WHITEPAPER SERIES

Anti-Spam Capabilities of Internet Exchange Version 3.1

Version 1.0 **September 1998**

Hong Kong Computer Center, 20/F
54-62 Lockhart Road
Wan Chai
HONG KONG
Tel: +852 2520-0300
Fax: +852 2648-5913

The Peak Tower, 15/F
107 Alfaro Street
Salcedo Village, Makati City
PHILIPPINES
+63 (2) 811-3999
+63 (2) 811-3939

USA Support/Sales: +1 (408) 481-9985
USA Fax: +1 (888) 562-3561

Email: info@ima.com
Website: http://www.ima.com

Please send all comments and suggestions related to the Whitepaper Series to *doc@ima.com*

TABLE OF CONTENTS

INTRODUCTION

Ever since the Internet moved itself into mainstream society several years ago, people have been assaulted with various forms of electronic solicitations. The Internet's version of the postal system's mass mailing has become particularly offending for many Internet users. This type of unsolicited email, referred to as *spam*, differs considerably from its postal system counterpart.

Unlike unsolicited postal mail, spam is very cheap to produce and to disseminate. In addition, the delivery costs, and sometimes local storage costs, are shouldered by the recipient of the message, rather than the sender. This significant difference in economics makes spam mail not only a nuisance to the recipient, but also costly at the same time.

BACKGROUND

Due to the ease and low costs associated with spam, many people and organizations have jumped onto the spammers bandwagon to send their message as far and wide as possible. The early spammers were not very sophisticated with their methods, resulting in their identities and email addresses becoming known easily. This resulted in offended recipients retaliating against the spammers by complaining to their Internet service providers, and/or sending spam mail in return (also known as mail bombing). In some cases the resulting email bombing practically shut down several Internet service providers, and resulted in providers disconnecting offending spammers.

> **Note:** it is our position that mail bombing is just as anti-social, and potentially illegal as spamming and is not a recommended course of action.

As more and more spammers became uneasy about the prospect of sending out say a million spam messages, only to have half of them returned, and more than likely be the target of multiple mail bombings, better techniques had to be developed. About this time it was apparent that for any spammer to keep on sending junk mail, it was imperative that they be able to conceal their identity. To this day, this is the underlying premise by which most spammers operate, and fortunately provides us with some tools for the identification and prevention of much of this undesirable message traffic.

Present day spammers use many different techniques, usually in tandem, in order to get his message out to the widest audience without his identity being detected. Some of these include the modification of the message headers, providing bogus message envelopes, and the routing of their traffic through innocent, unsuspecting third parties.

RFC-822 Header Masquerading

Internet electronic mail is made up of the message body, which may or may not contain attachments, and the message header. The message header, whose format is defined by the Internet standard RFC-822, contains information related to the message, such as the subject, submission Date, original recipients, etc. (see Figure 1). These information are presented to the user directly by their mail user agent (UA). Common UA's include Lotus cc:Mail, Microsoft Outlook Express, and Eudora from Qualcom.

Date: Wed, 02 Sep 1997 19:49:00 –0700 (PDT)
To: Rommel Fajardo <rommel@ima.com>
From: Jules Hernandez <jhernandez@inod.com>
Subject: Class Reunion
Cc: Patricia Rosero <tricia_r.l@usa.net>
Jim Morisson <jim_morisson@doors.com>
Tim Kehres <kehres@ima.com>

Figure 1. Basic Internet Email Headers

In order to conceal from the recipient the identity of the true sender of the message, the *From:* header is usually forged to point somewhere other than the true sender. Spammers that have somehow managed to retain any decency usually set this address to a non-existent domain, so as not to cause mail bombing of an innocent bystander. Unfortunately, this practice is not universal, so it is always advisable to be extremely cautious if any kind of retaliatory action is contemplated.

At the same time that the spammer covers up the sender address, bogus *To:* and *Cc:* addresses are usually employed.

Other information present in the RFC-822 headers includes trace information that are recorded as messages move from one machine to the next in the course of being delivered to their final recipient (See Figure 2). Most of the trace information can be found in the RFC-822 *Received:* message header, which records the transit machine name and a time stamp. Most *Received:* headers also contain the name of the machine that sent the message.

Return-Path: <rommel@ima.com>
Received: from pusa.ima.com by pimail.ima.com (8.8.7/1.14.5) with ESMTP id NAA18988; Wed 02 Sep 1998 13:35:17 GMT
Received: from cc:Mail by pusa.ima.com (IMA Internet Exchange 3.1 beta release) id 0002A35; Wed 02 Sep 1998 16:49:55 GMT
Message-Id: <00002A35.C21379@ima.com>
Date: Wed, 02 Sep 1998 16:55:47 +0800 GMT
From: Rommel Fajardo <rommel@ima.com>
To: Jim Morisson <jim_morisson@hotmail.com>
Subject: Class Reunion
Cc: Patricia Rosero <tricia_r.l@usa.net>
Jerry Garcia <jgarcia@gdead.com>

Figure 2. Extended Internet Email Headers

Since these *Received:* headers contain information that can be used to trace the origin of a message, many spammers do whatever they can in order to cover their tracks.

While it is not possible for them to forge headers produced by downstream message transport agents, they can attempt to cover up the first hops of a message. What spammers usually do is to manually create several sets of headers, and then to inject the message into the message transport agents, usually through a dialup connection with a local ISP, making it appear that the message originated with their forged first *Received:* trace. A person with enough experience to interpret these headers, however, can usually determine the ISP and time of submission if he or she is explicitly searching for this information. Once identified, the ISP can be contacted. If the ISP is willing to cross-reference its dialup connection logs, the spammer can then be identified.

SMTP Envelope Masquerading

Email messages flow through the Internet with a separate message envelope and message content. All the information needed to transport the message is contained in the envelope, including the destination and sender addresses. The message content consists of the message body and the RFC-822 headers as described in the previous section. Other than the insertion of trace information into the message as they flow from one system to the next, the message content is usually not touched or modified in any way by message transport agents.

Messages are conveyed from one system to the next using the Internet Simple Mail Transfer Protocol or SMTP (see Figure 3). This is a simple language that two computers use to exchange electronic email messages. Within the Internet, a machine that wants to send email to another machine establishes a TCP connection to port 25 of the destination machine or the host. The host runs a SMTP server daemon that listens for connection requests on TCP port 25. The initial dialog between two computers exchanging email identifies the sending computer, the message sender, and the recipients of the message.

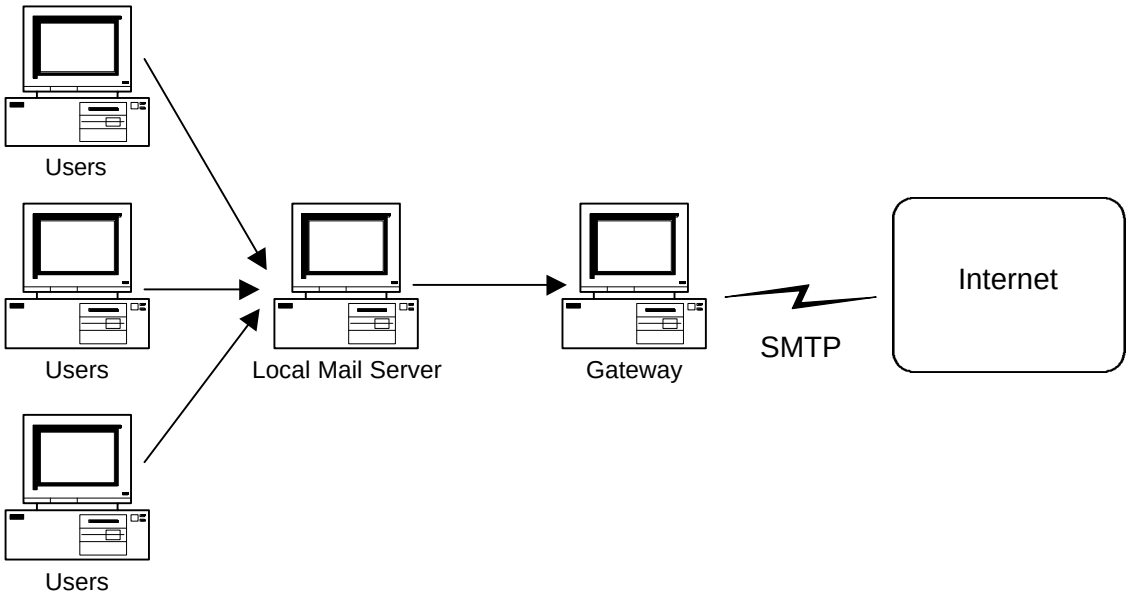

Figure 3. SMTP transports email messages on the Internet

Today's spammers usually try to cover their tracks by either forging the system name he is sending from and/or the sender address. The first exchange in any SMTP session is known as the *HELO* command, where the connecting site sends its machine name. Until very recently, most email system designers have adhered to the letter of the law as set down by the various Internet standards. These standards indicate that while it is legal to try to verify the supplied name based upon the known connection address of the remote site, it is illegal to deny service should this information is not validated. Recent changes to many systems around the Internet, however, now allow for the dropping of connections when the supplied information is obviously bogus.

The second phase of the SMTP dialog identifies the sender of the message. The SMTP *MAIL FROM* command is used to supply the envelope sender address. The supplied sender address for spam mail is almost always bogus.

SMTP Relaying Using Third Parties

Another technique that spammers use to conceal their identities and reduce costs is to shift the burden of the final message delivery to unsuspecting and unprotected sites across the Internet. This practice, known as third-party relaying, is perhaps one of the more controversial, as many consider this practice to be an unethical, immoral, and in many jurisdictions, illegal theft of service. Although third party relaying has some legitimate uses (such as in debugging mail connectivity), it provides spammers with a tool for increasing the number of junk messages they can send. And in cases where the spammer employs relaying while using forged headers to point to the relay site as the source of the junk mail, most of the wrath of the spam recipients is typically focused on the relaying parties, with some of them being blacklisted by the rest of the Internet community.

The method of performing the relaying is quite simple. During the SMTP transaction, the envelope information is conveyed first, followed by a single copy of the message. What most spammers usually do is to create a message envelope with many - perhaps hundreds or more recipients - and then in a single message transfer, move the responsibility of delivery to the envelope recipients to the relay site. Once received, the relay site will then attempt to deliver the message to all recipient addresses found in the envelope one at a time. This can result in significant CPU as well as network bandwidth utilization at the relay site.

Due to practical limitations found in many of the common Internet message transfer agents, arbitrarily long envelope recipient lists are not usually employed for reliability reasons (not all sites can effectively handle them). Instead what most spammers usually do is batch a spam run into many smaller batches. Say 100,000 messages are to be sent out via a relay - the spammer can group these into 1,000 messages with each message constructed with 100 recipients in the envelope.

This practice can be especially damaging for a relay that performs parallel delivery of messages and is not configured to handle high message loads. In such cases, the spam injected to the system can practically paralyze the normal delivery of other messages until it is flushed or removed from the system.

SPAM PREVENTION TECHNIQUES

With all the tools available to the modern spammer, it can seem like a formidable task to protect oneself from such abuse. The good news is that many of the loopholes used by spammers can be closed. The down side is that as many different methods are employed by the spammers, at least as many different protective measures need to be used to combat spam.

Connection Control

When an SMTP connection is established between two systems, the electronic, or IP address of the sending site is known to the called site. This address cannot be forged, as it is required in order for communication between the parties to proceed. At this point, decisions can be made regarding whether or not to accept or proceed with the connection and resulting data transmission.

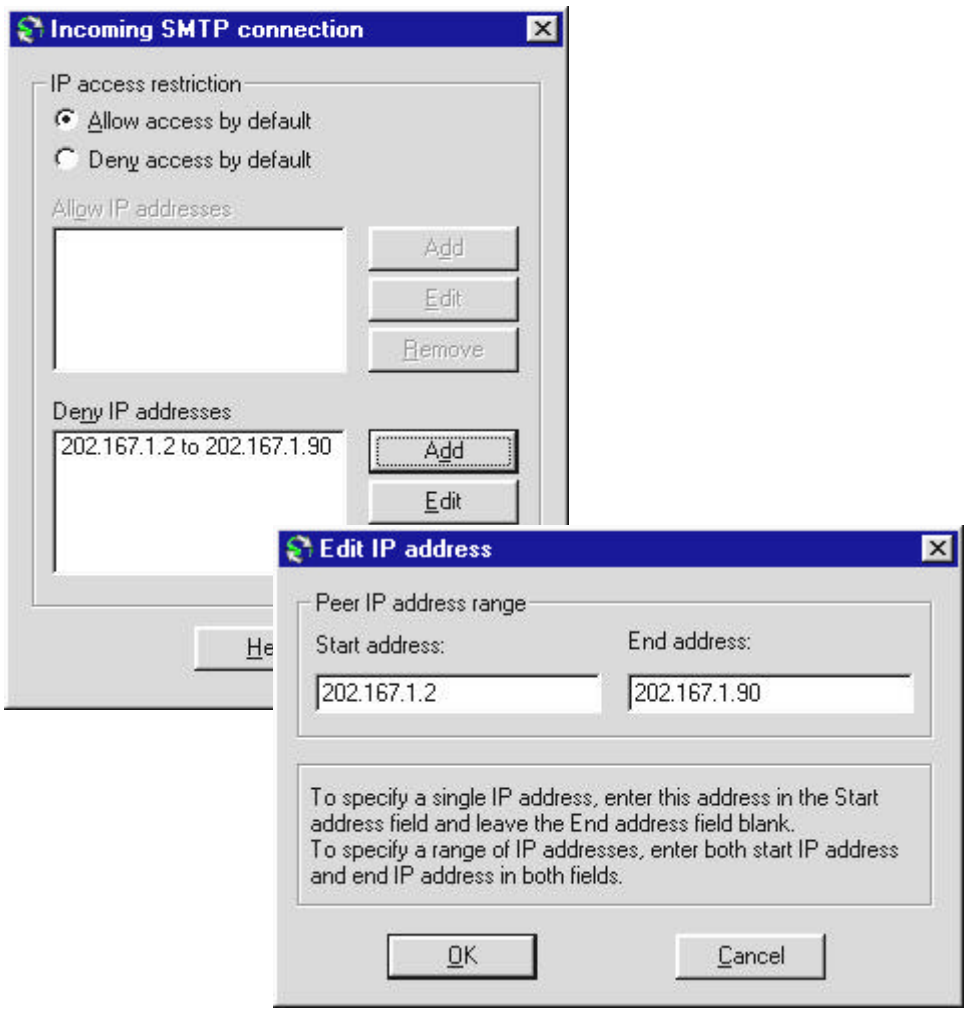

Figure 4. Internet Exchange's GUI for configuring **Incoming SMTP Connection**

IP Address Filtering

If the IP address of the offending spamming organization is known, it is possible within Internet Exchange to make a list of IP addresses such that SMTP connection requests from these addresses will be denied at the SMTP session establishment phase (see Figure 4). This has the advantage of terminating the spam message(s) before any network or CPU resources have been consumed. The downside is that the IP address of the sending site must be known beforehand. If access to the organization's router is available, this is another place where IP access control can be installed.

Remote Site Name Verification

The initial SMTP command requires the connecting system to identify itself through the *HELO* command. Normally, this information is used only for logging purposes, and no verification is performed. Internet Exchange can be configured to perform a reverse address lookup within the Domain Name System based upon the known IP address of the connecting site (see Figure 5). If the supplied name and the name returned by the DNS do not match, Internet Exchange will assume that this is a fraudulent connection, and will terminate the session before any additional CPU and bandwidth resources are consumed.

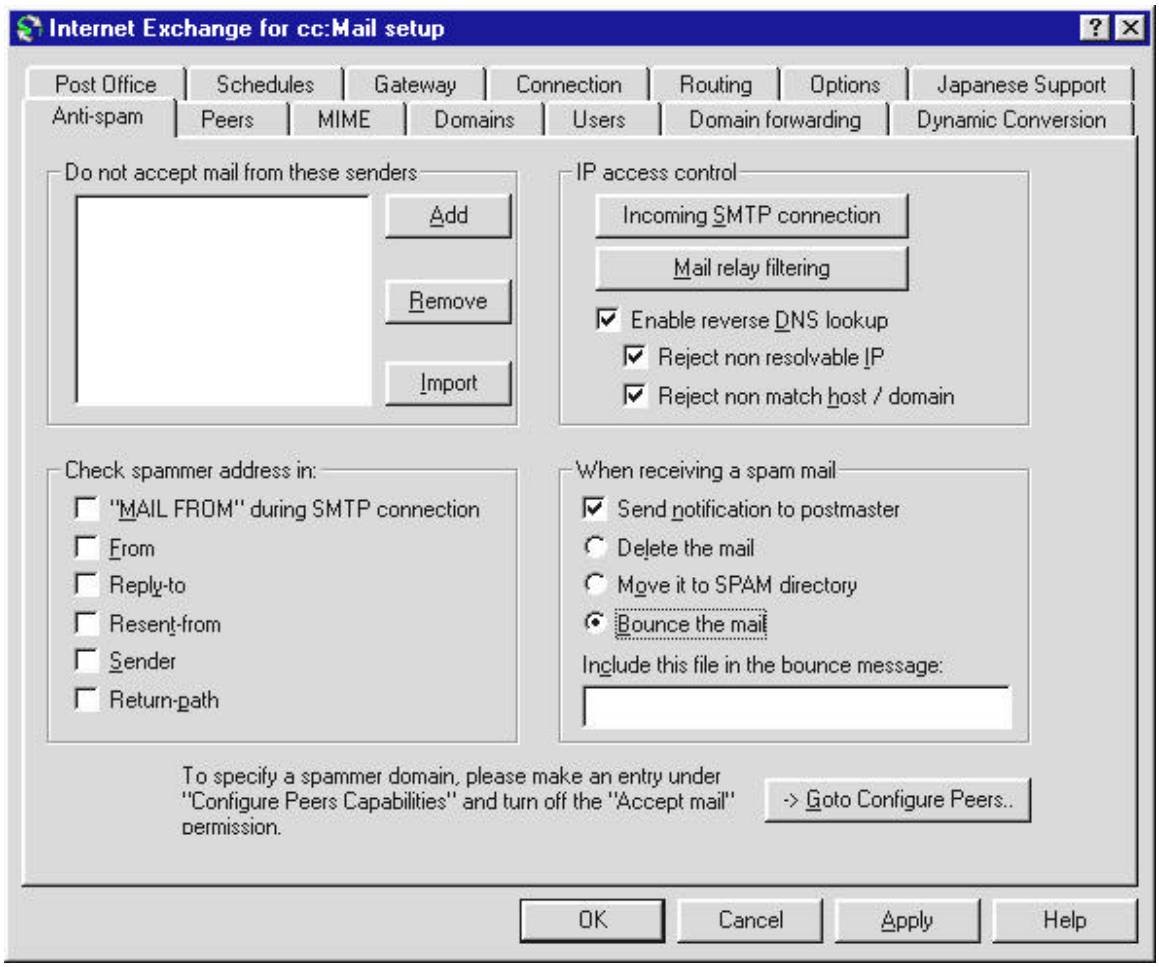

Figure 5. GUI for activating **Enable Reverse DNS Lookup** option

It should be noted that the reverse address lookup, even when using the caching DNS server within Internet Exchange, can take a noticeable period of time to complete for non-cached data.

SMTP Envelope Detection

After the initial SMTP greeting that identifies the calling SMTP system, the originating system identifies the envelope sender in the *MAIL FROM* command. If the spammer has not forged the envelope sender, then this will point back to the spammer. Most of the time, while the envelope sender will be forged, it may have been forged to a well-known address. If this address is known and consistent, Internet Exchange may be configured to terminate the SMTP session upon receipt of blacklisted envelope senders.

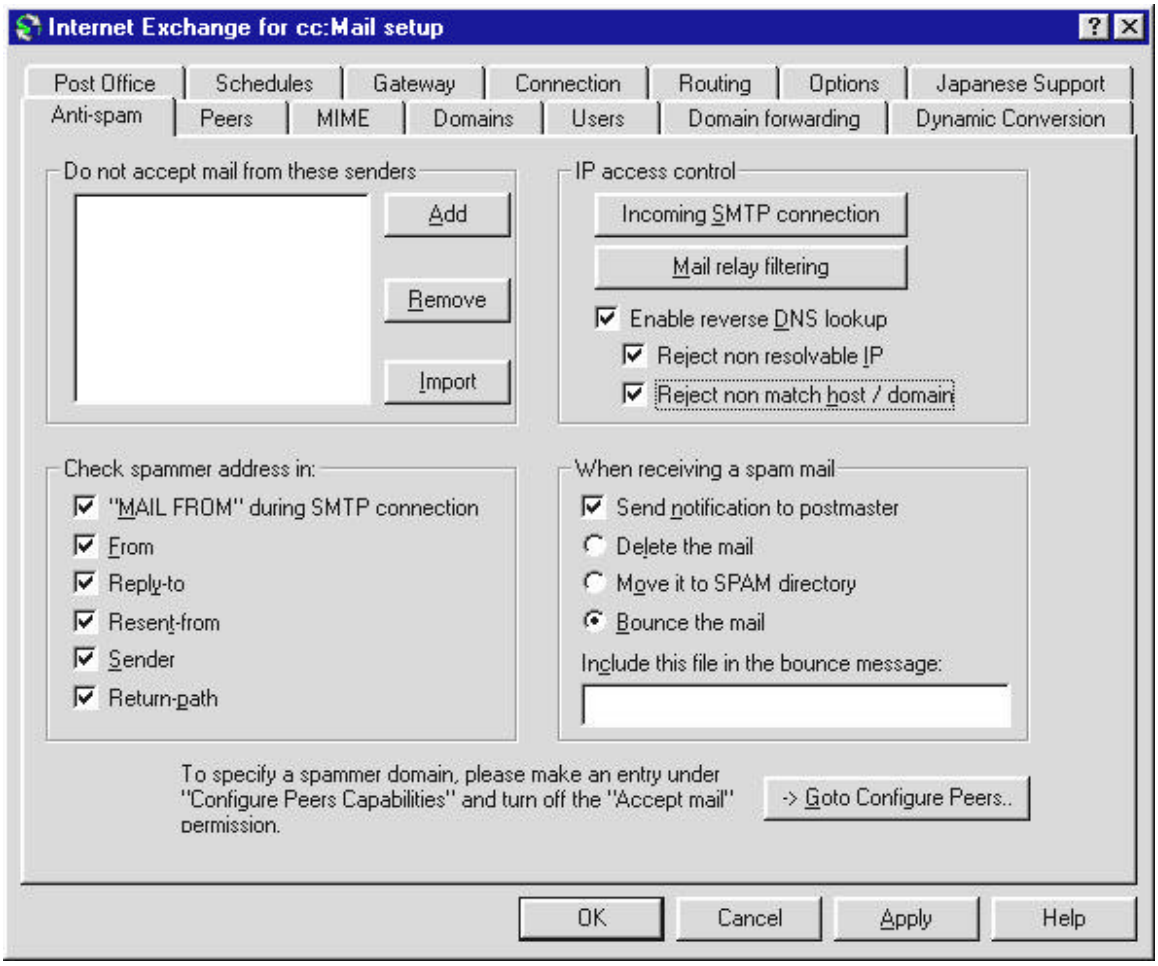

Figure 6. GUI for allowing the checking of RFC-822 headers in received messages

RFC-822 Header Detection

If a spammers message gets beyond the IP address control and the SMTP envelope detection techniques, it is possible to perform some tests based upon data in the RFC-822 message header. Internet Exchange allows the local administrator to screen incoming messages based upon addresses found in the following RFC-822 header

fields: *From:, Reply-To:, Resent-From:, Sender:,* and *Return-Path:.* These addresses are also configured in Internet Exchange's **Anti-spam** configuration screen (see Figure 6).

Mail Relay Filtering

Internet Exchange Version 3.1 and beyond offer facilities to prevent spammers from using Internet Exchange as a spam mail relay. Two different methods can be used to configure the system. If the sites that you wish to keep out are well known, including their network addresses, these addresses can be configured into a local blacklist control list using the (see Figure 7). If the option **Allow access by default** is selected, the gateway accepts all IP addresses except for those mentioned in the **Deny IP address** list. SMTP connection requests from those addresses contained in the **Deny IP address** list will then be rejected.

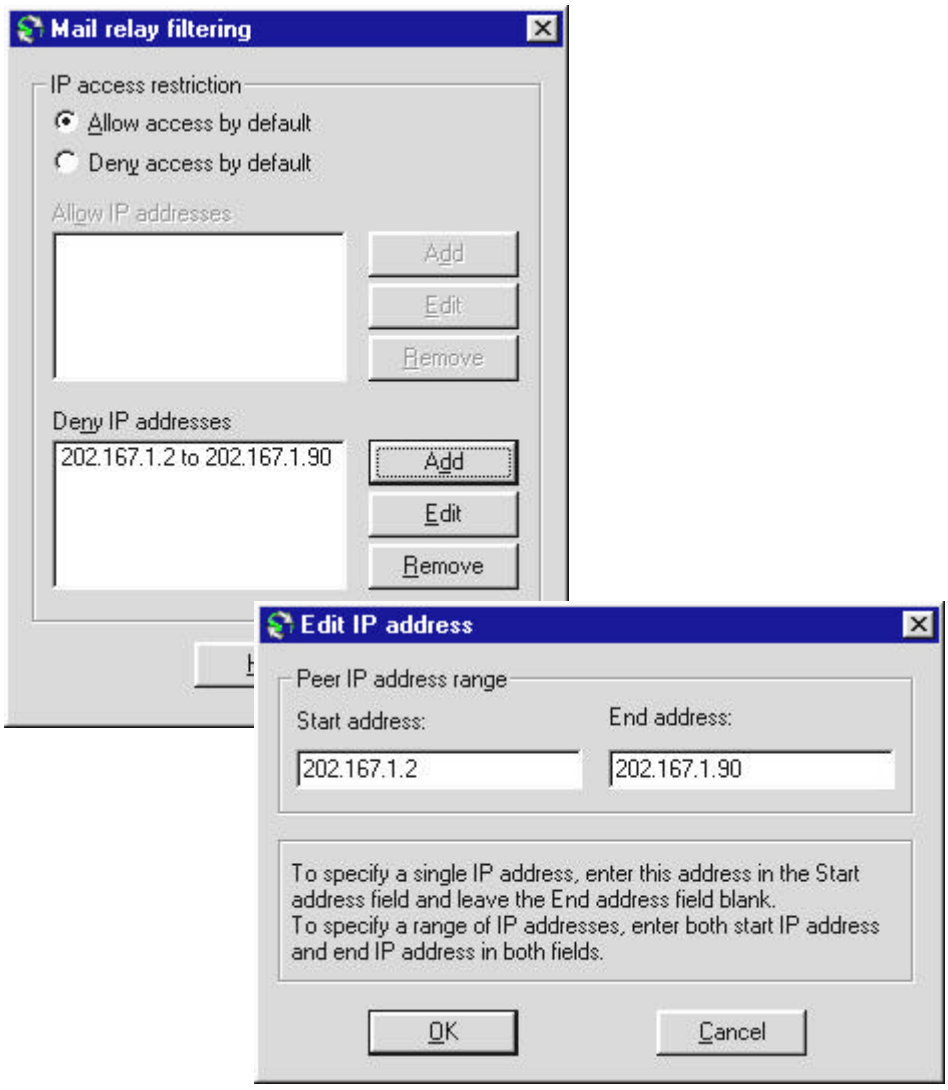

Figure 7. GUI for making a list of denied IP addresses

More often than not, however, the identity and network addresses of the potential attackers are not known beforehand. In this case, the best defense is to block the

relaying of all SMTP traffic, other than SMTP traffic from networks that you grant explicit access to. This is actually the recommended configuration for Internet Exchange. If mail relaying is needed, say for local MTA's or user workstations, these machine or network addresses can be configured into an allowed control list, which will permit Internet Exchange to continue to relay email for these special sites or machines (see Figure 8).

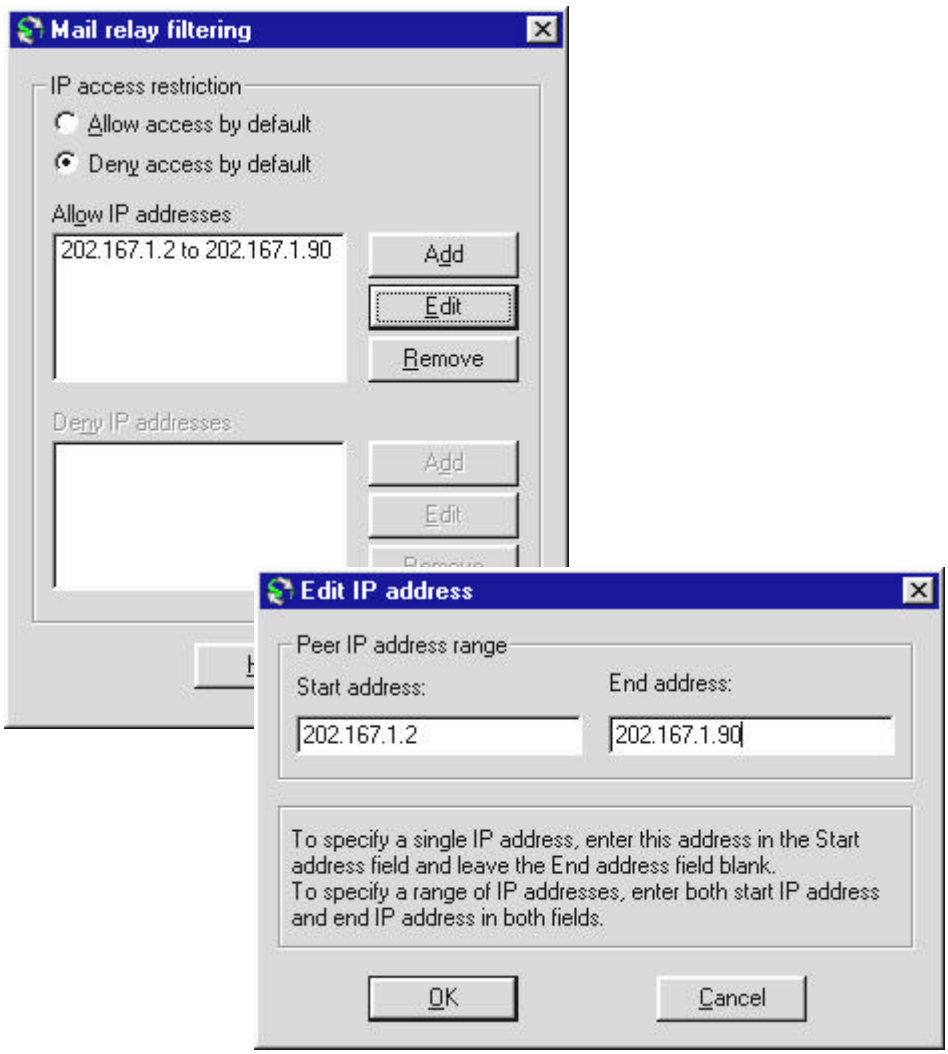

Figure 8. GUI for making a list of allowed IP addresses

Independent Spam Black Lists

Another approach for the preventing unwanted mail relaying is to rely upon spam blacklists maintained by other organizations. One such system is the *Mail Abuse Protection System - Realtime Blackhole List* (http://maps.vix.com). While there is no current support within Internet Exchange for automatic reference to any such lists, some Internet MTA's, such as *Sendmail,* have recently added support where connections will be dropped as soon as a blacklisted site is encountered.

Such systems, although effective, have certain drawbacks. One is that you are depending upon an external organization to decide for you what constitutes spam and what does not. While most of the time their selections will be quite accurate, there is no guarantee that all blacklisted sites are actual spammers. In addition, there is no way that the system can identify a potential spamming organization until after that particular organization has sent out junk mail.

Another drawback (or advantage, depending upon your point of view) of using such systems is that sites that have been hijacked and victimized by spammers who used them as relays often get blacklisted themselves. Anti-spam systems based on spam blacklists will block legitimate email originating from these sites until the sites' administrators convince the blacklist maintainer to remove them from the blacklists. The intent here is to encourage site administrators to take the proper precautions in defending the integrity of sites. However, this is considered as a drastic measure by some.

CONCLUSION

Like most technologies, the Internet has its downside. It provides society with a powerful tool for disseminating and gathering vital information, but at the same time, it can be used by unscrupulous individuals or organizations to send unsolicited email or spam mail to a large portion of the Internet community. This may seem harmless to some, but to serious Internet users, this practice is an invasion of personal email resources. Access to the Internet is not free (and in some areas very expensive), and spamming uses precious bandwidth resources which Internet service providers can ill-afford, given the already high volume of traffic being handled by the Internet. In addition, spamming contributes to reduced productivity in many organizations as recipients of spam mail spend considerable time in trying to sort which mail are junk and which are legitimate.

In response to the growing public clamor to curb the proliferation of spam mail, the U.S. government, in cooperation with organizations such as the Coalition Against Unsolicited Commercial Email (CAUCE) and Network Abuse Clearinghouse (NAC), has unveiled plans to implement new regulations that will penalize senders of unsolicited mail. However, the government will not find it easy to implement such laws. Certain groups, such as the Direct Marketing Association (DMA), are likely to lobby against the enactment of any anti-spam email law. Thus, it will take some time before an effective counter-measure against the proliferation of junk mail on the Internet can be put in place. And even when such laws are already implemented, it does not guarantee that spamming will stop entirely.

Fortunately, there are already several tools in the market that have been proven to be capable of effectively deterring spammers. One such tool, Internet Exchange 3.1, enables gateway administrators to protect their sites against spam mail by using several built-in functions, including SMTP connection control, IP address filtering, mail relay filtering, remote site name verification, and RFC-822 header detection. Using these functions, gateway administrators can easily configure Internet Exchange to ban known spammer sites from gaining access to the gateway, thereby protecting the integrity of their email systems.