Understanding SPAM - Terminology, approach, and best practises

Article ID	176356
Last Modified	10 November 2014
Categories	
Products	MailGate All Versions;

��? Spam represents both an enormous administrative nightmare and a costly productivity drain for virtually every enterprise, and IT organizations are under intense pressure from everyone from the CEO on down to do something about it. And for good reason ? without robust inbound spam protection, it is estimated that nearly 80% of email received by a business user is spam. Multiply that figure by the number of employees in your organization, and it?s clear that spam is much more than a nuisance ? it?s a serious threat to your fiscal well being. And a significant proportion of this email spam includes pornography, offensive, or hate-based content that enterprises cannot allow into the organization without risk of creating a hostile workplace and facing costly litigation.

MailGate?s layered spam filtering effectively blocks more than 99 percent of all inbound spam ? with virtually zero false positives ? so users don?t have to wade through oceans of unwanted email to get to legitimate business messages. And because MailGate filters and blocks messages at the gateway, it can dramatically reduce the load on corporate email servers, improving network performance and overall enterprise security.

This technote provides an introduction to understanding spam capture within MailGate. It provides a definition of the various terms used in spam capture, a discussion of the practical considerations of catching spam, and some specific recommendations for configuring MailGate. It should also serve as a reference document to help answering specific questions raised by end users whose e-mail is protected by MailGate.

WHAT IS SPAM?

Spam is commonly defined as either unsolicited bulk e-mail, or unsolicited commercial bulk e-mail. In simple terms it is e-mail which is sent out to multiple recipients who did not ask for it, and typically have no way of stopping the sender from sending it to them. Usually it is trying to sell them something or scam them in some way. Sometimes it has no direct commercial motive, but may be trying to ?sell? a political or religious idea.

Within MailGate, Spam is also referred to as Junk e-mail.

WHAT ISN?T SPAM?

Some messages, while offensive, are not spam. For example, an e-mail sent from one individual to another which contains rude and offensive language will not necessarily be marked as spam by MailGate. Administrators can set up separate policies to filter for offensive words and content in non-spam messages.

WHAT IS BROADCAST OR BULK E-MAIL?

Axway classifies a second type of e-mail in addition to spam. This is bulk or broadcast e-mail. For simplicity we will refer to this as broadcast e-mail for the remains of this note. Broadcast e-mail differs from spam in that it is in some way solicited. It is likely to be something that the recipient has directly or indirectly signed up for, even though they may not remember doing so. It may be marketing messages from legitimate companies who they have previously done business with, e-mail updates from news organizations, or other regular subscription e-mail messages. In broad terms, a good way to think about broadcast e-mail is that it should be e-mail for which the unsubscribe process will work ? i.e. you can take action to stop receiving it.

WHAT IS A REGULAR E-MAIL?

A regular e-mail is anything that is not a spam or broadcast e-mail ? i.e. the normal person-to-person e-mail traffic that people exchange in the working day.

WHAT IS A FALSE POSITIVE?

A false positive is a regular e-mail message that has been incorrectly classified as broadcast or spam. The avoidance and elimination of false positive is a critical part of any spam filtering technique.

CONFIDENCE LEVELS

Each message which is classified by the Axway Spam Analysis Engine (SAE) will be assigned a confidence level. This is an assessment of the confidence the engine has in the spam or broadcast classification it has given to a particular message.

The confidence levels are:

High ? message is almost certainly classified correctly as spam or broadcast, and highly unlikely to be regular e-mail
Medium ? message is very likely to be spam or broadcast, and very unlikely to be regular e-mail
Low ? message is probably spam or broadcast, but a small amount may be incorrectly classified regular or broadcast e-mail

The Spam Analysis Engine builds up the confidence level by compiling information from a broad range of heuristic and complex analysis of a messages structure and contents. The ratings from the various heuristics and analyses are summed together to form an overall score, and **high**, **medium** or **low** confidence level is assigned depending on whether this score reaches certain threshold values.

CLASSIFICATIONS

Messages classified as spam are further broken down into sub-categories depending on the nature of the content. MailGate separates Broadcast (or Bulk) at the highest level.

The following spam classes are specified:

Adult ? Pornographic, sexual, or shocking in content, or advertises products, services, or Web sites that are pornographic or adult-oriented (even if the message itself contains no obscene language, pornographic images, etc.)
Scam ? E-mail attempting to scam, con or mislead the recipient. Messages receiving this classification include phishing, advanced-fee fraud messages, and other attempts at identity theft, direct or indirect money extraction.

All spam messages that do not fall in any of the categories above will not receive a specific classification tag within the Spam Analysis Engine, and therefore can be identified by the fact that they lack any of the other classification tags.

AXWAY SPAM ANALYSIS ENGINE

�� The Axway Spam Analysis Engine uses a ?cocktail? of advanced technologies to block spam, fraud and phishing messages, with a focus on minimizing false positives. Its four main components are the Dynamic Anti-Spam (DAS), Recurrent Pattern Detection (RPD), and Adaptive Image Filtering (AIF) sub-engines. Complementary techniques include content analysis, rule-based heuristics, and outbreak detection. With the integrated Cyren Recurrent Pattern Detection? (RPD) system, the engine can even block spam outbreaks coming from large numbers of unknown ?zombie? machines, which are not blocked by conventional reputation-based systems. RPD technology is also used to analyze real-time outbreak information while ignoring individual message content, allowing spam trends to be pro-actively detected and stopped at the gateway.

All of this is backed up by Axway?s Message Protection Lab (MPL): using a combination of detailed human review and artificial intelligence-enhanced techniques to analyze more than 400 million messages every day and to identify new spam trends and tactics, the lab tracks threats round-the-clock from worldwide sources in more than 10 different languages. Axway experts then quickly create new proactive and reactive measures to combat them, including meticulous content, image, and attachment analysis, and rules-based heuristics. Similarly to the way antivirus engines work, automated DAS updates are published to Axway MailGate Anti-Spam filters as often as every half an hour, 24x7. And to minimize false positives in a global business environment, the Lab analyzes both spam and legitimate email gathered internationally and provided by enterprise customers.

HOW THE AXWAY SPAM ANALYSIS ENGINE WORKS

Each of the technologies and heuristics used can contribute to the confidence level of high, medium or low that the Spam Analysis Engine assigns to a particular message. Broadly each heuristic contributes either in a positive or negative way to the overall confidence a message is spam, depending on how strong an indicator that heuristic is that a message is or is not spam based on the millions of messages processed by the lab. Considering the main principles of spam (to send as many relatively small messages as quickly as possible) and to boost performance, the engine filters messages by size and scans for spam messages that are up to 200 - 500 KB in size (configurable, according to the specifics of each organization).

Because of the constantly changing nature of spam and the techniques used to combat it, Axway do not publish details of the individual heuristics used and how they contribute to the confidence rating. In addition, Axway does not make this information public because of its potential usefulness to spammers wishing to subvert the Spam Analysis Engine.

THE AXWAY SPAM HEADER

To each message, scanned by the Spam Analysis Engine, a custom spam header is appended. The X-TMWD-Spam-Summary header gives explicit information about the engine that has categorized the message and the category/confidence given that will help Axway address false positives or false negatives when needed. The header components are message markers that help MPL re-classify the message.

Customers should mind that the intended usage of the X-TMWD-Spam-Summary header is troubleshooting and MPL re-classification only; Axway **strongly discourages for this header to be used in custom policies**. We may change the header formatting at any time and without any notice, in our effort to enhance the anti-spam service and any custom policies using this header will be affected.

CONNECTION MANAGEMENT

MailGate also includes techniques to identify machines sending spam at the connection level. This allows a significant percentage of undesirable e-mail to be dropped during the initial negotiation between the sending machine and MailGate, meaning it never needs to be accepted and processed. While this is not the focus of this note, some of the techniques that are available here are:

? **DNS Block List (DNSBL)** ? uses third-party databases of known spam sources to reject connections. (part of the Edge defense license)
? **Manual IP address blocking** ? rejects e-mail from a particular machine known to send spam. (part of the core license)
? **Recipient Verification at the Relay** ? rejects mail to unknown recipients in the initial negotiation. (part of the core license)
? **?Edge? protection** ? analyzing traffic patterns to detect Denial of Service, Directory Harvest or other malicious attacks. (part of the Edge defense license)
? **IP Reputation** - Cyren IP Reputation service correlates, identifies, classifies and tags at the source spam, malware or phishing outbreaks across multiple IP addresses, even if they have been recently hijacked and have no existing negative reputation. (part of the IP Reputation license)
? **DKIM** - DomainKeys Identified Mail (DKIM) is an email validation system designed to detect email spoofing by providing a mechanism to allow receiving mail exchangers to check that incoming mail from a domain is authorized by that domain's administrators. (part of the Edge defense license)
? **SPF** - Sender Policy Framework (SPF) is a simple email validation system designed to detect email spoofing by providing a mechanism to allow receiving mail

exchangers to check that incoming mail from a domain is being sent from a host authorized by that domain's administrators. (part of the Edge defense license)
? **BATV -** Bounce Address Tag Validation (BATV) is a method for determining whether the bounce address specified in an E-mail message is valid. It is designed to reject backscatter, that is, bounce messages to forged return addresses. (part of the Edge defense license)

CUSTOM ALLOW AND BLOCK LISTING

Every company or organization has partners with whom they exchange mail on a regular basis. Similarly, a lot of customers are often targeted by spammers or other malicious or offensive sources. MailGate allows the administrators to maintain custom allow and block lists and assigning custom spam or legitimate classification to a sender or a domain, or even configure MailGate to automatically classify individual senders as legitimate based on past mailing behavior. These features are the Domain, Sender and Auto classifications, respectively. In addition, end-users with access to the Personal Quarantine Manager can configure their own Allow & Block lists.

WHY DO WE HAVE A BROADCAST CLASSIFICATION?

It is not always easy for recipients to distinguish between what is unsolicited spam e-mail, and e-mail which is sent in bulk but has in some way been directly or indirectly solicited. Axway classifies e-mail that it believes to be legitimate broadcast e-mail so that administrators can tag this e-mail appropriately and let the recipients know that this is likely to be from a legitimate source. This helps avoid multiple questions and feedback to the administrators on why such e-mail has not been classified as spam or junk. Of course, should this classification be incorrect, either the recipient or administrator can report the e-mail for reclassification to Axway?s Message Protection Lab to the spamlab.fp@axway.com address.

WHY DOES SOME SPAM STILL GET THROUGH?

Sending spam e-mail is a highly profitable business and the spammers are constantly trying to figure out new ways of defeating spam filtering applications. While Axway is constantly evolving and developing our techniques for detecting and dealing with spam, the spammers are constantly trying to figure out how to get their messages through. Even the best filters such as Axway?s (which recently achieved a 99% capture rate with no false positives in independent testing) will not catch every spam message. Even with all of the reactive and pro-active techniques at our disposal, a new spam variation may be introduced which temporarily defeats our filters. There is always a balancing act between identifying the last few spam messages and the risk of incorrectly identifying a regular message as spam and generating false positive. It is simply not possible for any spam filter to catch every single spam message without introducing an unacceptable level of false positives.

WHY DO WE GET FALSE POSITIVES?

In the same way that it is impossible to capture 100% of all spam messages, it is also not possible to totally eliminate false positives and still provide effective spam filtering. At Axway we do everything we can to avoid classifying regular messages as spam or broadcast, and we are constantly tuning our filters to ensure we catch spam and only spam. Our analysts work to identify what is unique about particular spam and why it is different from regular e-mail, and craft very precise heuristics to distinguish between the two. New heuristics will not be introduced unless they can be encoded in this way, and we will always take the decision to miss a small amount of spam rather than raise the risk of false positives. Considering also that virtually every regular message is different, we cannot totally eliminate false positives. Despite having the lowest false positive rate achieved in independent testing, on very rare occasions a regular e-mail will have enough in common with a spam message to incorrectly trigger the spam filters. We cannot totally remove this risk, but the risks can be mitigated using the end-user reporting feature on MailGate. This capability allows individual users to receive a summary of the messages sent to them which have been identified and stopped as spam, and release any messages which they would like to receive. Axway also provides a feedback mechanism by which false positives can be sent to Axway so the filters can be updated to avoid similar false positives going forward.

��? As discussed in the previous section, though we may get very close, we will never achieve 100% accuracy. Some small amount of spam will slip through the net. There will also be very rare occasions when messages are misclassified. This section gives details of how to submit missed and misclassified messages to Axway so that classifications are corrected in the spam filters. We rely on this feedback to help us constantly improve our filters. Please do not use the e-mail addresses below to submit questions or queries relating to spam messages. Any questions or queries should be submitted to support@axway.com.

HOW TO REPORT MISSED SPAM

Any spam messages which managed to get through Axway?s filters should be submitted to spamlab.miss@axway.com. Please create a new e-mail message to spamlab.miss@axway.com and add any original spam messages as attachments. In this way, all of the information about the spam message is preserved. Simply forwarding messages directly to spamlab.miss@axway.com without attaching them to a new message removes some of the header information and makes it much harder for our lab to analyze the messages.

HOW TO REPORT FALSE POSITIVES

False positives can be submitted to Axway by sending them to spamlab.fp@axway.com. As with missed spam messages, it is preferable to create a new message and add the false positive message as an attachment rather than forwarding it directly. MailGate allows administrator to submit false positives directly to the lab from the administration interface, and it also allows end-user to submit them directly to the lab from the end-user interface. Note that e-mails incorrectly classified as broadcast should also be submitted to the spamlab.fp@axway.com address.

GETTING FEEDBACK ON SUBMISSIONS

All messages reported to Axway?s Message Protection Lab are checked against the latest filters and the heuristics are updated to correct the misclassification or to catch a new spam trend. Administrators can sign up to daily or weekly reports to receive summarized feedback of how messages submitted from their domain have been handled and classified by the lab. If you want to subscribe to this report, please contact Axway Global Support at support@axway.com.

WILL USERS BE ALLOWED TO RELEASE MESSAGES?

The first consideration when setting up policies on MailGate is whether users will have permission to release their own spam from quarantine. The product can be configured to send recipients a regular summary of those messages which have been caught as spam, and allow them to release some or all of these messages without the need for intervention by an administrator. If you chose to enable this feature, you can use more aggressive policies in quarantining spam, since recipients will be able to see what has been caught and release messages which have been caught in error.

WHAT TO DO WITH BROADCAST MESSAGES?

It is generally recommended that Broadcast messages are let through to the recipients, and appropriately tagged by MailGate. There are three reasons for this recommendation:

1. Many users wish to receive broadcast message, and blocking messages such as subscribed news updates can lead to significant administrative overhead in dealing with requests to release and modify filters to allow these e-mails through.

2. Broadcast e-mail can often be confused with spam by recipients, particularly when they either forget they have signed up for them, or the relationship is indirect via a partner. Clearly marking broadcast e-mail helps recipient?s understand the difference between spam and broadcast, and let?s user take the appropriate action to remove themselves from the distribution.

3. Broadcast messages are the most similar to regular business traffic, and therefore this is where any false positives are likely to occur.

Axway recommends adding an annotation similar to that below to the front of each Bulk or Broadcast message.

HOW AGGRESSIVE TO BE WITH SPAM?

MailGate allows administrators to control how aggressively the spam policies are applied. The product also allows control of which messages appear in user reports when the Personal Quarantine feature is used. Each administrator can set the system up in the most appropriate way for their organization, but the following settings have been shown to provide a good basic set for most users.

RECOMMENDATION 1: BE AGGRESSIVE OR VERY AGGRESSIVE WITH ADULT MESSAGES, AND DON?T INCLUDE THEM IN END USER REPORTS

The chance of a regular e-mail being classified as adult is extremely low. You can therefore afford to be aggressive or very aggressive in catching them, and do not need to make these messages available for end-user release. These e-mails can be extremely distasteful and offensive, and most organizations do not want the recipients to be aware of their existence or provide capability for users to release them deliberately or accidentally.

RECOMMENDATION 2: BE AGGRESSIVE WITH SCAM MESSAGES. CHOSE WHETHER TO INCLUDE THEM IN THE USER REPORT, DEPENDING ON THE NATURE OF YOUR BUSINESS

Scam messages are very rarely classified as false positives, and an aggressive approach can be taken with them. Since they are deliberately intended to mislead the recipient, it can be counter productive to include such messages in end user reports because the recipients could release them from quarantine in the belief that they are legitimate and then become of a victim of the scam. It is therefore not normally recommended that these messages are not incorporated into end user reports. However, where the individual recipient or organization is in regular recipient of critical communications from financial institutions, it may be desirable to include such messages in end user reports so that this critical communication is not lost in the rate case of a false positive.

RECOMMENDATION 3: BE AGGRESSIVE WITH UNCLASSIFIED SPAM MESSAGES, AND INCLUDE THEM IN END-USER REPORTS IF YOU USE THEM

There are a huge number of different types of spam messages, and only a subset of these will be classified as Adult or Scam messages. The remainder will pass through MailGate without a classification, and can be considered as general or unclassified spam. As with the other categories, Axway makes every effort to avoid false positives in this type of spam, and the number of false positives should be extremely low. However, due to the general and broad nature of unclassified spam, it is in this category that false positives are most likely to occur, and for this reason it is recommended that these messages are included in end user reports if you are using them.