Facing down the masterminds of unsolicited Internet mail

THE SPAM BUSINESS

Spammers charge real money for their dubious services, and hundreds of advertisers are willing to pay. We'll show you some innovative techniques for controlling and containing spam, including strategies for slowing down spam bots, keeping spammers from getting your address, and separating spam from legitimate email.

BY JOE CASAD, ULRICH BANTLE, AND TOBIAS EGGENDORFER

According to the email service provider Postini [1], of some 524 million messages the provider handled worldwide in a period of 24 hours, 88 percent were spam (345 million messages), including 2 million "special offers," 650,000 get-rich-quick schemes, and 2 million messages with sexual content. Just 46 million legitimate emails actually reached their targets.

Despite the best efforts of the experts, the spam glut isn't going away. Most organizations focus on containing the problem to prevent losses in admin time and user productivity. We'll show you some of the latest strategies for fighting spam in this month's cover story. We'll start by examining some techniques for keeping spammers from getting your address in the first place. Then we'll show you how you can throw the spammers off your trail with a tarpit. We'll also review some anti-spam appliances and services, and we'll describe a custom solution for a user-trainable spam filter that operates from the server side.

Know the Enemy

The origins of the term "spam" are not entirely clear. The term was originally coined on Usenet, where it referred to unsolicited advertising. When the phenomenon hit email, people soon starting calling UCE (Unsolicited Commercial Email) spam. Nowadays, most people simply refer to any kind of unsolicited mail as spam.

The anti-spam project Spamhaus [2] estimates that 200 spammers generate 80

percent of all spam in the USA and Europe. As spamming organizations are typically run by groups rather than individuals, Spamhaus assumes that there are somewhere in the region of 600 professional spammers in this field. You'll find a top ten list of the world's most notorious spammers at the Spamhaus website [3].

Although most users despise spam, many companies still resort to it. One reason for the continued existence of spam is that marketing managers can't resist the extremely low cost. Spammers typically charge between US$ 100 and US$ 200 for a spam drop (EUR 80–160). The cost is so low that companies can pay it with hardly a dent in their budgets. Spammers find a steady supply of customers, even though the "messages" go to unknown email addresses in a totally untargeted way.

Spammers operate on the fringes of the legal system, sometimes passing themselves off as legitimate businesses, even though they use tools such as email worms and viruses to build webs of hijacked robot computers for their dirty work. As Spamhaus puts it, "…some countries do little to deter spammers from operating within their borders. These countries become safe havens for the spam operations that plague everyone else, including their own nationals. Countries with the highest number of spammers operating within their networks are usually those with poor or non-existent spam laws."

Spamhaus rates the Unites States as the country with by far the biggest spammer population, but you'll notice from the Spamhaus top ten list that China and Russia are also major spam distribution spots. According to research by the anti-malware Kaspersky Lab, Russian spammers offer a variety of packages with varying numbers of addresses, ranging from 100 to 3.7 million addresses, without any target-group restrictions. Most advertisers opt for the maximum number of addresses, regardless of the audience.

The companies that get involved with spam typically don't care whether the spamming action returns the desired results. In the Kaspersky survey, none of the respondents had actually measured the effectiveness of their spam investment. Some respondents guessed that spamming accounts for something like 0.01 to 0.05 percent of their turnover.

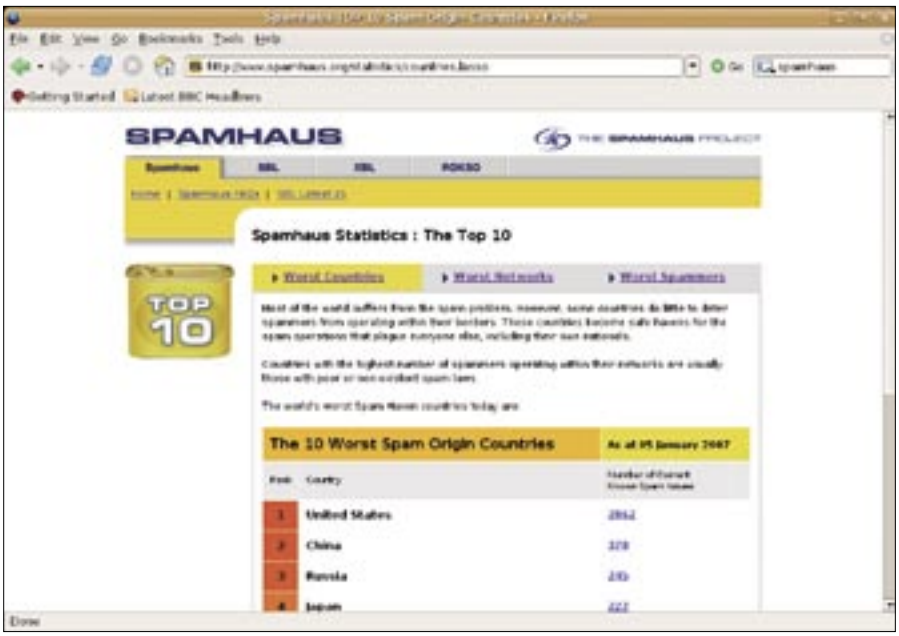

Figure 1: Spamhaus maintains lists of the worst spam countries, the worst spam networks, and the world's most notorious spammers.

The Best Defense

The computer industry has developed a broad collection of strategies for dealing with the spam glut. Anti-spam forces depend on tools such as:

- **Email blacklists and whitelists:** These lists contain the email addresses of known spammers (blacklists), and of known, legitimate mail senders (whitelists). Whitelists reduce the occurrence of false positives. Blacklists are typically ineffective, as spammers will tend to spoof sender addresses.
- **IP-based white and blacklists:** Using a similar approach, these lists catalog spammers by IP address. This technique was useful in the days when open relays were the main distributors of spam. Today, blacklists are typically too aggressive – sometimes blocking all dynamic IP addresses and even whole Asian countries. Unfortunately, this automatically knocks out many legitimate sources.
- **URL blacklisting:** Many spam mails advertise specific websites. If a offending URL appears in a message, the message is assumed to be spam.
- **Content filters:** Content filters analyze the mail content and try to separate spam from ham by detecting typical spam phrasing. The favorites are expressions like "click here" and "unsubscribe," but also "Viagra." Spammers combat strategy this by disguising telltale words.
- **Lazy HTML / Webbug:** This is a special kind of content filter that looks for mails with images the mail client is supposed to download off the Internet. The images are usually served up by a server-side script that evaluates a specific GET parameter at the same time. Spammers use this technique to detect whether a message has been read, thus verifying their address lists.
- **Bayesian filters:** In contrast to manually configured content filters, which include a static list of pertinent words, the Bayesian filter attempts to generate a list based on probability theory. To do so, it analyzes spam and ham mails and bases the probabililty that a message is unsolicited advertising on the frequency of specific expressions. Spammers try to confuse these filters by adding lists of random words to their messages. This explains the jumble of words that typically accompanies today's spam.
- **Image filters:** Image filters attempt to analyze image content. Early filters were restricted to simple color detection logic for common skin tones used in pornography.
- **Checksumming filters and collaborative filters:** Collaborative filters compare messages reaching multiple accounts and attempt to discover simi-

Translate MultiCore Cluster Power into Parallel Application Performance

Get your applications ready for scalable processing

Right the first time:

Intel® C++ and Fortran Compilers for Linux*

Highly optimized compilers designed to handle the most demanding cluster applications. The compilers provide advanced support for threading through OpenMP* and auto-parallelism capabilities to take advantage of performance features available in the Multi-Core Intel® Processors

Intel® Cluster Toolkit for Linux*

New Release, Version 3.0!

Provides a single bundle of all Intel® Cluster Tools to assure highest performance on small and large Intel processor-based clusters, and to efficiently develop, optimize, run, and distribute parallel applications with:

- **Intel® Trace Analyzer and Collector**
  Analyze performance bottlenecks and understand detailed runtime behavior of distributed applications to maximize computational performance and scalabilit
- **Intel® Math Kernel Library Cluster Edition**
  Highly-optimized, threaded routines (BLAS, LAPACK, ScaLAPACK, FFT) for high performance science, engineering, and financial applications.
- **Intel® MPI Library**
  Develop MPI-2 standard applications for all major cluster configurations and network architectures with only one single high performance MPI library
- **Intel® MPI Benchmarks**
  Measure with MPI standard benchmarks (former PMB)

> ***"Intel's MPI and Cluster Tools provide us the best cluster development environment. Using Intel Trace Analyzer and Collector, we were able to shorten MPI communication time by half by finding and removing bottlenecks with non-blocking and blocking communication patterns."***
>
> *Dr. Takahiro Koichi*
> *Computational Astro Physics Laboratory*
> *RIKEN, Japan*

Polyhedron Software Ltd.
Tel. +44 (0)1865.300579
Fax +44 (0)1865.300232
http://www.polyhedron.com/intel/index.htm
sales.intel@polyhedron.com

Grey Matter Ltd.
Tel. +44 (0)1364.654100
Fax +44 (0)1364.654200
http://www.greymatter.co.uk
maildesk@greymatter.co.uk

QBS Software Limited
Tel. +44 (0) 8456.580 580
Fax +44 (0) 8902 7600
http://www.qbssoftware.com
sales@qbssoftware.com

Software

larities. The logic is convincingly simple; if many users receive the same message, it is likely to be spam. Of course, this technique does endanger legitimate newsletters, although whitelists can help prevent incorrect classifications. More cautious filters wait until a user identifies a message as spam. These filters additionally apply other criteria to distinguish spam and ham. To comply with data protection legislation, the central filter just receives checksums of the actual messages. The checksum algorithm has to be resilient against minor changes in the mail content, as the approach is well-known to spammers, who add random text to junk mail.

- **Graylisting:** The graylist strategy delays acceptance of incoming mail, with the target feigning a temporary error in the SMTP dialog. At this point, the sender will already have transmitted his IP, along with the source and target addresses. The mail server stores this information and accepts the message if the sender attempts to reconnect. The idea is that the worms spammers use do not have full-fledged SMTP engines and thus view the temporary error as a permanent condition. Unfortunately, spammers now fully understand this system and can easily work around graylisting systems.
- **SPF, Sender ID, DK:** Sender Permitted From, or Sender Policy Framework, as well as Sender ID, or Yahoo's domain key, create a DNS entry defining the sources from which a specific domain is allowed to receive email. Besides the uncertain patent situation, all of these approaches have one major disadvantage: most spam is now sent by computers identified by DNS entries as responsible for the domain. Registering domains and creating DNS entries is part of the spammer's daily life; after all, spammers are constantly on the move to avoid URL filters and abuse reports.

Figure 3: Microsoft's Sender ID proposal received a cool response from the Debian Project.

You can expect to see more anti-spam techniques as computer systems change and the spam story continues.

More to Come

Despite the big arsenal of anti-spam strategies, spam continues to flood inboxes around the world. Spammers have become quite sophisticated, and they are every bit as resourceful and creative as the good guys. The stock spam campaigns of this past summer show how aggressive and sophisticated spamming methods have become.

Spammers have now turned to new techniques to evade fingerprinting technologies employed by spam filters, for example, introducing animated GIFs to tout their wares. Once a filter has recognized the patterns in the spam message and created a digital fingerprint, the integrated image changes size, color, or position to avoid detection. Minor adjustments that the recipient would never notice from just reading the mail, such as tilting the image by a single pixel, can throw off the spam radar.

So the arms race continues. Don't expect a permanent solution to the spam problem anytime soon. The battle will go on as long as advertisers are willing to pay for it and the worldwide email infrastructure is unable to contain it. The best you can do is filter what you can and try to stop the spammers from getting your address. Read on for more on how you can fight back. ■

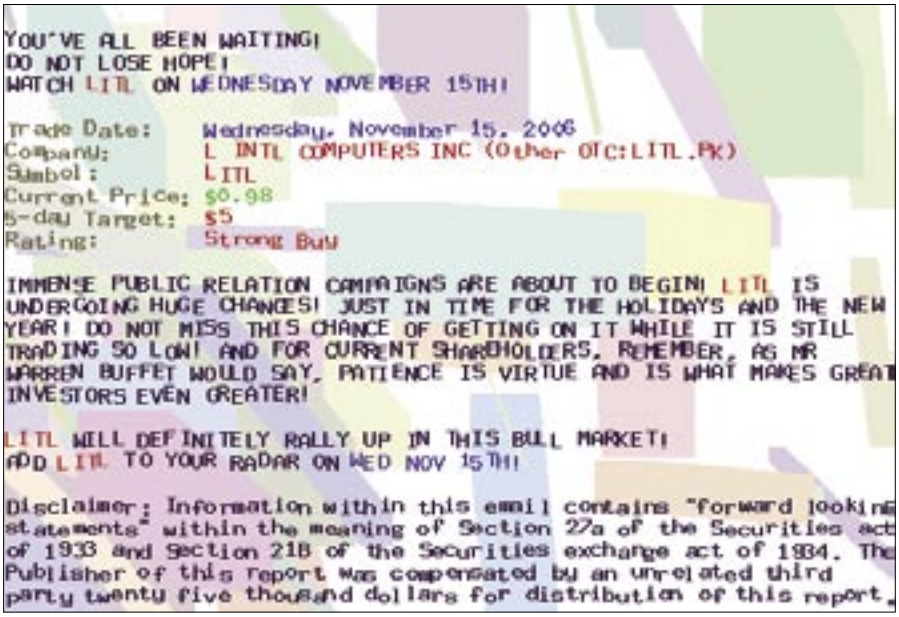

Figure 2: Attempts to trick spam filters are not always as obvious as this.

INFO

[1] Postini statistics: *http://www.postini.com/stats/*

[2] Spamhaus: *http://www.spamhaus.org*

[3] Top ten spammers: *http://www.spamhaus.org/statistics/spammers.lasso*

[4] Kaspersky Lab: *http://www.kaspersky.com/de/*