



realtimepublishers.com<sup>tm</sup>

# *The Administrator Shortcut Guide<sup>tm</sup> To*



# Blocking Spam with Sender Validation



**SpamLion<sup>TM</sup>**  
Anti Spam Gateway

*Alan Sugano*

## Introduction

**By Sean Daily, Series Editor**

Welcome to *The Administrator Shortcut Guide to Blocking Spam with Sender Validation!*

The book you are about to read represents an entirely new modality of book publishing and a major first in the publishing industry. The founding concept behind [Realtimepublishers.com](http://Realtimepublishers.com) is the idea of providing readers with high-quality books about today's most critical IT topics—at no cost to the reader. Although this may sound like a somewhat impossible feat to achieve, it is made possible through the vision and generosity of corporate sponsors such as SpamLion, who agree to bear the book's production expenses and host the book on its Web site for the benefit of its Web site visitors.

It should be pointed out that the free nature of these books does not in any way diminish their quality. Without reservation, I can tell you that this book is the equivalent of any similar printed book you might find at your local bookstore (with the notable exception that it won't cost you \$30 to \$80). In addition to the free nature of the books, this publishing model provides other significant benefits. For example, the electronic nature of this eBook makes events such as chapter updates and additions, or the release of a new edition of the book possible to achieve in a far shorter timeframe than is possible with printed books. Because we publish our titles in “real-time”—that is, as chapters are written or revised by the author—you benefit from receiving the information immediately rather than having to wait months or years to receive a complete product.

Finally, I'd like to note that although it is true that the sponsor's Web site is the exclusive online location of the book, this book is by no means a paid advertisement. Realtimepublishers is an independent publishing company and maintains, by written agreement with the sponsor, 100% editorial control over the content of our titles. However, by hosting this information, SpamLion has set itself apart from its competitors by providing real value to its customers and transforming its site into a true technical resource library—not just a place to learn about its company and products. It is my opinion that this system of content delivery is not only of immeasurable value to readers, but represents the future of book publishing.

As series editor, it is my *raison d'être* to locate and work only with the industry's leading authors and editors, and publish books that help IT personnel, IT managers, and users to do their everyday jobs. To that end, I encourage and welcome your feedback on this or any other book in the Realtimepublishers.com series. If you would like to submit a comment, question, or suggestion, please do so by sending an email to [feedback@realtimepublishers.com](mailto:feedback@realtimepublishers.com), leaving feedback on our Web site at [www.realtimepublishers.com](http://www.realtimepublishers.com), or calling us at (707) 539-5280.

Thanks for reading, and enjoy!

Sean Daily

Series Editor

Introduction.....	i
Chapter 1: Spam and Spam Filtering Methods.....	1
A Brief Summary of Spam .....	2
Why Does Spam Exist? .....	2
How Spammers Get Your Email Address .....	3
Anti-Spam Legislation.....	4
Traditional Spam Blocking Methods.....	4
Keyword Searching.....	4
Advantages.....	5
Disadvantages .....	5
ORDB Checking.....	6
Advantages.....	7
Disadvantages .....	7
Whitelists .....	8
Advantages.....	8
Disadvantages .....	8
Blacklists.....	9
Advantages.....	9
Disadvantages .....	9
MX Record Lookup .....	9
Advantages.....	9
Disadvantages .....	10
Heuristics and Bayesian Filtering.....	10
Advantages.....	11
Disadvantages .....	11
Sender Validation.....	11
Advantages.....	15
Disadvantages .....	16
Summary.....	17

## Copyright Statement

© 2004 Realtimedpublishers.com, Inc. All rights reserved. This site contains materials that have been created, developed, or commissioned by, and published with the permission of, Realtimedpublishers.com, Inc. (the "Materials") and this site and any such Materials are protected by international copyright and trademark laws.

THE MATERIALS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. The Materials are subject to change without notice and do not represent a commitment on the part of Realtimedpublishers.com, Inc or its web site sponsors. In no event shall Realtimedpublishers.com, Inc. or its web site sponsors be held liable for technical or editorial errors or omissions contained in the Materials, including without limitation, for any direct, indirect, incidental, special, exemplary or consequential damages whatsoever resulting from the use of any information contained in the Materials.

The Materials (including but not limited to the text, images, audio, and/or video) may not be copied, reproduced, republished, uploaded, posted, transmitted, or distributed in any way, in whole or in part, except that one copy may be downloaded for your personal, non-commercial use on a single computer. In connection with such use, you may not modify or obscure any copyright or other proprietary notice.

The Materials may contain trademarks, services marks and logos that are the property of third parties. You are not permitted to use these trademarks, services marks or logos without prior written consent of such third parties.

Realtimedpublishers.com and the Realtimedpublishers logo are registered in the US Patent & Trademark Office. All other product or service names are the property of their respective owners.

If you have any questions about these terms, or if you would like information about licensing materials from Realtimedpublishers.com, please contact us via e-mail at [info@realtimedpublishers.com](mailto:info@realtimedpublishers.com).

## Chapter 1: Spam and Spam Filtering Methods

Spam—everyone hates it, and it has reached epidemic proportions over the past year. Some estimates list the spam rate as high as 70 percent of all Internet mail traffic. Spam clogs up Internet WAN lines and consumes a significant amount of a user’s day. If you have reached the point at which spam is annoying enough to do something about, this guide will help you do so by focusing on the following topics:

- Existing anti-spam technologies
- Enterprise-wide spam solutions
- Spam filtering topologies
- Spam product selection, implementation steps, cost justification, Return on Investment (ROI), and integration with existing mail packages
- Spam filtering add-ons, estimated costs, and implementation pitfalls

Installing spam filtering software on a single workstation is a fairly simple task; however, implementing an enterprise-wide spam filtering solution requires careful evaluation and planning. You can expect difficulties—particularly false positives—when implementing any spam solution. In addition to being prepared for these considerations, you need to be aware of and plan for ongoing maintenance, which can be a hidden cost when implementing a spam solution.

There are many methods to block spam:

- Keyword filtering
- Open Relay Database (ORDB) checking
- Whitelists
- Blacklists
- Mail Exchange (MX) record lookups
- Heuristics
- Sender validation

However, only sender validation holds the promise of blocking 100 percent of spam. Sender validation has been around for quite some time and has had success in the Post Office Protocol (POP3) market. The concept of sender validation is very simple. If a user that is on your “approved senders” list sends you a message, you get the message. If the user is not on the list, the message is quarantined. Most sender validation spam solutions deal with individual POP3 mailboxes. Although these individual solutions work well, such has not been the case for past network enterprise deployments of sender validation. Sender validation has been criticized as an undesirable solution for fighting spam in enterprise environments. To avoid any problems and benefit from the 100 percent blocking power of sender validation in an enterprise environment, simply select a vendor that has a mature sender validation solution. In this guide, we’ll examine how to avoid the pitfalls of sender validation and implement this solution to cut spam to zero.

Fortunately, implementing an anti-spam solution is one of the easiest IT projects to cost justify. Imagine the productivity savings each user will experience if their spam is cut to zero! Typically even a small company can recoup their investment of an anti-spam solution in as little as 2 months. For larger companies, the cost recovery is even faster. Thus, sender validation is a solution that sells itself. Before we jump into how to begin saving money through sender validation, let's briefly establish a foundation of spam history and terminology.

## A Brief Summary of Spam

Everyone knows what spam is; we've all received it. It's the automated mass email of advertisements and other annoying email messages. Just as important as defining what spam is, is to define what spam is not: a virus, identity theft, or instant messaging.

### *Why Does Spam Exist?*

Spam exists because it works. When compared with snail mail junk mail campaigns, spam has significantly lower costs. Consider the example that Table 1.1 shows.

<b>Traditional Mail Campaign</b>	<b>Spam Email Campaign</b>
Cost per piece \$1.37	Cost per piece \$0.001
Mail 10,000 pieces	Email 1,000,000
Total cost is \$13,700	Total cost is \$1000
Hit rate is 2 percent	Hit rate is .02 percent
Total hits of 200 at \$68.50 each	Total hits of 200 at \$5 each

**Table 1.1: Traditional mail vs. a spam email campaign.**

From this very simple example, you can see that spam campaigns typically cost much less per hit than a traditional direct mail piece. But because the hit rate is much lower (in this example 100 times lower) than a direct mail piece, spammers must send out significantly more pieces to achieve the desired number of hits. Thus the reason that spammers use the "shotgun" approach in their mail campaigns—the cost per piece is almost zero, so spammers can afford to send their message to any email address on which they can get their hands. They don't bother trying to target their lists for specific groups that might be interested in the product. Spam is all about volume; the more messages sent, the better chance of receiving a hit.

Although spammers closely guard their hit ratios, they are making money. However, they must annoy a significant amount of the population to get the desired number of hits—before I implemented a spam solution, I typically received 200 to 300 messages per day. Spam has grown to a point at which both end users and organizations are willing to invest in a solution to stop the spam and recoup valuable lost productivity.

## How Spammers Get Your Email Address

Spammers use *email harvesting* to continually get new email addresses. They use harvesting spiders/programs such as Atomic Harvester III, Email Marketing, and Text Bomber that monitor the Internet looking for new email addresses to gather. These programs are capable of gathering email addresses on specific Web sites, can target users in specific geographic areas, can target users in specific newsgroups and chat rooms, and can spoof IP addresses of bulk email servers.

 For more information about the capabilities of spam harvesting programs, check out <http://www.emilemail.com/>.

One of the more covert harvesting programs uses EMAIL\_ID, which will capture your address when you simply visit a site by tricking your browser into giving your name and email address. If the security level on Microsoft Internet Explorer (IE) is set to the default level, you should receive a warning message before this information is submitted.

Spammers might also attempt to guess your email address by using a dictionary/directory attack. This type of attack simply runs down a list of names and tries each one until it gets a hit. When a hit is determined, the spammer exploits the entire domain name by following the naming convention (for example, <first\_initial><last\_name>@<domain\_name>) for email addresses in the domain. Dictionary attacks are common on hotmail.com, msn.com, and other widely used email domains because of their mail volume and number of users. Spammers hit these sites continuously 24 hours a day, 7 days a week with dictionary attacks. When a hit is identified, the email address is recorded, and this list is sold to other spammers. These sites are continuously under attack, so you are almost guaranteed to receive spam if you set up a mail account here.

If your Internet connection slows suddenly you might be under a dictionary or Denial of Service (DoS) attack. Examine the log in your firewall to attempt to identify the source of the attack. If possible, use your firewall to block the IP address(es) from which the attack originates, and contact your ISP and ask them to block the IP address at the backbone to prevent further problems.

In a recent Federal Trade Commission (FTC) study, 86 percent of email addresses that were posted on Web pages, chat rooms, and message boards received spam. One email address received spam 9 minutes after a message was posted in a chat room!

 For more information, refer to <http://www.ftc.gov/bcp/online/pubs/alerts/spamalert.htm>.

Thus, never have a direct link from your Web page to a real person's email address. Use a generic email address such as *info@<domain\_name>*. Spammers tend to leave these generic addresses alone, and if they do receive spam, the address can be easily changed. Alternatively, your company can create a Web form (rather than use a generic email address).

 I've had mixed success submitting an opt-out request to spam mail. If the spam appears to come from a legitimate source, I've had better luck with the opt-out request. Be aware, however, that replying to a spam mail verifies to spammers that they've reached a real person. Use the opt-out feature at your own risk.

## Anti-Spam Legislation

As a result of the spam problem, 30 states within the United States have passed laws that make it illegal to send spam, but enforcing the laws inter-state and even within the same state (not to mention internationally) is difficult if not impossible. In June 2003, the Burns-Wyden bill passed. This bill legislates that spammers can face up to 1 year in prison and a maximum fine of 1 million dollars. Although anti-spam legislation will help, it probably will not solve the problem. Law enforcement has higher priorities within the IT industry such as catching virus creators and cyber terrorists. Thus, rather than wait for a legal remedy to this problem, the only effective solution is to use a spam blocking tool.

## Traditional Spam Blocking Methods

There are quite a few anti-spam software packages on the market, most of which use a combination of spam blocking methods to reduce the amount of spam in a user's mailbox. However, spammers are constantly developing new methods of bypassing spam filters. Thus, except for sender validation solutions (which don't necessitate ongoing updates), spam filtering solution vendors must develop additional methods of blocking spam to keep up with the spammers. Let's examine these methods and their advantages and disadvantages.

### ***Keyword Searching***

For keyword searching, the anti-spam software looks for specific words or phrases in an email. In you're in the market for this type of solution, look for a package that supports keyword phrases, keyword conditions, and keyword searching in either the subject or body of the email message. Keyword searching can reduce the amount of spam by performing a search for words that are likely to be included in the spam message (for example, Viagra, refinance, and mortgage). Phrase searching with conditions will give you more flexibility to search for items such as "need cash" and "refinance." This functionality provides a finer degree of control when searching for keywords and should help reduce false positives. Some spam filter vendors allow you to update your keyword searches based on the most current spam messages on the Internet.

## Advantages

If the message you receive has a consistent word or phrase, keyword searching is an effective method of blocking spam. It is very useful for blocking other unwanted messages that contain viruses, such as the Sobig.F worm, that use the several phrases as the following email message shows:

Re: Approved  
Re: Re: My details  
Re: Thank you!  
Re: That movie  
Re: Wicked screensaver  
Re: Your application  
Thank you!  
Your details

 Although anti-spam software can block unwanted messages that contain viruses, do not rely anti-spam software as your only virus email scanner. Purchase a virus scan option with the anti-spam package or install a dedicated email virus scanner on your email server.

## Disadvantages

Unfortunately, this method requires that you receive at least one email with a consistent keyword before you can block future messages with a keyword search. You must manually maintain the keyword list as new spam messages are received, unless the spam filtering vendor supplies updates for you. In addition, this method has the potential to consume considerable resources on the server—for example, if you perform searches on the message body versus the subject line or add keywords to the search list, more resources will be consumed on the server. On a heavily loaded server, some messages can get through the keyword search. Smart spammers randomize the words in the subject and message body in an attempt to bypass the keyword filter. Finally, keyword searches have the potential to cause many false positives depending on the type of mail your company receives.

## ORDB Checking

A mail server configured as an open relay allows spammers to bounce messages off the mail server to send the spammers' messages. Some packages can perform an ORDB check to determine whether a message was received from a mail server that is identified as an open relay (see Figure 1.1).

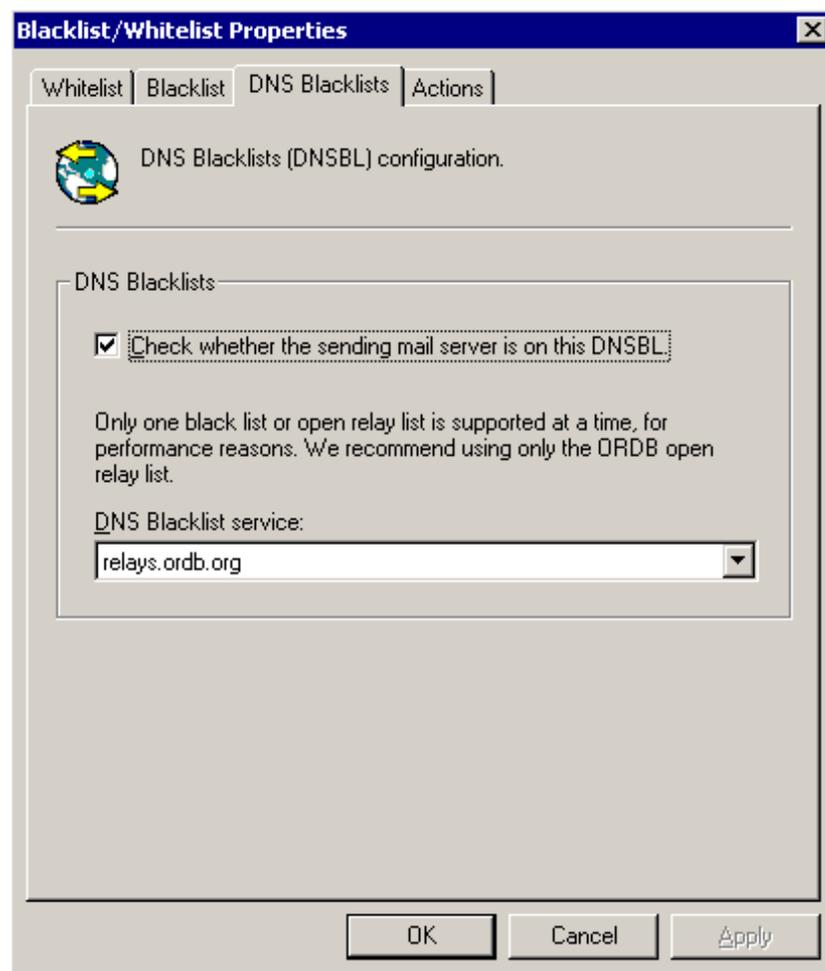


Figure 1.1: A DNS Blacklists screenshot from GFI Mail Essentials.

 If your server is an open relay, it is simply a matter of time before it is listed in one or more of these databases:

<http://abuse.easynet.nl/blackholes.html>

<http://www.delink.net/>

<http://dnsbl.njabl.org/>

<http://dsbl.org/main>

<http://ordb.org/>

For a comprehensive list, refer to <http://www.declude.com/junkmail/support/ip4r.htm> and <http://www.moensted.dk/spam/>.

### How to Determine Whether Your Server is an Open Relay

Many of these sites can test whether your server is an open relay. When you bring up a new mail server, it is a good idea to test it to verify that the server is not open. If your server is marked as an open relay, you must first close it, then submit it for retesting. If you are running an earlier mail package (for example, Microsoft Exchange 5.0, Novell GroupWise 5.2) that cannot be shut down as an open relay, take a look at the anti-spam features of your firewall. Some firewalls have anti-spam features built-in to their Simple Mail Transfer Protocol (SMTP) daemons that can help close the open relay. Another option is to upgrade your email software to a version that does not allow open mail relaying. An irony of the ORDB is that it provides a convenient list for spammers to relay their messages—the opposite of what the ORDB is trying to prevent. After the server has been retested and is no longer an open relay, it should be removed from the database. If the server is marked as an open relay, it might be listed in multiple databases. If such is the case, you must submit a removal request from each database on which the server is listed. The response time for a removal request varies depending on the database list.

Sometimes after the server is removed from an ORDB, the server might still have difficulty sending mail messages to one or more domains. If all else fails, you can change the external IP address and MX record of the server to bypass this problem. Typically, the ORDBs only list specific IP addresses rather than ranges of IP addresses. Thus, changing your mail server address is a simple workaround if your mail server is identified as an open relay (even though it is not open anymore).

A quick way to test this workaround is to change the outside address of your firewall, then try to send mail to the problem domain(s). If you are successful, issue an MX record change to the ISP that hosts your domain. If this workaround does not work, don't bother with the MX record change—the sending problem lies somewhere else, possibly with DNS, the mail server, or the message is infected with a virus. Be aware that you will temporarily take down your incoming mail while you run this test until you either update your MX record and it propagates throughout the Internet or change the firewall back to its original address. For this reason, it is a good idea to have extra IP addresses when ordering your DSL, T1, or broadband connection from your ISP even if you plan to use Network Address Translation (NAT) on the firewall. If you decide to use this approach, make absolutely sure that your relay is closed before changing the IP address; otherwise, you will end up on the ORDB again.

### Advantages

Checking whether an email message came from a server marked as an open relay can block as much as 50 percent of your spam. Another benefit is that once this method has been configured, there is no on-going maintenance.

### Disadvantages

Checking an ORDB consumes bandwidth because a lookup must be performed for each received message. If you use this method, rely on one of the larger ORDBs such as ordb.org. Open relay checking can potentially generate false positives because the relay might already be closed. Unfortunately spammers are getting smarter in their relaying methods. In the past they would find an open relay and exploit it until it was marked as an open relay. Now they hop from server to server and relay a smaller number of messages. This process makes it very difficult to identify the mail server as an open relay. Going forward, this anti-spam method will become less effective.

Some mail filtering services maintain their own “real-time” open relay list that is continually updated. When a mail server appears to deliver spam, the relay is tested periodically to verify that it is still sending spam. Once the server stops sending spam, it is automatically taken off the open relay list. This approach was developed to catch the technically savvy spammer that hops from open relay to open relay to avoid detection.

### **Whitelists**

If a message is received from an email address or domain on a whitelist, the message is delivered to the user. If you’re shopping for this functionality, look for a package that can support entire domains with one whitelist entry such as `*@<whitelist_domain.com>` (instead of separately listing individual users in the whitelist). This feature is very useful for users who correspond with multiple users in another company regularly. Of course, you don’t want to open an entire domain—such as `*@aol.com`, `*@yahoo.com`, and `*@hotmail.com`—from which users will receive spam on a regular basis, but for other domains, this feature can save a lot of administrative time.

Typically, a whitelist entry overrides conflicting configurations. For example, if a message is received from a user that is on the whitelist, but the message originated from a server marked as an open relay, the message is allowed through. Some software packages can automatically add users to a whitelist when an internal user sends mail to that person. However, this feature can be undesirable, especially if a user decides to opt-out of a mailing list. By replying to the mailing list message, the opt-out address is automatically added to the whitelist. If you decide to turn on the auto-add whitelist feature, make sure your users do not reply to such opt-out email messages. Alternatively, IT staff can simply remove an unwanted address from the whitelist. Many solutions offer the choice of per-person or company-wide whitelists, which enable administrators to decide whether users’ auto-add feature will affect other users.

### **Advantages**

Preloading a whitelist of approved senders will reduce the number of false positives when implementing anti-spam software. For this reason, preloading this list is an integral part of any whitelist implementation. At least, give the whitelist system time to “learn” who your users send mail to before turning on the spam-blocking feature. If the whitelist overrides other filter values, you can use the whitelist and blacklist in combination to filter out spam. (I’ll discuss blacklists in the following section.) For example, you can block an entire domain, such as `*@hotmail.com`, in the blacklist, then selectively list email addresses in the hotmail.com domain for messages from senders you want to pass through the spam-filtering software.

### **Disadvantages**

If you do not implement the auto-add whitelist feature, this list must be maintained manually. Even if you preload a whitelist, expect to receive several false positives when implementing anti-spam software. If you’re running Microsoft Outlook, you can export all the email addresses in the contacts list for each user, consolidate the list, format it based on the spam-filtering software requirements, then import the list into the whitelist. The number of manually added whitelist entries should taper off after the package is up and running for a few weeks—especially if you enable an auto-add feature. Both the whitelists and blacklists are responsible for the majority of the ongoing maintenance for anti-spam packages that use these methods of blocking spam.

## **Blacklists**

Blacklists work just the opposite of whitelists—if a message is received from an email address or domain on a blacklist, the message is rejected by the server. Blacklists have the same drawbacks as keyword searching, because you usually have to receive a spam message before you can block it (unless, of course, you already blocked the entire domain).

### **Advantages**

If spam is consistently sent from a single email address, blacklists are an effective spam-fighting tool. However, more than 75 percent of spam is from a one-time use address, blacklists alone will help to protect against only a quarter of the spam. As I previously mentioned, you can use the blacklist and whitelist combination to block an entire domain, then only let selected messages through the spam filter. If you implement a server-based spam-filtering solution, you need to enter the blacklisted address only once on the server; after the address is blacklisted, all mail received from this address is automatically blocked at the server level.

### **Disadvantages**

The biggest disadvantage of blacklists is ongoing maintenance. As new messages appear, the administrator must add the sender's name to the blacklist. Most spammers use "throwaway" email address such as spam123@yahoo.com. Once the email service recognizes the sender as a spammer, the account is deactivated. However, because many spammers don't even bother acquiring an email account in the first place—a recent study showed that more than 76 percent of spam is from nonexistent accounts—deactivation of a spammer's account is of little consequence. Maintaining a list of all these one-time use accounts causes exhaustive and excessive blacklist checking by the server—particularly considering that most spammer addresses are only used once. Thus, with blacklists, you're always one step behind the spammer, so a subscription to a blacklisting company is required to make this method an effective spam-fighting tool.

### **MX Record Lookup**

MX record or a reverse DNS record lookup performs a DNS query on the sender's domain. If the sender's domain matches the MX record IP address of the server, the mail is accepted. If the IP address does not match, the message is rejected.

### **Advantages**

This approach can work well if a sender's domain name has been spoofed by a spammer. In such a case, the server would know that the message is not coming from the legitimate contact.

## Disadvantages

This approach has the potential to create many false positives. The following scenarios can cause a false positive:

- Incorrect or missing reverse DNS record—Many companies do not bother to have a reverse record created when establishing the MX record for their mail server.
- Multiple mail servers—Larger companies or ISPs can have multiple mail servers for their domains. When the server performs a reverse lookup, the server might not get all mail server IP addresses for the domain, which can cause a false positive because the IP address of the sender's server might not match the IP address of the reverse lookup.

For these reasons, I suggest using other methods for spam blocking.

## *Heuristics and Bayesian Filtering*

Heuristics and Bayesian filtering is one of the more recent methods developed to block spam. The software gathers statistics about the type of message received, then makes a judgment call about whether the message is spam. To make this determination, some software packages use a point scoring system and others use custom algorithms. This method can be a very effective weapon against spam.

Heuristics and Bayesian filtering works like a blackjack player who is counting cards. A card counter knows that the deck is in his or her favor when a series of low cards appears because this means that the deck is “ten rich,” increasing the probability that the dealer will bust if the dealer must draw a card. Heuristics and Bayesian filtering similarly looks at words in email messages that are already marked as spam, then compares how often key words appear in an incoming email to estimate the probability that the message is spam. Generally, more recent data is more heavily weighted and email keywords are continually updated with new and current information. This system gives heuristics and Bayesian filtering the advantage of becoming somewhat self-maintaining.

If you're considering an heuristics and Bayesian filtering solution, consider a filter that looks at outgoing email to reduce the amount of false positives. For example, if you work for a refinance company and the word mortgage appears quite frequently in your outgoing emails, you want to ensure that messages that contain mortgage aren't blocked. In this particular case, the word mortgage will not have such a heavy weight for incoming mail because it occurs quite frequently in the company's outgoing mail. This analysis of outgoing email will reduce the amount of false positives.

Because heuristics and Bayesian filtering typically takes the whole message into account, it can usually catch misspelled words such as s\*e\*x or v-i-a-g-r-a. In fact, these misspelled words almost guarantee that the message is spam because a legitimate email will most likely never spell words in this manner.

## Advantages

The biggest selling point for heuristics and Bayesian filtering is that this solution is very low maintenance. Heuristics and Bayesian filtering constantly gathers information about incoming mail and updates statistics on an ongoing basis. Because it typically only looks at mail sent and received by the company, the statistics are custom-tailored for the company's email. Usually these statistics are more heavily weighted on the most recent data. Some companies claim they can block out as much as 99 percent of spam with a very low percentage of false positives by using heuristics and Bayesian filtering.

## Disadvantages

Heuristics and Bayesian filtering is only as good as the engine/algorithm making the spam judgment call. Typically, the entire message is evaluated, which results in an additional load on the email server assuming the heuristics and Bayesian filtering engine is installed on the same machine as the mail server. On a heavily loaded server, this spam-blocking method can cause performance issues.

In addition, after the heuristics and Bayesian filtering analysis, each message is typically assigned a probability ranging from 0 to 100 percent that the message is spam. This probability must be fine-tuned over time. Set the threshold too high, and too much spam gets through. Set the threshold too low, and you generate many false positives. Refer to the software documentation for a recommended initial setting, then fine-tune this setting based on your company's requirements. Because every company's email is different, you must use trial and error to determine the best setting for your company. Also, because heuristics and Bayesian filtering has the potential for generating false positives, look for a package that also supports a whitelist or some other method of receiving a legitimate message that was incorrectly marked as spam.

## Sender Validation

At a basic level, sender validation works by letting mail through if the sender is on an approved list and rejecting the mail if the sender is not on the list. Think of sender validation as an "intelligent whitelist." Once a sender is placed on the approved list, the mail server will accept mail from this address. The concept is simple, but it is the management of the approved list and a smooth validation process that are keys to a successful sender validation anti-spam solution. Most corporate sender validation packages work like the flowchart that Figure 1.2 shows.

## Sender Validation Process

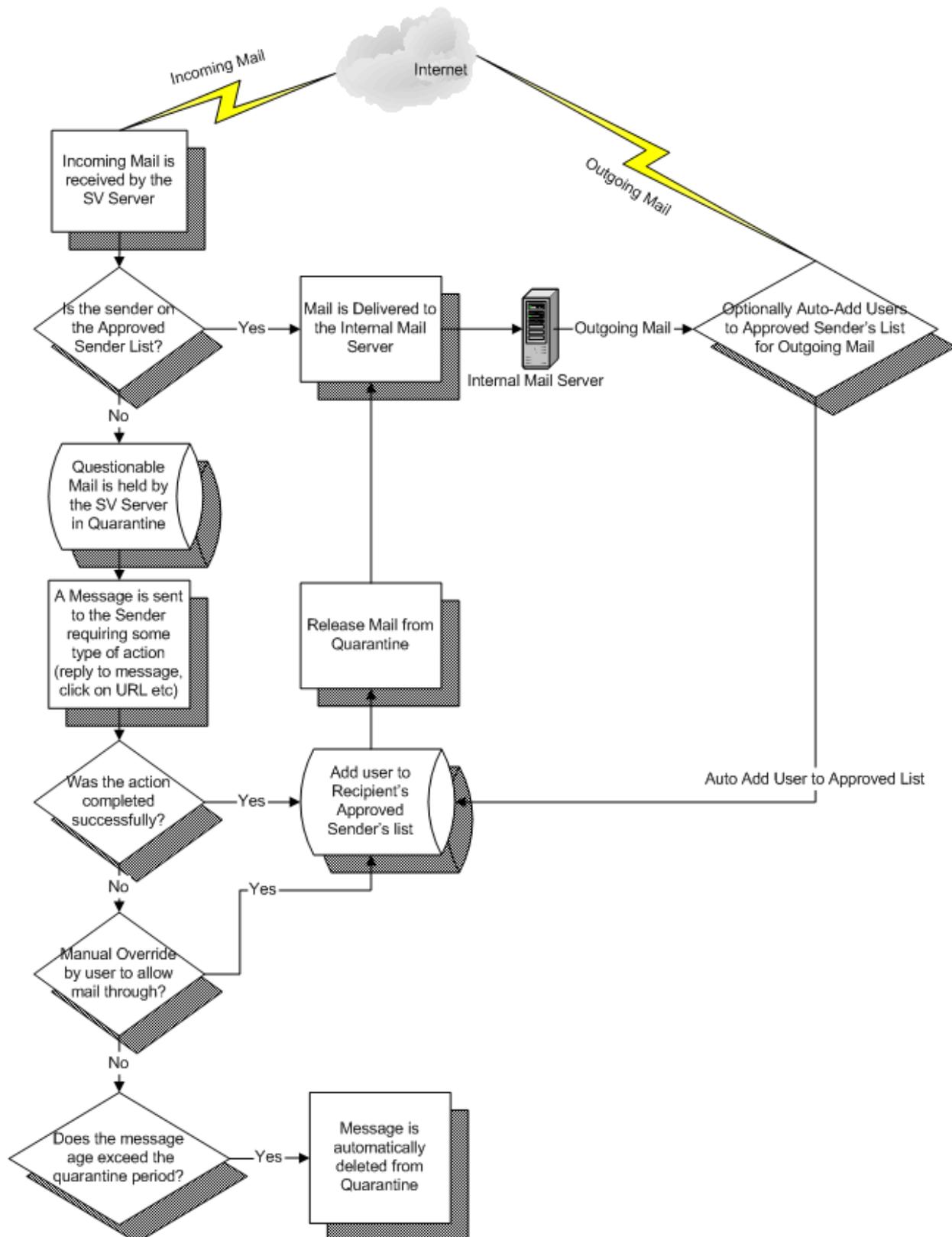


Figure 1.2: A flowchart that illustrates the sender validation process.

### Sender Validation Solutions of the Past

Historically, desktop/POP3 versions of sender validation have had the most success in the spam world. They are the easiest to implement and evaluate for the individual user; however, they are generally impractical within a business environment. The advantage of such solutions is that, compared with an internal enterprise solution, the software development cycle for a desktop sender validation solution is relatively short. It is fairly easy to evaluate and install because the sender validation server and software are usually located off-site.

However, some sender validation POP3 solutions are not online all the time, which can cause mail delivery problems. Depending on the solution, the validation process can be cumbersome and take a long time (days) to complete. It is possible to encounter a deadlock situation when both the sender and receiver have a sender validation spam protection for their mail. Sometimes the sender validation solution will not send an NDR to a legitimate sender, so the sender assumes their mail was delivered, but it wasn't. Make sure you can manually add senders to the approved list to avoid a deadlock situation. Some of the earlier attempts at sender validation were built by amateurs and lack the stability and features of a proper corporate sender validation solution.

Early sender validation enterprise solutions (circa 1995) lacked the stability and functionality that corporate users required. The advantage of such solutions is that the sender validation server is internal, so it is always online. However, many of the early attempts at sender validation were immature products, which resulted in a bad reputation for sender validation. These immature products were costly to evaluate because they typically required a dedicated internal server and complete implementation of the package just to evaluate it (a significant investment in both hardware and time for IT staff to purchase, install, configure, and sometimes develop the software for the sender validation server). With some early sender validation solutions, it was possible to encounter a deadlock situation between companies. Like the POP3 solution, you sometimes had the problem of lost NDRs, so a legitimate sender assumed you received their message when you actually had not. Some of the earlier sender validation solutions were developed in-house by internal IT departments or individuals, which had mixed success rates. Some of these sender validation systems have matured and evolved into today's systems.

If you're considering a sender validation solution, look for the following features to ensure a successful implementation:

- Auto-learn outbound communication—Look for a sender validation package that monitors outbound communication. Ideally the package provides the option to auto-add a user to the approved list when an outgoing mail is sent to the user. This feature can dramatically reduce false positives. However, even with the auto-learning feature, expect false positives with any “first-contact” message sent through an sender validation solution. Typically, this happens only the first time a user purchases an item online and the new vendor sends a receipt or other notification of the purchase. If the vendor's address is not in the approved sender's list, the user will receive a false positive. The false positive typically occurs because the e-commerce vendor does not go through the validation process, and the user has not done prior business with that vendor. Once the first interaction has been manually approved by the user, this one-time false-positive situation does not recur.

- Pre-load approved list—The sender validation software should allow an administrator to upload a predefined list of approved senders prior to package implementation. Doing so will reduce the number of false positives and messages in quarantine when the package is first implemented. This reduction in false positives makes this feature mandatory for any successful sender validation implementation. Alternatively, look for a solution that provides an auto-learn feature, which could be enabled for a period of time (a few weeks to a month) with spam protection off. As I mentioned earlier, such an opportunity would allow the solution to “learn” who the company communicates with before activating the protection (although loading a customer and vendor contact list provides for a much faster deployment).
- Removal of deadlocks—Ensure that the sender validation solution offers some method for removing deadlocks (for example, through a user override of a sender’s address). A deadlock can occur if two users within companies that are using sender validation software send each other a message. Both systems send and wait for the other system to respond to the mail, creating a deadlock situation. Make sure that the deadlock removal process has been thoroughly tested in a corporate mail environment.
- Approved list management—Users should have the option to manually add and delete items from the approved list.
- One time validation—Once a sender is on the approved list, they should not have to validate again.
- Approved list—Any sender on the approved list should have their message delivered.
- Flexible quarantine—To reduce maintenance costs, the package should have the option to auto-delete a message in quarantine after a user-defined period of time if the sender does not validate. This feature will also reduce the storage requirements on the sender validation server. Messages that receive a non-delivery report (NDR) during the validation process should automatically be deleted from quarantine.
- Flexible validation—The validation process varies from package to package. Some sender validation packages require the sender to click an HTTP link, require a reply with certain text in the message, or require the user to enter a pass phrase to get on the approved senders list. Regardless of the method, look for a package that has flexibility in its validation process. Make sure that the validation process is compatible with any mail client and mail server. The validation process should make it extremely difficult for an automated mail system to complete the validation process. The user should be notified that certain mails are pending validation and have the option to override the sender validation filter. The entire validation process should be easy to use and understand by the sender to reduce the sender’s confusion and the number of false positives.
- Backup flexibility—The sender validation approved list is a key component to this spam-blocking method, so ensure that the sender validation package configuration can be easily backed up on a regular basis to allow for a graceful recovery in the event of a hardware failure. For larger implementations, look for flexibility in the database engine (such as the ability to use SQL Server or Oracle as the database back end). This feature is especially important for implementations in which the sender validation server will handle a very large number of users (20,000+ users).

- Load balancing/fault tolerance—Very large companies and those for which email is a mission-critical application should look for a package that supports load balancing and/or fault tolerance/failover between multiple servers. Even without this feature, make sure that it is easy to bypass the sender validation filtering server (typically an IP address change on the firewall) in case of a complete hardware failure on the server.

 Fault tolerance is a concern with any anti-spam solution. Using a secondary MX record to the ISP's backup-relay server is an excellent solution for fault tolerance.

 Vendors such as SpamLion and MailFrontier provide sender validation packages that offer all these necessary features.

## Advantages

There are quite a few advantages of sender validation, especially when compared with other spam-filtering methods:

- More effective—Some sender validation implementations experience 100 percent reduction in their automated spam. Even if you don't achieve 100 percent reduction, sender validation will be significantly more effective than other spam-filtering methods.
- False positives—Sender validation packages that have an auto-learn feature will result in a lowered false positive rate compared with other methods. In addition, once a sender is on the approved list, a false positive will not be repeated. This functionality places a light load on the server for approved senders because the server simply performs a lookup on the sender's email address rather than a battery of tests to determine whether the message is spam.
- Low maintenance—Once sender validation is implemented, there is a lower maintenance rate than with other filtering solutions (particularly compared with a whitelist and blacklist implementation). Maintenance is lowered even further if a sender list is preloaded as part of the implementation process and the approved sender's list remains relatively constant.
- Easy deployment—Some sender validation solutions eliminate the need to deploy additional software at the desktop level; thus, if there is no deployment on the desktop, there is no ongoing software maintenance necessary at the desktop level. This feature makes future upgrades easier, because you only need to upgrade the server.
- Available as a service—Some sender validation corporate solutions are available as a service rather than an internal dedicated server solution, giving you the flexibility to implement the solution as a service or on a dedicated internal server.
- No compatibility issues—Most of the sender validation solutions are compatible with all types of mail servers. Their validation process has been well tested for a variety of users and corporate environments.

- User flexibility—Most sender validation solutions can be enabled for all or only a portion of users. Thus, these solutions can coexist with non-protected people or other spam-filtering solutions.
- Security and serviceability—Internal sender validation solutions do not expose the internal mail server to the Internet. This feature allows the IT staff to patch the sender validation server (the “external server”) without taking down the internal mail server.
- Client independent—Most sender validation solutions work at the server level, so the end user gets the benefit of sender validation regardless of the client (MAPI, wireless, Web-based, POP3) used to access the mail server.

## Disadvantages

To truly benefit from sender validation, you need to be aware of the disadvantages of this spam-blocking method:

- Significant initial cost—The sender validation solutions that have the most flexibility and features typically require a dedicated server, while other spam-filtering solutions can be installed on an existing mail server. Dedicated server sender validation solutions require more setup time because the OS must be installed on the dedicated server. Some sender validation vendors are in the process of developing solutions that install on an existing mail server. Depending on the current load of your internal mail server, you might want to implement a dedicated server solution anyway. With increased mail traffic, storage requirements, virus scanning, instant messaging, and advanced groupware features, your server might already be severely taxed. On a cost-per-user basis, sender validation is less expensive for larger companies because the cost of an internal server is spread out over a larger number of users. For smaller companies, a service-based sender validation solution might be less expensive than a dedicated internal server solution. Although sender validation can have a higher initial cost for larger companies, the total cost of ownership (TCO) over the first year is lower than other solutions because of the lower ongoing maintenance costs and greater overall effectiveness.
- One-time false positives—Many e-commerce vendors do not respond to customer inquiries, which can lead to a false positive of the sender validation sender when conducting an initial purchase. An easy workaround for this issue is to train end users to add the e-commerce vendor’s email address to their approved list when they first purchase an item from a new vendor. Of course, these vendors only need to be validated the first time, so this disadvantage is only an issue for purchases from new vendors.
- Legitimate senders don’t validate—Some email users are not familiar with the sender validation process and therefore do not validate, causing their messages to bounce back. If the mail is legitimate, usually the skeptical sender will call the recipient to see why the message bounced. The recipient can then ensure the legitimacy of the validation process or simply add the senders name to the approved list. Because of this issue, the validation process should only be necessary once, be simple, and easy to understand even for the novice user.

- End-user training—Typically, sender validation solutions have a Web-based interface to manage messages in quarantine and the approved list. Most of these interfaces are easy to use; however, some resources should be budgeted for end-user training. This training can be a short seminar or a simple instruction manual about how to use the sender validation software. End users are usually open to learning about how to manage their message quarantine and approved list rather than having to deal with an overwhelming amount of spam.
- Spammers can validate—Although highly impractical and unlikely for the spammer, it is theoretically possible for them to go through the process of validation and get on a user's approved list. If such should occur, users should have the ability to remove an address from their approved lists.
- Dedicated server—If the sender validation implementation requires a dedicated internal server, this server is one more resource that must be maintained by the IT department. It must be backed up on a regular basis and receive the same care as any other server on the network.
- Constant new mail senders—If your company constantly receives mail from different users rather than repeat customers, sender validation is probably not the correct solution for your company. In such cases, ongoing maintenance will probably be higher with a sender validation solution than with other spam-filtering methods.

## Summary

The spam problem becomes more of an issue everyday—spam exists because spammers make money doing it. The cost per piece of spam is dramatically lower than a traditional direct mail campaign. However, as we explored in this chapter, there are many spam-blocking methods available and being developed to combat this growing problem.

Each of these spam-fighting strategies has advantages and disadvantages. Often a combination of these strategies can provide a satisfactory solution for blocking spam. Among all of these methods, only one can potentially eliminate 100 percent of spam—sender validation. Although sender validation got a bad rap in the 1990s as a result of homegrown systems that lacked features and functionality and had poorly designed user interfaces, the current crop of sender validation solutions are ready for the corporate environment and have been fully tested and refined.

Spam robs a tremendous amount of time and resources from end users, IT staff, mail servers, WAN links, and storage requirements. The good news is that almost any solution will save your company money. Obviously, you want the best solution, and sender validation is a good fit for most companies. Once you've decided that sender validation is the right technology for your company, you must evaluate the advantages and disadvantages of each sender validation package. In the next chapter, we'll take a look at finding the right sender validation package for your company as well as how to evaluate the package you choose to ensure that it is the best solution for your environment.