Diederich, and Wolfgang Nejdl. Mailrank: using ranking for spam detection. In Proc. of CIKM.
- [4] S. Garriss, M. Kaminsky, M. J. Freedman, B. Karp, D. Mazi'eres, and H. Yu. Re: Reliable email. In Proc. of NSDI, 2006.
- [5] Shuang Hao, Nadeem Ahmed Syed, Nick Feamster, Er G. Gray, and Sven Krasser. Detecting spammers with SNARE: Spatio-temporal network-level automatic reputation engine. In USENIX Security, 2009.
- [6] Z. Li and H. Shen. SOAP: A social network aided personalized and effective spam filter to clean your e-mail box. In Proc. of IEEE INFOCOM, 2011.
- [7] A. Mislove, A. Post, P. Druschel, and KP Gummadi. Ostra: Leveraging trust to thwart unwanted communication. In Proc. of NSDI, 2008.
- [8] M. Sirivianos, K. Kim, and X. Yang. Introducing social trust to collaborative spam mitigation. In Proc. of IEEE INFOCOM, 2011.
- [9] M. Perone. An overview of spam blocking techniques. Technical report, Barracuda Networks, 2004.
- [10] A. Gray and M. Haahr. Personalized, Collaborative Spam Filtering. In Proc. of the Conference on Email and Anti-Spam (CEAS), Mountain View, CA, USA, July 2004.
- [11] J. Golbeck and J. Hendler Reputation Network Analysis for Email Filtering. In Proc. of the Conference on Email and Anti-Spam (CEAS), Mountain View, CA, USA, July 2004.
- [12] Isode: Benchmark and comparison of Spamassassin and m-switch anti-spam, Technical report, Isode, April 2004.
- [13] An MCL-Based Approach for Spam Profile Detection in Online Social Networks Faraz Ahmed, Muhammad Abulaish, IEEE 2012, Computing and Communication.
- [14] Preventing Unwanted Social Inferences with Classification Tree Analysis Sara Motahari, Sotirios Ziavras, Quentin Jones, IEEE, 2009

- [15] Detecting Spammer on Twitter Fabr'icio Benevenuto, Gabriel Magno, Tiago Rodrigues, and Virg'ilio Almeida
- [16] Spamassassin. http://spamassassin.apache.org/
- [17] http://royal.pingdom.com/2013/01/16/internet-2012-innumbers/
- [18] YouTube. http://www.youtube.com.
- [19] http://projects.puremagic.com/greylisting.

