7-2-2010

# System Analysis of SPAM

Nisar A. Shah
*University of Kashmir*, nassgr@yahoo.com

M. Tariq Banday
*University of Kashmir*, sgrmtb@yahoo.com

Recommended Citation
Shah, Nisar A. and Banday, M. Tariq, " System Analysis of SPAM" (2010). *All Sprouts Content*. 251.
http://aisel.aisnet.org/sprouts_all/251

芽|Sprouts Working Papers on Information Systems ISSN 1535-6078

# System Analysis of SPAM

Nisar A. Shah
University of Kashmir, India

M. Tariq Banday
University of Kashmir, India

## Abstract

Increasing reliance on the electronic mail (e-mail) has attracted spammers to send more and more spam e-mails in order to maximizing their financial gains. These unwanted e-mails are not only clogging the Internet traffic but are also causing storage problems at the receiving servers. Besides these, spam e-mails also serve as a vehicle to a variety of online crimes and abuses. Although several anti-spam procedures are currently employed to distinguish spam e-mails from the legitimate e-mails yet spammers and phishes obfuscate their e-mail content to circumvent anti-spam procedures. Efficiency of anti-spam procedures to combat spam entry into the system greatly depend on their level of operation and a clear insight of various possible modes of spamming. In this paper we investigate directed graph model of Internet e-mail infrastructure and spamming modes used by spammers to inject spam into the system. The paper outlines the routes, system components, devices and protocols exploited by each spamming mode.

**Keywords:** Spam, Anti-spam, Spam filter, Mail Server, MUA, MTA.

**Permanent URL:** http://sprouts.aisnet.org/8-47

**Reference:** Shah, N.A., Banday M.T. (2008). "System Analysis of SPAM," . *Sprouts: Working Papers on Information Systems*, 8(47). http://sprouts.aisnet.org/8-47

# System Analysis of SPAM

**Prof. N. A. Shah**
*Deptt. of Electronics and Instrumentation Technology*
*University of Kashmir, Srinagar, India*
Email: nassgr@yahoo.com

**M. Tariq Banday, *Lifetime Member*, CSI**
*Deptt. of Electronics and Instrumentation Technology*
*University of Kashmir, Srinagar, India*
Email: sgrmtb@yahoo.com

## Abstract

*Increasing reliance on the electronic mail (e-mail) has attracted spammers to send more and more spam e-mails in order to maximizing their financial gains. These unwanted e-mails are not only clogging the Internet traffic but are also causing storage problems at the receiving servers. Besides these, spam e-mails also serve as a vehicle to a variety of online crimes and abuses. Although several anti-spam procedures are currently employed to distinguish spam e-mails from the legitimate e-mails yet spammers and phishes obfuscate their e-mail content to circumvent anti-spam procedures. Efficiency of anti-spam procedures to combat spam entry into the system greatly depend on their level of operation and a clear insight of various possible modes of spamming. In this paper we investigate directed graph model of Internet e-mail infrastructure and spamming modes used by spammers to inject spam into the system. The paper outlines the routes, system components, devices and protocols exploited by each spamming mode.*

## Keywords

Spam, Anti-spam, Spam filter, Mail Server, MUA, MTA.

## 1. Introduction

E-mail has emerged as a free, valuable and crucial worldwide business tool but its availability is put at risk [10] by the kinds of unsolicited content that are fed into it. Growing volumes of Spam, malware and virus infections received via e-mail are the major concerns for both e-mail users and its service providers [1]. Spam e-mails are making adverse effects on the Internet bandwidth as it constitutes most of the e-mail traffic. Further, the volume of Spam received by users is creating storage problems on email servers resulting in lower performance levels and as such demand for larger memory mailboxes is rapidly increasing. Spam e-mails are serving as a carrier for various other online crimes and abuses that include: carrying out of phishing attacks, delivering viruses and worms, financial loss or even identity theft. Several technological and legal anti-spam measures have been proposed [2] but mainly filtering and blocking approaches are currently employed as they do not need any infrastructural change in the e-mail system. Spam is injected at various places into the e-mail system by spammers using a variety of techniques and tools that include spoofing, botnets, open proxies, mail relays, untraceable Internet connections, and bulk mail tools called mailers. In order to devise an effective anti spam procedure it is thus essential to have a clear insight of Internet e-mail infrastructure and the spamming modes used by the spammers for spamming.

The rest of this paper is organized as follows: In section 2 we present the basic e-mail communication model followed by Directed Graph model in section 3. In sections 4 and 5 we respectively make the mail path analysis and mail categorization. In section 6 we deduce the spamming modes, outline the protocol groups exploited and list places where anti-spam measures can be applied for a particular spamming mode. Finally we conclude in section 7.

## 2. E-mail Communication Model

A simple communication model as shown in figure 1 consists of four components along the path of an e-mail message [3]. They are sender client, sending server, receiving server and receiving client all on the Internet. E-mail clients are client computers running Outlook Express, Office Outlook, Eudora or other similar mail client application while e-mail servers are server computers running server software e.g. Exchange server or Sendmail Server. In Web based e-mail services such as in case of mail.yahoo.com and gmail.google.com clients and servers are combined and are integrated behind a web server. The purpose of devices used is mentioned hereunder:

**Sender's Client:** The sender composes an e-mail message, generally using a mail client program on local machine. The mail once, composed is not immediately sent out over the Internet; it is held in a buffer area called a spool. This allows the user to be "unattached" for the entire time so that a number of outgoing messages can be created. When the user is done, all of the messages can be sent at once.

**Sender's SMTP Server**: When the user's mail is ready to be sent, it connects to the internetwork. The messages are then communicated to the user's designated SMTP server. The mail is sent from the Senders client machine to the senders SMTP server using SMTP. It is also possible for the sender to work directly with senders SMTP server; in this case sending is simplified.

**Recipient's SMTP Server:** The sender's SMTP server sends the mail using SMTP to the recipient's SMTP server over the internetwork. There, the e-mail is placed in the recipient's incoming mailbox (or inbox). This is comparable to the outgoing spool that exists on sender's client machine. It allows the recipient to accumulate e-mail from many sources over a period of time and retrieve them as per their convenience.

**Recipient's Client:** In certain case the recipient may access its mailbox directly on the recipients SMTP server. More often, however, a mail access and retrieval protocol, such as Post Office Protocol (POP3) or Internet Message Access Protocol (IMAP), is used to read the mail from the SMTP server and display it on the recipient's local machine using an e-mail client program, similar to the one used to compose the message at the senders client.

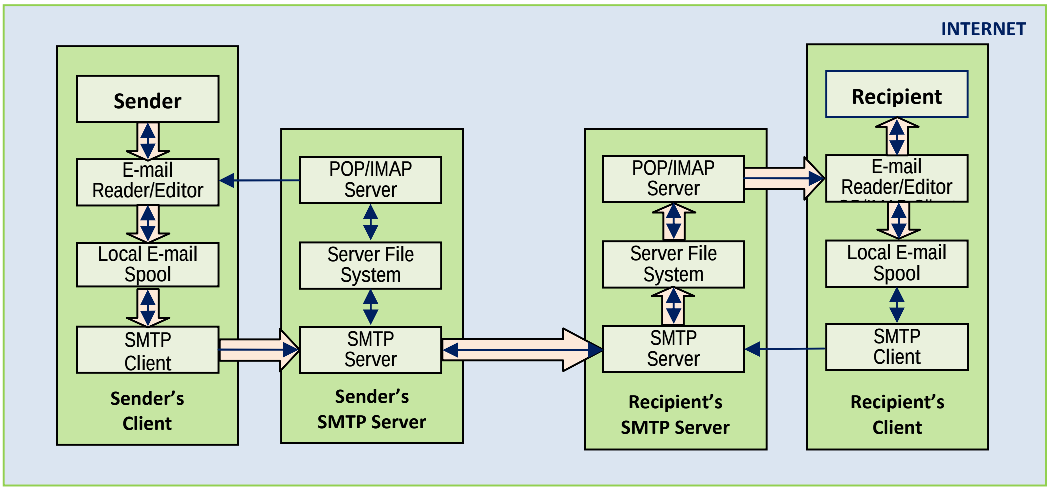

Figure 1: E-mail Communication Model [3]

*Each device consists of a number of different elements, which communicate as indicated by the dark and thin arrows. The large arrows show a typical email transaction*

Internet pioneer Jon Postel formalized the technical specifications for transferring e-mail with the Simple Mail Transfer Protocol (SMTP) [4, 9] which has undergone several revisions and has been adapted as a Request For Comments (RFC) by IETF (Internet Engineering Task Force) [5] which is responsible for defining and maintaining e-mail standards. SMTP is an application layer protocol for TCP/IP based Internet infrastructure which sets conversational and grammatical rules for exchanging e-mail between computers. The SMTP is simple in content and requirements. It minimizes information that must be included in the exchange and leaves functions such as authentication to other protocols and applications. This simple architecture makes SMTP easy to implement and use, but the spammers have misused these advantages and have exploited its underlying trust to target recipients with spam, hide their own identities, and conceal their tracks [11]. The IETF offers protocols that add security features to SMTP, but these have not been widely adopted. The backwards-compatibility challenge and the need for widespread, if not universal, adoption of any such solution, impede the effort to revise SMTP to overcome the treats to the current e-mail system. Thus far, e-mail software vendors have not sought to fix the spam problem within SMTP; rather, their solutions treat the protocol as given and use various other anti-spam procedures.

## 3. DG Model of E-mail Internet Infrastructure

The directed graph model of e-mail Internet infrastructure [6] as shown in figure 2 is based on the types of the communicating entities called e-mail nodes.

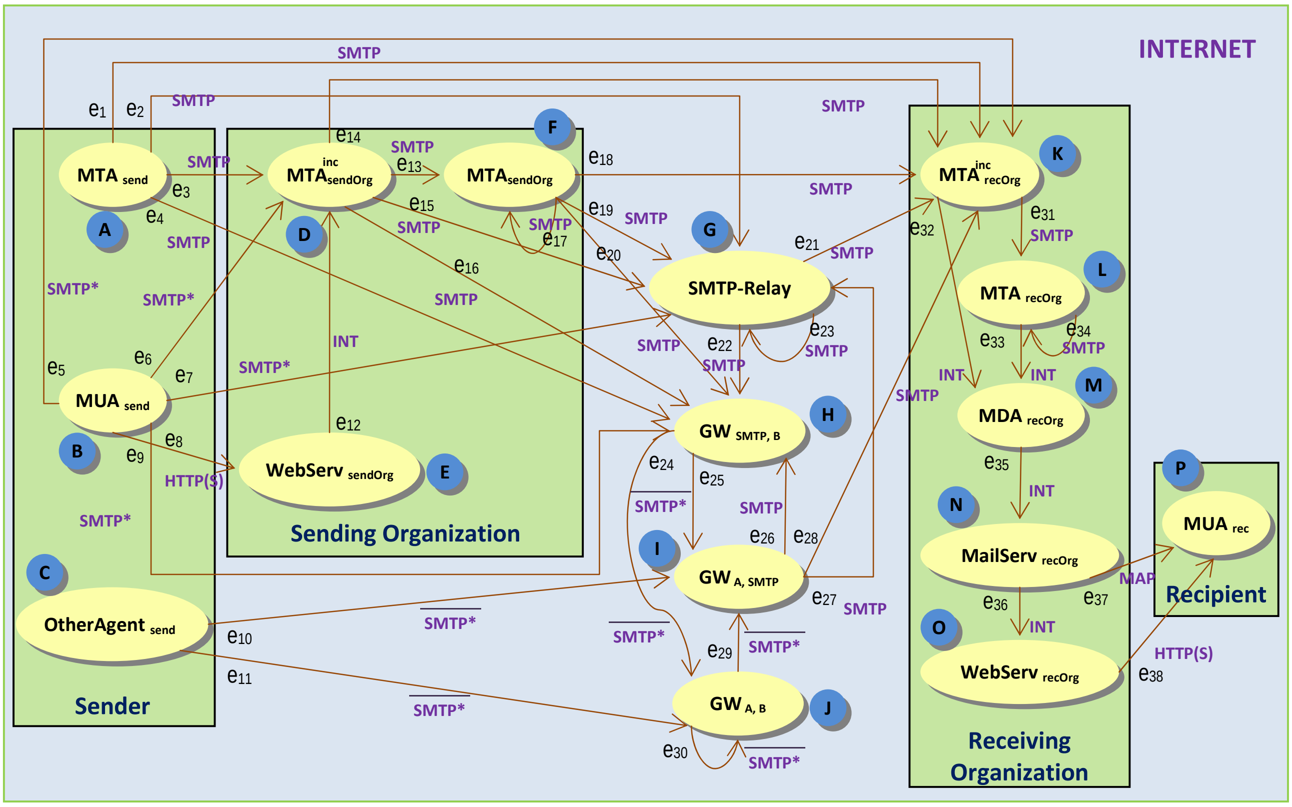

Figure 2: Directed Graph Model of E-mail Internet Infrastructure [6]

In this, e-mail communication is modeled as a directed graph of **V** vertices and **E** edges. Each vertex corresponds to an e-mail node which is a software unit involved in e-mail delivery process and works on the TCP/IP application layer. Nodes working on the lower layers such as routers and bridges which represent options to send e-mail without using SMTP are not considered in this model as almost all e-mail communication uses SMTP directly or indirectly. The vertices are grouped into five sets $\mathbf{V_{set1}}$ through $\mathbf{V_{set5}}$ depending on the component it belongs to from the distinct components namely *Sender*, *Sending Organization*, *Receiving Organization*, *Recipient* along the *Internet*. The nodes corresponding to each component are listed in table 1.

**Table 1: List of Nodes in Directed Graph Model**

| ***Node Name*** | ***Node Definition*** |
|---|---|
| ***$V_{set1}$* (Sender Node Set)** | |
| $MTA_{send}$ | Senders Mail Transfer Agent can only establish SMTP connections with ESPs incoming $MTA^{inc}_{sendOrg}$, $SMTP-Relay$, $GW_{SMTP,B}$ or $MTA^{inc}_{recOrg}$. |
| $MUA_{send}$ | Senders Mail User Agent can establish SMTP[*] connections with $MTA^{inc}_{sendOrg}$, $SMTP-Relay$, $GW_{SMTP,B}$, $MTA^{inc}_{recOrg}$ or an HTTP(S) connection with $WebServ_{sendOrg}$. |
| $OtherAgent_{send}$ | Other agents can establish $\overline{SMTP}*$ i.e. the connection other than SMTP based with gateways when such connections are possible i.e. with $GW_{A,SMTP}$ or $GW_{A,B}$. |
| ***$V_{set2}$* (Sending Organization Node Set)** | |
| $MTA^{inc}_{sendOrg}$ | Sending organizations incoming Mail Transfer Agent can establish SMTP connections with $MTA_{sendOrg}$, $SMTP-Relay$, $MTA^{inc}_{recOrg}$, or $GW_{SMTP,B}$. |
| $MTA_{sendOrg}$ | Mail Transfer Agent sending organization can establish SMTP connections with $MTA^{inc}_{recOrg}$, $SMTP-Relay$, $GW_{SMTP,B}$, or other$MTA_{sendOrg}$. |
| $WebServ_{sendOrg}$ | Sending Organizations Web Server can establish ESP internal protocol based connection with $MTA^{inc}_{sendOrg}$. |
| ***$V_{set3}$* (Internet Node Set)** | |
| $SMTP-Relay$ | SMTP Relay can establish SMTP connections with $MTA^{inc}_{recOrg}$, $GW_{SMTP,B}$ or other SMTP Relay. |
| $GW_{SMTP,B}$ | Gateways making $\overline{SMTP}*$ connections with $GW_{A,SMTP}$ , or $GW_{A,B}$. |
| $GW_{A,SMTP}$ | Gateways making SMTP connections with $GW_{SMTP,B}$ $SMTP-Relay$ or $MTA^{inc}_{recOrg}$. |
| $GW_{A,B}$ | Gateways making $\overline{SMTP}*$ connections with $GW_{A,SMTP}$ or other $GW_{A,B}$ . |
| ***$V_{set4}$* (Receiving Organization Node Set)** | |
| $MTA^{inc}_{recOrg}$ | Receiving Organizations Mail Transfer Agent making SMTP connection with $MTA_{recOrg}$ or ESP internal protocol based connection with $MDA_{recOrg}$. |
| $MTA_{recOrg}$ | Receiving Organizations Mail Transfer Agent making SMTP connection with other $MTA_{recOrg}$ or ESP specified internal connection with $MDA_{recOrg}$. |
| $MDA_{recOrg}$ | Receiving Organizations Mail Delivery Agent making ESP specified internal connection with $MailServ_{recOrg}$. |
| $MailServ_{recOrg}$ | Receiving Organizations Mail Server making mail access protocol MAP based connection with $MUA_{rec}$or ESP specified internal connection with $WebServ_{recOrg}$. |
| $WebServ_{recOrg}$ | Receiving Organizations Web Server making HTTP(S) based connection with $MUA_{rec}$. |
| ***$V_{set5}$* (Recipient Node Set)** | |
| $MUA_{rec}$ | Recipients Mail User Agent does not make any outgoing connection considering forwarding of email to be treated as a new sequence. |

All ***MUA*** nodes are simply software packages that normally allow end users to compose, create or read e-mail. They may be used to send e-mail to the receiving ***MUA*** directly or indirectly. ***MTA***s are in effect postal sorting agents that have the responsibility of retrieving the relevant mail exchange (MX) record from the Domain name Servers (DNS) [7] for each e-mail to be send thus mapping the distinct e-mail addressee's domain name with the relevant IP address information. They may also be used to compose and create e-mail messages. Node named ***OtherAgents*** are software packages that send e-mail message through gateways. ***WebServ*** nodes are the e-mail servers that provide the Web environment for composing or sending or reading an e-mail message. SMTP-Relays [8] are the nodes that perform e-mail relaying. Relaying is the process of receiving e-mail message from one SMTP e-mail node and forward it to another one. This scenario takes care of the mailbox exchange forwarding rule and indirect mail delivery using Local Mail Transfer Protocol (LMTP) (RFC 2033). Gateway nodes are used to convert e-mail messages from one application layer protocol to other. Gateway nodes named $GW_{SMTP,B}$ accept SMTP e-mails and transfer it with a protocol other that SMTP and $GW_{A,SMTP}$ performs the inverse process at incoming and outgoing interfaces. Gateway nodes $GW_{A,B}$ do not use SMTP either for incoming or outgoing interfaces. A process called Proxy may be done at these nodes when incoming and outgoing interfaces use same protocols.

Each edge of the graph connecting two e-mail nodes represents possible e-mail flow between them using a particular set of protocols. In table 2, we list the groups of protocols used in e-mail flow between two possible e-mail nodes along with the protocols in each group and the edges using protocols from that group for flow of the e-mail.

**Table 2: Protocol Groups, Protocols in each group and edges using a particular protocol set**

| ***Protocol Group*** | ***Protocols in group*** | ***Edges*** |
|---|---|---|
| $SMTP$ | SMTP protocol (RFC 821), SMTP service extension protocols ESMTP including Service Extension for Authentication (RFC 2554), Delivery by SMTP Service Extension (RFC 2852), SMTP Service Extension for Routing Enhanced error (RFC 2034) and SMTP Service Extension for Secure SMTP over Transport Layer Security (RFC 3207). | $e_1$, $e_2$, $e_3$, $e_4$, $e_{13}$, $e_{14}$, $e_{15}$, $e_{16}$, $e_{17}$, $e_{18}$, $e_{19}$, $e_{20}$, $e_{21}$, $e_{22}$, $e_{23}$, $e_{26}$, $e_{27}$, $e_{28}$, $e_{31}$, and $e_{34}$. |
| $SMTP^*$ | All protocols in SMTP set and all SMTP extensions for e-mail submission from MUA to e-mail node with SMTP incoming interface. E-mail node can be MTA defined in RFC 2821, MSA defined in RFC 2476. Using MSA various methods can be applied for ensuring authenticating user that include IP address restrictions, secure IP and POP authentication. | $e_5$, $e_6$, $e_7$ and $e_9$. |
| $\overline{SMTP*}$ | All Internet application protocols except those specified in $SMTP^*$ group, all propraitory application protocols used on the Internet (also used for tunneling), all Internet protocols on the transport and network layers such as TCP/IP as it is quite possible to send e-mail without the use of application layer protocols. | $e_{10}$, $e_{11}$, $e_{24}$, $e_{25}$, $e_{29}$ and $e_{30}$ . |
| $HTTP(S)$ | HTTP (RFC 2616), HTTP over SSL and HTTP over TLS (RFC 2818). | $e_8$ and $e_{38}$. |
| $INT$ | ESP specific Protocols and procedures for internal e-mail delivery. | $e_{12}$, $e_{32}$, $e_{33}$, $e_{35}$ and $e_{36}$. |
| $MAP$ | All email access protocols used to transfer e-mails from the recipient e-mail server to MUA that include IMAP version 4 (RFC 1730) and POP version 3 (RFC 1939). | $e_{37}$. |

Using the discussed model different types of e-mail delivery including delivery of spam e-mail can be described in terms of a set of directed paths. The graph provides a framework for analyzing anti-spam measures and shows all possible spam delivery routes.

## 4. Mail Path Analysis

Each option of sending a legitimate e-mail represents an option to send bulk e-mail as well as spam. The directed graph shown in figure 2 has 38 edges labeled $e_1$ through $e_{38}$ representing a possible e-mail flow and connecting 16 e-mail nodes labeled A through P. Two e-mail nodes are connected by an edge if and only if the Internet e-mail architecture allows e-mail flow between the corresponding node types. As can be noted from the graph, the edges $e_{17}$, $e_{23}$, $e_{30}$ and $e_{34}$ at e-mail nodes F, G, J and L respectively designate loops in the graph. A loop at node F indicates a possible chain of more than one Mail Transfer Agents. The loop at node G designates a possible use of more than one SMTP Relay. A possibility of the use of more than one Gateway converting one protocol based e-mail to another is indicated by a loop at node J. The use of multiple Mail Transfer Agents at receiving organization is mentioned by a loop at L. Table 3 provides an approximation of various paths from an e-mail node to the target node i.e. node P.

**Table 3: Paths from any node to P**

| *Node* | *Adjacent Nodes* | *Possible Paths* |
|---|---|---|
| *A* | *K, G, D, H* | *AK..., $A\bar{G}$…, AD…, AH…* |
| *B* | *K, D, G, E, H* | *BK…, BD…, $B\bar{G}$…, BE…, BH…* |
| *C* | *I, J* | *CI…, $C\bar{J}$…* |
| *D* | *K, F, G, H* | *DK…, $D\bar{F}$…, $D\bar{G}$…, DH….* |
| *E* | *D* | *ED…,* |
| *F* | *F, K, G, H* | *$\bar{F}$K…, $\bar{F}\bar{G}$…, $\bar{F}$H…* |
| *G* | *G, K, H* | *$\bar{G}$K…, $\bar{G}$H…* |
| *H* | *I, J* | *HI…, $H\bar{J}$…* |
| *I* | *K, H, G* | *IK…, IH…, $I\bar{G}$…* |
| *J* | *J, I* | *$\bar{J}$I…* |
| *K* | *L, M* | *$K\bar{L}$…, KM…* |
| *L* | *L, M* | *$\bar{L}$M…* |
| *M* | *M, N* | *MN…* |
| *N* | *P, O* | *NP, NO…* |
| *O* | *P* | *OP* |
| *P* | *None* | *None* |

*Note: Dots in paths represent all paths from the last node mentioned in the path e.g. MN… means the paths MN, MNP and MNOP. Also a bar on a node label represents a possible loop e.g. $\bar{G}$ means one or more gateways in the path.*

From table 3, it is evident that the arrangement of nodes and edges of graph shown in figure 2 makes numerous paths for e-mail transaction. The paths include both direct and indirect paths. Direct paths make a connection between a sending node and a receiving node through an Internet Service Provider (ISP) (not shown in the graph) which merely forwards the TCP packets. Indirect paths establish connections between a sending node and a receiving node through intermediate nodes like nodes of the (E-mail Service Provider) ESP and/or Internet E-mail Service Nodes (IESN).

## 5. Mail Classification

A complete e-mail transaction is one that originates from any one of the possible starting node {A, B, C, D, F} and terminates by delivering e-mail to any one of the possible receiving node {K, L. N}. An e-mail send to the first MTA (node K) can be considered as delivered and thus we can safely consider delivery up to node K for further analysis. Out of the five participating components i.e. Sender, Sending Organization also called E-mail Service Provider (ESP), Internet E-mail Service Nodes (IESN), Receiving Organization and the Recipient, last two do not affect the delivery process. Hence e-mail classification can be made on the basis of participation of different types of nodes belonging to first three components i.e. Senders Nodes, ESPs Nodes and IESNs nodes. This classification is shown in table 4.

This classification shows that their exist as many as 14 unique ways of sending e-mail or spam which differ from one another in terms of the paths followed, protocols used and the types of the e-mail nodes used. Security system violations of ESPs caused by various infections like viruses, Trojan horses, worms, etc. would create more ways to send spam.

Table 4: Mail Classification

| S. No. | Participating Node(s) | Originating Node(s) | Protocol Group(s) | Path(s) |
|---|---|---|---|---|
| 1. | *MTA of ESP* | **D or F** | SMTP | $DK, D\bar{F}K, \bar{F}K$ |
| 2. | *MTA of ESP then Relay(s)* | **D or F** | SMTP | $D\bar{F}\bar{G}K, D\bar{G}K, \bar{F}\bar{G}K$ |
| 3. | *MTA of ESP then Relay(s) & Gateway(s)* | **D or F** | SMTP, $\overline{SMTP *}$ | $D\bar{F}\bar{G}H\ldots K,\ D\bar{G}H\ldots K,\ D\bar{F}H\ldots\bar{G}K,\ DH\ldots\bar{G}K,\ \bar{F}\bar{G}H\ldots K,\ \bar{F}H\ldots\bar{G}K$ |
| 4. | *MTA of ESP then Gateway(s)* | **D or F** | SMTP, $\overline{SMTP *}$ | $D\bar{F}H\ldots IK, DH\ldots IK, \bar{F}H\ldots IK$ |
| 5. | *Senders MTA or MUA* | **A or B** | SMTP, $\overline{SMTP *}$ | $AK, BK$ |
| 6. | *Senders MTA or MUA then Relay(s)* | **A or B** | SMTP | $A\bar{G}K, B\bar{G}K$ |
| 7. | *Senders MTA or MUA then Gateway(S)* | **A or B** | SMTP, $\overline{SMTP *}$ | $AH\ldots IK, BH\ldots IK$ |
| 8. | *Senders MTA or MUA then Relay(s) & Gateway(S)* | **A or B** | SMTP, $\overline{SMTP *}$ | $A\bar{G}H\ldots K, B\bar{G}H\ldots K, AH\ldots\bar{G}K, BH\ldots\bar{G}K$ |
| 9. | *Other Agents then Gateway(s)* | **C** | SMTP, $\overline{SMTP *}$ | $CI\ldots IK, C\bar{J}I\ldots IK, CIK, C\bar{J}IK$ |
| 10. | *Other Agents then Gateway(s) and Relay(s)* | **C** | SMTP, $\overline{SMTP *}$ | $CI\ldots\bar{G}K, C\bar{J}I\ldots\bar{G}K, CI\ldots\bar{G}\ldots IK, C\bar{J}I\ldots\bar{G}\ldots IK$ |
| 11. | *Senders MTA or MUA & then MTA(s) of ESP or Web Server of ESP* | **A or B** | SMTP, SMTP*, HTTP(S), INT | $AD\bar{F}K, ADK, BD\bar{F}K,\ BDK,\ BED\bar{F}K, BEDK$ |
| 12. | *Senders MTA or MUA & then MTA(s) of ESP or Web Server of ESP then Relay(s)* | **A or B** | SMTP, SMTP*, HTTP(S), INT | $AD\bar{F}\bar{G}K,\ AD\bar{G}K,\ BD\bar{F}\bar{G}K,\ BD\bar{G}K,\ BED\bar{F}\bar{G}K,\ BED\bar{G}K$ |
| 13. | *Senders MTA or MUA & then MTA(s) of ESP or Web Server of ESP then Gateway(s)* | **A or B** | SMTP, SMTP*, HTTP(S), INT | $AD\bar{F}H\ldots IK, ADH\ldots IK, BD\bar{F}H\ldots IK, BDH\ldots IK, BED\bar{F}H\ldots IK, BEDH\ldots IK$ |
| 14. | *Senders MTA or MUA & then MTA(s) of ESP or Web Server of ESP then Relay(s) & Gateway(s)* | **A or B** | SMTP, SMTP*, HTTP(S), INT | $AD\bar{F}\bar{G}\ldots I\ldots K, AD\bar{G}\ldots I\ldots K, AD\bar{F}H\ldots\bar{G}\ldots K, ADH\ldots\bar{G}\ldots K, BD\bar{F}\bar{G}\ldots I\ldots K, BD\bar{G}\ldots I\ldots K, BD\bar{F}H\ldots\bar{G}\ldots K, BDH\ldots\bar{G}\ldots K, BED\bar{F}\bar{G}\ldots I\ldots K, BED\bar{G}\ldots I\ldots K, BED\bar{F}H\ldots\bar{G}\ldots K, BEDH\ldots\bar{G}\ldots K$ |

## 6. Spamming Modes

The mail classification shown in table 4 can be used to deduce modes for spamming by grouping those entries which use nodes that belong to the same participating component i.e. Sender, ESP or IESN. Thus obtained spamming modes are presented in table 5.

Table 5: Spamming Modes

| Mode | Participation | Protocols Exploited |
|---|---|---|
| 1. | ESP | SMTP |
| 2. | ESP & IESN | SMTP and $\overline{SMTP *}$ |
| 3. | Sender | $\overline{SMTP *}$ |
| 4. | Sender & IESN | SMTP and $\overline{SMTP *}$ |
| 5. | Sender & ESP | SMTP, HTTP(S), INT, and SMTP* |
| 6. | Sender, ESP and IESN | SMTP, HTTP(S), INT, and SMTP* |

Spamming done using modes 1 or 2 represent ESPs being involved in spamming either directly or indirectly owing to some security violations due to viruses or worms. Option 3 represent spamming directly to MTAs of receiving organization without use of ESPs or IESN using Internet service providers (ISP) that in this case simply forward TCP packets of the sender on either port 25 or 587. The spamming option 4 represents spamming using indirect means by making use of one or more types of IESN. Spammers in this option are not restricted to port 25 and 587 only. In Option 5, spammers use ESPs services for sending spam. Option 6 represents spammers exploiting ESPs by using ESPs to forward spam to intermediate nodes i.e. IESNs. This option of spamming is unlikely to occur without the use of support of the ESPs.

Besides the spamming modes identified in table 5; infected ESPs on sender or receiving side make other spamming modes also possible. These include situations in which spammer is sending the spam directly to internal MTA or MDA or Mail Server. A receiving ESP may itself send spam to the mailboxes on its Mail Server for its financial gain; however the chances for such a form of spamming is low.

The possibility of detecting spam and filtering it can be performed at various places in the system. Depending on the spamming mode used for spamming different possibilities for its detection exist. These possibilities are outlined in table 6.

**Table 6: Possible detection places**

| ***Place of Detection*** | ***Spamming Modes*** | | | | | |
|---|---|---|---|---|---|---|
| | ***I*** | ***II*** | ***III*** | ***IV*** | ***V*** | ***VI*** |
| *Recipient* | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| *Receiving ESP* | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| *IESN* | × | ✓ | × | ✓ | × | ✓ |
| *ISP* | × | × | ✓ | ✓ | ✓ | ✓ |
| *Sending ESP* | × | × | × | × | ✓ | ✓ |

A ✓ indicates the possibility of designing a spam detection technique and a × marks no such possibility. Possibility of detecting spam at recipient and receiving ESP exist for all modes of spamming. However, detecting techniques working at these places are considered to be least efficient because they cannot save network recourses from being misused by spammers. Detecting schemes that are closes to the sender would prove to be more successful.

## 7. Conclusion

Spam originates from an illegitimate Sender or in some cases from ESPs for financial gains. There are numerous possible routes for its flow through the Internet that besides exploiting other protocols, mainly exploits the trust build into the SMTP protocol and its extensions. Spamming with some reasonable assumptions can be grouped in six distinct classes depending on the type of the participating components. Spam detection and filtering can be performed at various places in the system; however any detection measure that is close to the sender would prove to be a more successful prevention. Such a prevention measure requires a wide scale change in the existing SMTP based e-mail and its adoption by ESPs

## References

[1] M. Siponen and C. Stucke, “Effective Anti-Spam strategies in companies: An international Study”, procedings of the 39th Annual Hawaii International Conference on System Science (HICSS’06), vol. 6, pp. 127c-136c, Jan, 2006.

[2] M. Tariq Banday and Jameel A. Qadri, “SPAM – Technological and Legal Aspects”, Kashmir University Law Review (KULR), vol. XIII, no. XIII, pp. 231-264, 2006.

[3] Charles M. Kozierok, “TCP/IP Guide: A complete, Illustrated Internet protocols reference”, ISBN 81-7366-464-1, Oct 2005.

[4] Klensin, J., “Simple Mail Transfer Protocol”, RFC 2821, Apr 2001.

[5] IETF, “Internet Engineering Task Force”, http://www.ietf.org.

[6] G. Schryen, “A Formal Approach towards Assessing the Effectiveness of Anti-spam Procedures”, proceedings of 39th Hawaii International Conference on System Science, vol. 6 pp. 129a-129a May 2006.

[7] D. Atkins and R. Austein. “Threat analysis of the Domain Name System (DNS)”, RFC 3833, Aug 2004.

[8] P.J. Sandford, J. M. Sandford, D. J. Parish, "Analysis of SMTP Connection Characteristics for Detecting Spam Relays", International Multi-Conference on Computing in the Global Information Technology - (ICCGI'06), pp.68, 2006.

[9] P. Tzerefos, C. Smythe, I. Stergiou, S. Cvetkovic, “A comparative study of Simple Mail Transfer Protocol (SMTP), Post Office Protocol (POP) and X.400 Electronic Mail Protocols, pp. 545 – 554, Nov 1997.

[10] Surmacz, R. Tomasz, “Reliability of e-mail delivery in the era of spam”, International Conference on Dependability of Computer Systems, DepCoS-RELCOMEX’07, pp. 198 – 204, Jun 2007.

[11] Z. Duan, K. Gopalan, X. Yuan, “Behavioral characteristics of spammers and their network reachability properties”, IEEE International Conference on Communications, art. no. 4288706, pp. 164-171, 2007.

芽|Sprouts

Working Papers on Information Systems | ISSN 1535-6078

*Office*:
Sprouts
University of Amsterdam
Roetersstraat 11, Room E 2.74
1018 WB Amsterdam, Netherlands
Email: admin@sprouts.aisnet.org