GFI WHITE PAPER

How to keep spam off your network

What features to look for in anti-spam technology

A buyers' guide to anti-spam software, this white paper highlights the key features to look for in anti-spam software and why.

Introduction

This paper helps you identify key features needed to effectively deal with spam.

The growth and cost of spam

The Radicati Group, a US research firm, estimates that 52% of current global email traffic is spam mail and predicts this will reach 70% by 2007. Similarly, the European Union estimates that 50% of all email messages are spam.

This means that employees must dedicate part of their work time to dealing with spam, which results in a decrease in productivity (and an increase in frustration!). Loss of productivity is the main cost of spam, particularly as so many spam mails are received per day. There is also the cost of bandwidth wasted by spam, as well as other storage and network infrastructure costs. Furthermore, with the influx of spam and its deletion, an important message could accidentally be trashed along with the unsolicited mail in the rush to clear one's inbox of junk mail

Ferris Research calculated that if an employee receives just five spam mails a day and spends 30 seconds on each, he will waste 15 hours a year on junk mail - now multiply that by the hourly rate of each employee in your company and you will have a very conservative idea of the cost of spam to your organization. The Radicati Group reported that spam cost IT around US$49 per mailbox in 2003, and expects this to skyrocket to US$257 per mailbox in 2007.

It is essential to put a stop to spam to save time, money and bandwidth. One step towards achieving this is to advise your network users to keep their email address private (no postings to message boards etc.). However, apart from applying common sense, you also need to deploy an effective server level anti-spam tool.

Choosing the correct anti-spam software

Many software packages are available on the market to help you combat spam; but not all are incisive enough in dealing with spam. A number of key features/issues that you should look for are discussed below.

Server-based or client-based?

Battling spam at client level is much more time-intensive than at the server level. It requires you to deploy anti-spam software to all workstations on your network and involves frequently returning to those workstations to update the anti-spam rules on each of them. It also means that your email infrastructure is being taxed by spam, as your server message stores are filling up with useless emails waiting for deletion. What's more, it also

involves time on the part of your users, who have to identify spam or update their rule sets: This is the very thing you are trying to oppose in your bid to block spam!

In addition it does not have the information and resources that a server-based anti-spam software has - it does not allow you to perform sending server checks, for example. To block spam effectively, you need to have a server-based anti-spam product, because it offers these advantages:

- Installation at the gateway eliminates the deployment and administration hassle involved with desktop-based products.

- Far cheaper to license.

- Prevents spam from even entering your email infrastructure, meaning that your email stores are not full of spam messages.

- Server-based anti-spam software has more information, and can do more to detect spam effectively.

Bayesian filtering technology

A few years ago, most anti-spam products simply used a list of keywords to identify spam. A good set of keywords could catch plenty of spam. However, nowadays Keywords-based spam catching generates too many false positives and requires too much manual updating.

It's now widely acknowledged by leading experts and publications that the best way to catch spam is using a Bayesian filter. A Bayesian filter uses a mathematical approach based on known spam and ham (valid email). This gives it a tremendous advantage over outdated spam technology that just checks for keywords or relies on downloading signatures of known spam.

More information about Bayesian filtering can be found in the whitepaper Why Bayesian filtering is the most effective anti-spam technology at http://www.gfi.com/whitepapers/why-bayesian-filtering.pdf.

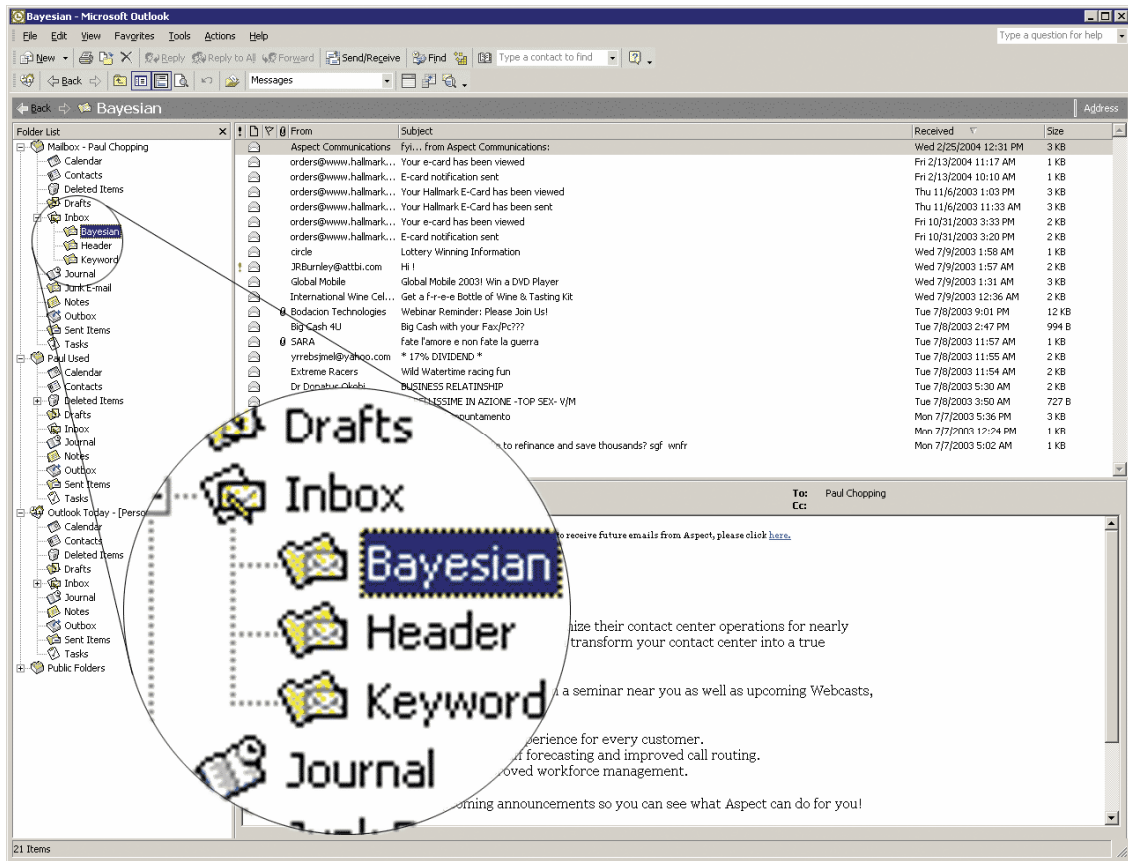In short, Bayesian filtering has the following advantages:

- Looks at the whole spam message, not just keywords or known spam signatures

- Learns from your outbound mail (ham) and therefore reduces false positives greatly

- Adapts itself over time by learning about new spam and new valid mail

- Dataset is unique to company, making it impossible to bypass

- Multilingual and international.

Tailored ham data file for Bayesian filter

It is very important that the Bayesian filter uses a dataset that is custom-created for your installation: the ham data MUST be collected from your outbound mail (this way, the Bayesian filter is tailored to your company through an initial training period). Some anti-spam software uses a general ham data file that ships with the

product. An example is the Outlook spam filter or the Exchange Server Internet Message filter. Although this technology does not require the initial learning period, it has 2 major flaws:

- The ham data file is publicly available and can thus be hacked by professional spammers and therefore bypassed. If the ham data file is unique to your company, then hacking the ham data file is useless. For example, there are hacks available to bypass the Microsoft Outlook 2003 spam filter.

- Secondly the ham data file is a general one, and since it's not tailored to your company it cannot be as effective as a customized one. You will suffer from noticeably higher false positives. For example, a financial institution might use the word "mortgage" many times over and would therefore get a lot of false positives if using a general ham data file.

Reviewing spam is easy if it's stored in a subfolder of a user's mailbox

Automatically updated spam data file for Bayesian filter

The spam data file of the Bayesian filter must be constantly updated with the latest spam by the anti-spam software. This will ensure that the Bayesian filter is aware of the latest spam tricks, resulting in a high spam detection rate (note: this is achieved once the required initial two-week learning period is over). Choose an anti-spam product that will collect this spam data for you and allow you to automatically download these updates!

Spam handling to efficiently review spam

Inherent in anti-spam technology is the fact that there will be false positives, i.e., mail being flagged as spam even though it is not actually spam. Therefore good anti-spam software should provide an easy way for users to review mail marked as spam in a fast and efficient manner.

To save administrators time and hassle, anti-spam software had best include an option to direct mail identified as spam to individual users' junk mail folders. In addition, the software should sort the spam into different folders depending on what identified it as spam. This quick access to mail marked as spam greatly helps the user review his/her spam efficiently. Some anti-spam products require the user to login to a web-based system and review their mail one by one – in practice; this is cumbersome for the user and will lead to the feature being rarely used.

Flexible whitelists to reduce false positives

Anti-spam software must have an efficient way to automatically build extensive Whitelists. Whitelists should identify all valid business partners, so that their mail is never flagged as spam. Good anti-spam software should include the facility to automatically create and update these whitelists.

How GFI MailEssentials tackles spam

GFI MailEssentials' approach to spam detection is based on the following key methods and technologies:

- **Tackles spam at the server level** - GFI MailEssentials installs on your Exchange 2000/2003 Server, or in front of your mail server (if using Exchange 5.5 or another mail server). It detects spam BEFORE it reaches your mail server. This way, spam does not tax your email infrastructure, and any spam detection rule updates need only be deployed on the GFI MailEssentials machine. Whitelists (domains/email addresses you always wish to receive mail from) and blacklists (domains/email addresses from which you do not want to receive mail) can be used at server level.

- Analyzes the content of the mail using **Bayesian filtering** and uses ham data specific to your company. The spam data is automatically updated by downloading the latest spam data from the GFI website. For more information on Bayesian filtering, check this white paper at http://www.gfi.com/whitepapers/why-bayesian-filtering.pdf.

- **Reduces false positives through an automatic whitelist** - GFI MailEssentials includes a patent-pending automatic whitelist management tool. This unique technology means that all business partners are automatically added to your whitelist - without any need for administration - and their mail will not be passed through the spam filter, greatly reducing false positives.

- **Flexible spam handling** - After a mail is found to be spam, it can be forwarded to a sub folder in the user's mailbox. If they find a valid email (for example, a newsletter which they wish to receive), users can add the sender to the whitelist.

- GFI MailEssentials includes **keyword checking capabilities** so that administrators can further tune their anti-spam filters. For added protection, Bayesian filtering is supplemented by a number of **other spam detection technologies**, including intelligent mail header analysis and by checking senders against custom blacklists and public blacklists such as ORDB or SpamHaus

About GFI

GFI is a leading software developer that provides a single source for network administrators to address their network security, content security and messaging needs. With award-winning technology, an aggressive pricing strategy and a strong focus on small-to-medium sized businesses, GFI is able to satisfy the need for business continuity and productivity encountered by organizations on a global scale. Founded in 1992, GFI has offices in Malta, London, Raleigh, Hong Kong, and Adelaide which support more than 200,000 installations worldwide. GFI is a channel-focused company with over 10,000 partners throughout the world. GFI is also a Microsoft Gold Certified Partner. More information about GFI can be found at http://www.gfi.com.