SPAM AND OPEN RELAY BLOCKING SYSTEM

NOR HAFIZAH BINTI KHADZIR

A report submitted in partial fulfillment of the
requirements for the award of the degree of
Bachelor of Computer Science (Computer Systems & Network)

Faculty of Computer Systems & Software Engineering
University College of Engineering & Technology Malaysia

NOVEMBER, 2005

ABSTRACT

Spam or Unsolicited Bulk Email (UBE) is an electronic mail message that is sent to a large number of recipients who have not requested the message. Spam often contains pornographic, offensive, deceptive and carrying viruses. There are many types of anti-spam had been used to capture the spam but it is still appearing at user mailbox. This is because the settings of the spam filter had been known by spammer. The aim of this project is to develop spam filter that not only can be set by the system but also by the user. Spam and Open Relay Blocking System are anti-spam software, that will filter e-mail by header and sender that had been set either by user or the system. SORBS may help user through the setting so that user will not lose any white mail. SORBS will be applied at personal computer or notebook. It is through the use of SORBS methods that spam can be reduced from entering the user mailbox. Besides, setting that had been done and only known by the user make spammer take time to attack the user e-mail filter.

ABSTRAK

"*Spam*" adalah surat elektronik yang telah dihantar dalam kuantiti yang banyak kepada penerima yang tidak mengkehendakinya. "*Spam*" kadang-kala mengandungi kandungan yang berunsurkan pronografi, agresif, penipuan dan membawa virus. Terdapat pelbagai jenis anti-spam yang digunakan untuk menyekat kemasukan "*spam*" tetapi "*spam*" masih lagi dapat dikesan pada peti simpanan surat elektronik penerima. Masalah ini timbul disebabkan syarat penapisan anti-spam tersebut telah diketahui oleh menghantar "*spam*". Tujuan projek ini dihasilkan adalah untuk membina penapis "*spam*" yang bukan sahaja sistem yang menentukan syarat penapisan malah pengguna juga dapat melakukan syarat penapisan mengikut kesesuaian mereka. Spam and Open Relay Blocking System adalah perisian anti-spam yang menapis surat elektronik berdasarkan pengirim dan tajuk surat elektronik. Perisian SORBS diaplikasikan pada komputer peribadi atau komputer bimbit. Dengan menggunakan SORBS, masalah "*spam*" dapat dikurangkan daripada memasuki peti simpanan surat elektronik. Ini kerana syarat penapisan anti-spam hanya diketahui dan dilakukan oleh pengguna dan ini menyukarkan penghantar "*spam*" untuk menyerang perisian tersebut.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURE

ABBREVIATION

ACK	-	Acknowledgement
AOL	-	American Online
DNS	-	Domain Name Server
EDU	-	Education
GMAIL	-	Google mail
GOV	-	Government
IMAP	-	Internet Message Access Protocol
IP	-	Internet Protocol
ISP	-	Internet Service Provider
MAPS	-	Mail Abuse Prevention System
MIL	-	Military
MSN	-	Microsoft Network
ORG	-	Organization
POP	-	Post Office Protocol
POP3	-	Post Office Protocol version 3
RAM	-	Random Access Memory
RBL	-	Real time Black hole List
RCP	-	Receiver
SDLC	-	System Development Life Cycle
SMTP	-	Simple Mail Transfer Protocol
SORBS	-	Spam and Open Relay Blocking System
SVM	-	Support Vector Machine
TCP/IP	-	Transmission Control Protocol / Internet Protocol
UBE	-	Unsolicited Bulk E-mail
US	-	United State

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Spam or unsolicited mail means mail that had not been accepted to receive by the recipient. The spam had been created by spammer to send a thousand of unsolicited e-mail messages to business mail with aim to slow down e-mail system and young or inexperienced e-mail user with sexual material.

Until now, spam is still being a big issue for e-mail service. This can be referred to the survey conducted by Halverson Organization at Appendix A [1]. The research results until 2004, show that spam e-mail is not in the small quantity but it has thousand of it. According to a survey conducted by Trend Micro Incorporated, until August 2005, Malaysia spam report is 0.44 per cent out of 42.12 per cent for Asia spam report [2].

Although every e-mail had been providing with anti-spam filter, spam is still appear in receiver inbox. This is because the anti-spam filter can not detect the spam mail if the spammer had changed the pattern of the e-mail. Therefore, anti-spam filter that will be used should be flexible to make sure the spam cannot enter the receiver inbox.

1.2 Problem Statement

One of the principles that are important to effective construction of solutions is the diversity principle. The diversity means spammers employ a lot of tricks to throw identification off course [3].

This sometimes misleads the filters, especially when spammer made up a new trick. This will cause the filter let the spam through while others will pick up on it. Successful spam is something that requires a lot of effort from the spammer. The way spammer's attacks are by analyzing a specific filter type and exploit a weakness in its rules. When the general population of users takes advantage of range different filters, it will not be worthwhile for spammers to target one type. Targeting several types of filters is very complex.

Diversity in spam identification techniques will enhance the overall stopping power of spam identification. By using Spam and Open Relay Blocking System, user can make their own setting besides some setting been provide by system. This will make the spammers more difficult to change trick because only the user who user the filter know how the setting had been done. This will reduce number of spam from entering the mailbox.

1.3 Objectives

The objectives of the project are:

i. To develop prototype of Spam and Open Relay Blocking System at personal computer which have these characteristics:
    a. Filter mail by From Address (sender address).
    b. Filter mail by Subject (e-mail header)

1.4 Scopes

The scopes of the project are:

i. To develop anti-spam that can be apply at personal computer and user can control it.

ii. Spam and Open Relay Blocking System using in window environment.

iii. Spam and Open Relay Blocking System will filter all e-mail in user inbox and delete spam mail according to system and user setting.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter contain all information for this project by gathering result about e-mail protocol, SMTP Basics, analyzing the e-mail message, about spam, protection that should be use to secure mail, review of spam solution activity, research on related work, spam statistic, use "whois" to track down a spammer and take action against spammers. All information will be used while developing the system.

2.2 Basic E-mail Protocol

The e-mail technology that we use today was developed before the spam problem was a problem and before the widespread use the Internet. There are three (3) main protocols in use to send and receive basic email. E-mail is send using Simple Mail Transfer Protocol (SMTP) that was proposed in 1982 [4]. Post Office Protocol (POP) to receive e-mail proposed in 1984 [5]. In 1996, Internet Message Access Protocol (IMAP) was proposed [6]. Figure 2.1 shows e-mail flow from a sender to a recipient.

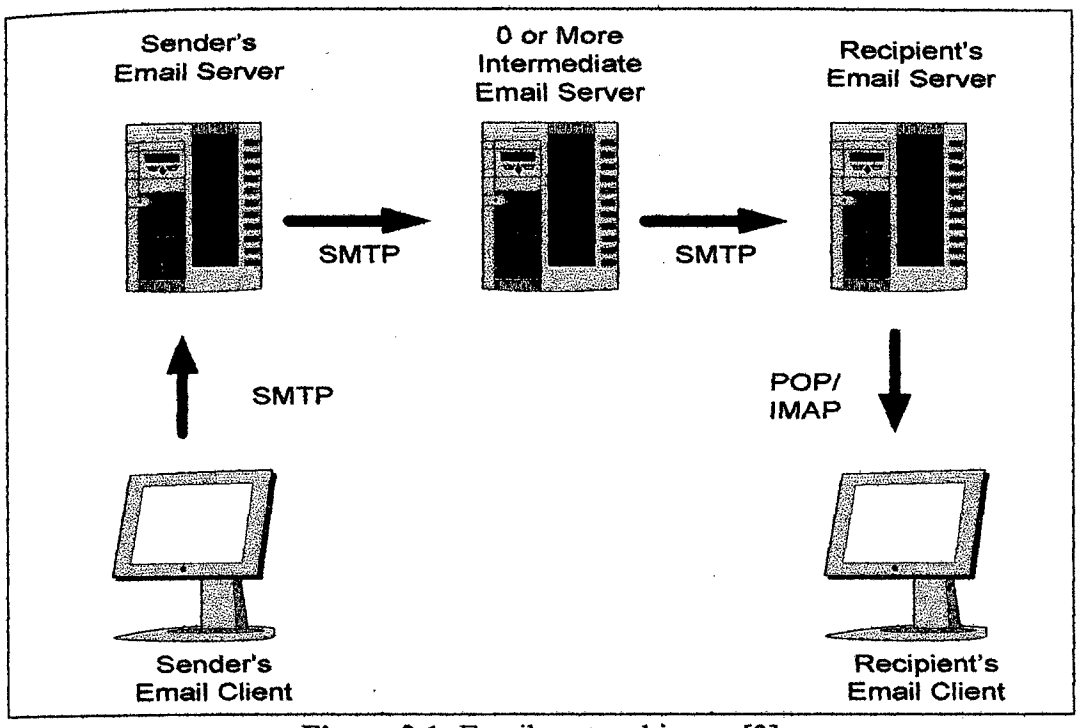

Figure 2.1: Email protocol in use [3]

Figure 2.1 show that sender commonly composes the email in an email "client" (end-user application). Email client is the software use by email users such as Yahoo Mail, Hotmail, GMail and etc. These three (3) protocols are implemented at clients. The client transmits the email message to the Sender's Email Server via SMTP when sending an email. An email server generally is computers that transmits email messages through the Internet to other email servers and make particular email messages available to users for download as appropriate. The email server is usually associated with and maintained by the sender's Internet Service Provider (ISP) or organization.

Then, the server transmits the message to Recipient's Email Server via SMTP same way the sender's email client transmitted the message to it. However, sometime it maybe not the Recipient's Email Server if the recipient or his/her ISP or organization has set up some sort of email forwarding system. Therefore, if there is Intermediate Email Server, the message gets transmitted in the same way via SMTP until it reaches the final Recipient's Email Server.

When the message reaches Recipient's Email Server, the message is deposits in a place designated for recipient. The message can be picking up by recipient using either POP or IMAP protocols. The difference between these two (2) receiving protocols is that with IMAP the message remains on the server. Meanwhile for POP the message is removed from the server. Users that use Web mail clients, e.g. Hotmail, would not necessarily use either POP or IMAP but would connect directly with the last server via a Web application.

These protocols have continued to perform effectively as describe while email traffic has increased by many order of magnitude. The Internet is also very different place now (in both size and nature) from 1980's and the protocol would be designed differently if the protocols designer know future problem. With regards to the spam problem in particular, the significant architectural issues lie within SMTP. POP and IMAP are just used to receiving email that already exists, therefore, are not relevant to the spam problem.

2.3 SMTP Basics

SMTP is a protocol that been used to send e-mail from sender to receiver. Figure 2.2 show that there are six (6) step process sending email via SMTP. First, connection is made to the server through a variety of means (connect) and server sends back an acknowledgement (ACK). Second, the HELO command is send to identify the sender's computer, followed by an acknowledgement from the server. Third, the MAIL FROM command sends to identify the sender, followed by an acknowledgement from the server. Fourth, the RCP TO command send to identify the receiver followed by an acknowledgement from the server. Fifth, the DATA command is send including the message itself, followed by a final acknowledgement from the server. Finally, the QUIT command is send, followed by the server closing the connection (close).

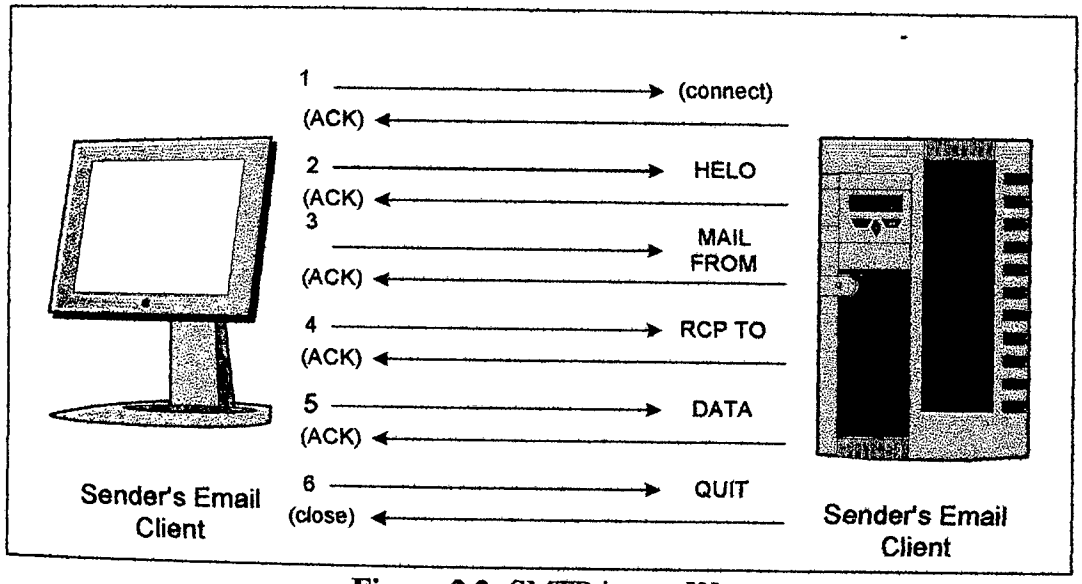

Figure 2.2: SMTP in use [3]

2.4 Analyze the E-mail Message.

First look at the full header of an e-mail message to begin the process of tracking the spammer. Some common methods to view the full headers include:

i. **Outlook Express**
   a. Go to File> Properties>Details> Message Source

ii. **Outlook**
   a. Go to View> Option> Internet Headers

iii. **Netscape Messenger**
   a. Go to View> Message Source

```
Received: from eyou.com [218.6.2.239] by ipswitch.com
(SMTPD32-7.12) id A5C32DAB0140; Thu, 17 Oct 2002 05:41:23 -0400
From: "surplusandcloseouts.com" <marketinging@eyou.com>
To: <jsmith2-NUL@ipswitch.com>
Subject: Cash in your Surplus Inventory
Sender: "surplusandcloseouts.com" <marketinging@eyou.com>
Mime-Version: 1.0
Content-Type: text/html; charset="ISO-8859-1"
Date: Thu, 17 Oct 2002 17:41:28 +0800
Reply-To: "surplusandcloseouts.com" <surplus@tgtrading.com>
Content-Transfer-Encoding: 8bit
Message-Id: <20021017054100.SM00226@eyou.com>
X-RCPT-TO: <jsmith2-NUL@ipswitch.com>
Status: U
X-UIDL: 329328911
```

Figure 2.3: Example of an e-mail header

Figure 2.3 shows that received lines go in reverse order starting at the top so the top line is the last server to handle the message. In this case the message was sent from eyou.com using IP address of 218.6.2.239 and was received by the server ipswitch.com. When an e-mail is passed through a server that does not belong to either the sender or the recipient it is called relaying.

Record the hostname and the IP addresses. In the example, eyou.com although this may not be involved, the spammer could have easily forged the hostname 218.6.2.239 an IP address is harder to forge than a hostname so this will probably lead to some useful contact information; tgtrading.com, while this could be

forged as well, the reply-to is often more likely to be valid, especially if it matches a hostname in the body of the email. www.surplusandcloseouts.com URL was given to yield useful information.

2.5 Spam

E-mail use for business purpose is fast on the rise as companies increasingly recognize electronic messaging as an efficient means of communication that is quicker and cheaper than more traditional method such as sending message by using letter that take more time and this can interrupt the company productivity. According to Mike Elgan, e-mail is a wonderful medium because of the flexibility, automatable, asynchronous and faster [7].

E-mail carrying offensive messages or confidential corporate information can create immense inconvenience and expense for company that has not equipped its mail server with the appropriate tools. The same goes for spammers who use e-mail system at work to send thousand of unsolicited e-mail messages / Unsolicited Bulk Email (UBE). **Unsolicited** means that the recipient has not granted verified permission for the message to be sent. **Bulk** means that the message is send as part of a larger collection of messages, all having substantively identical content [8].

Technical definition of spam first, the recipient's personal identity and context are irrelevant because the message is equally applicable to many other potential recipient. Second, the recipient has not verifiably granted deliberate, explicit and still – revocable permission for it to be sending. Third, the transmission and reception of the message appears to the recipient to give a disproportionate benefit to the sender [8].

In business field, spam not only consuming bandwidth and slowing down e-mail system, but also frustrating time-waster, forcing employees to sift through and delete mounds of junk mail. It also proves irritating and offensive to recipients who

feel their privacy has been invaded. However, there is a third aspect to spam – it constitutes a security hazard. [9]

Spam not only disrupts business email users but can affect the way email and internet services are used in the home. As the email is increases, so does the spam message that users receive. Spam writers often target young or inexperienced email user with much spam being sexually explicit material that may reach children or that is otherwise offensive to recipients who did not request to receive such material.

Spam is not limited to sexually explicit e-mail but it is also deceptive. Deceptive spam includes email that user misleading information either about who sent the email or what the email is regarding to trick the recipient into opening the mail because they think it's something that it isn't [10]. Spammer often provide false information in the 'from' such as from financial institution, instructing customers to enter their account details and internet banking password for verification. The spammer will take this opportunity access to customers account.

The action of spammers over the past year proofed that current technologies are not enough. Knowing that only a small percentage of their output will get past today's filters, spammers have responded by significantly cranking up the volume of emails they send. So networks are burned with even more junk than before. About 90 per cent of e-mail users receive spam or unsolicited commercial mail at least once a week, a recent survey conducted by the Gartner Group shows. The research results, issued in June 1999, revealed that almost half surveyed were spanned six or more times a week. According to some surveys, email traffic now consists of nearly four Spam messages for every legitimate one [11].

Therefore, Spam does not only attack business but also home users can receive unwanted sexually explicit and fraudulent emails.

2.5.1 Spam Characteristics

There is a connection between legitimate e-mail marketing and spam. Table 2.1 shows that spam has some or all the following characteristics:

Table 2.1: Spam characteristics [12]

	Spam characteristics	Explanation
i.	Address is not valid	User will get delivery error message if user try to reply spam e-mail.
ii.	Forged headers	Use to hide the origin of the e-mail that can cause difficulty to trace spammers.
iii.	Recipient identity is irrelevant	The e-mail is valid for other recipient.
iv.	Dictionary attack addresses	At 'To' address line is examined; user can see different variants of the recipient's e-mail address. For example, if e-mail address happened to be SaraLee@hotmail.com, user would see SaraLee@yahoo.com, SaraLee@aol.com, Sarahly@hotmail.com etc.
v.	Subject line has no bearing on the content of the e-mail	Spammer use programs to randomly generate characters in the subject line in order to by pass spam filter.
vi.	E-mail content is of dubious nature	The header topic would cover get rich schemes, body enhancement etc.
vii.	Unsubscribe does not work in spam e-mail	Trying not to unsubscribe from spam e-mail is often the case that the link does not work or opens up an advertisement's website.
viii.	May contain hidden scripts	If spam contains HTML, it may contain hidden JavaScript that can open up website and activate advertisement popup windows.

2.5.2 Top Spam E-mails

Table 2.2 shows a sampling of the top spam e-mails currently roaming the Internet:

Table 2.2: List of top spam e-mails [12]

Top spam e-mails	Explanation
URGENT AND CONFIDENTIAL	Nigerian spam asking e-mail recipients for bank information in order to deposit large amounts of money from Nigeria
GET A FREE PASS TO THOUSANDS OF XXX SITES	This advertising pornographic websites
Protect Your Computer Against viruses for $9.95	Offer for cheap anti-virus software program
Verification Department	Credit card scam e-mail supports to be from credit card company offering the recipient the opportunity to receive a credit card, irrespective of credit history or employment status.
Online Auction Marketing Secrets	Spam scam that claims to reveal secret to online auctioning success.
Printer Cartridges-Save up to 80% - free shipping offer	Selling printer cartridge with enormous savings
$100 FREE, Please Play Now	Spam e-mail is an offer to join an online to receive free credit $100

2.5.3 Spamming Techniques

Spamming is not difficult to do. There are many websites offer clients necessary tools in order to send "Bulk e-mail". Beside, it also contains millions of valid e-mail addresses. The spammer can conceal their identity using open relay, open proxies and zombies [13]. In the future there will be a new trend where

spammers team up with virus writers so that infected machines act as spam relays. Table 2.3 shows the technique that usually use by spammer to attack user e-mail.

Table 2.3: Spamming techniques

Spamming technique	Explanation
i. Dictionary attack	The spammer uses software to create every conceivable variation of a person's name. They would use common first name and common last name and combine them.
ii. Spambots	Software that written in C for speed and portability with the primary purpose to crawl through the Internet looking for e-mail addresses. There are also known as spiders, deviant web crawlers, harvesters or robots.
iii. Spoofing	Spammer makes the e-mail coming from other than spammer and the message appear as it coming from someone you know or a company that you trust. Spammers will driving around hotspot location such as university campus and send spam after they had found unsecured access point.
iv. Bandwidth theft	Spammer's now targeting wireless network and unsecured wireless access point to send out bulk e-mail.
v. Bypassing filters	Designed to look for occurrences of certain words in an e-mail message and the spammers task is to get these words to the recipient without having the e-mail known as spam. The techniques are imbedded words in the image, using spaces and symbols between letters of words e.g. PORN would be P O R N and combination of letters and numbers e.g. VIDEO TAPE would be VIDEO T4PE.