Antispam Security Best Practices

First, the bad news. In the war between spammers and legitimate mail users, spammers are winning, and will continue to do so for the foreseeable future. The cost for spammers to send an unsolicited campaign to a harvested email database is next to nothing, particularly when compared to other forms of advertisements. A recent Forbes study indicates a 30 second prime time television advertisement prices at 1.8 cents per viewer; a full page color magazine ad about 0.9 cents per reader. A highway billboard costs 0.21 cents per car. Direct mail is the most expensive, at over 50 cents per third-class letter mailed. Spam can be sent at one one-hundredth of a cent per recipient. Simply put, spam is a low cost, sensible advertising medium, and since recent studies suggest that over 90% of all email is spam, spammers tend to agree…

If spammers are profiting, who's losing? Legitimate mail users who only want to read messages they intend to receive. Costs in combating spam include: antispam security software, service or hardware along with updates and maintenance, bandwidth consumed by receiving, processing and filtering spam and lost productivity when spam is received.

This document is intended to provide an overview of the most popular and effective options currently available to combat spam, making a final recommendation on how organizations can most effectively protect themselves from spam.

Antispam Scanning – Fight the Battle Here or Abroad?

First, a run down of the path a spam message takes to reach your inbox.

1. The message is sent out through an SMTP[1] mail server
2. The message is received by the SMTP mail server that hosts your mail
3. The message is retrieved by an email client, typically using IMAP[2] or POP3[3]

The first decision organizations must make is at what point in the mail delivery process should antispam filtering occur? Options include: at the receiving mail server, at the email client, or introducing an additional gateway layer in the mail delivery process intended solely for filtering.

- **Email Client**

  Some type of user interaction with your antispam system is beneficial, as it allows your organization to address your particular requirements. For example, allowing users to add or remove email addresses to white or black lists is a helpful practice, as well as allowing users to inform your antispam system if a message was not flagged as spam when it should have been, or vice versa. It is also recommended to allow users to view messages sent to their address that were marked as spam in their email client software or web mail services, as it allows users to locate messages that may have been inadvertently flagged as spam. While these examples of user interaction make good sense, relying solely on email client

antispam scanning is not recommended because spam messages are processed by the mail server and consume additional bandwidth (antispam processing and auto updates take place on each individual machine as opposed to one central source), it doesn't allow for sufficient or easy access to administrative control and overview, and it often puts too much responsibility in the hands of each user.

- **Email Server**

  Antispam processing at the mail server is most often recommended because it provides one central source for scanning where all messages are received. It allows administrators to control and oversee antispam processing, adjust configurations and thresholds when necessary to improve scanning or correct errors, and most mail servers with built in antispam features can reject spam before they are received on the mail server, preventing the possibility of spam reaching the user and consuming additional bandwidth. Many mail servers also include advanced antispam security that can catch most spam with minimal maintenance. The greatest concern with mail server scanning is the strain it can put on your mail server, which can affect mail server performance and stability. If your mail server software is running on a server with adequate system specifications and your email traffic is light to normal, most mail server antispam systems will not greatly affect performance. If your email traffic is heavy it's possible that server side scanning could impair mail server performance, in which case load balancing is often recommended: running multiple mail servers to distribute the work load and prevent a single server from bogging down. One option in heavy traffic environments is to host one server that is solely responsible for SMTP traffic, along with antispam, and possibly, antivirus scanning, then run a second server where your user accounts are hosted. Regardless of the amount of traffic you receive, scanning at the mail server is recommended.

- **Hosted ASP Service**

  Hosted ASP[4] services solve many problems, such as concerns over antispam processing bandwidth and hardware/software requirements, but they also raise different concerns. Hosted antispam services act as a gateway between the delivering mail server and the receiving mail server, filtering messages before they reach your mail server. This type of service makes good sense for organizations with limited IT resources or expertise and either have a serious problem with spam or receive a large amount of messages. While hosted services can remove a large percentage of spam before they reach your mail server, they also introduce a new point of failure, often reduce supervision and control over antispam filtering, and route your mail stream through a 3rd party, a practice many businesses are leery of. Price can also be a concern, since most hosted services are more costly than alternative solutions.

Technologies – The Good, the Bad and the Bayesian

It's best to begin this section with a disclaimer: no one antispam technology filters all spam, and there is no one technology in development that will be an all-encompassing solution. Does that mean mail administrators should wave a white flag in defeat to the spammers of the world? Certainly not! Antispam technology has evolved greatly over the past decade to combat similar advances by spammers, and with the appropriate technologies enabled spam can be a minor inconvenience instead of a major issue. Here is a run down of some of the most effective antispam technologies currently available, along with their strengths and weaknesses.

Black and White Lists

Ensure that email received from your approved contacts bypass antispam filters, and mail sent from known spammers or contacts you don't want to receive email from are denied. This technology is fundamental for any antispam security system, as it makes sure you receive email from people you know and want to communicate with, and rejects messages from contacts who have spammed you before. The only problems with this technology is it's only as effective as your black and white list maintenance and it does not protect against spammers who use domains fraudulently. If you regularly or automatically update these lists with accurate contacts it is highly effective, and generally can reduce antispam processing bandwidth significantly by automatically approving or rejecting messages based on the address.

DNSBL's and SURBL's

Reject messages based on known spammer domains or IP addresses either in the body of the message or from the message sender. Domains or IP addresses that have been used for spam are added to DNSBL[5] and SURBL[6] lists. Messages containing these domains or sent from these IP addresses are then detected as spam. This technology can be quite effective in rejecting messages from known spammers, eliminating the need for more resource intensive message scanning. Some of the more popular and effective list providers are: Spam Haus, Outblaze and Spam Cop. The only potential risk with this technology is that if you're using a list that is either too aggressive or not aggressive enough in identifying spam sources, you might have false positives or false negatives. The aforementioned lists, however, are widely used and generally considered reliable.

Greylisting

This technology rejects the initial message receipt request from an unknown address, and then accepts the following retry. Greylisting is effective because legitimate mail servers are configured to automatically retry messages that fail in the first attempt, per RFC[7] compliance regulations, while spammers most often will not resend messages when they are rejected. Greylisting significantly reduces bulk spam, and beyond the extra bandwidth required for retries (which is minimal), there is little downside to using it.

Heuristic Analysis

One of the most effective antispam technologies, Heuristic analysis (often associated with the SpamAssassin open source solution) scans messages for thousands of traits commonly found in spam messages. Each spam trait has a unique score or weight associated with it. Each message is then assigned a total score based on the total amount of spam traits and their associated scores. Mail server administrators are able to assign a threshold score to determine what score constitutes spam (e.g., a lower threshold will result in more messages labeled as spam, and higher thresholds result in fewer messages labeled as spam). This technology is difficult for spammers to bypass due to the combination of thousands of spam rules and traits; spammers may avoid some rules and traits, but will likely be caught by others. It also allows administrators to customize their tolerance of spam by assigning a threshold score that filters a higher or lower percentage of spam. The only downside to this technology is the amount of bandwidth and processing it requires scanning each message. In most cases the performance hit is negligible, but in high traffic environments it is a consideration.

Bayesian Filtering

A technology that statically calculates probabilities, Bayesian considers both the good and bad tendencies of messages to reach a percentage likelihood of messages being spam. Bayesian uses databases of spam messages and legitimate messages and their tendencies to learn what constitutes spam and legitimate mail. This is a time-tested technology that has grown less effective over time, but still filters enough spam to make it useful. Bayesian filtering is only as effective as its databases of spam and non-spam messages, so the more accurate and up to date the database, the higher rate of performance. Many mail administrators and/or users train Bayesian filters locally, while some email servers provide predetermined databases and automatic updates to keep Bayesian databases as current as possible.

Intrusion Prevention / Tarpitting

Tarpitting detects when mail is unsuccessfully sent to unknown users. If the number of unsuccessful attempts exceeds a specified limit the sender’s ip address is added to a black list where mail is rejected from this IP address for a specified period of time. Spammers commonly send to a wide variety of invalid user accounts on a domain in successive attempts, in hopes of landing a valid address. This technology detects these attempts and prevents them from continuing their invalid deliveries. Tarpitting targets an exact type of spam behavior and effectively blocks it. There is no known downside to tarpitting.

Razor / Community

Razor, or Vipul’s Razor, is a community-based technology that allows knowledgeable users to report new types of spam and prevent Razor users from receiving the same messages. Similar to antivirus security products, Razor ensures that known spam is

rejected. When the Razor community of users detects a spam message a unique hash mark is assigned to the message. When Razor users receive messages with this same hash mark the message is rejected. This is a more recently developed technology that is expanding in use, and since it is user community based, its effectiveness is also increasing. While this is a recommended technology that detects a great amount of spam, it does suffer from the flaw of not being able to detect new spam messages not yet reported by Razor users. It also doesn't allow organizations to determine their definition of spam, as messages reported by the Razor community as spam might be a legitimate message to you.

Challenge / Response

A technology loved by some and hated by others, no one can question that it significantly reduces spam. Challenge / Response systems send a challenge message to a message sender that requires the sender to take an action - click a link, enter a password, reply with subject line intact, etc. Once the sender follows the instructions in the Challenge their message is approved, and many systems will then add their email address to a white list of approved senders. This technology is quite effective since most spam bots are not capable of or will never respond to a challenge message. The downside is that many email users consider this intermediary step to be a hassle, and many businesses fear they may lose legitimate messages due to their customers not following this extra step to authenticate. One common reason for users not responding to challenges is that challenge messages can be reported as spam by the message sender's antispam system, in which case the user never sees the Challenge or responds to it. Challenge/Response systems are most effective when used in a layered approach where only questionable messages that cannot be identified by other antispam technologies - that do not require sender interaction - receive a challenge as a last ditch effort to prove whether this message was sent by a spam bot or a live person.

Sender Identification

An antispam technology spawned from our parents telling us to never talk to strangers, the premise of this technology is messages are accepted from senders whose identity can be verified and penalized or rejected by those that can't be identified. There are many variations of this technology currently available, including: SPF[8] and Domain Keys[9]. Since the vast majority of spammers use a fraudulent domain or mail server to send spam, if the sender's identity can be authenticated against a legitimate source, most spam can be eradicated using this technology. The primary issue with this technology is that it requires the participation of organizations to prove that they are legitimate, and unless this inclusion is mandatory or absolutely necessary, the technology is not completely reliable. The other issue is that there is no one agreed upon technology to provide this one solution. This technology may prove to be more effective than any other, but those days are ahead of us.

Conclusion – When in Doubt, Layer

Despite the growing (and growing) amount of spam that dwarfs legitimate mail, with the appropriate technologies in place organizations can minimize the affect that spam has on their email communications. All organizations that host their own mail server are encouraged to scan messages at the mail server, as the benefits of centralized, effective scanning at the point of receipt far outweigh the risks. Client side interaction, consisting primarily of reporting false positives and negatives and maintaining a private or global black and white list, is also beneficial.

When considering the technologies to use to combat spam, it's best to remember that, while some technologies are more effective than others, all of the aforementioned technologies can contribute to an organizations' antispam strategy. While no one can predict the future of spam, the one thing we can be sure of is that new types of spam will continue to emerge, most designed to override existing technologies. By implementing a wide variety of layered technologies mentioned in this article, you can protect your organization from most types of spam and ensure that you are prepared for the evolution of spam.

Author: Don Hoyt
Deerfield.com Messaging Analyst
donh@deerfield.com
(989)-732-8856, ext. 273

Glossary

[1] SMTP (Simple Mail Transfer Protocol) - the standard for sending emails over the Internet, SMTP is a relatively simple email transfer protocol that transmits text messages on port 25.
[2] IMAP (Internet Message Access Protocol) – an application layer Internet protocol that allows email clients to access their inboxes on a remote server without downloading the messages onto their machines. This protocol allows for quick email client access and viewing because nothing is downloaded to the clients, the messages stay on the server.
[3] POP3 (Post Office Protocol version 3) – an application layer Internet protocol that allows email clients to download messages to their client from an email server.
[4] ASP (Application Service Provider) – an Application Service Provider is a business that hosts a computer based service for other businesses or individuals from a remote location.
[5] DNSBL (DNS Blacklist) – DNSBL's are lists of IP addresses reported as conducting an activity that others may wish to avoid.
[6] SURBL (SPAM URI Realtime Blocklists) – a technology used to locate domain names associated with spam within the body of an email message, and mark the message as spam.
[7] RFC (Request for Comments) – documentation regarding new Internet technologies and research that form a type of concept or idea.
[8] SPF (Sender Policy Framework) – an extension of the SMTP protocol that allows software to locate and reject messages sent using forged email from addresses. Owners of domain names can specify what machines are able to send email using their domain. Software that checks SPF records can then determine if the sending machine has authority to send email using a domain, and if not, flag the message as spam.
[9] Domain Keys – an email authentication system used to verify the domain name of the email sender and the integrity of the message. This technology creates a digital signature of new outbound messages, then the receiving mail server uses cryptography to verify if the message sender is who they say they are, and if the message was intercepted.