IPSWITCH

IPSWITCH Collaboration Suite

Spam Filters Need the Human Touch

Whitepaper

Executive Summary

Readers of this or any other whitepaper on spam need no proof that spam is a problem. They know that; they want to know what can be done about it. This paper answers that question for organizations that use or distribute email.

Spam is an unsolicited email message, either human or systems generated, that is received by another. The purpose? To get the receiver to do something: typically, to make a purchase. The vast majority of mailbox users do not want the message. Phish, on the other hand, is a special kind of spam designed to steal personal or corporate information. Most spam irritates, phish is dangerous. No one wants phish but some may want certain spam. That Rolex replica, for example, looks very tempting.

Given this subjective nature of spam, the most effective and accurate filter is a mailbox user with a delete button. Unfortunately, while this approach catches 100% of the spam with no false positives, it does so at an unacceptable cost for an enterprise with many mailboxes. The total read-and-delete time, not to mention the resulting user frustration, is huge. Moreover, the bandwidth needed to distribute the spam plus the potential for spreading corrupting content is unacceptable.

Phish users with a delete button frequently do not know that the message from an apparently legitimate site (sometimes called a spoof) is inherently dangerous. The consequences of supplying the information it asks for can be identity theft.

It follows that spam must be caught before it enters an organization's network. There are three ways an organization might do this:

1. Hope that legislation will make it illegal
2. Hope that litigation will make it unprofitable
3. Use technology to stop it, no matter how much is sent

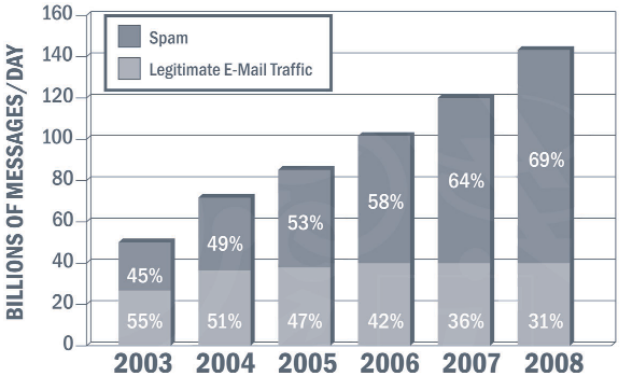

Chart 1. Source: *The Radicati Group*

Legislation and litigation continue to make great eight second sound bites, as we predicted, but to date have had little or no effect on the volume and nature of spam. In fact, as can be seen from Chart 1, spam as a percentage is actually increasing. Some sources actually believe the percentage is higher than that cited by The Radicati Group. Technology, supported by human analysis, remains the best solution for handling any volume of spam for small and medium businesses, enterprises and service providers.

This paper first examines the importance of including human analysis in the spam filtering process. It then discusses some ways it can be done. Next, it describes how Mail-Filters builds human analysis into its solutions. Finally, it recommends the most important question an IT manager must ask before buying a filter, and draws some conclusions.

10 Maguire Road
Lexington, MA 02421
(781) 676-5700

Zekeringstraat 17
1014 BM Amsterdam
The Netherlands

The Importance Of Human Analysis In The Spam Filtering Process

The need for human analysis when filtering spam cannot be overemphasized. It is the only way to catch most spam without generating high false positives. Formula based filters, without significant end user intervention, have false positive rates that are typically measured as one per hundreds of messages instead of one per million as they should be.

To see why formula based filters have a false positive problem, let us compare them to another detection problem: an automated airport metal detector. At an airport, if the detector is too sensitive, everyone gets stopped. If it is not sensitive enough, undesirable objects get through. The same applies to a spam filter.

The airport uses human intervention to handle this problem. The sensitivity of the detector can be set fairly high to catch as much metal as possible. An inspector provides the human analysis by using a metal-detecting wand to detect false positives - innocent people who set off the alarm with hip replacements, for example.

It follows that just as there are some objects that technology thinks it can detect, but only a human can recognize for sure, there is some spam that can only be recognized by a human. What did the US Supreme Court justice say? "I can't define pornography; I just know it when I see it."

Some Ways to include Human Analysis in The Spam Filtering Process

If spam filters must rely on some level of human analysis, "to know spam when they see it", how should that human analysis be achieved? There are typically two ways: have the user provide the human analysis, as formula based filters do, or build the human analysis into the filter. Let's consider the less desirable way first.

Have the user provide the human analysis

This is rather like the old quality control approach where the buyer provided the final inspection on an automobile. It took the load off the manufacturer, but created buyers' remorse. When a filter requires user intervention, those users can provide it in one or more of the following ways:

1. Set the filter to be fairly lenient so that some spam gets through, and then delete what they don't want.
2. Set the filter to be rigorous, so that it misidentifies some legitimate e-mail, and then search through the folder where the suspected spam has been diverted to retrieve the false positives.
3. Select the type of filter where they have to learn enough about spam and spammers to write rules and maintain the filter so that it can recognize the evolving spam.

The problem with any of these three user interventions is that it requires more work of the user than should be necessary. Deleting uncaught spam or retrieving misidentified good messages places a load on the user that will only increase as the number of spam messages increase and spammers become more sophisticated. As for teaching the filter what is spam, only a devoted anti-spam user is willing to treat the anti-spam filter like a temperamental car to be worked on constantly to keep it running. To most users the spam filter is a tool, not a hobby. This problem of overloading the user leads to the second more desirable option.

Build human analysis into the filter

Building human analysis into the filter not only reduces the work a user has to do; it actually improves the filters' ability to catch spam while avoiding false positives. It does require more work on the part of the filter supplier, as we shall see. This is why formula based filters are attractive to developers. But Mail-Filters designed its filters with human analysis built in, which is why those filters perform as accurately as they do.

Human Analysis Built Into the Mail-Filters Solutions

Mail-Filters has analyzed billions of spam messages. Those messages came in many different forms and those forms changed constantly as the spammers come up with new ways to defeat spam-filtering techniques. It became obvious that a formula could not be effective without considerable effort on the part of the end user. These observations led Mail-Filters to the development of a dual technology approach. The

10 Maguire Road
Lexington, MA 02421
(781) 676-5700

Zekeringstraat 17
1014 BM Amsterdam
The Netherlands

first technology is a database to identify spam, and the second is a scanner to identify spammer tricks. Human analysis is built into both.

The Mail-Filters Bullet Signatures Database is comprised of Bullet Signatures. These are small, targeted, and lethal spam signatures handcrafted by human editors and based on key characteristics of the message that are often repeated from one spam message to another. Bullets are not checksums, which are easily defeated by today's spammers, and are not based on the whole message or single key words.

Bullet Signatures are continuously updated by the Mail-Filters editors, usually within minutes of new spam hitting the Internet, to maintain their effectiveness and accuracy. Just the latest updates are transmitted to minimize download time.

But just updating the database is not enough. Spammers are ingenious. They use many tricks to fool a spam filter. This is where the Mail-Filters StarEngine comes in.

The Mail-Filters StarEngine (Spammer Tricks Analysis and Response engine) is designed to catch tricks identified by the Mail-Filters editors. The StarEngine has counter measures that look for tricks such as falsified information in the headers and other places in the message plus other unique identifying characteristics of spam. Tricks handled by the StarEngine include:

Hash Busting - The random insertion of characters or words intended by spammers to fool signature based filters.

Snow-Flaking - An effort to make each HTML email unique, like a snowflake, by inserting invisible characters or HTML comments into messages.

Header Forging - Falsifying the header information on which many filters rely to identify spam

IP Hopping - Constantly changing the IP information to confuse filters based solely on lists

Misspelling - Deliberately misspelling words such as $ex, or V1agra

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

HTML Email - Often used by spammers to display graphics and increase response rate. HTML email is difficult for many filters to scan.

These tricks can be effective when used on certain filters, particularly those without built-in human analysis. A comparison of how vulnerable certain types of filters are to spammer's tricks helps explain why spam-catching performance varies among filters.

A Comparison of Filters with and Without Human Analysis

To understand spammers' tricks and the kind of filters they are effective against Mail-Filters has created the "Spam Filter Vulnerability Chart" below. It shows the seven tricks described above and the major types of filters that are vulnerable to those tricks. Is it self serving to say the Mail-Filters StarEngine with

Spammer tricks	Types of spam filters				
	Signature –based (with hash or check sum values)	Formulas (with weighted factors)	Bayesian	RBL	Mail-Filters StarEngine (with Bullet Signatures)
Hash busting	Vulnerable				Not vulnerable
Snow flaking	Vulnerable				Not vulnerable
Header forging		Vulnerable		Vulnerable	Not vulnerable
IP hopping				Vulnerable	Not vulnerable
Misspelling		Vulnerable	Vulnerable		Not vulnerable
Embedded content	Vulnerable	Vulnerable	Vulnerable		Not vulnerable
HTML	Vulnerable	Vulnerable	Vulnerable		Not vulnerable

10 Maguire Road
Lexington, MA 02421
(781) 676-5700

Zekeringstraat 17
1014 BM Amsterdam
The Netherlands

Bullet Signatures is not vulnerable to any of these tricks? Yes it is, but it is also true. It is the reason our filters constantly outperform other filters as demonstrated by objective tests conducted by companies who have chosen to partner with us.

An Additional Vulnerability

In addition to being vulnerable to spam messages that use the tricks described above, all the filter types, except the Mail-Filters solutions, are vulnerable to foreign spam. Engineered from the start to catch foreign spam, the Bullet Signature database catches spam in symbol-based languages such as Chinese and Japanese.

Given how vulnerable filters are to spammers' tricks, it pays a filter buyer to ask a certain question before making an investment.

The Question A Filter Buyer Should Ask

In spite of the importance of human analysis in filtering spam, the most important question is not "how does your filter provide for human analysis?" While critical to filter performance this analysis is only a means to an end. The important question is actually, "How much spam does your filter catch and how many false positives does it generate while doing it. Don't accept less than 95% spam caught while generating less than 1 in 1,000,000 false positives AT THE SAME TIME. Note: Many filters have adjustable sensitivity. They increase spam catching, but at the expense of more false positives.

If a filter supplier other than Mail-Filters answers this question to your satisfaction, then go ahead and ask the human intervention question. The reason: we are convinced that the only way other filters can catch high levels of spam without generating high levels of false positives is by requiring considerable human analysis on the part of administrators and end users. We do not think that is acceptable.

Conclusion

Until spamming becomes uneconomical spammers will continue to do it. They will find it worthwhile to use time and ingenuity to trick filters into letting their messages through, knowing someone will fall for their sales pitch.

Human analysis is vital to accurately identifying spam, and that intervention should be built into the filter; it should not occur after the filtering is done.

10 Maguire Road
Lexington, MA 02421
(781) 676-5700

Zekeringstraat 17
1014 BM Amsterdam
The Netherlands