ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Managing Mass-Mailing System in Distributed Environment

V.Srikanth
M.Tech,
SS InfoTech, Hyderabad.

Abstract:

We investigate the problems in mailing system. This is the new approach to protect the mails and detecting the spam mails. This system is helpful for detect spam mails and filtering mails using key serves. we implements the server when get the message, a key server encrypt and pass the message to another server and encrypt message using key servers. The present system is difficult to identifies the accuracy spam mails. Here we are developing one frame work for spam detection.

Introduction:

E-mail is one of the most important communication ways in the Internet. Sometimes an organization work depends on mail system functionality, and somewhere mail transfers may form a large part of total network traffic. A special case we have when organization use one mail domain in a large and geographically distributed network (e.g. central office and regional departments), because it becomes impossible to keep even internal mail messages inside one local network. The goal of the work was to design a fast, convenient and reliable distributed mail system.The paper starts with a short overview of the mails filtering and detecting the spam mails. This project is develop in pop3 protocols. Pop3 protocols can not be support the any child protocols.

I.Pop protocols :

On certain types of smaller nodes in the Internet it is often impractical to maintain a message transport system (MTS). For example, a workstation may not have sufficient resources (cycles, disk space) in order to permit a SMTP server [RFC821] and associated local mail delivery system to be kept resident and continuously running. Similarly, it may be expensive (orimpossible) to keep a personal computer interconnected to an IP-style network for long amounts of time (the node is lacking the resource known as “connectivity”).

Despite this, it is often very useful to be able to manage mail on these smaller nodes, and they often support a user agent (UA) to aid the tasks of mail handling. To solve this problem, a node which can support an MTS entity offers a maildrop service to these less endowed nodes. The Post Office Protocol – Version 3 (POP3) is intended to permit a workstation to dynamically access a maildrop on a server host in a useful fashion. Usually, this means that the POP3 protocol is used to allow a workstation to retrieve mail that the server is holding for it. POP3 is not intended to provide extensive manipulation operations of mail on the server; normally, mail is downloaded and then deleted. A more advanced (and complex) protocol, IMAP4, is discussed in. For the remainder of this memo, the term “client host” refers to a host making use of the POP3 service, while the term “server host” refers to a host which offers the POP3 service. We implement this project in pop3 environmental.

2. Spam classifications :

In this digital age, which is the era of electronics & computers, one of the efficient& power mode of communication is the email. Undesired, unsolicited email is a nuisance for its recipients; however, it also often presents a security threat. For ex., it may contain a link to a phony website intending to capture the user’s login credentials (identity theft, phishing), or a link to a website that installs malicious software (malware) on the user’s computer.Installed malware can be used to capture user information, send spam, host malware, host phish, or conduct denial of service attacks as part of a “bot” net. While prevention of spam transmission would be ideal, detection allows users & email providers to address the problem today [1].Spam filtering has become a very important issue in the last few years as unsolicited bulk e-mail imposes large problems in terms of both the amount of time spent on and the resources needed to automatically filter those messages [2]. Email communication has come up as the most effective and popular way of communication today. People are sending and receiving many messages per day, communicating with partners and friends, or exchanging

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

files and information. E-mail data are now becoming the dominant form of inter and intra-organizational written communication for many companies and government departments. Emails are the essential part of life now just likes mobile phones & i-pods [2]. Emails can be of spam type or non-spam type as shown in the Fig.1. Spam mail is also called as junk mail or unwanted mail whereas non-spam mails are genuine in nature and meant for a specific person and purpose. Information retrieval offers the tools and algorithms to handle text documents in their data vector form [3].The Statistics of spam are increasing in number. At the end of 2002, as much as 40 % of all email traffic consisted of spam. In 2003, the percentage was estimated to be about 50 % of all emails. In 2006, BBC news reported 96 % of all emails to be spam. The statistics are as shown in the following table I. Spam can be defined as unsolicited (unwanted, junk) email for a recipient or any email that the user do not wanted to have in his inbox.It is also defined as "Internet Spam is one or more unsolicited messages, sent or posted as a part of larger collection of messages, all having substantially identical content."There are severe problems from the spam mails, viz., wastage of network resources (bandwidth), wastage of time, damage to thePC's & laptops due to viruses & the ethical issues such as the spam emails advertising pornographic sites which are harmful to the young generations [5]. Some of the existing approaches to solve the problem from spam mails could be listed as below ….Email is the most widely used medium for communication world wide because it's Cheap, Reliable, Fast and easily accessible. Email is also prone to spam emails because of its wide usage,cheapness & with a single click you can communicate with any one any where around the globe. It hardly cost spammers to send out 1 million emails than to send 10 emails. Hence, Email Spam isone of the major problems of the today's internet, bringing financial damage to companies and annoying individual users [4].

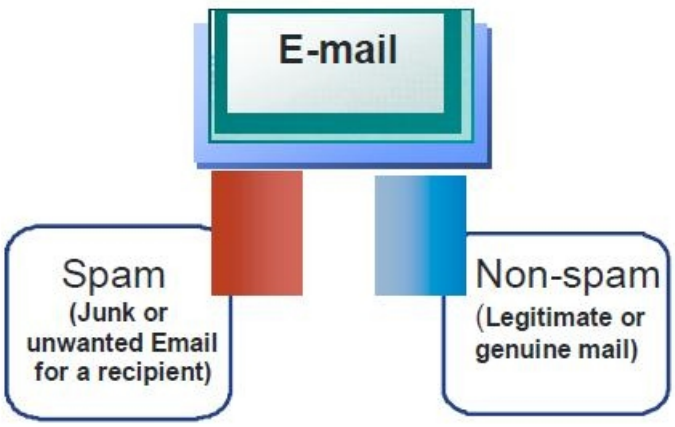


TABLE I. STATISTICS OF THE SPAM MAILS

Daily Spam emails sent	12.4billion
Daily Spam received per person	6
Annual Spam received per person	2,200
cost to all non-corporate Internet users	$255 mill
Spam cost to all U.S. Corporations in 2002	$8.9 billion
Email address changes due to Spam	16%
Annual Spam in 1,000 employee company	2.1 million
Users who reply to Spam email	28%

3. Spam Filtering Methods:

Though the first spam was sent in 1978 it began to be written about it as a problem in scientific literature only from 1982. One of the first papers where this problem is considered is the Peter J. Denning's article[4]. The first mathematical apparatus applied to spam filtering systems is the Bayes' algorithm, which was used first by Sahami et.al in 1996 and then by other researchers [5-8]. Bayes' classifier relies on famous Bayes theorem and the first papers about it could be met as early as 1960 [9]. During more than 40 year history Naive Bayes Classifier (NBC) was used for the solution of very different type of tasks: from classification of texts in news agencies till primary diagnosis of diseases in medicine. For the problems where NBC is appliedthere is usually selected presence or absence of words in the text as a characteristic, i.e. the set of characteristics T is a set off all words in documents. Hereby, if the word i is present, the weight of charac-teristics t1iw otherwise . In case of e-mail filters where spam classification is used, there taken into account the area where the word had been met: heading, subject and body of the e-mail. wi=1 , otherwise wj = 0Beginning from the publication of Gary Robinson [10], in some filters (for example, Spam Assassin) there came to be used the method of overlapping probabilities sug-

Our method Suggested method :

The probabilities that a message contain the given word : Let's suppose the suspected message contains the word "replica". Most people who are used to receiving e-mail know that this message is likely to be spam, more precisely a proposal to sell counterfeit copies of well-known brands of watches. The spam detection software, however, does not "know" such facts, all it can do is compute probabilities.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

The formula used by the software to determine spam mails.

$$\Pr(S|W) = \frac{\Pr(W|S) \cdot \Pr(S)}{\Pr(W|S) \cdot \Pr(S) + \Pr(W|H) \cdot \Pr(H)}$$

where:

• is the probability that a message is a spam, knowing that the word "replica" is in it;
• is the overall probability that any given message is spam;
• is the probability that the word "replica" appears in spam messages;
• is the overall probability that any given message is not spam (is "ham");
• is the probability that the word "replica" appears in ham messages.

4. Pop3 Implementation :

The Post Office Protocol (POP) is designed to allow a workstation(PC) to dynamically access a maildrop on a server host. POP3 is the version 3 (the latest version) of the Post Office Protocol. In other words, POP3 allows a workstation to retrieve mail that the server is holding for it. POP3 transmissions appear as data messages between stations. The messages are either command or reply messages. There are several different technologies and approaches to building a distributed electronic mail infrastructure. Among them: POP (Post Office Protocol), DMSP (Distributed Mail System Protocol), and IMAP (Internet Message Access Protocol). Of the three, POP is the oldest and consequently the best known. DMSP is largely limited to a single application, PCMAIL, and is known primarily for its excellent support of "disconnected" operation. IMAP offers a superset of POP and DMSP capabilities, and provides good support for all three modes of remote mailbox access: offline, online, and disconnected.

POP was designed to support "offline" mail processing, in which, mail is delivered to a server, and a personal computer user periodically invokes a mail "client" program that connects to the server and downloads all of the pending mail to the user's own machine. The offline access mode is a kind of store-and-forward service, intended to move mail (on demand) from the mail server (drop point) to a single destination machine, usually a PC or Mac. Once delivered to the PC or Mac, the messages are then deleted from the mail server.

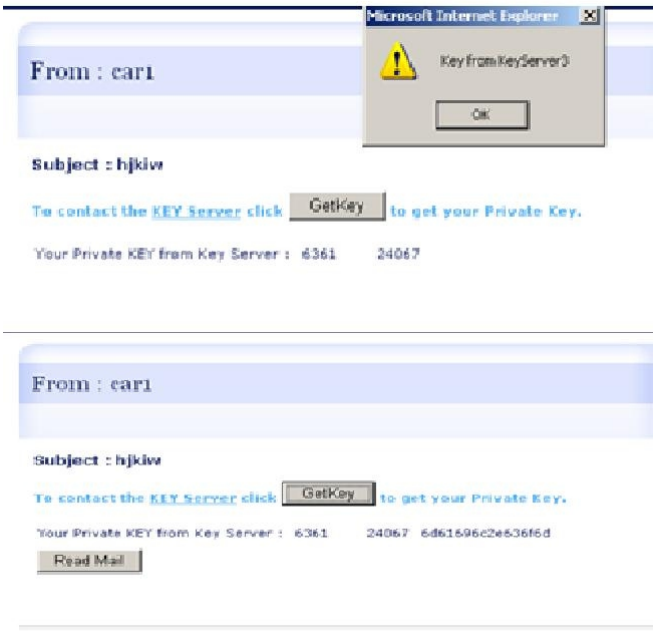
POP3 is not designed to provide extensive manipulation operations of mail on the server; which are done by a more advanced (and complex) protocol IMAP4. POP3 uses TCP as the transport protocol.

5. Mail Filtering Techniques :

However, any large-scale collaborative anti-spam approach is faced with a fundamental and important challenge, namely ensuring the privacy of the e-mails among entrusted e-mail entities. Different from the e-mail service providers such as Gmail or Yahoo mail, which utilizes spam orham(non-spam) classifications from all its users to classify new messages, privacy is a major concern for cross-enterprise collaboration, especially in a large scale. The idea of collaboration implies that the participating users and e-mail servers have to share and exchange information about the e-mails (including the classification result). However, e-mails are generally considered as private communication between the senders and the recipients, and they often contain personal and confidential information.

Therefore, users and organizations are not comfortable sharing information about their e-mails until and unless they are assured that no one else (human or machine) would become aware of the actual contents of their e-mails. This genuine concern for privacy has deterred users and organizations from participating in any large-scale collaborative spam filtering effort.To protect e-mail privacy, digest approach has been proposed in the collaborative anti-spam systems to both provide encryption for the e-mail messages and obtain useful information (fingerprint) from spam e-mail. Ideally, the digest calculation has to be a one-way function such that it should be computationally hard to generate the corresponding e-mail message. It should embody the textual features of the e-mail message such that if two e-mails have similar syntactic structure, then their fingerprints should also be similar.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal



CONCLUSION:

In this paper we are protecting the spam mails and filtering mails. We are suggesting the one method that's support efficient techniques for spam filtering. We are going to implanting the pop3 protocols which can not use any smtp protocols.

REFERENCES:

[1]Sudipto Guha, Adam Meyerson, Nina Mishra, Rajeev Motwani and LiadanO'Callaghan, "Clustering Data Streams,"IEEE Trans.s on Knowledge & Data Engg., 2003.

[2] Enrico Blanzieri and Anton Bryl, "ASurvey of Learning- Based Techniques of Email Spam Filtering," Conference onEmail and Anti-Spam., 2008.

[3]Jain A.K., M.N. Murthy and P.J. Flynn, "Data Clustering : A Review,"ACM Computing Surveys., 1999.

[4]Tian Zhang, Raghu Ramakrishnan, Miron Livny, "BIRCH: An Efficient Data Clustering Method For Very Large Databases," Technical Report, Computer Sciences Dept., Univ. of Wisconsin-Madison, 1996.

[5] Porter. M, "An algorithm for suffix stripping", Proc. Automated library Information systems, pp. 130-137, 198.