SECURITY

Spam Bashing: The 98 Percent Solution

Drew Robb

There's no reason to suffer from spam—Here's how to get rid of it.

There's been good news and bad news recently in the legal fight against spam, or unsolicited commercial email (UCE). On the upside, U.S. District Court Judge Charles R. Rolle entered judgments on Dec. 17, 2004 totaling more than $1 billion against three bulk emailers. The bad news is that the defendants never showed up in court and it is questionable whether any money will ever actually be collected from them.

"Although spam is the responsibility of relatively few people, successful legal actions are a drop in the ocean," said Gartner, Inc. research director Ant Allan. "The problem is that even domestic operators can move their operations off shore—the bulk of such email is now originating outside the Western countries."

Meanwhile, the ratio of spam to legitimate email continued to climb in 2004. Email security vendors MX Logic of Denver and FrontBridge Technologies of Marina Del Rey, CA, reported levels of 86.88 and 88 percent spam respectively, for their managed service clients, on December 31, 2004—one year after the enactment of the Controlling the Assault of Non-Solicited Pornography and Marketing (CAN-SPAM) Act.

Some customers notice even higher levels. "We get about 60,000 emails a week, more if we do a trade or tickets go on sale," said David Curry, director of information services for the Seattle Mariners baseball team. "On some of our accounts, 90 to 95 percent of the emails were spam, and it took a long time to go through and delete those emails."

This is not to say that all is lost. Far from it. Anti-spam software, appliances and services continue to improve their ability to identify unwanted messages. So even if the volume of spam continues to rise, less of it is hitting user in-boxes.

A survey published in 2004 by The Radicati Group of Palo Alto, CA showed that a 10,000-user organization could cut its annual spam-related productivity losses from $30 million to less than $5 million by deploying email filtering technology, in the process cutting spam's share of email traffic from 48 to 13 percent.

According to Gartner's Ant Allan, "The most effective controls are technology controls. Even the crudest can get rid of the majority of spam coming into the organization, and the best of them can do in the high 90 [percents]."

Types Of Filtering Techniques

It would be nice to write about a foolproof method to completely eliminate spam, but the only proven way to achieve that is by shutting down the email server. We want email, but not a flood of junk, so we must acknowledge the constant technological race between the anti-spam vendors and the spammers, who are always figuring out ways to circumvent the controls. It takes a multilayered approach and enough humility to recognize that some or all of these efforts may prove ineffective in the very near future, as the perpetrators develop yet another workaround.

While the exact mix of methodologies varies among vendors, most use an assortment of the following filtering techniques:

- **Anti-spoofing—**One spamming technique is to make email look like it is coming from the local domain so it gets past filtering software. Anti-spoofing detects and eliminates such emails.
- **Blacklist—**These are lists of Simple Mail

Anti-spam Groups

- Coalition Against Unsolicited Commercial Email (www.cauce.org)
- Anti-Spam Research Group (asrg.sp.am)
- Mail Abuse Prevention System Realtime Blackhole List (www.mailabuse.org)
- Open Relay Database (www.ordb.org)
- SPAMCOP (spamcop.net)
- SPAMHAUS (www.spamhaus.org)
- SPEWS (www.spews.org)

Drew Robb is a Los Angeles-based freelance writer specializing in technology and engineering. He can be reached at 323/660-4862.

Transfer Protocol (SMTP) or IP addresses from which all email will be blocked. These lists can be maintained locally, or a company can subscribe to real-time blacklists (RBLs) maintained by non-profit organizations. The sidebar "Anti-spam Groups" lists several such RBLs.

■ **Challenge/response system—**When a message is received from someone not on a "white list" of approved senders, an email is sent back to that person, requiring a response before the email goes through. One variety is called CAPTCHA—Completely Automated Public Turing Test to tell Computers and Humans Apart—which requires the sender to perform some task that automated software cannot do. Carnegie-Mellon University's CAPTCHA website (www.captcha.net) has several examples of such tests.

A Turing Test, first described by Alan Turing in 1950, is a method for determining if a computer can think like a human being. See the University of California at San Diego's Cognitive Science Department's Turing Test page for more information (cogsci.ucsd.edu/~asaygin/tt/ttest.html#new).

■ **Checksum—**A method of creating a signature for known spam. If other email comes in with the identical signature, it is blocked. (Unfortunately, senders can get around this by adding random words to e-mail, thereby changing the signature.) Vipul's Razor (razor.sourceforge.net) is an example of a checksum mechanism.

Spam attacks keep changing, so filters and other tools to fight them are constantly evolving

■ **Complex dictionary checking—**A detection method to identify when spammers use variations such as v1agra or v!ag_a in an attempt to get around filters.

■ **Header Analysis—**The header of a message should give the originator, message routing, priority level, etc. Spammers often alter the message header to make the source harder to trace. Analyzing the header is one way to identify bogus emails.

■ **Heuristic Analysis—**Heuristic methods are capable of learning what to do based on experience. In this case, it means a class of methods whereby the software analyzes known spam and then uses those characteristics to analyze incoming email. Bayesian analysis is one type of heuristic analysis. This filter compares incoming email against profiles developed from batches of known good and junk email to determine the probability of it being unwanted. Companies can set a threshold for whether to keep or toss the email.

■ **Keyword Analysis**—An early approach which looks for certain words (Vioxx, mortgage, ink jet cartridge; sexual terms) that frequently show up in bulk emails. Anti-spam software generally comes with pre-built keyword lists, but companies should review these. Merck & Co., for example, would not want to block emails about its drug Vioxx and Bank of America would not want to refuse emails from people inquiring about mortgage rates.

■ **Lexical Analysis—**A more sophisticated way of analyzing an email's text than just using keywords. This can include use of phrases, Boolean operators or other techniques to analyze the content of a message.

■ **Quantity Checking—**This looks for a large volume of email coming from a single address and flags it for the administrator's attention. Either it is the source of a legitimate form of bulk email, such as a newsletter, or it is junk email.

■ **rDNS—**Reverse DNS lookups match the DNS and IP addresses on incoming mail as a way to catch spam. If they don't match, the messages can be flagged or blocked.

■ **Sender policy framework—**SPF is a way to verify the message envelope before sending the body of the message.

■ **Whitelist—**A list of known good email sources whose email is not blocked by the system. Companies can generate their own whitelists, or they can use a reputation service such as Habeas, Inc.'s Sender Warranted Email, where a trusted third party certifies the email source as legitimate.

Since most vendors use most of these techniques, what matters is how effectively their particular mix blocks unwanted email, without blocking the specific types your company wants to let through. Gartner ranked Brightmail (now Symantec), Postini, CipherTrust and FrontBridge as the leaders in spam filtering as of March 2004, but both the nature of spam and the tools to fight it are constantly evolving.

Software, Appliance Or Service

In addition to using different mixes of analytical tools, vendors differ in the overall approach they take. Broadly speaking, customers can choose between server software, a security appliance or a managed service:

■ **Software—**Vendors taking the software route include Computer Associates, Symantec and McAfee. The Seattle Mariners use Computer Associates eTrust software, which, according to IS director David Curry, removes 95 percent of the unwanted mail. It took him two days to set up the software package on a spare server, with most of that time spent discussing requirements and con-

Deleting spam messages one by one is tedious, and saps worker productivity

figuring the filter.

"It has a ton of options," he said. "I am impressed with how granular the administrative tool is." As an added bonus, he went live the day before the MyDoom virus hit, and out of 20,000 infected emails, none got through the filter.

■ **Appliances—**Appliance vendors include Barracuda Networks and CipherTrust. Cox Communications uses six CipherTrust IronMail appliances to filter 40 million messages coming into its Atlanta headquarters every month. Senior messaging specialist Franklin Warlick said he wanted to go with an appliance to stop the email at the gateway. It took about half an hour to install the appliances, and another day to tweak the settings. He then spent the next month adjusting the whitelists from the company's 35,000 employees.

"One person's newsletter is another person's spam," he explained. He said that having the appliances in place has kept him from having to upgrade the company's Microsoft Exchange infrastructure. Although the total number of mes-

Risky Business: The Folly Of 90-Day Email Retention

Thomas Bookwalter

Companies that trash their email too quickly are asking for trouble.

The motivation for considering a short retention policy for emails may be based on the concern that there might be "something" in the emails that could turn out to be damaging to the company (or possibly incriminating in regulatory investigations or litigation). Or it may be driven by a desire to limit or control the size of individual mailboxes on the mail server.

Either way, the conclusion is that it's safer not to have the emails at all; the underlying notion is that because the records are in emails, the rules of records retention do not apply.

Such a conclusion is a grave mistake, however.

Email Is NOT A Record Type

Email is not a record type. E-mail is a delivery system like FedEx or the U.S. Postal Service. When it comes to retention rules, what matters is the content of the message, not the way in which it was delivered. Retention rules are based on the purpose, use or content of the message.

If a given record would be retained when originated in paper form, that record must also be retained when it originates in email—and it must be retained for the same period of time as a similar physical record would be kept. The courts have determined that the rules that govern the retention of the paper records also apply to emails.

This should not be considered unusual, since emails now are used to perform such official tasks as:

- Filing official documents with state and federal agencies.
- Sharing working papers on developing strategic plans and financial reports.
- Dealing with product and service problems.
- Gathering customer information.
- Negotiating, finalizing and agreeing on contracts.
- Addressing employee health care, pension and disciplinary issues.
- Receiving job applications and resumes and offering employment.
- Informing customers and prospects about new products.

The State of California Records Information Management Handbook says it most eloquently:

Retention or disposition of email messages must be related to the information they contain or the purpose they serve. The content, transactional information, and any attachments associated with the message are considered a record (if they meet the agency's record management plan criteria). The content of email messages may vary considerably, and therefore, this content must be evaluated to determine the length of time the message must be retained.

One of the difficulties with email is arbitrary size limits on email user "mail boxes" which require users to purge or archive files or be restricted in their use of the system until the mailboxes are kept below the size limit. This may contribute to improper deletion of emails that are records. Education of Information Technology professionals on the records implications and proper training of personnel can ensure [that] good records management procedures are followed. Use of an electronic recordskeeping system also helps to manage this increasing source of records.

NOTE: Simply backing up the email system onto tapes or other media or purging all messages after a set amount of time are not appropriate strategies for managing email.

Problems Created By Short Retention Strategies

As the California guidelines suggest, one intent of a short email retention policy is to manage mailbox sizes on mail servers. Periodically, employees reach the mailbox limits and are asked to "clean up" their mailboxes. Wanting

sages has mushroomed from 8 million to more than 40 million per month, there is no added load on the messaging infrastructure, as the CipherTrust appliances block 38 million of those messages. Every two to four months the company receives updated filtering software.

■ **Managed Service—**Going with a managed service frees up IT staff, and keeps the spam completely out of the company, although the services generally offer less customization than spam-blocking products. Reebok had been using internal spam-blocking software, but it was only 30 percent effective and some users were receiving more than 300 spam emails daily. In June 2003, Reebok switched to FrontBridge Technologies managed services for its 4,000 users in the U.S., Western Europe and Asia. It took less than an hour to go through the configuration screens, and MIS director Rod Baker said they are now blocking 90 to 95 percent of the spam. It also has saved them from having to add storage.

Weight loss firm Jenny Craig uses Postini's

Filtering out spam also avoids adding email server and storage capacity

to keep the emails for reference, employees move the mailboxes from the mail server to their laptops or desktops. But central systems do not back up these files and cannot search them.

The problems that result are twofold. First, if a company is in litigation, the Federal Rule of Civil Procedures (FRCP) obligates the company to provide the names and addresses of all people that have information relevant to the case, and to surrender all related documents (even before discovery requests). If the emails are on desktops and laptops, employee names should be provided and the relevant files on their desktops collected for submission in the case. Obviously, this will be difficult and time-consuming—if not impossible—to do completely.

Second, if the emails are deleted and no copies remain within the company, there are new risks, as several recent court decisions have made clear:

■ In a 1988 case, the courts ruled that litigants were required to retain documents that they knew or should have known would become material at some point in the future.

■ A 1995 decision made it clear that data in computerized form is discoverable even if paper hard copies of the information have been produced.

■ A third ruling establishes a clear obligation to preserve records, noting that, "Spoliation is the destruction or material alteration of evidence or failure to preserve property for another's use as evidence in pending or reasonably foreseeable litigation."

■ In a fourth case, in light of a clear precedent for the obligation to preserve records, a $1 million penalty was assessed because the company persistently ignored the requirement.

■ Lastly, Sarbanes-Oxley Sections 802 and 1102 introduce felony penalties for record tampering that may apply to email that is not retained properly.

If a company strives to run an honest business, short email retention strategies not only introduce new risk but also reduce the company's abilities to manage its employee conduct and protect itself if there are problems.

Failure to properly manage records, email included, also sends a signal to the investment community that something might be amiss.

The courts are consistent in their opinions regarding email retention:

■ E-mail is not a record type, it is a delivery and storage method.

■ The content and use of the email is what determines its retention requirements.

■ Deleting emails or letting them "disappear" is regarded by the courts as "spoliation," the destruction of records.

■ Under new laws it is a felony punishable by fines and imprisonment of up to 10 or 20 years to remove or let disappear records that are required to be retained.

E-mails Have A Life Of Their Own

One final note: There is a common misconception that just because emails have been removed from one mail server, they are gone. First, every email has a sender and at least one receiver, often more. Once people receive an email, they often send it to others, save it for personal reference or protection. Some send it off to a personal mailbox outside the company. Just because an email has been deleted from the corporate mail server does not ensure that it is gone. If that email is presented as evidence against a company, not only does the question arise of why the company has no copy of its own—but how does the company prove the email has not been altered from the original? (see *BCR*, January 2002, pp. 46–48).

A company may intend for a short retention strategy to protect the enterprise. The truth is that a short retention strategy exposes the enterprise to more risk, not less.

Short retention strategies for emails and other documents are short sighted. Such strategies are Risky Business□

Thomas Bookwalter is president of FMDC, a records management and compliance consulting company. He has been personally involved in the design and implementation of numerous compliant records management solutions. He can be reached at Thomas@fmdc.com or by phone at 908/812-5000.

service. Director of technology Jeff Nelson said that prior to using Postini, his staff had to spend about three hours a day handling spam, in addition to the time the end users spent. He also likes the extra level of fault tolerance that comes from using a service provider: If his Exchange server goes down, the emails stay on Postini's server until he is back up again, rather than being bounced back to the senders as undeliverable.

What Works, What Doesn't?

Given the variety of tools available to catch spam, you will probably wind up selecting a product that incorporates a variety of methods, rather than individually selecting the techniques to be used. According to Teney K. Takahashi, market analyst for The Radicati Group, "Many vendors tout the 'funnel method,' in which message traffic is gradually funneled down by several stages of filters.

"Each concentrates on a different type of spam, before messages are ultimately delivered to end users," he continued. "For businesses where spam is a major issue, this approach, while generally more costly, is superior to purchasing any one product." All the major enterprise spam-fighting vendors use this type of multiple filter approach.

Many vendors suggest a "funnel method" that subjects email to a series of filters

He explained that spammers have been fairly successful at working around basic content filters and that real-time blacklists block some spammers, but offenders frequently change their IP addresses to avoid being blocked. Heuristic filters provide a greater degree of accuracy.

"Reputational filters, which use a network of many users to create a reputation for email senders, are becoming increasingly popular and effective," Takahashi continued. "However these filters require a large base of users to accurately and actively flag spam—a practice that is not always successful."

Gartner's Ant Allan said that filtering approaches themselves are becoming a commodity item. Any product should be able to remove greater than 90 percent of the spam, but the manageability needs to be examined as well.

"One thing to consider how easy it is to set up different rules for different groups of users," he said. "For example, you may want to be more relaxed in the rules for filtering email to marketing people, since they may want to see the direct-mail techniques other companies are using."

One other piece of advice from Allan is to push the blocking as far away from the end user as possible, so it doesn't tie up the premises network. This can include blocking it at the ISP, using an email service provider, or doing the spam filtering at the gateway, before it reaches the email server. You don't want to just use filtering on the desktop.

"The real benefit to an organization is reclaiming its infrastructure," he said. "If you are letting it get through to your Exchange server, you are halving the capacity of your email infrastructure. Don't just focus on effectiveness, but also pay attention to the enterprise class features for working with large populations." He cited desirable features including ease of managing large user populations and the ability to establish different message rules for different user groups.

Conclusion

There is no simple solution to eliminating spam, but the fight is worthwhile. Reclaiming infrastructure availability is an easily measurable result. A less quantifiable benefit is the reclaiming, not only of employee time, but also of attention, concentration and peace of mind. Every piece of spam, even if it only takes a few seconds to scan the header and delete, distracts an employee from the work at hand and makes them think about something else.

Constant disruptions, even if brief, can kill productivity. We have doors, receptionists and security guards so salesmen can't just walk in and interrupt employees. We have caller ID, "no-call" lists, voice mail and message screening to reduce unwanted phone calls. Now it is time to do the same for email□

Companies Mentioned In This Article

Barracuda Networks (www.barracudanetworks.com)

CipherTrust (www.ciphertrust.com)

Computer Associates (www.ca.com)

FrontBridge Technologies (www.frontbridge.com)

Habeas, Inc. (www.habeas.com)

McAfee (www.mcafee.com)

Microsoft (www.microsoft.com)

MX Logic (www.mxlogic.com)

Postini (www.postini.com)

Symantec (www.symantec.com)