ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

Examine the E-Mail Spam Detecting and Email Spam Filtering Techniques

V.Charumathi MCA, S.Thirumal MCA, MPHIL.

M. Phil Scholar, Department of Computer Science, Arignar Anna Arts College, Cheyyar, India

Assistant and Head, Department of Computer Science, Arignar Anna Arts College, Cheyyar, India

ABSTRACT: All over the world, millions of people use email to communicate around the world. The spam is unwanted bulk of data that a user receives in the form of email without user knowledge. so, this spam may increase the memory storage, bandwidth, network resources and also be used for some attack. This attack may destroy uses identity or data. In this paper we discuss some of the email detection and spam filtering techniques are list based filtering, content based filtering, Bayesian filter, collaborative spam filtering etc., to reduce spam storage and bandwidth and also secure our data in effective manner.

I. INTRODUCTION

[5]Electronic mail (email) is a method of exchanging messages between people using electronics. Email operates across computer networks, which today is primarily the Internet.
[6] Email is much older than ARPANet or the Internet. It was never invented; it evolved from very simple beginnings. Early email was just a small advance on what we know these days as a file directory - it just put a message in another user's directory in a spot where they could see it when they logged in. Simple as that Just like leaving a note on someone's desk. Probably the first email system of this type was MAILBOX, used at Massachusetts Institute of Technology from 1065. Another early program to send messages on the same computer was called SNDMSG. Some early email systems required both user to be online at the same time, in common with instant messaging. Today's email systems are based on a store-and-forward model. Email servers accept, forward, deliver, and store messages. Internet email was extended by Multipurpose Internet Mail Extensions (MIME) to carry text in other character sets and multimedia content attachments. Ray Tomlinson picked the @ symbol from the computer keyboard to denote sending messages from one computer to another. So then, for anyone using Internet standards, it was simply a matter of nominating name-of-the-user@name-of-the-computer **. [9]** E-mail message comprises of different components: E-mail Header, Greeting, Text, and Signature .This was particularly useful in parts of the world where telephone costs to the nearest email system were expensive. (often this involved international calls in the early days) With connection charges of many dollars a minute, it mattered to be able to prepare a reply without being connected to a telephone, and then get on the network to send it.
It was also useful because the "offline" mode allowed for more friendly interfaces. The first important email standard was called SMTP, or simple message transfer protocol.

Electronic spamming is the use of electronic messaging systems to send an unsolicited message (spam), especially advertising, as well as sending messages repeatedly on the same site. Spam in email started to become a problem when the Internet was opened up to the general public in the mid-1050s. [1]Spam was created by Hornell in 1037. It was originally named “Hornell Spiced Ham”, but was eventually changed to the catchier name, “SPAM”. Usually they come in the form of advertisement, sometimes even containing explicit content or malicious code. According to the statistics from ITU (International Telecommunication Union), 70% to 80% of emails in the internet are spam’s which have become worldly problem to the information infrastructure. In order to control the problem are used in lot of anti-spam methods.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

II.EMAIL SERVICE PROVIDERS

[10]The following table shows the popular email service providers:

S.No.	Service and Description
1.	**Gmail** Gmail is an email service that allows users to collect all the messages. It also offers approx 7 GB of free storage.
2.	**Hotmail** Hotmail offers free email and practically unlimited storage accessible on web.
3.	**Yahoo Mail** Yahoo Mail offers unlimited storage, SMS texting, social networking and instant messaging to boot.
4.	**iCloud Mail** iCloud Mail offers ample storage, IMAP access, and an elegantly functional web application.
5.	**ATM Mail** ATM Mail is a free email service with good spam protection.
6.	**Mail.com** and **GMX Mail** Mail.com and GMX Mail offers reliable mail service with unlimited online storage.
7.	**Shortmail** Shortmail offers easy and fast email service but with limited 500 characters per message.
8.	**Inbox.com** Inbox.com offers 5 GB of free online storage. IMAP is not supported by Inbox.com
5.	**My Way Mail** My Way Mail offers clean and fast free email service but lacks in secure messaging.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

III. EMAIL SPAM

[4]Spam is abuse of electronic messaging system to send unsolicited bulk messages. Today large volumes of spam emails are causing serious problem for the users, and internet services. Such as, It degrades user search experience, It assists propagation of virus in network, It increase load on the network traffic, It wastes the resources such as bandwidth, storage, and computation power, It also wastes the user time and energy. While the most widely recognized form of spam is email spam, the term is applied to similar abuses in other media: instant messaging spam, Usenet newsgroup spam, Web search engine spam, spam in blogs, wiki spam, online classified ads spam, mobile phone messaging spam, Internet forum spam, junk fax transmissions, social spam, spam mobile apps, television advertising and file sharing spam.
[8]General advices to avoid spam's are use the spam filter, Never reply the spam, Don't post your email address on your web site, and Never buy anything from spam.

A. TECHNIQUES USED BY SPAMMERS

[7]In this section, we will discuss the different techniques used by the spammers.

- **Domain Spoofing** − The spammer sends an email on behalf of a known domain so the receivers think that they know this person and open it.
- **Poisoning Filters** − A filter can be poisoned by adding text with the same color of the background to reduce the scoring of the filters.
- **Directory Harvesting** − In directory harvesting, spammers generate email addresses by using known email addresses from corporates or ISP (Internet Service Provider).
- **Social Engineering** − Spammers send promotional emails to different users such as offering huge discounts and tricking them to fill their personal data.
- **Junk Tags** − Spam Words can be hidden by including invalid HTML tags within the words.
- **Invalid words** − Special characters are inserted in the spam words. For example: V!AGRA.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

IV.LITERATURE REVIEW

A. SPAM FILTERING TECHNIQUES

- **Blacklist:**

[1]Black list is the form of rule based filtering that uses one rule to decide which emails are spams. Black list are the list of IP address of machine or record of email addresses that have been previously used to send spam. When incoming message arrives, the spam filter checks to see if it's IP or email address is on the black list, if so, the message is considered spam and rejected. Blacklist can be used for on both large scale and small scales [1].

Advantage is it can block substantial amount of email.
Disadvantage is a blacklist provider can block an entire net block range instead of just an individual IP.

- **White list:**

[11]A white list is a list or register of entities that provide a particular privilege, service, mobility, access or recognition. Entities on the list will be accepted, approved and/or recognized. White listing is the reverse of blacklisting.

Disadvantage:[3]It is difficult to insert all possible sender address in the list.

- **Black holes:**

[1]Spam Black holes work hand in hand with Blacklist. The way Black holes work is someone posts message on websites, Usenet, forum etc., showing their email address. The email address they use is generally a machine account that detects who sent the spam and the IP address of to a DNS Blacklist.

Advantage is the email is received from one of these addresses the sending server can added to a Blacklist stopping it from sending any more messages.
Disadvantage is it can't see any disadvantages to using Black holes in order to detect spam, they are important as they enable blacklist to be updated with computers that are sending unwanted emails.

- **Grey lists**

[1]The grey list system, the receiving mail server initially rejects messages from unknown users and sends a failure message to the originating server. If the mail server attempts to send the message second time- a step most legitimate server will take – the grey list assumes the message is not spam and let it proceed to the recipient's inbox. At this time grey list filter will add the recipient's email or address to a list of allowed senders. Though grey list filter require fewer system resources than some other types of spam filters, they also delay mail delivery, which could be inconvenient when you're expecting time sensitive messages.

Advantage: [2] Gray list filters require fewer system resources
Disadvantage:[2] delay mail delivery, which could be inconvenient when you are expecting time-sensitive message.

- **Content Based Filter**

[1]Content Based Filter is the most commonly used group of methods to filter spam. Content filter act either on the content, the information contained in the mail body, or on the mail headers (like "Subjects") to either classify, accept or reject a message.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

- **Bayesian Filter**

[1]It is considered to be a more advanced form of Content Based Filter, employ the laws of mathematical probability to determine which messages are legitimate and which are spam. Bayesian filter learn from both good and spam emails, result in an adapting and efficient anti_spam approach.

Bayesian email filter take the advantage of Bayesian theorem is

P (spam/word) = [P (word/spam) P (spam)] / p (word)

Theorem in the context of spam, says that the probability that an email is spam, given that it has certain words in it, is equal to the probability of finding those certain words probability that any email. The characteristics a Bayesian filter can look the words in the body of the message, its header(sender and message path) and HTML code, word pair, Phrase, and Meta information.

Advantages are it can be trained on per user basis, the spam that user receive is often related to the Online user's activities, it will assign high probability based on the user's specific patterns, the word Probability is unique to each user and can evolve over time with corrective training whenever the filter Incorrectly classifies an email.

Disadvantage is that they need to be trained properly in order for them work to work at their most effective and the training leads more time.

- **Collaborative filtering:**

  It is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users.
- **Word-Based Filtering**:

  A Word-Based Filtering is the simplest type of content-based filter. It simply block any email that contain certain terms.
- **Reverse Lookup:**

[3]In reverse lookup also known as a reverse DNS(DOMAIN NAME SYSTEM).The host is associated with a given IP address. By using this routine, the receiver can confirm the identity of the domain name of the sender.

Disadvantage: This technique is not effective for the mobile users and the users with invalid IP address.

V.CONCLUSION

Now a day's spam is a biggest problem in the web data. Spam causes a number of problems of both economical and ethical nature, which results in particular in the attempts of legislative definition and prohibition of spam. In this paper we discuss about email and spam detection techniques for rectifying problem like bandwidth, network resources, storage space and Trojan.

REFERENCES

1. Savitha Teli, Santoshkumar Biradar "Effective Spam Detection Method for Email", IOSR Journal of Computer Science (IOSR-JCE) e-ISSN 2278-0061,p-ISSN:2278-8727.
2. Akshay p.Gulhane, Shraddha A.Jalan, Sakshi Gudadahe, Ajinkya R.Bijwe,"Spam Filtering Methods for Email Filtering",International Journal Of Computer Science And Applications",vol 6,no:2,Apr 2013,ISSN:0974-1011.
3. Alireza Nemaney Pour, Raheleh Kholghi and Soheil Behnam Roudsari," MINIMIZING THE TIME OF SPAM MAIL DETECTION BY RELOCATING FILTERING SYSTEM TO THE SENDER MAIL SERVER",International Journal of Network Security & its Applications(IJNSA),Vol.4,No.2,March 2012.
4. https://en.wikipedia.org/wiki/Spamming
5. https://en.wikipedia.org/wiki/Email
6. http://www.nethistory.info/History%20of%20the%20Internet/email.html