# Spam : A Threat to Network Security in Digital Library and Information Centres

Subhajit Choudhury Biswanath Dey Prof. S. Kumar

**Abstract**

*The paper introduces with a brief history of spam. Defines the term spam as Electronic junk mail or junk newsgroup postings. The trick of spam operators. Enlightens with some tips of Spam handling. It also describes various tools of spam and a through investigation of spam filtering, its importance, and available software that can help for spam filtering and information security in an information system. Lastly, the importance of anti-spam filtering to libraries and information centres have been discussed. Concludes with suggestion for using anti-spam software.*

**Keywords :** Digital Library, Network Security.

## 1. INTRODUCTION

The history of spam starts in the year 1975 with RFC706 by Jon Postel which was *jj@cup.portal.com says: HELP ME!*

In 1980 a MUD is a multi-user-dungeon. That's a somewhat archaic term for a real time multi-person shared environment, which is to say a shared world where users can chat, move around and interact with locations and objects in the environment. MUDs were named that because the first reminded people of "adventure" or "Dungeons and Dragons" games that involved jointly exploring a cave or dungeon. Modern successors of the MUD include EverQuest and The Sims Online

In 1993, Richard Depew tried to make some changes in USENET. He advanced a somewhat controversial idea called retro-moderation, where newsgroups would be semi-moderated (that is to say, regulated so that not all postings would appear) through a moderator who canceled postings that broke the rules.

In April of 1994, the term was not born, but it did jump a great deal in popularity when two lawyers from Phoenix named Canter and Siegel posted a message advertising their fairly useless services in an upcoming U.S. "green card" lottery. This wasn't the first such abusive posting, nor the first mass posting to be called a spam, but it was the first deliberate mass posting to commonly get that name. They had posted their message a few times before, but on April 12, they hired an mercenary programmer to write a simple script to post their ad to every single newsgroup (message board) on USENET, the world's largest online conferencing system. There were several thousand such newsgroups, and each one got the ad. Quickly people identified it as "spam" and the word caught on. Future multiple postings soon got the appelation. Some people also applied it to individual unwanted ads that weren't posted again and again, though generally it was associated with the massive flood of the same message. It turns out, however, that the term had been in use for some time before the famous green card flood.[1]

## 2. DEFINITION

Electronic junk mail or junk newsgroup postings. Some people define spam even more generally as any unsolicited e-mail. However, if a long-lost brother finds your e-mail address and sends you a message, this could hardly be called spam, even though it's unsolicited. Real spam is generally e-mail advertising for some product sent to a mailing list or newsgroup. In addition to wasting people's time with unwanted e-mail, spam also eats up a lot of network bandwidth. Consequently, there are many organizations, as well as individuals, who have taken it upon themselves to fight spam with a variety of techniques. But because the Internet is public, there is really little that can be done to prevent spam, just as it is impossible to prevent junk mail. However, some online services have instituted policies to prevent spammers from spamming. [2]

### 2.1 Types of Spam

There are two main types of spam, and they have different effects on Internet users.

a. **Cancellable Usenet spam** is a single message sent to 20 or more Usenet newsgroups. Usenet spam is aimed at "lurkers", people who read newsgroups but rarely or never post and give their address away. Usenet spam robs users of the utility of the newsgroups by overwhelming them with a barrage of advertising or other irrelevant posts. Furthermore, Usenet spam subverts the ability of system administrators and owners to manage the topics they accept on their systems.

b. **Email spam** targets individual users with direct mail messages. Email spam lists are often created by scanning Usenet postings, stealing Internet mailing lists, or searching the Web for addresses. Email spams typically cost users money out-of-pocket to receive. Many people - anyone with measured phone service - read or receive their mail while the meter is running, so to speak. Spam costs them additional money. On top of that, it costs money for ISPs and online services to transmit spam, and these costs are transmitted directly to subscribers. [3]

## 3. HOW SPAMMERS OPERATE

Many spam emailers use tricks to get you to read their messages. For example, they use the "Subject:" line to entice to open the message. Because of the tricks spammers use to send the email to anyone, the email address may not even be visible in the "To:" line of the message, and almost never see the email addresses of the other people they sent the message to. The worst thing about spam, though, is that the spammers use tricks that help disguise the origin of their messages.

One of the spammer's most common tricks is to relay messages through the email server of an innocent third party. This tactic doubles the damages: both the receiving system and the innocent relay system are flooded with spam. And for any mail that gets through, often the flood of complaints goes back to the innocent site because it was made to look like the origin of the spam. Many spammers send their spam from a free account from a large ISP such as AOL, Yahoo!, or Hotmail, then abandon the account and open a new one to use for the next assault. Another common trick that spammers use is to forge the headers of messages, making it appear as though the message originated elsewhere. This is called spoofed email. There are some pieces of information in the full headers that the spammer cannot forge, but even after technical investigation into the source of the message, most often the resulting information leads to a dead end, usually an abandoned account or an innocent mail relay server. [4]

### 4.a Techniques of Spam Handling

Spam can be caught and removed from incoming e-mail stream, it's time to review just how it's caught at each step along the way. Though the names for the tools and technologies used will vary widely from one vendor or developer to another, nearly all automated spam handling occurs using one or more of six message-handling techniques.

**4.a.1 Pattern matching/text or content filters :** By searching incoming e-mail for matches to specific patterns in message headers, subject lines, message bodies and so forth, inspection of the character data in the message itself allows a great deal of spam to be identified as such and rejected.[6]

**4.a.2 Whitelist/user verification filters :** By comparing the sender of a message against a user-specified list of senders from whom incoming e-mail will be accepted, it's possible to separate senders from whom the user is willing to accept e-mail (except when it's infected with malware). Other senders can then be subjected to a verification test whereby they must respond to an e-mail that requires some interaction and intelligence on the recipient's part to prove that there's a real person on the other end of the message chain, not just a spam broadcasting program.

**4.a.3 Blacklist/address or domain blocking :** Numerous parties operate services that identify sender addresses (and sometimes, entire domains) where spam is known to originate. By placing such addresses or domains on a blacklist, e-mail that originates from them can be summarily blocked at any point in the path between sender and receiver. Most server-based spam-handling tools make some use of this approach, but it's best used in conjunction with one or more of the other approaches mentioned here.

**4.a.4 Rule-based message ranking/text or content filters :** Instead of accepting or rejecting messages outright on the basis of their content, particularly where certain text patterns are concerned, this technique assigns various weights to certain patterns and rejects only those messages whose combined weights exceed some specific threshold. Rules can be added or adapted over time and weights changed to reflect the actual composition of message traffic as it changes over time, which makes it more effective than simple pattern-matching approaches.[7]

**4.a.5 Statistical message analysis/text of content filters :** Interestingly, statistical analyses of identified spam as compared to possible spam using Bayesian probability models for which words or phrases are most likely to identify spam versus those most likely to identify desirable mail produces better spam identification than rules-based approaches. These filters are often called Bayesian filters because of their mathematical underpinnings. They are becoming increasingly popular in many anti-spam handling tools and services.

**4.a.6 Message screening for malware, virus, worm, Trojan, etc. :** If a message contains identifiable malware of some kind, it's probably not a good idea to deliver it no matter what's in the message body (though some tools, in the interests of free speech, simply strip out infected content or attachments and send the message body on unaltered, except to note that something infected was removed). This kind of screening and handling should be applied to all e-mail no matter what kinds of other spam handling techniques may be applied.[8]

**4.b Technology : Anti-Spam Filter Software :** In this section some anti-spam software have been discussed in brief for use in library and information centers.

**4.b.1 Email Protect from Content Watch :** EmailProtect is an outstanding tool that lets you integrate your email filter preferences easily. EmailProtect comes with predefined categories that intelligently label spam several ways; settings let you filter incoming email by category, email address, server, domain, and even by specific key words. These detailed options allow you to fine-tune your blocking so you only block what you consider spam; this is one of few filters that allows such customization. EmailProtect also has a white list for those who want to receive email from specific address regardless of content.

**4.b.2 SpamEater Pro from High Mountain :** It works independently of your inbox. It supports any of the standard POP3 mailboxes, and with the use of 3rd party applications you can configure SpamEater Pro to work with Hotmail, Yahoo, and other web based accounts. SpamEater Pro supports an online blacklist databases, these databases are maintained and updated daily. It also has a built-in rule set that will also allow you to customize your own set of rules for your particular needs.

**4.b.3 Qurb :** Qurb is an excellent Spam filtering application. Once installed, your inbox will only accept email from email addresses on your approved senders list. All other email will be quarantined in the Qurb folder where you can review them at your convenience. When messages arrive from an unknown sender, Qurb will notify you and redirect them to your Qurb folder.

**4.b.4 ChoiceMail One by DigiPortal :** It is an anti-spam program that puts spam control in your hands. Instead of using the traditional spam filtering tools, ChoiceMail One assumes everything is spam until one grant permission to let it through.

**4.b.5 Spam Killer :** Spam Killer does have several filter options to help fine-tune your spam filtering needs and allows easy updates from its main menu. It also supports Hotmail and Exchange accounts, not just POP3. But it does lack a lot of the necessary tools you would like to see it have to help in fine-tune and eliminate the junk mail and spam one get daily. Spam Killer does have one annoying feature included in Spam Killer that you are unable to remove, and that's the McAfee Security Center, which is a part of the Spam Killer program.

**4.b.6 Spam Buster :** Spam Buster is an average spam filtering package that looks at the header of your email and the size in order to quickly determine if it's Spam or not. Spam Buster show poor results in cleaning out your junk mail and spam. Spam Buster also lacks the automatic updates like most spam filtering packages, no reporting capabilities and cannot import an address book.

**4.b.7 Matador :** Matador 3.5's rich feature set will allow the user to fine-tune the filter categories to the desired specifications in order to block (or unblock) Spam email. Matador 3.5 reporting capability is very good; reports can viewed by Junk by Day, Reasons, Total History, or by Total Reasons. Matador 3.5 provides the user with the ability to challenge an email by sending email verification to test whether it is a legitimate email or junk.

**4.b.8 SpamNet by CloudMark :** It is a very good spam-filtering package that relies on a peer-to-peer community to fight spam. As it contribute to the community of fighting spam, it will develop a rating of trustworthiness.

**4.b.9 Spam Agent :** Spam Agent is a very good Spam filtering product that will allow you to set up your own custom filters to aide in the removal of those unwanted emails. Spam Agent can filter by sender, subject, recipient, and may specify by attachment type and more. It is very easy to install and use, and Spam Agent will work in the background and monitor the incoming emails while in continue of work.

**4.b.10 iHateSpam :** It is one of the few packages that will allow you to set threshold levels to control how strict you want to monitor your incoming email. The default setting is "Average" which will quarantine most junk mail and reduce the chances of catching Spam. One can raise the level to "All Spam" which will quarantine all junk mail. Or, if anyone want minimal filtering, which can set it to "Some Spam".

## 5. CONCLUSION

Spam is flooding the Internet with many copies of the same message, in an attempt to force the message on people who would not otherwise choose to receive it. Most spam is commercial advertising, often for dubious products, get-rich-quick schemes, or quasi-legal services. Spam costs the sender very little to send — most of the costs are paid for by the recipient or the carriers rather than by the sender.

Library and information Centres are totally depending much on internet through various network. The use of email as communication media is a day-to-day inseparable phenomena in these centres. Apart from the library staff for their regular use, the users are also allowed to access the email from library and information centers. The chances of getting the spam on these centres are very high and obvious. It is observed that 10 of every 13 e-mail messages qualify as one form of spam or another. In percentage terms, that means nearly three-quarters, or 72.3 percent, of all e-mail messages that travel the Internet are unwanted garbage. The awareness of the spam and use of the anti-spam software has become a necessity for these cnetres.

## 6. REFERENCES

1. Brad, Templeton; "Origin of the term "spam" to mean net abuse"; in http://www.templetons.com/brad/spamterm.html <http://www.templetons.com/brad/spamterm.html> accessed on 16/09/2005 at 12.46 pm
2. Jupiters Corporation Ltd, Webopedia, Spam <http://www.webopedia.com/TERM/s/spam.html> accessed on 16/09/2005 at 3.40 pm
3. Choice, Australian Consumers' Association, "Spamming-a online epidemic" <http://www.choice.com.au/viewArticle.aspx?id=100899&catId=100518&tid=100008&p=2> accessed on 16/09/2005 at 3.57 pm
4. http://familyinternet.about.com/od/emailsafety/a/spam101operate.htm accessed on 16/09/2005 at 4.06 pm
5. www.pcmag.com/article2/0,1895,849550,00.asp accessed on 16/09/2005 at 4.10 pm
6. http://www.ecr6.ohio-state.edu/spam_handling.html
7. http://www.nenie.org/misc/spam.html
8. Media Tec Publishing Inc. "Choosing Anti-Spam Software" <http://www.certmag.com/articles/templates/cmag_tools_generic.asp?articleid=1001&zoneid=92> accessed on 16/09/2005 at 5.05 pm
9. www.SpamResearchCenter.com accessed on 16/09/2005 at 5.42 pm
10. http://www.qurb.com/get/ accessed on 16/09/2005 at 6.03 pm

## ABOUT AUTHORS

**Sh. Subhajit Choudhury,** IIT Guwahati, North Guwahati, 781 039, Assam. Registered for Ph.D (Library and Information Science), did M. Phil (LIS), MLIS, BLIS. Working in IIT Guwahati for last 09 years. Well versed in modern technology and computers. Published 10 articles including one in IFLA 2004.

**E-mail :** subhajit.lib@gmail.com, subhajit@iitg.ernet.in

**Mr. Biswanath Dey,** is working as a Computer Engineer at IIT Guwahati since last five years and doing his Ph.D. at Dept. of Computer Sc. & Engg., IIT Guwahati. Has two papers in his credit. Did B.E.(CSE) from Jorhat Engineering College, Jorhat and M.Tech.(IT), Gold Medalist from Tezpur University, Tezpur.

**E-mail :** bdey@iitg.ernet.in

**Prof. S. Kumar,** Reader & Head, S.S. in Lib. & Inf. Sc. (Faculty of IT), Vikram University, Ujjain (Madhya Pradesh) (India) With 34 years of teaching experience to Post Graduate Classes and 1 year to M.Phil. 16 years experience in guiding Ph.D. research, 6 students awarded Ph.D. Have 90 published papers in journals, conferences and seminars volumes. Presented papers in 49th FID (Jaipur) 1998 and International Congress on Digital Libraries (New Delhi) 2004. Besides 20 papers presented in other International & National Conferences and Seminars. Won Award of Commendation for paper presentation in National Conference in ILA. Have 2 books and more then 90 articles to the credit. Paper presented in IFLA World Library Conference, Argentina, 2004 as Speaker by co-author. Two papers have been accepted in IADIS International Conference, 2004 Spain and CISTA 2004 Joint meeting of the International Conference on Information System Analysis and Synthesis and 10th International Conference on Information System Analysis and Synthesis (ISAS 2004), Orlando, USA. Paper also published in IFLA World Library Conference, OSLO, 2005.

**E-mail :** sslisvikram@yahoo.com