

# NATIONAL ASSOCIATION OF REALTORS®

April 2003

## Reducing Spam Email

- Background (Why do I get spam?)
- Prevention – options for the user
- Prevention – options for the company
- Qualities good anti-spam tools should have

### ***Background (Why do I get Spam?)***

#### ***What is spam?***

Spam is junk email that is sent to you by someone who has no prior or existing relationship to you.

Also called unsolicited commercial email (UCE), unsolicited bulk email (UBE), or junk mail, spam is defined by the fact that the recipients did not solicit the mail or divulge their email address for receiving such mail. Yet each day, thousands of spam programs scan web pages, newsgroups, and other online documents to harvest email addresses in bulk. In addition, there are programs that generate millions of emails using combinations of letters and numbers, such as JSmith, placed in front of the realtors.org portion of the address. Enough are generated that many match real e-mail accounts.

#### ***Where does the term spam come from?***

Most people call unsolicited email “spam” which entered the slang vocabulary following a Monty Python skit. Hormel Foods, the maker of SPAM luncheon meat, at first didn’t appreciate this use of its trademarked brand name and initially fought it vigorously. The company has since relented but still insists that “SPAM” in all caps is the luncheon meat whereas “spam” in lowercase is the nuisance email garbage.

#### ***Why is spam so prolific?***

Unlike direct mail or telemarketing, email marketing has very low marginal cost. As a result, despite extremely low response rates, spammers can make a profit fairly easily. The more emails a spammer can send, the greater his profit while the cost remains nearly constant.

#### ***What types of spam are there?***

Nuisance email messages come in many forms and originate from many different sources. Some are clearly fraudulent, while others just want to sell you something, convince you to visit a website, or promote an idea. Here are a few categories:

1. Confidence games, pornography and unethical senders
2. Chain letters, hoaxes and urban legends
3. Legitimate offers from legitimate senders
4. "Occupational spam" from your colleagues

If you try to unsubscribe from "the bad guys" or numbers 1 & 2, your request will only serve to validate your address, and you will receive more spam. Spammers pay a premium for validated addresses. If it is from one of "the good guys", or numbers 3 & 4, they should be practicing ethical permission-based marketing, where you can safely unsubscribe and the sender should politely stop sending messages.

### ***How much spam is there?***

Junk-mail filterer Postini, Inc., which processes about 40 million messages per day through its hosted service, reported in January that spam as a percent of overall e-mail volume grew by more than 150% in 2002 and that the average user's email volume is polluted with 60% spam. Brightmail, a similar service, estimates that 40% of email is spam.

Joyce Graff, a VP at the Gartner Group, says that the volume of spam is 10 times greater than last year (2002) and 16 times greater than two years ago (2001).

MessageLabs, a managed e-mail security service provider, says that managers who receive 50 or more e-mails a day spend at least 10 minutes each hour dealing with spam. They go on to say that nearly 20% of all mail server crashes are due to an influx of spam.

San Francisco-based market research company Ferris Research has estimated that unwanted commercial email cost U.S. corporations \$8.9 billion in 2002.

Ferris computed the cost of spam by calculating its costly effects in three areas: loss of worker productivity; consumption of bandwidth and other technical resources; and use of technical support time. The researcher found that lost productivity accounted for 40% of the drain.

## ***Prevention – options for the user***

### ***How do I reduce the spam I get?***

The key to reducing spam is to keep your email away from the spammers. The programs and scripts that generate spam feed on addresses that are harvested from many sources. The easiest source is the web itself. Spiders and software robots scour the Web constantly and can easily identify an address by the required @ sign. The first step in reducing spam is to hide your address from these spiders. Do not post your address on

any Web site. If you must post it on a site, use the “human-readable” equivalent. For instance, “JDoe at realtors dot com”.

You may not always be aware when your address gets posted to a site. Many mailing lists are archived to the Web. Meetings and conferences often have Web pages that list presenters and attendees- and usually their email addresses.

Another source of nuisance email uses information that you may have given out while conducting an online transaction. Anytime you purchase or register a product, you’ll likely be asked to provide several pieces of information. Some of that is essential to the transaction. You obviously can’t have an item shipped without providing your physical address. Nor can you receive any necessary follow-up information without including your email address.

Online registration systems generally go far beyond asking for such essential information by requesting data related to income, interests, and other demographics. Such details can then be used to classify you for future marketing efforts, either by the company to which you provided the information or to other companies or organizations that may obtain it. The more information you supply, the wider it will be disseminated among marketing types. Only provide the information that is essential and required for the transaction at hand. Don’t fill in the optional demographic fields unless they are required.

Keeping multiple email accounts – one public, another private – might help you deal with spam. You would have one email address that you never publish that you share only with your co-workers, friends and family. This would be used for business purposes and for that which requires a timely response. Your public account is the one you publish openly.

If you do get spam, do not reply or try to remove yourself from the email list. Doing so can validate your address actually causing an influx of more spam as it can be sold for a premium at that point.

Lotus Notes comes with filters called rules to detect spam and delete it, or at least move it to a folder where you can peruse it at your leisure. You can create rules that look for a domain (Greatdeals.com, for instance) or for certain words that wouldn’t be used in common business conversation.

Using this approach catches less and less as time goes on. It also can’t be used to catch phrases that could be construed as inappropriate or phonetically spelled words.

## ***Prevention – options for the company***

### ***How can corporations stop spammers?***

- “Blacklists” compile and distribute IP addresses and domain names of known spammers.
- “Whitelists” can be built by companies to identify legitimate senders.
- Content-analysis tools look for keywords.

- Behavioral-analysis tools look for patterns such as large numbers of recipients or blind copies.
- Address-validation tools do reverse Domain Name System lookups to ensure the sender isn't trying to cloak his identity.
- Header Analysis checks for consistency between the sender and the "from" fields as well as tactics used by known spammers.
- Digital fingerprints developed with algorithms and heuristics identify and block or filter common spam patterns.
- New products can scan for graphics such as skin tones to combat pornography – these tools are very new and still being perfected.
- Probe-networks consist of dummy accounts set up through various ISPs and corporate clients to attract spammers. These companies monitor these networks to detect new tricks and continually evolve. New rules are distributed and updated continually.

## ***Qualities good anti-spam tools should have***

### ***Boundary Defense***

Good spam defense begins with a good message transfer agent (MTA) that understands boundary defense tactics. It's like hiring a security guard. A good "security guard" MTA should have the ability to:

- Identify and abort a denial-of-service attack (unauthorized use of a domain to send out spam in such a large quantity that it shuts down the network.)
- Identify and abort a dictionary attack or address-harvesting attack (computer-generated permutations of possible email attacks, such as "asmith@realtors.org" and "bsmith@realtors.org." The software then records which addresses are "live" and adds the addresses to the spammers list. These lists are typically resold to many other spammers. )
- "Non-accept" a message (simply decline to accept it, rather than receiving it at all.)

### ***Header Analysis***

A good anti-spam agent will first analyze the header, looking for characteristics typical of spam messages. If the header alone gives sufficient evidence that the message is spam, the message can be "non-accepted" even before the body is taken in.

The header analysis should look for things such as:

- Validity of the sender (using reverse DNS lookup)
- Consistency between the sender and the from fields

- Tactics used by known spammers that are highly unlikely to be found in normal messages

### ***Content Analysis***

Each product has a slightly different approach to content analysis. The technique should be evaluated against the goals of the organization.

Content analysis includes one or more of the following capabilities:

- A set of rules to search for known spammer tactics
- A set of rules to search for known chain letters, hoaxes and urban legends
- The ability to look for words and phrases in a targeted “words list” (for example, racy content or financial services)
- The ability to do contextual analysis
- The ability to “tune” the product for the environment

Content analysis, when implemented properly, will reduce the volume of unsolicited and undesired e-mails hitting enterprise mail servers, but won’t completely eliminate spam. In fact, spam protection measures may even hinder the usefulness of e-mail by unintentionally blocking desired messages (i.e., false positives).

### ***Heuristics***

Vendors will sometimes tout their heuristics, or algorithms, whose intention is to predict which messages might be tomorrow’s spam based on characteristics in the header or body. The term heuristics is most often applied to viruses, where the algorithms try to find variances of previously seen viruses. The term is less precise when applied to spam. The tactics used in these algorithms must be examined carefully.

### ***Sensing or Reporting***

In some cases, deciding whether a message is spam or not can only be done by watching the sender’s behavior on the Internet. Some reporting companies put decoy e-mail accounts in all the places spammers love to harvest addresses to identify spammers and then creating rules or filters to block those senders. Several of the new products are attempting to create consortia or user groups to develop and share anti-spam rules.

### ***Blacklists and White Lists***

Most blacklists and filtering methods are indiscriminate in what they block. Blacklists may incorrectly list domains as spam sources, which could keep legitimate emails from reaching their destinations. And because filtering is largely based on keyword matches, emails that some people may not consider spam could also be intercepted. Using blacklists as the only anti-spam tactic is entirely unsatisfactory. Used as just one data point in a point system, however, blacklists can be helpful. An enterprise can also create a white list of domains that are always allowed to receive e-mail, no matter what the content is.

### ***Defining Actions***

A good spam-control product should define actions to be taken, depending on which rules were tripped, such as:

- Nonaccept
- Return to sender
- Forward a copy to the supervisor
- Quarantine
- Report egregious messages to a central reporting point

The more information the rules engine can provide, the more thorough the actions can be.

### ***False Positives***

What people fear the most is that an anti-spam agent will falsely identify valuable business messages as spam and delay their delivery, creating uncomfortable situations with recipients. Here are two examples:

- A reporter sent an announcement of a friend's wedding to 100 of his friends who worked for various companies. The message was full of excitement and CAPITAL LETTERS and, for emphasis, many exclamation points !!!!! as well as an invitation to give money to a charity in honor of the happy couple. The message was identified as spam by tools in place at nearly half the companies.
- A bookstore specializing in mystery books was corresponding with one of its customers. It had no trouble receiving the customer's messages, but the store's replies were not getting through. Finally, the customer called her Internet service provider (ISP), and found out that the ISP was blocking messages from mysterylovers.com because it had the word "lover" in the domain name.

Those are typical problems encountered when using a poor spam control tool, one that leaps too quickly to the wrong conclusions. They usually stop only 20 percent to 30 percent of the spam, and catch too many false positives.

### ***Better Tools***

The better tools use a kind of "point system," where each spam tactic identified in the message earns one or more points. The message is not declared to be spam, however, until a certain threshold of points is reached. In each of the examples above, the headers would have included no spammer tactics, which should have allayed concern, but the tools jumped too rapidly to the conclusion that it was spam. The tools used too few criteria and did not look at the mitigating factors.

It is particularly helpful if the spam-control product puts the total number of points into the header (in an X-field) and the reasons why the message scored high. The action system, or another person or process, can then make further decisions based on this analysis.

Probably the greatest concern is that an anti-spam tool might catch a message from an irate customer who is using inappropriate language. Such a message, which earns some points for language, but not enough header points to be spam, might be put into a quarantine pile for review and special handling by a supervisor, rather than being sent to the next-available help desk person.