CERT-In

Indian Computer Emergency Response Team

Enhancing Cyber Security in India

An Overview of SPAM: Impact and Countermeasures

Valsa Raj Uchamballi Sabyasachi Chakrabarty Basudev Saha

Department of Information Technology
Ministry of Communications and Information Technology
Govt. of India

Issue Date: March 16, 2005

Index

1. Introduction

Electronic Mail is used as a means of communication using computers on the Internet. Email can also be sent automatically to a large number of addresses. It's one of the most widely used facilities on the Internet. Electronic mailing systems were not designed with security as primary focus. Initial participants of the E-Mail infrastructure were limited and were based on trust relationships among them. The design shortcomings and flaws were later exploited by malicious users in various ways. Malicious activities with E-Mailing infrastructure included various forms of unsolicited mail, Identity theft, Denial of service etc.

Any e-mail message sent to multiple e-mail addresses where the recipients have little or no direct prior relationship to the topic of the communications is known as Unsolicited Bulk E-Mail (UBE). UBE is usually, but not always, spam. An UBE with contents which are commercial in nature is termed as Unsolicited Commercial Email (UCE). The contents of UBE or UCE may range from scams, false antivirus notifications, pornographic material, illegal medicines, chain letters, and religious or political spam.

Though there are various definitions for spam, broadly a message is termed as spam if - **[Ref 1]**, **[Ref 2]**

- The recipient's personal identity and context are irrelevant because the message is equally applicable to many other potential recipients
- The recipient has not verifiably granted deliberate, explicit, and still-revocable permission for it to be sent
- The transmission and reception of the message appears to the recipient to give a disproportionate benefit to the sender.

The key characteristics of spam are

- The unsolicited nature of the communication.
- The commercial focus of the communication.
- The significant volumes in which these messages are typically sent.

Spam is not just restricted to electronic mail but has now spread to other form of electronic communication. Spam over instant messaging (SPIM) and spam over internet telephony (SPIT) is widely prevalent. Spam on mobile devices is also a growing menace.

Origin of the word *Spam*

The original word *Spam* was coined in 1937 by the Hormel Corporation as a brand name for its blend of *spiced ham*. In 1970, a BBC comedy show featured a cafe that had a menu which featured food items with lots of *Spam*. The cafe was also filled with Vikings who periodically sang "Spam, spam, spam, spam ... lovely spam, wonderful spam ..."

Thus the meaning of the term is associated with something that keeps repeating and repeating to great annoyance.

Spam has been growing rapidly over the years. According to some estimates, 64% of the total Internet mail in the year 2004 was spam, while other estimates put the figure at 77% and some others even put estimates at 88%. [Ref 31, 32, 33]

In a survey, Message labs [Ref 34] identified 9.2 billion or about 73% of total Internet mail traffic as spam. A quarter of this spam in personal email accounts is considered to be pornographic in nature.

Spam has both direct and indirect costs. Direct costs are in the way of financial losses due to scams. However, they affect indirectly by reducing employee efficiency, consuming network resources, causing annoyance, and in the spreading of viruses/worms and other scams like phishing. Spam also adds a cost to the service provider and the end user in the form of filtering software, added bandwidth and storage. These costs, direct and indirect when combined, make spam a growing issue for companies and have a significant negative effect on business.

The worldwide cost of spam is rising enormously. According to a study by Radicati Group and Message Labs, the estimated worldwide cost to businesses due to spam is US$ 20.5 billion [Ref 33]. In fact, spam could be costing an average company US$ 4.1 million a year in lost productivity, according to a report published by IDC [Ref 32]. Another study by Forrester Research, 2003, found spam to be responsible for nearly US$20 billion in lost time and expenses worldwide [Ref 35].

History of *Spam*

The 25th anniversary of the earliest documented email mass mailing was on May 3rd 2003 and the 10th anniversary of the term Spam being applied to a USENET post was March 31st, 2003.

This first Spam ever written was in 1978 by a marketer at Digital Equipment Corporation. The mail was intended to be sent to every email address on the ARPANET. **[Ref. 19]**

One of the first mass mailings on Usenet was in January 1994 from Clarence L. Thomas IV with the subject was *Global Alert For All: Jesus is Coming Soon*. It was a long email about the end of the world. **[Ref. 20]**

Spam in the modern sense began in 1994 when two lawyers from Phoenix named Cantor and Siegel posted an advertisement for their services in the upcoming *US Green Card Lottery* **[Ref. 21]**. They posted their advertisement to every single newsgroup on USENET. Thousands who received the advertising message were not happy and one Usenet user wrote, "Send coconuts and cans of spam to Cantor & Co." Ever since the term spam, has been used to describe unsolicited messages. **[Ref. 22]**

Motives for spam can vary from being political to financial. Advertising of particular product through spam is on the rise. Spam is used for financial gain by baiting users in phishing attacks, or for other scams like the infamous Nigerian Scam **[Ref: 23]**. Spam can also be

used as a medium for distribution of malicious content. The motives for spam can also be political; it can target specific countries or even used for espionage. The ability to distribute messages in a short time to a huge audience with almost no cost to the spammer makes spamming so prevalent. Many organizations hire spammers to promote their product. There are cases of spammers earning money by selling database containing millions of e-mail addresses.

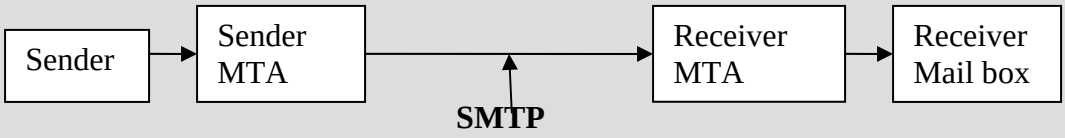
2. Spam techniques

2.1 SMTP Problem Areas

Simple Mail Transfer Protocol (SMTP) is the de facto protocol being used by modern mailing systems to exchange electronic mails over the Internet. SMTP is defined in RFC 821 **[Ref: 6]**.

Simple Mail Transfer Protocol

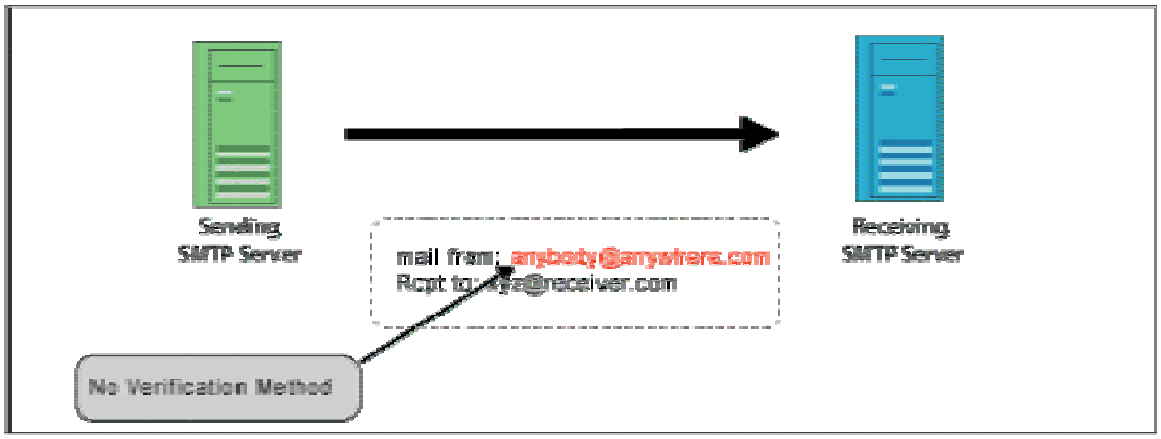
SMTP works between two communicating Message Transfer Agents (MTA) on TCP port 25 using a defined set of commands. The sending MTA accepts mail from end users and delivers the mail to MTA of the recipient domain using SMTP protocol. A typical command sequence between two communicating MTA is as follows




```
S: HELO sender.com
R: 250 receiver.com Hello sender.com, pleased to meet you

S: MAIL FROM: <sender@sender.com>
R: 250 OK

S: RCPT TO: <receiver@receiver.com>
R: 250 OK

S: DATA
R: 354 Start mail input; end with <CRLF>.<CRLF>
S: …..The message text goes here...
S: ...text...
S: <CRLF>.<CRLF>
R: 250 OK
```

The Sender MTA locates the IP address of the receiving MTA through a MX query on the DNS server. Sendmail was one of the first MTA to implement SMTP.

The claimed source identity in a spam message can either be authentic or false. Some users or organizations send bulk mail with their original identity, without forging the sender name and domain name.
However, in most cases, the sender of the spam forges the user name and domain name. Spammers also try to impersonate as legitimate users or organizations. The forging of identity or impersonation of identity in electronic mailing systems is done by exploiting a weakness in the SMTP protocol and its implementation. SMTP does not include any sender authentication to verify the authenticity of the sender.



Along with the flaws of SMTP protocol, various other flaws like open relays and open proxies are used by spammers to send spam. Various mass mailing worms also exploit these flaws. Filtering of sender addresses can be effectively used to reduce this kind of unsolicited bulk email. Legislative controls can also help in minimizing such spam.

2.1.1. Open relay

Internet was designed for redundancy. In its original design, an SMTP server was designed to accept mail destined for any other domain and forward it appropriately. This facility is known as relaying. However, this facility began to be misused by spammers.

An SMTP server is said to be “relay-enabled” if it accepts mails even if the “RCPT TO:” domain address is not within its defined list of served domain names, and processes the mail for delivery.

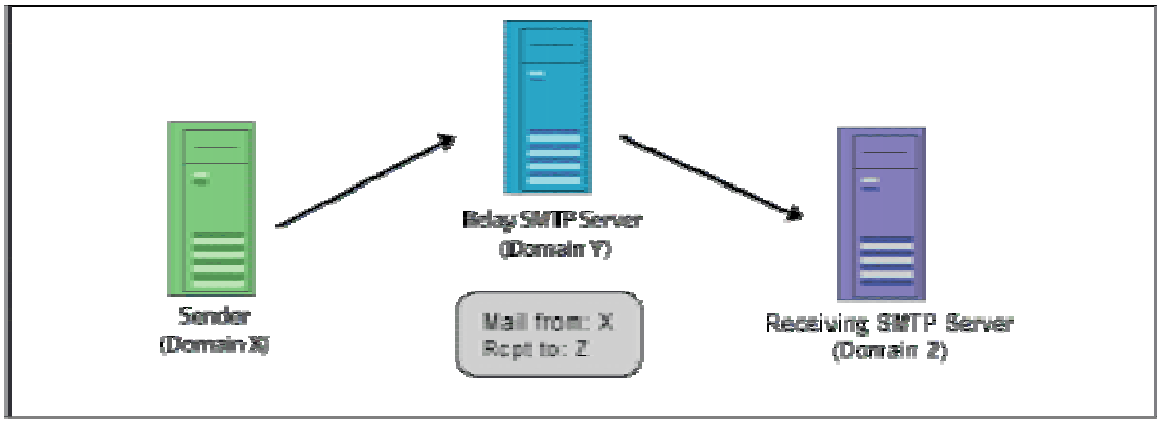


Fig 2

Email clients (eg.: Eudora, Outlook Express, Mozilla Thunderbird) communicate and submit their mail to an intermediate mail server for delivery. The email clients usually use SMTP to communicate to the relay-enabled server, which in turn forwards the mail for delivery, to the recipient.

However, to prevent misuse of this relaying by spammers, the mail servers use various mechanisms to authenticate such mail submission from email clients.

When there is no authentication mechanism to prevent such misuse of the relay and any sender can send his mail to another domain using the Mail relay, it is called an 'Open relay'. An Open Relay doesn't restrict any client from forwarding mails to another domain through it.

Earlier versions of most SMTP server software had relay 'Open' by default and thus open relays were widely prevalent. The Open relays are used by spammers to forward spam anonymously and to bypass spam filters.

2.1.2. Stealth & open proxies

A Proxy Server should accept requests only from its own clients by either forcing a client to connect only from a range of IP addresses, or by using authentication. Any proxy server that doesn't restrict its client base to its own set of clients and allows any other client to use it is known as an *Open Proxy*. An *Open Proxy* will accept client connections from any IP address and allow it to make connections to any Internet resource. *Open Proxy* servers act as a blind intermediary to any other network addresses without any authentication.

Open Proxies are used by spammers for mass mailing using forged identities. An Open Proxy can be used by a spammer to anonymously connect to a mail server. Further, any mail that a spammer sends shall appear to originate from the *Open Proxy* server. The use of an *Open Proxy* can also be used to bypass filters based on both domain name and IP address.

2.1.3. Mass Mailing Worms

Mass-mailing worms are a general class of malicious code which propagate through email. They generate high volumes of traffic that clog networks and overload mail relays. A lot of spam originates from home computers infected with various mass mailing worms. According to one estimate, about Four-fifths of spam is a result of these mass mailing worms. [Ref: 37]

The majority of mass-mailing worms target Windows systems and most of these worms entice an unsuspecting user to open an attachment containing malicious executable payload. Some worms also exploit vulnerabilities in widely used mail user agents (eg. Microsoft Outlook Express), to execute on the target system, without the need for a user to open any malicious attachments.

Though most mass mailing worms spread through emails, they also spread through other means including "default shares" of Windows. Newer mass mailing worms generally contain their own SMTP engine and are capable of bypassing filters or detection mechanisms deployed at the outgoing mail server.

The first such mass mailing worm known as Melissa was launched in March 1999. Since then various mass-mailing worms have been released, causing a lot of damage. Some of the worms that caused significant damage in the recent past are Sobar, Netsky, Klez, Beagle, and Mydoom.

2.1.4. 'Spam zombie' machines

Zombie machines are normally computers with broadband connectivity taken over by malicious users. They are said to be compromised by Remote Access Trojans (RAT) and can be remotely triggered by hackers to launch a massive attack against a target.

Spammers exploit these systems by turning them in to 'Spam zombie' machines, programmed to send out spam. According to some estimates, about one-third of spam originates from these zombie computers. **[Ref: 24]**

In the recent past, a trojan proxy server called *Mitglieder* appeared in the first week of 2004 infecting through IRC channels and turning infected machines into zombie proxy servers. It was widely exploited for spamming.

Interestingly, zombies have also been used against spammers. Lycos was in the news for creating a sort of a zombie army in a bid to bring down the world's most notorious spammers by distributing a screensaver called '*Make love not spam*' which launched a distributed denial-of-service attacks on spammers' Web sites. However, the distribution of the screensaver was discontinued after protests from the Internet community, on this proactive move against the spammers.

2.2 How do spammers get hold of email addresses?

The Internet is full of resources which can be used, directly or indirectly, as databases of email addresses. These include publicly available mailing list archives, USENET, web pages,

member directories and search engines. Spammers obtain email addresses, using various techniques including manual and automated software also known as crawlers, which crawl the web searching out email addresses. The email addresses are obtained by spammers from various other sources. They include online Yellow pages, whois records and other public postings.

2.2.1. Search engines

Queries on Internet search engines can be used to search email addresses. Spammers use manual and automated crawlers to harvest email addresses from search engines. Software that can extract and collect email addresses from different search engines is also available.

2.2.2. Addresses Posted in the Public Domain

Email addresses posted on websites are collected from the HTML pages by spammers using manual and automated crawlers. Spammers use programs which spider through web pages, looking for email addresses, e.g. email addresses contained in *MAILTO:* HTML tags. Email addresses published on the Internet usually receive a huge amount of spam.

On mailing lists and USENET, the email address is displayed in the *FROM:*, *REPLY TO:* , header of each message, and thus easily identified and harvested by spammers. Many on-line discussion forums are also archived on the Internet and addresses from these can be harvested by spammers using manual and automated methods.

Directories of email addresses are published on various white and yellow pages which are available online. Email addresses are harvested by spammers from these directories, using both manual and automated crawlers. Some of these crawlers can even extract the actual email address even if there is attempt to hide the address by jumbling or appending words.

A widely used technique to fight harvesting of email addresses from websites is the 'poison' CGI script. The script creates a page with several bogus email addresses with a link to itself. Crawlers of spammers' visiting the page would harvest the bogus email addresses and follow up the link. In this case, they enter an infinite loop thus polluting their lists with bogus email addresses.

2.2.3. Readymade lists

Readymade lists of email addresses are available for purchase on the Internet or otherwise. Spammers are also said to sell or exchange readymade email address lists collected through various means.

Many commercial agencies maintain a list of contacts including their email addresses. The list could include addresses collected through legitimate means like subscription to a magazine or a company's customer list. This list could fall into the hands of spammers who could use it to send spam.

2.2.4. Chat rooms

Chat rooms are another major source of email addresses for spammers. Email addresses can be easily obtained in some IRC channels and other chat rooms. Spammers harvest these email addresses, knowing that these are 'live' addresses. IRC bots are also used to send messages interactively to IRC and to chat rooms to harvest email addresses.

2.2.5. Brute force attacks

Dictionary attacks on both username and domain names are also performed by spammers. Brute force attacks involve trying every possible combination of email–ids. Some spammers guess email addresses, send messages and verify the list based on return error messages. Common email addresses such as postmaster, hostmaster, root [for UNIX hosts], etc. are common targets for spam. Brute forcing email servers for valid email address, though a tedious process, is popular among spammers. Shorter email ids are more susceptible to brute force attacks.

2.2.6. Domain Name Registration

Usually, every domain has three contact points - administration, technical, and billing. The contact points include the email address of the contact person.

The list of domains are usually made available to the public by the domain registries. As the contact points are freely available, e.g. using the 'whois' command, spammers harvest the email addresses from the contact points for lists of domains. These email addresses are usually valid and thus are a tempting target.

3. Spam Control

Spam can be controlled through the combination of the following methods:

1. Spam filters
2. Best practices implemented at various levels

3.1. Spam Filtering Technologies

Spam can be identified and controlled at different levels by the use of spam filters. It can be either at the gateway level, the server level or at the client level. All filtering tools are primarily based on the following technologies.

- **Content based Filtering**
- **Source address-based Filtering**

3.1.1. Content based Filtering

Content based filtering does a detailed inspection on the contents of an email message and helps in identifying spam messages. Content based filtering can be implemented through simple text pattern matching, or through statistical probability indication.

3.1.1.1. Pattern matching

Content based filtering depends on predefined patterns of text and rule-based rankings. They evaluate a large number of patterns against a candidate message. Some matched patterns add to a message score, while others subtract from it. The incoming e-mail is evaluated based on simple strings found in specific header fields, the header in general, and in the email body itself. If the score of a message exceeds a certain threshold, it is filtered as spam, otherwise it is considered a legitimate mail.

3.1.1.2. Hash matching

A database of hashes of known spam messages is stored. Each new email received is hashed and compared with the above database. If the hash matches any of the stored hash values, it is identified as spam.

3.1.1.3. Statistical Classification Engines

Spam filtering can also be performed using statistical classification of the contents of the message. It is said to be one of the most effective spam fighting methods. These engines assign a spam indicative mark to words or chunk of words based on previously identified spam messages. New incoming messages are verified against these to generate spam indicative probability. Thus, based on the prior appearance of certain words, or chunk of words, statistical classification engines determine the probability that the new email message is spam. Statistical engines build spam indicative probabilities of words automatically, with minimal human intervention. One of the most popular statistical classification engines is Bayesian filter.

> *Bayesian filtering* is named after English mathematician Thomas Bayes, who developed a theory of probability inference. Bayesian filtering is predicated on the idea that a 'Spam-indicative probability' can be generated for each word using statistical analysis. This can again be used to determine the overall "Spam probability" of a message based on its contents. Bayesian filters can train themselves to identify new patterns of spam and can also be adapted by the human user to adjust to the user's specific parameters for identifying spam. Bayesian filters are said to return less than one percent of false positives.

3.1.1.4. Heuristic Filters

Some filtering mechanisms implement a combination of the above techniques i.e. pattern matching and Bayesian Filters.

Heuristics are a series of rules used to score the spam probability of an email. These are human-engineered rules by which a program analyses an email message for spam-like characteristics. These rules may look for multiple uses of certain phrases incorporating hundreds of rules in order to catch spam. A message might get a certain number of points for containing a certain phrase; more points if it contains a URL link, and even more points if the message includes a phrase for un-subscription

request link. Depending on the parameters established, reaching a certain score would classify the message as spam

Heuristic engines are quite effective though the spammers are using more sophisticated techniques to get around this kind of filtering. Spammers reverse-engineer heuristic rules and create messages that can bypass the filters.

3.1.2. Source address-based filtering

Source-based filtering is overtaking content-based filtering as the main method for blocking spam. Spammers are always figuring out new ways to get around content filters, but hiding the source IP address and its behavior, is more difficult.

3.1.2.1. White list/verification filters

A white-list contains a list of source addresses, which the recipient wants to receive mail from and are approved by the receiver as not being sources of spam. A white-list filter works based on the white-list of explicitly approved source addresses. Mails originating from sources, defined in the white list, are forwarded on to the mailbox and the rest tagged as spam.

The disadvantage is that they place an extra burden on legitimate senders. An effort is also required to receive mails from new users.

3.1.2.2. Blacklists

A blacklist is almost opposite of a white list. It contains a list of source addresses which are known to be originators of spam. Local blacklists are usually prepared and maintained by administrators manually. However, global blacklists of well known spammer IPs, and domain names are maintained by different organizations, and are termed as distributed black lists.

Reputation analysis of mail sending IP addresses is another method to filter spam. Amount of spam received from an IP address is monitored and after a certain threshold it is recognized as a spam source and future attempts of that IP address to send mail will be blocked. This is a method to maintain blacklists.

3.1.2.3. Realtime BlackHole Lists (RBLs)

Realtime BlackHole Lists (RBLs) can be queried by SMTP engines to verify any incoming messages. RBLs contain IP addresses which are known to be originator of spam. RBLs are normally implemented by using a protocol similar to DNS and are popularly known as DNSRBLs. RBLs are maintained by many different RBL operators, and organizations can simply subscribe to them.

Some of the organizations maintaining RBLs are rbls.org. [Ref 33], spamhaus [Ref: 26], spamassassin [Ref: 27], spamcop [Ref: 28], mail-abuse [Ref: 29]. Using the RBLs, mail servers can reduce a significant amount of possible spam.

Spam is also identified through Mail Header analysis. An e-mail message contains routing information which can be analyzed to determine discrepancies in the format, because many spammers try to hide their tracks by placing invalid information in the header. The discrepancies may indicate a possible spam.

Possibility of false positive (i.e. blocking legitimate emails) is a major concern in spam filtering. The disadvantage of source address based filtering is that they can be cumbersome and time-consuming; requiring constant list maintenance in order to be effective. Spammers also use hundreds of new addresses, acquire new zombies or relay machines to route their spam. Sometimes spammers spoof legitimate senders addresses adding more difficulties to address-based filtering.

3.1.3. Challenge-Response

Like white lists, challenge-response systems allow mail only from previously authenticated address lists. Mails from new addresses are kept on hold and a challenge mail is sent back to the sender. If the sender successfully replies to the challenge, the mail is accepted as authentic mail and delivered to the receiver's inbox. The challenge message may contain some simple questions, which can only be replied by human users.

This mechanism nullifies the possibility of spam being sent from an automated spamming tool like different worms, zombie machines etc.

Though it is an effective method to block massive spam being generated by automated tools, worms etc, it puts significant overhead to the legitimate mail senders. Mails from automated systems like webpage re-mailers cannot be delivered to such systems. Even legitimate emails may not be delivered because the challenge is not completed. Deadlock situation may also arise when both the recipient and sender use challenge response systems to authenticate each other.

Anti-Spam Honeypot

Anti-Spam honeypots are being used to monitor spammer activity. These honeypots simulate the services of an open relay to attract spammers and detect their Identity. Some honeypots even identify robot email harvester [Ref 37] by hosting scripts with fictious e-mail addresses.

3.2. Spam control Best Practices

Best practices need to be implemented to the following :

1. SMTP Server
2. ISPs
3. email users

4. Mass mailers / Advertisers

3.2.1. Best practices: SMTP Server implementation

SMTP MTA servers should implement all the anti- spam features as described in *RFC2505*: *Anti-Spam Recommendations for SMTP MTA*. [Ref:1] The SMTP Servers should also be configured so as to perform the following:

- *Should not relay unauthorized mails (Should not be an Open Relay)*

Authentication mechanism should be used to prevent misuse of relay SMTP servers. The server MUST be configured to restrict unauthorized use as an Open Mail Relay.

- *Separate ports for submission and relay of messages*

The separation of ports used for submission and for relay of messages, both of which normally run on port 25, should be enforced. The submission of mail by internal e-mail clients (e.g. MS-Outlook, Eudora etc) to SMTP Servers should be separated from the port used for relay of messages. TCP port 587, defined as the standard mail submission port in RFC 2476 [Ref: 2], should be used for submission of messages by email clients. The separation of ports can allow a site to implement different policies for the two different types of services.

- *Implement client authentication before mail submission*

Authentication of email clients allowed to relay through an SMTP server should be performed. If possible, usage of e-mail clients should be restricted to a limited set of IP addresses. Authentication of email clients can be performed using any of the following-

  - POP before SMTP
  - SMTP-Auth with TLS (RFC 2554) [Ref:3]

- *Disable SMTP commands like VERIFY (VRFY)*

Some SMTP commands like VRFY and EXPN can expose user information to people probing a system for harvesting information in preparation for sending spam. These SMTP commands can be disabled on the SMTP server. Various other SMTP commands like HELP should also be disabled.

- *Prevent remote mails to local groups*

A group can be created on mail servers comprising several users on the server. A mail sent to the group is delivered to all users who are members of the group. Due to the threat of misuse of these groups by spammers, delivery of mail to groups should be disabled from remote users.

- *Define maximum number of recipients per message*

A single email message can be addressed to a number of recipients. A limit on the maximum number of recipients that may be addressed by a single message should be enforced, to prevent misuse by spammers.

- *Reject NULL sender identity*

Spammers sometimes try to send mail with no sender information. The Mail Server should be so configured that it rejects mails with NULL sender identity.

- *Maintain statistical analysis*

Extensive logging should be performed on the mail server. To study trends and identify spam, statistical analysis of the following should also be performed.

  - Top SMTP connections from hosts/IPs
  - Top "From Addresses"
  - Top "Recipient Address"

3.2.2. Best practices: Internet Service Providers

Internet Service Providers should implement security recommendations as defined in *RFC 3013*: Recommended ISP Security [Ref:4] . The following best practices should also be adopted by ISPs to combat spam –[Ref 35]

- *Close All Open Relays*

Monitor all SMTP servers in the network and stop any 'open relay' in the network. The Service Provider MUST ensure that its' email systems will not relay email for any unauthorized third party.

- *Control Open Proxy Servers*

All proxy servers of the ISP should accept requests from only its own clients by either forcing a client to connect from only a range of IP addresses, or through the use of authentication.

- *Control hosts which can act as an SMTP server for mail transaction*

Every ISP should control the setting up and the activity of all mail servers on its network. It must ensure that all email generated within its own network can be attributed to a particular customer or system.

- *Control home SMTP traffic*

The submission of mail by internal e-mail clients (e.g. MS-Outlook, Eudora etc) to SMTP Servers should be enforced on port 587, the standard port for mail submission as defined in RFC 2476 : Message Submission [Ref:2]. If possible, outgoing connections on port 25 should be blocked for home users. This will help in eradicating a majority of spam generated by worms and viruses from infected client machines.

- *Implement Rate Limits on Outbound Email Traffic*

ISPs should monitor traffic patterns on port 25 at the gateway. Rate limiting should be implemented, as per the average usage. Sudden spikes in traffic on port 25 should be investigated as it could indicate activity of worms.

- *Monitor and control web based E-Mail submission forms*

Email submission forms are available on websites for performing various functions like re-mailing website contents. These forms can be used to send anonymous mail. To prevent their misuse by spammers by the use of automated tools, the use of such forms should be monitored and controlled. To prevent misuse of automated tools, a text or number could be displayed on the web page in an image format, the entry of which can be mandated in the form for further processing.

- *Detect and contain virus/worm infected machines*

Worm and virus infected client machines are a major source of spam. ISP should employ mechanisms to control virus and worms within its network.

- *Logging of sources of mails generated within their network*

An ISP should ensure that all email generated within its network can be traced to its source. It should also ensure that the immediate source of email which arrives from other networks can be determined. Statistical analysis of SMTP connections from hosts and IP addresses should also be performed, to study trends and help identify spam.

- *Disseminate information and Educate Customers*

The most effective method of preventing spam is through awareness and education of users to proactively stop the propagation of spam.

3.2.3. Best Practices: Email Users

Email users should take precautions to prevent their e-mail addresses falling into the hands of spammers. Email users should adhere to the following.

- *Disguise e-mail addresses posted in a public electronic place.*

Email addresses posted in a public web should be disguised through simple means to avoid being harvested by manual or automatic crawlers. They can be disguised by simple means such as replacing the symbol @ in the email id with some other symbol. Thus, "example@domain.com" could be written as "example [at] domain dot com". The email-id can be also be published on the web page as an image. In addition, the 'MAIL TO:' tag in source of the HTML pages should not be used so as to avoid the email address from being harvested by manual and automatic crawlers.

Some service providers have options where characters can be added to the end of the email address. For example, AOL members can add characters to the end of their address as it appears in Usenet posts. Thus, the email-id: emailfaq@aol.com can become emailfaq@aol.com.anytext.

- *Carefully subscribe to commercial e-mail news letter*

Users should be extremely careful while giving out their e-mail addresses to unknown commercial sites.

- *Use multiple e-mail addresses*

Avoid using the same email-id for different purpose like business, personal, etc. Users should consider creating multiple e-mail addresses or accounts.

- *Disposable email address*

A solution for giving out email addresses to commercial sites can be had by setting up a disposable email address or forwarding address. By using a forwarding address one does not have to give the primary address to merchants or post it online. Instead a forwarding/disposable address can be created, and the mail sent to this address is delivered to the primary mailbox. If one receives unwanted e-mails on one of the forwarding addresses, one can delete that particular address, while not affecting the primary address. Forwarding addresses are available from various vendors like Bigfoot.com, NetAddress.com, Yahoo, etc.

- *Use a filter*

Enable spam filtering options in the email clients. Third party spam filtering software available for identifying and filtering spam on clients should be used.

- *Short e-mail addresses*

Short email addresses are easy to guess, and susceptible to brute-force attacks. Studies have shown that shorter email ids receive more spam. Infact, some free email address providers no longer accept short email addresses. Users should therefore use long email ids.

3.2.4. Code of Ethics for mass mailers/advertisers

Advertising via the use of Mass-Messaging 'Spam' is financially very attractive, as very little cost is used in sending spam However, as a responsible Internet user, advertisers should

avoid this option considering the various risks associated with it. The risks associated with mass mailing by advertisers range from bad publicity to legal ramifications to ISP service discontinuity. *RFC 3098: How to Advertise Responsibly Using E-Mail and Newsgroups* [Ref:5] provides a framework through which an advertiser, the recipients and the Internet community can coexist in a productive and mutually respectful fashion.

4. Long Term Measures to control spam

Various technical controls discussed to minimize spam are not sufficient to control it completely. Despite all the controls, spam is growing every year. Spammers are continuously innovating to beat the controls. The Internet can not be completely spam free unless and until the basic flaws of SMTP protocol are eradicated. Restructuring of SMTP protocol to address the flaws is considered to be the long term measure to control spam. Various frameworks and methodologies are being proposed by different research groups to address the SMTP design flaws.

SMTP Protocol Restructure

The proposed frameworks primarily focus on verifying the source identity of the e-mail message sender. This includes both verification of claimed "from domain name" and end user name. Eliminating domain spoofing will help legitimate senders protect their domain names and reputations, and help recipients more effectively identify and filter junk e-mail and phishing scams.

Towards this, an Internet Engineering Task Force (IETF) working group named MARID ('MTA Authorization Records in DNS') was formed in 2004. Various proposals were put forward and were examined by the group. Significant proposals among those are discussed below. Though the MARID working group has officially closed, some of the proposals are being carried forward.

The proposals basically fall in two categories: Path based authentication and Signature based authentication.

Path Based authentication

Path based authentication broadly follow the Lightweight Mail Authentication Protocol (LMAP) specification of Anti spam Research Group (ASRG) which tries to augment SMTP to validate the sender machine of an email.

The two major path based authentication proposals are:

- SPF - Sender Policy Framework by Meng Wong
- Sender ID

Both the proposals use reverse DNS lookup to authenticate the sender. The proposals originated from other proposals like Designated Mail Protocol (DMP) [Ref 38] and Reverse Mail Exchanger [Ref 39], which were implementations of the LMAP specification.

Signature based Authentication
The two major signature based authentication proposals are:

- Identified Internet Mail (IIM) by Cisco
- Domain Keys by Yahoo

4.1. SPF (Sender Policy Framework)

Sender Policy Framework is an extension of SMTP designed by Meng Weng Wong of POBOX. The framework suggested a mechanism for E-Mail source Identity validation.

SPF was submitted as input along with other proposals to the IETF MARID workgroup. The latest version of SPF is known as the Marid Proposal [Ref 42].

SPF allows an Internet domain to specify which machines are authorized to send e-mail for that domain. This information is specified using 'reverse MX' entries in Domain Name Service (DNS) records. The receiving mail servers implementing SPF then treat as spam any email that claims to originate from a domain but fail to originate from the servers authorized to send mail for that domain.

The SPF operates at the level of the SMTP transaction. It verifies i) The MAIL FROM: parameter of the incoming mail, ii) the HELO/EHLO parameter of the sending SMTP server and iii) the IP address of the sending SMTP server.

SPF has some limitations. It only validates the domain of the envelope sender (listed as "Return-Path: " in e-mail headers). Thus, domains that share mail senders could forge each others' domain. SPF does not validate that a given email actually came from the claimed user. SPF also breaks the inter-system SMTP forwarding (where an agent forwards email to someone else without changing the "from" address).

4.2.Sender ID

Sender ID is a proposed specification developed within the MARID IETF Working Group between May and October 2004.

The Sender ID framework is an extension of previous SPF proposal. It consists of two halves: the SPF Classic half, and the Purported Responsible Address (PRA) half.

This combined specification is the result of Microsoft's Caller ID for E-Mail proposal, Meng Wong's Sender Policy Framework (SPF), and a third specification called the Submitter Optimization.

Purported Responsible Address (PRA) is determined by an algorithm which determines the source of the email based on the email headers. While SPF is intended to work in the MTA level, PRA is checked at MUA levels. It is possible to authenticate this email by tracking the PRA.

The PRA is as correct as the email headers. "Purported" reveals the claimed source of a message from the headers and not necessarily where it actually came from.

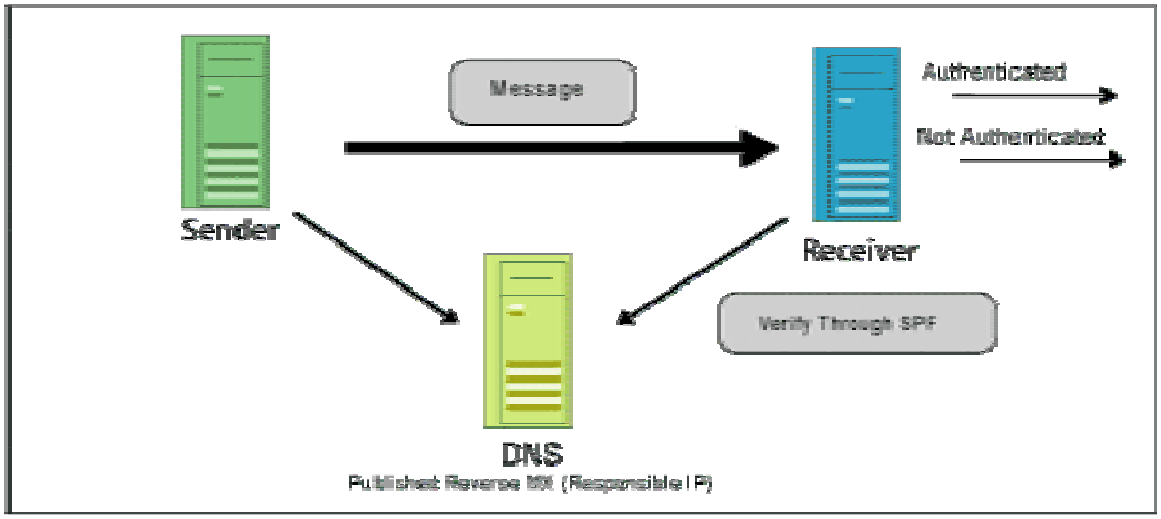


Fig 3

4.3. Domain Keys by Yahoo

Domain Keys framework is a proposal by YAHOO submitted as an Internet-Draft [Ref 43] for publication to the Internet Engineering Task Force (IETF).

Domain Keys uses the concept of digital signatures. There are two steps to signing an email with Domain Keys:

On the sender side, the domain owner generates a public/private key pair to use for signing all outgoing messages. The public key is published in the DNS, and the private key is made available to their Domain Key-enabled outbound email servers. When each email is sent by an authorized end-user within the domain, the Domain Key-enabled email system automatically uses the stored private key to generate a digital signature of the message. This signature is then pre-pended as a header to the email, and the email is sent on to the target recipient's mail server.

On the receiving side, the receiving email system extracts the signature and fetches the public key from DNS for the claimed From: domain. The Public key is used to verify the signature. If the domain is verified the email is delivered to the user's inbox.

4.4. Identified Internet Mail (IIM) by Cisco

Identified Internet Mail (IIM) has been proposed by Cisco Systems as a signature-based mail authentication standard designed to address spam. IIM allows a recipient to authenticate messages and verify their authorization

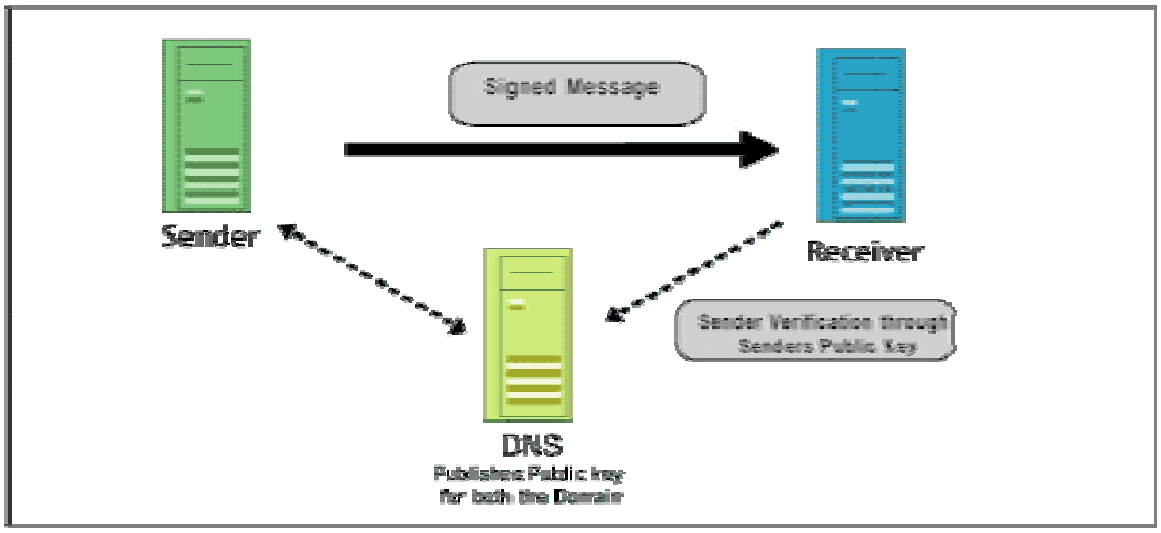


Fig 4

IIM verifies that the message sender is authorized to send messages using a given e-mail address and that the original message was not altered in any consequential manner. Identified Internet Mail messages are in standard RFC 2822 format with additional headers called IIM-Signature and IIM-Verification.

An Internet-Draft of IIM (draft-fenton-identified-mail-00) was submitted to the IETF [Ref 44]. Cisco Systems has released an open source implementation of Identified Internet Mail [Ref 45].

5. Legislative measures to fight spam

Technical controls alone can never eradicate spam completely. There is a need to define spam as a crime and bring it under the purview of legislation. The United States and various other countries have already adopted anti-Spam laws. Under these laws, commercial unsolicited mails are prohibited and rules have been framed for sending commercial e-mails. Many of the acts cover various form of electronic messaging like e-mail, short message service (SMS), multimedia message service (MMS), instant messaging (IM) etc.

Adoption of opt-in or opt-out method in spam legislation is an important issue. An opt-out list refers to email advertising lists in which recipients are signed up without their knowledge/consent and they may request to be removed from the list, where as opt-in list restrict sending mail only to those who have chosen to receive mail.

A spam law should enforce the following:

- UCE messages must not be sent
- Commercial electronic messages must include accurate sender information

- Commercial electronic messages must contain a functional unsubscribe facility
- Address-harvesting software & harvested lists must not be supplied, acquired or used.
- Penalties & compensation for spamming

Though various countries have spam legislation, not all of these have helped in controlling spam. However, it has given service providers the ability to use law to fight spammers and file suits against them.

Some of the major spam legislation across the world is as follows.

US CAN-SPAM Act 2003	[Ref: 16]
Australian spam Act	[Ref: 46]

India, as on date doesn't have an anti-Spam Act, though some prosecution has already taken place using existing legislation. One of the most successful spam laws to date is said to be the Australian spam Act that went into effect in the year 2004. Australian spam Act enforces three key elements – Consent, Identify and Unsubscribe. Australian spam Act enforces the concept of opt-in lists.

Nearly a year after the passage of the US Federal Can-Spam Act, the law has not been able to curb spam, according to a report published by MX Logic, an anti-Spam company[Ref: 47]. The survey for the year 2004 showed widespread and flagrant disregard for the U.S. law that went into effect on Jan. 1 2004. The fact that 42% of world spam originated from United States [Ref 48] gives an indication that the CAN-Spam Act has probably not had the desired effect of stopping spam.

6. Conclusion

None of the spam control discussed in this paper is sufficient to make Internet spam free. Industry wide collaboration, cooperation and knowledge sharing is required towards this. Initiative to adhere to the best practices by all service providers and end users is also required.

7. References

1. RFC 2505 - Anti-Spam Recommendations for SMTP MTAs
   http://www.ietf.org/rfc/rfc2505.txt

2. RFC 2476 - Message Submission
   http://www.ietf.org/rfc/rfc2476.txt

3. RFC 2554 - SMTP Service Extension for Authentication
   http://www.ietf.org/rfc/rfc2554.txt

4. RFC 3013 - Recommended ISP Security

http://www.ietf.org/rfc/rfc3013.txt

5. RFC 3098 - How to Advertise Responsibly Using E-Mail and Newsgroups or - how NOT to $$$$$ MAKE ENEMIES FAST! $$$$$
   http://www.ietf.org/rfc/rfc3098.txt

6. RFC 821 – Simple Mai Transfer Protocol
   http://www.ietf.org/rfc/rfc821.txt

7. RFC 2822 - Internet Message Format
   http://www.ietf.org/rfc/rfc2822.txt

8. http://www.dslreports.com/forum/remark,2976699~mode=flat

9. Anti-Spam research Group
   http://asrg.sp.am

10. Open Relay Database
    http://www.ordb.org

11. Centre for Democracy & Technology
    http://www.cdt.org

12. Coalition Against Unsolicited Bulk Email, Australia
    http://www.caube.org.au/

13. CAUCE, The Coalition Against Unsolicited Commercial Email
    http://www.cauce.org/

14. http://www.mail-abuse.org
15. CAN-SPAM Act
16. CERT-In: Open Proxy Servers:
17. Sender Id Home Page:
    http://www.microsoft.com/mscorp/twc/privacy/Spam/senderid/default.mspx
18. IETF-MARID
    http://www.ietf.org/internet-drafts/draft-ietf-marid-core-03.txt
    http://www.ietf.org/internet-drafts/draft-ietf-marid-pra-00.txt
    http://www.cdt.org/speech/Spam/030319Spamreport.shtml

19. Reflections on the 25th Anniversary of spam
    http://www.templetons.com/brad/Spam/Spam25.html

20. First Mass Mailing
    http://www.mailmsg.com/Spam_history_002.htm

21. The term 'Spam' has 10th anniversary
    http://washingtontimes.com/upi-breaking/20040412-062338-4511r.htm

22. First Commercial spam
http://www.mailmsg.com/Spam_history_001.htm

23. Nigerian Scam
http://www.snopes.com/crime/fraud/nigeria.asp

24. Hijacked PCs blamed for a third of spam
http://news.zdnet.co.uk/internet/security/0,39020375,39118252,00.htm

25. Spamhaus
www.Spamhaus.org

26. Spamassassin
27. Spamcop
28. mail-abuse
29. Spamhaus spam definition
http://www.Spamhaus.org/definition.html
30. Kelkea spam definition
http://www.kelkea.com/support/Spam_def.html

31. http://www.ripe.net/ripe/docs/Spam.html#toc3

32. "The True Cost of spam and Value of Anti-Spam Solutions Study, 2004"
http://www.idc.com/getdoc.jsp?containerId=pr2004_04_12_112824

33. RBLS
http://rbls.org/
http://www.email-policy.com/Spam-black-lists.htm

34. http://e-com.ic.gc.ca/epic/internet/inecic-ceac.nsf/en/h_gv00170e.html
35. http://www.mxlogic.com/PDFs/IndustryStats.pdf
36. A New Tool In The spam War
http://www.securityfocus.com/columnists/291
37. http://www.theregister.co.uk/2004/06/04/trojan_Spam_study/
38. www.pan-am.ca/dmp
39. www.ietf.org/internet-drafts/draft-danisch-dns-rr-smtp-04.txt
40. LMAP
41. ASRG
42. *Marid Proposal*
http://www.ietf.org/internet-drafts/draft-ietf-marid-protocol-03.txt
43. 'draft-delany-domainkeys-base-01.txt'
http://www.ietf.org/internet-drafts/draft-delany-domainkeys-base-01.txt

44. http://www.ietf.org/internet-drafts/draft-fenton-identified-mail-01.txt
45. http://sourceforge.net/projects/identifiedmail/.
46. http://www.noie.gov.au/publications/NOIE/spam/Act2003/index.htm
47. http://www.mxlogic.com/news_events/01_03_05.html

48. http://news.com.com/U.S.+leads+the+dirty+dozen+spammers/2100-7349_3-5503344.html
49.

8. Glossary

Phisher	
Spoof	To deceive for the purpose of gaining access to someone else's resources
SMTP	Simple Mail Transfer Protocol
ASRG	Anti Spam Research Group
LMAP	Lightweight MTA Authentication Protocol
IIM	Identified Internet Mail
DNS	Domain Name Service
MTA	Message Transfer Agent
PRA	Purported Responsible Address
SPF	Sender Policy Framework
MUA	Message User Agent
IETF	Internet Engineering Task Force
MARID	MTA Authorization Records in DNS
RFC	Request For Comments
SMS	Short Message Service
MMS	Multimedia Messaging Service
UCE	Unsolicited Commercial Email
UBE	Unsolicited Bulk Email
IM	Instant Messenger
IP	Internet Protocol
MAPS	
ISP	Internet Service Provider
IRC	Internet Relay Chat
POP	Post Office Protocol
TLS	Transport Layer Security