

## **How Filtering Techniques Can Screen Out Spam**

Spammers continually devise ways to escape antispam filters. This article looks at the current state of spam activity, some of the principal filter evasion tactics used by spammers, and the key antispam filtering technologies that are available today.

---

Apart from making money, spammers' keenest interest is to evade antispam filters. Never has that been truer than now. Gone are the days when organizations could block unsolicited email using homegrown approaches and static keyword filters. Today spammers are continually raising the stakes by devising ways to escape filtering. This article will look at the current state of spam activity, some of the principal filter evasion tactics used by spammers, and the key antispam filtering technologies that are available to you today.

### **The state of spam today**

How big of a problem is spam? The latest edition of the Symantec Internet Security Threat Report paints a pretty sobering picture. For the period between July 1 and December 31, 2004, Symantec reported a 77 percent growth in spam for companies whose systems were monitored for it. The weekly totals of spam rose from an average of 800 million spam messages per week to well over 1.2 billion spam messages per week by the end of the reporting period. All told, spam made up more than 60 percent of all email traffic observed by Symantec during this period.

While there is little doubt that spam is an annoyance to users and administrators, it is also increasingly a serious security concern, as it can be used to deliver Trojan horses, viruses, and phishing attacks. In addition, high volumes of spam can create DoS (Denial-of-Service) conditions where email systems become so overloaded that legitimate email and network traffic are unable to get through.

### **Getting around the filters**

Not surprisingly, large-scale spammers have become more adaptable and sophisticated over the years. This is partly a matter of simple economics. According to some estimates, for a spammer to bring in \$1 million a month, all that is required is a \$20 purchase from one out of every 2,000 "spammees" -- a mere 0.05% response rate.

Given such favorable economics, spammers will cycle through fake domain names and alter email subject lines so precisely and efficiently that by the time older antispam tactics can discern a pattern, the damage is done and a new attack with different characteristics has already been launched. Mass mail software even allows spammers to run mail through preprogrammed checklists, evaluating whether it will likely be blocked by spam filters.

Increasingly, content modification using HTML has become the spammer's most powerful antifiltering technique. Spammers choose HTML because it attracts attention,

enables tracking (spammers can verify whether a targeted email address is valid), and allows them to insert bogus tags in order to circumvent filtering.

Spammers have also proven adept at disguising the external appearance of URLs so that recipients are fooled into believing that the URLs belong to a legitimate organization. The success of email “brand spoofing” attacks is testimony to the power of this tactic. In such attacks, spammers create fraudulent emails and disguise URLs, purporting to originate from legitimate organizations in order to entice recipients to provide private and financial information.

While sending out bulk email is a fairly simple matter, spammers need a mechanism to conceal their identity, thus avoiding source blocking by IP address. One method involves the misuse of open proxy servers. Open proxy servers are misconfigured or virus-infected computers that allow traffic for virtually any network service to be channeled through a host computer.

Spammers routinely identify and hijack insecure proxy servers, whose owners may have no idea that their systems have been misappropriated. By some accounts, two-thirds of all spam emanates from these hijacked servers.

Spammers are also coming up with new ways to hijack computers, as shown by such mass-mailing computer worms as Sasser, Netsky, and SoBig. While these viruses didn’t have especially malicious payloads, they did install mail programs on victims’ computers, setting the stage for an immense network through which spam could be relayed.

### **Today’s filtering technologies**

Although many antispam solutions claim to work right out of the box, they actually offload much of the spam-fighting burden on administrators and end users. The question for many IT departments then becomes: How much time do we want to spend fighting and managing spam? After all, ongoing filtering and training costs can quickly spiral out of control.

In Symantec’s case, the effectiveness and accuracy of its spam filters are made possible by its multiple logistics and operations centers. These centers work globally, evaluating mail for new variations of spam and issuing filters to identify and capture similar messages.

Spam analysis at these centers begins with the Probe Network, an array of more than 2 million decoy email addresses and domains, also known as spamtraps or honeypots. This global network of email accounts attracts and collects large quantities of spam -- tens of millions of spam messages pass through the Probe Network every month. As messages come into the centers, automated processes and expert technicians go to work, analyzing incoming spam and developing countermeasures.

While no one has so far developed a silver bullet against spam, a variety of filtering techniques have been created to keep spammers at bay – all of which must constantly be

evaluated and updated if enterprises are to keep ahead of spam. Among the techniques that Symantec utilizes:

- **Reputation filtering.** Reputation-based blocking is a filtering technique that examines the quality or reputation of the sending source or mail server. It involves monitoring hundreds of thousands of email sources to determine how much mail sent from these addresses is legitimate and how much is spam. It also tracks data such as mailing patterns, the presence of open proxy or unsecured mail servers, volume of messages sent, and complaints.
- **Heuristics.** Heuristic filters analyze the header, body, and envelope information for incoming messages, checking for the presence of distinct spam characteristics. Each message is assigned an overall score, which is then compared to a threshold that determines whether the message is spam or not. Heuristic filters, once they are trained to determine what spam and legitimate mail looks like, can be very effective at identifying new spam. The drawback of many heuristic filters is that they can create a substantial administrative burden. If not properly trained and weighted for accuracy, they can also produce significant numbers of false positives (legitimate email messages that are incorrectly identified as spam).
- **Header filters.** To proactively identify first-time spam, header filters consist of regular “expression-based” filtering rules that exploit commonalities or trends present in spam messages. Examples of telltale spam characteristics that a header filter would address include traces of information left in messages by spammer tools and modified time zones.
- **Attachment filters.** Message attachments have long been a favorite tool of spammers. By attaching a deceptively named file or image to an email, spammers tempt recipients to click through and open the file. Filters based on a particular MIME attachment (for example, a specific pornographic image used in a real-time spam attack) can stop that attachment from reaching users. Attachment signatures therefore make it unnecessary to block entire categories of attachments.
- **URL filters.** Continually evolving URL-based filtering technologies aim to reverse spammers’ new methods of URL masking and obfuscation techniques. For example, URL filters can stymie spammers’ attempts to encode URLs with extraneous characters.
- **Language identification.** It’s estimated that between 10% and 20% of all spam is written in languages other than English. As multilingual spam becomes a larger problem, antispam solutions must take into account the language in which messages are written. Solutions need to contain language identification abilities and heuristics that apply only to particular languages.
- **Custom filters.** An effective antispam solution should provide customization tools that allow administrators to be more aggressive in targeting unwanted mail.

That way administrators can create filters to proactively block or handle mail that does not necessarily meet the criteria of spam. For example, administrators could filter email from marketing lists that generate user complaints or use excessive bandwidth.

### **What to look for**

Accuracy is the single largest differentiator when it comes to antispam products. Accuracy in this regard refers to the false positive rate. For example, if a company received 100,000 messages a day, even a 1% false positive rate means 1,000 messages are mistakenly blocked every day. Obviously, such a rate is too high. In addition to evaluating antispam solutions based on the filtering technologies discussed above, enterprises are encouraged to look for solutions that:

- Produce low or no false positives
- Have a track record, through product reviews or customer validation, of having extremely low false positive rates
- Employ a balanced mix of technologies to guard against overaggressive filtering
- Have safeguards for preventing, detecting, and resolving suspected false positives
- Provide quarantine options to let users ensure that legitimate messages are not lost.

### **Conclusion**

The lifecycle of spam continues to evolve. Spammers are escalating the battle with new filter evasion and dissemination techniques.

In response, Symantec's research and development groups continually adapt filtering technologies to challenge spammers and screen out their attacks. These technologies, backed by a comprehensive spam analysis infrastructure, enable Symantec to provide an accuracy rate of 99.9999%.

While selecting the right antispam product can be daunting, organizations cannot afford to ignore the flood of spam targeting their servers and employees. The costs in terms of lost IT resources, employee productivity, and legal liability are simply too great. Effective spam protection is no longer an option – it's a necessity.

### **Related links**

Symantec Brightmail AntiSpam

<http://enterprisesecurity.symantec.com/products/products.cfm?ProductID=642>

White Paper: Filtering Techniques in Symantec Brightmail AntiSpam 6.0

[http://symantec.gmms.gettyimages.com/stagingArea/Email\\_export01\\_27\\_2005\\_09\\_49\\_00/0/0/10355860\\_SBMASv6\\_FiltTch\\_wp.pdf](http://symantec.gmms.gettyimages.com/stagingArea/Email_export01_27_2005_09_49_00/0/0/10355860_SBMASv6_FiltTch_wp.pdf)

Webcast: Stop Spam and Email-Borne Threats with Symantec's New Hosted Mail Security Solution

<http://enterprisesecurity.symantec.com/content/webcastinfo.cfm?webcastid=147>

Webcast: The Spam Problem Solved: Bechtel Corporation Case Study

<http://enterprisesecurity.symantec.com/content/webcastinfo.cfm?webcastid=143>

Symantec Antispam Center

<http://enterprisesecurity.symantec.com/content.cfm?articleid=2492>