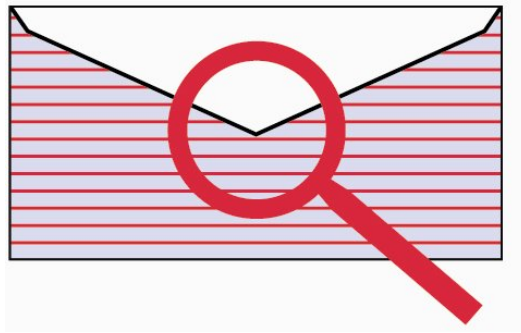


Mail-Filters Technology

How does it work? Why is it better?



A White Paper

April, 2005

Introduction

Who is Mail-Filters?

Mail-Filters, founded in 2001 in San Mateo, CA. has a mission to eliminate spam for companies, service providers and OEMs. Ultimately, companies and end-users just want spam to go away – with minimum effort on their part and without missing any legitimate e-mail. While most solutions focus on spam catching and not false positives, from our inception, we have focused on eliminating false positives, while still catching spam. Today, the Mail-Filters technology consistently catches over 95% of spam with less than 1 in 1,000,000 false positives – the best combination on the market.

Mail-Filters has developed proprietary technology that KNOWS whether a message is spam, and doesn't use formulas, AI, or algorithms to formulate an educated GUESS if a message is spam because GUESSING causes false positives.

Most customers have relatively few requirements:

- A solution must catch most spam, while generating few, if any, false positives
- Spam messages should be deleted or quarantined at the perimeter, before getting to the production mail server, in order to conserve bandwidth and e-mail server resources and stop malicious code that some spam creates.
- The anti-spam solution can't slow down e-mail noticeably – e-mail queues cannot backup with messages.
- The solution must not consume end-user or administrator time for maintenance and updates.

There are other system requirements (platform support, installation ease, quarantine review, etc.), but the four above are cited by almost every customer.

GUESS Technology: It is no doubt desirable, from a vendor's perspective, to write a program just once using a formula or algorithm to determine if a message is spam. This avoids the ongoing effort of updating and maintaining the filter. This was the approach Mail-Filters first investigated for its solutions, but found false positive rates were too high, often in excess of 1%.

Indeed, many competitors have taken this approach and their solutions can catch a lot of spam. Unfortunately, spammers are very good at disguising their messages to look like the regular mail customers want to receive. For GUESS solutions to catch a lot of spam and not generate false positives they have to take a "firm stand on the fence" approach. They place a lot of good mail in a "maybe this is spam" folder. Unfortunately the user or administrator is left with the task of sorting through it to make the final decision, with little or no help.

Another way some filters address the false positive problem is by providing a kind of tuning. These solutions claim both high catch rates and low false positive rates – however they don't

say that one rate is achieved at the expense of the other. In other words, their solutions can have EITHER a high catch rate OR a low false positive rate, but not both at the same time.

Thus, there are a lot of competitors on the market (roughly 200 of them) claiming 95% catch rate with no, or almost no, false positives. When customers try them out, their experience is disappointing. In a recent survey, 70% of businesses had deployed anti-spam solutions, but 70% of those are unhappy with the results. So why are so many companies not changing to get better solutions? Because 74% had already tried at least 2 solutions, but got similarly poor results. They now feel that all anti-spam companies claim performance they can't deliver.

Bottom Line: Customers need their anti-spam solution to KNOW whether a message is spam, to stop getting fooled by new spammer tricks, and to stop GUESSING whether a message is spam and getting it wrong 1-3% of the time. Just think what that false positive rate is for a typical business user: one a day or more. If that false positive is from an important customer, vendor or business colleague, and it goes undetected the results could be disastrous.

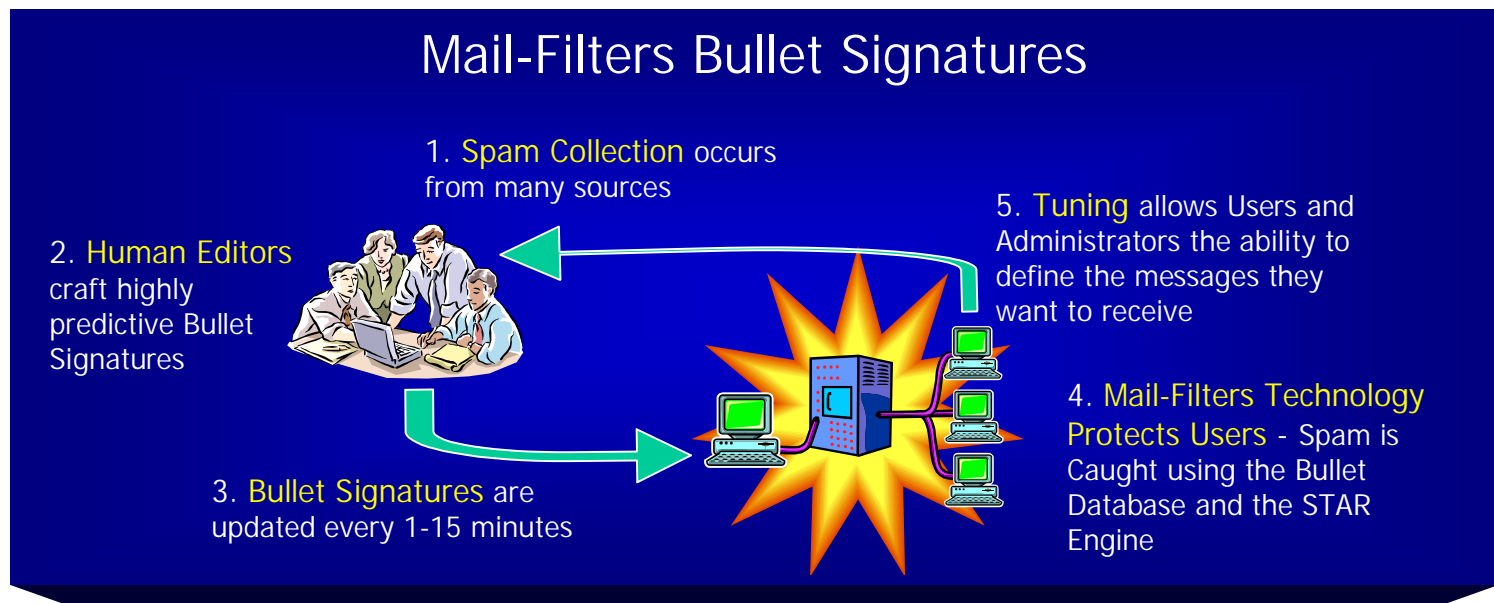
Mail-Filters technology is different. This white paper will explain how Mail-Filters technology works, why it's different from other technologies, and why it delivers 95+% catch rate while generating less than 1 in 1,000,000 false positives.

The Mail-Filters KNOW Technology

Mail-Filters has invented two new proprietary approaches for defending against spam attacks: the Mail-Filters Bullet Signature Database that enables the detection of spammers and the messages they send; and the StarEngine, used to neutralize spammer tricks.

The Bullet Signature Database

Mail-Filters Bullet Signatures are small, targeted, and lethal spam signatures handcrafted by human editors. Bullet signatures are constantly updated to maintain effectiveness and accuracy. From a high level, here's how they work:



Spam is Collected From Many Sources

How is spam collected? Mail-Filters collects spam from many different sources; most of it comes from real customers working to help themselves and the community at large to get rid of spam.

What about honeypots? Many competitors, like BrightMail, use honeypots (or a PROBE network), to gather spam. The theory is that spammers will scan the Internet for e-mail addresses, and then use those addresses to spam. A couple of years ago, this was very true, but now if you use only honeypots to collect spam, you are missing a significant portion of available spam because spammers gather e-mail addresses in many different ways. These include:

- Spiders that gather e-mail addresses found on websites and newsgroups. This type of e-mail address is the least desirable to spammers.
- Dictionary attacks that find e-mail user addresses by sending messages to a company using common e-mail names, then tallying the bounced messages. If they send 1000 and 985 bounced, they know they have 15 good e-mail accounts.
- E-mail addresses that are found or purchased from companies or individuals that have a list of legitimate e-mail addresses, either by doing business with them or enticing someone into giving them their e-mail address.

In other words, the honeypot technique collects only one type of spam. If a spammer doesn't use names gathered by spiders, the honeypot will never see it and the filter that depends on honeypots is defeated.

Human Editors Craft Bullet Signatures

Why are Human Editors important? Trained editors can figure out if a message is spam or not. While most messages are obvious, it takes training to figure out if a message was solicited. Most newsletters fall into this category. For example, many users will tell their spam filter administrator that newsletters from travel sites are spam. Unfortunately, once messages from this site are marked as spam, e-tickets don't get delivered.

It is obvious that accurate identification of spam is critical to the process. There have been competitive solutions that use a community of humans to identify spam or spammers and have a system that automatically apprises the community of this determination. The problem has been that these individuals do not have the training to determine what is and isn't spam for the entire community and false positives occur.

How is a Bullet Signature Created? Bullet signatures are crafted from different attributes of a message – the underlying structure, if you will. Our editors look for the specific combination of attributes that will enable this signature to catch this spam and usually other spam messages from this same spammer. The signature will almost always have bits of

information from the body of the message, and it might also include items from the headers, like the sending IP address or domain name. The Bullet Signature is a combination of attributes that our editors determine won't be repeated in valid e-mail. Since we have our trained experts creating these Bullets, you would expect them to be accurate – and they are.

Why can't the Bullet Signature process be automated? We believe virus makers and spammers are smart and motivated to get around defenses. As anti-virus manufacturers still rely mostly on human analysis and response mechanisms to be accurate, so must anti-spam software. It takes human experts to see the subtle wrinkles spammers are creating to get around spam filters. There are many automated tools our editors use to quickly craft the most effective Bullet Signature, but the ultimate arbiter of whether a message is spam or not and the actual crafting of our Bullet Signatures falls to our human editors.

How are Bullet signatures different from other signature-based systems? All signature-based systems are accurate – they only identify known spam. However the Mail-Filters Bullet signature system is significantly different from other signature systems. Other signature-based (or “Fingerprint”) systems use a checksum or hash value that is calculated based on the characters and where they are in the message. The exact same messages will have exactly the same checksums.

This “Fingerprint” approach was based on the observation, a couple of years ago, that spammers would craft one message then send it to millions of users. If you could collect one of those messages, calculate the checksum, and then let everyone know about the checksum value, then you would have an effective and accurate anti-spam filter. BrightMail based their filtering technology on this technique. Unfortunately, spammers realized what was happening and they crafted tricks to change the checksum for each message they sent – without losing the overall message. Sometimes spammers add random characters or words, changing the checksum. The result of this is that the checksum databases got huge and their effectiveness declined.

To be clear, Mail-Filters does NOT use checksums or hash values in their Bullet signatures. The spammer can pull all sorts of tricks to defeat checksums and they will have no effect on Bullet Signatures.

Does the Mail-Filters technology require a Bullet Signature for every spam message? No. Because Bullet Signatures identify the underlying characteristics that spammers use to build their messages, a single Bullet Signature may detect many different spam messages from the same spammer. This fact allows the Mail-Filters human editors to concentrate on new spammers or tricks without needing to look at the many different variations sent by the same spammer.

Updates to the Bullet Signature Database

What about updates – how often do they occur? The Bullet Signature database is updated as often as once a minute to maintain effectiveness and accuracy.

How do updates occur? Updates occur as an HTTP outbound request. This is an authenticated connection where either a full Bullet Signature database or only the adds, changes, and deletes from the current database are downloaded. Signatures are never pushed down to a server for security reasons.

Tuning

What if someone gets a message they regard as spam? If any user receives a message they think is spam, they need only forward it to spam@mail-filters.com. There is an automated attendant that creates a rule for just that end user in that company, but more importantly, the spam message gets nominated for inclusion in the global Bullet Signature list.

What if someone wants to get a message in the future that was marked as spam? If, on the rare occasion a message is marked as spam but is really valid e-mail, then users just forward the message to notspam@mail-filters.com and a rule will automatically be created for that user.

What kind of false positives occur? The type of false positive the Mail-Filters technology typically gets is a newsletter that has accepted advertising from a spammer. It turns out that spammers use the same characteristics in their online advertising as they do in their spam messages – thus the newsletter is marked as spam. A quick forward of the newsletter to notspam@mail-filters.com takes care of that issue.

Who participates in the Tuning? Any end user or administrator can forward messages. It helps the entire community in their battle against spammers. An administrator can set a rule for the entire domain or company by forwarding the message and putting the word DOMAIN or COMPANY in the subject line. Tuning is one of the major ways spam is collected by our human editors and used in crafting Bullet Signatures.

The Mail-Filters StarEngine

The Mail-Filters StarEngine (**S**pammer **T**ricks **A**nalysis and **R**esponse Engine) looks for spammer tricks such as falsified information in the headers and other places in the message and other unique identifying characteristics of spam.

What tricks are neutralized? Spammers now know that they have to do specific things to get around certain filters. They have quite an arsenal of tricks to try to get their messages read. Here are a few:

The Mail-Filters STAR Engine

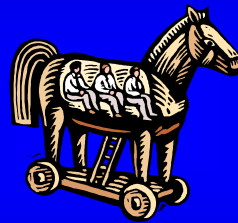
Spammer Tricks Defeated by the Mail-Filters STAR Engine



Hash Busting:
Random insertion of characters or words used to defeat signature based filters.



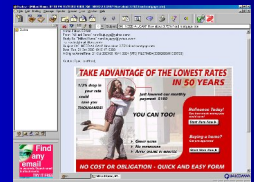
Snow-Flaking:
An effort to make all HTML e-mail unique – like a snowflake – it inserts invisible characters or HTML comments into messages.



Embedded Content:
Usually an HTML message that displays content pulled from a web page based on an embedded URL in the message. Gets by most spam filters.



Foreign Language:
Most filters only work in their native language so most foreign language spam gets through.



HTML E-Mail:
Often used by spammers to display graphics and increase response rate. HTML e-mail is difficult for many filters to scan.

Remember, spammers select their tricks to get around specific types of filters. For example, if SpamAssassin has a rule that looks for the word SEX in the subject line, a spammer might spell it \$EX or \$\$EX or S_E_X or \$ E X – you get the idea. A SpamAssassin administrator would have to be very dedicated to find and create rules for all the variations of words spammers use today. The following chart illustrates what tricks a spammer might use to get around certain types of filters.

Spammer Tricks	Mail-Filters StarEngine + Bullet Signatures	Ordinary Signatures (Checksums or Hash Values)	Linguistic Formulas (SpamAssassin, etc.)	Bayesian	RBLs
Hash Busting	Not Vulnerable	Vulnerable			
Snow-Flaking	Not Vulnerable	Vulnerable			
Misspelling	Not Vulnerable		Vulnerable	Vulnerable	
IP Hopping	Not Vulnerable		Vulnerable		Vulnerable
Embedded Content	Not Vulnerable		Vulnerable	Vulnerable	
Foreign Lang. Spam	Not Vulnerable		Vulnerable	Vulnerable	
HTML Spam	Not Vulnerable	Vulnerable	Vulnerable	Vulnerable	

How does the StarEngine keep up with the spammers? Most tricks are handled through the Bullet Signature Database, but some tricks are neutralized through the StarEngine. Updates to the StarEngine occur about every 1 to 2 months and are automatically updated, without necessary intervention from administrators. There is no doubt we are in an arms race between the forces of good (spam filters) and evil (spammers). As spammers come up with new tricks, our human editors and engineers work together to respond to the threat.

Why are other filter technologies more vulnerable to tricks? No one thought five years ago that spam would be the problem it is today. The analysis done prior to a year ago had solutions architected with one of two assumptions in mind:

- A spammer creates one message and sends it to millions of users.
- You can look at a spam message in the Inbox, sometimes without even opening it, and see that it's spam.

Algorithms or techniques were created using these assumptions. Spammers have taken advantage of these assumptions and make their messages look as innocuous as possible, then utilize their tricks to get them around spam filters. Today's solutions, architected with last year's assumptions, have trouble maintaining their catch rate. In reaction to their lower effectiveness, those filter's aggressiveness is then turned up, causing more false positives.

What about Foreign Language Spam? Mail-Filters is designed from the ground up to support all languages, including languages with double-byte character sets like Japanese, Korean, Chinese, and Arabic. The Bullet Signature database is about 1/3 non-English spam and there are companies using the Mail-Filters technology across the globe on 6 continents.

The Mail-Filters Advantage

Mail-Filters has crafted proprietary technology to catch spam. But more importantly, the architecture of the Mail-Filters technology is flexible enough to win the arms race with the spammers. The Mail-Filters technology is proven, having been used by businesses since March of 2002. It is being used by OEMs such as Toshiba, Sybari, FilterLogix, Elron, Nurivision, Solinus and Zix Corporation to solve spam issues for their customers.

Compare Mail-Filters to other technologies – Checksum signatures have been described above in some detail, but here are some other anti-spam technologies:

Linguistic Formulas – [SpamAssassin and others – sometimes called Heuristics or Artificial Intelligence] – this type of filter utilizes a formula to estimate if a message is spam by looking at the words used in a message. For example, if a message has the word FREE in the subject line, it might get 30 points, an exclamation point in the subject line might give it another 10 points, etc. – if the message reaches a preset threshold, it gets marked as spam. Basically, if the solution has a slider bar to change the threshold, it's usually this type of filter. Spammers attack this filter by misspelling words, making the message look like a regular message, and linking to HTML code from within a message. Unlike the Mail-Filters technology that requires no

administration, a linguistic formula-based solution requires constant updating of the rules to remain effective. Additionally, because of the computational effort required to calculate the message scores, the hardware resources required are much greater, usually about 5 to 10 times greater, than when using the Mail-Filters technology.

Statistical or Bayesian Filters – A statistical filter looks at a corpus of good mail and compares it to a corpus of spam. It then tries to determine the words or language used in each to look for differences. This is a good filter for an individual that has some time to keep up with it. Unfortunately, it begins to fail as an enterprise or more global solution as more users contribute to the good mail and spam increases. The differences become watered down and filter's effectiveness is lost. Additionally, there are now specific spammer tricks aimed at Bayesian filters including misspellings, random word insertions, HTML and Embedded Content. Foreign Language can also be a problem.

RBLs – Realtime Blackhole Lists are a collection of IP addresses known to be the source of spam. Many RBLs add IP addresses from companies based on one complaint and it might take weeks for a company to get off the list. Additionally, many RBLs not only want to stop spam, but also want the company whose IP address is listed to agree to a "Code of Conduct" that many companies will not agree to, even if they are not spammers. The result is that many RBLs create false positives and don't catch much spam. A recent test showed one popular RBL catching about 30% of the spam, with about a 30% false positive rate. While Mail-Filters technology supports the optional use of RBLs, customers do not typically turn it on.

Challenge / Response – This requires the sender to answer a question before the message is delivered if the sender is not in the recipient's white list, usually their e-mail address book. In a recent test, about 40% of legitimate senders did not respond. Most thought the Challenge / Response was too complicated or too much effort, especially if they wanted to ask about a product for sale. Others thought it was a trick and that the Challenge itself was spam or even a scam. We've even seen spam mimic a Challenge so that it would be opened. Challenge / Response has potential to work for many individuals, but in the business world where lots of messages are received from unknown senders, it is going to be tough to implement without losing valid communication.

Traffic Analysis – There are a couple of anti-spam solutions that try to analyze where spam is coming from based on traffic patterns. We've already seen spammers that are starting to send only 10 or 20 thousand messages from 50 to 100 Internet locations to overcome this method of detection.

What about combining technologies? We come back to the basic requirements of a spam solution – it's got to catch spam without false positives. Most users find the Mail-Filters performance alone meets their needs. But it has also been combined with filters with a relatively high false positive rate, one in 1,000 or so, to achieve the Mail-Filters rate of one in 100,000.

For more information visit the Mail-Filters website at www.Mail-Filters.com or contact Mail-Filters at 650-655-7700.

© 2005 Mail-Filters.com, Inc. All rights reserved.

Mail-Filters Technology

Mail-Filters, Mail-Filters.com Inc., PhishRepellent, SpamCure, SpamRepellent, Bullet Signature, STAR Engine, StarEngine SDK, StarEngine Enterprise SDK, StarEngine Desktop SDK, and StarEngine Embedded SDK are trademarks of Mail-Filters.com Inc. All other trademarks, product or service names, brand names, company names and/or logos appearing in this release are the property of their respective owners.

Mail-Filters.com, Inc.

411 Borel Ave. Suite 510
San Mateo, CA 94402
650-655-7700