MPmail Spam Filter Service Data Sheet

What is Spam?

Wikipedia defines spam as "the use of electronic messaging systems to send unsolicited bulk messages, especially advertising, indiscriminately". This definition reflects the experience of most people who have been bothered by spam. Unfortunately, technical application of the definition is difficult:

- How does a system know that an email is unsolicited, and that the recipient has not in fact given permission for it to be sent?
- Unsolicited does not necessarily mean unwanted - it is quite possible for an unsolicited email to be welcome, e.g. an enquiry from a new customer.
- What exactly is "advertising"?

Because of this, it is unsurprising that different people have different opinions as to whether a Particular email is legitimate or spam. In fact, it is not uncommon for one person to class a message as spam on one occasion and as a legitimate email on another. So how can a technical system deliver the required result with a high degree of reliability?

How is Spam Detected?

Due to the difficulty of technically implementing the correct definition of spam, many spam filter manufacturers use another definition, the "mass mail criterion". According to this, an email is classed as spam if it is sent to a large number of different addressees within a short space of time, possibly also by various senders. Unfortunately, this also includes certain types of legitimate email such as newsletters. On the other hand, individual or low volume emails with clearly abusive content are not detected via the mass mail criterion.

In light of the different definitions of spam, details of detection rates and false classification rates are difficult to compare. At most, they can be accurately predicted on an individual basis, i.e. for each individual user. For this reason, every user must ultimately decide for themselves which emails they want to receive and which they do not, and transmit this information to the spam filter. The spam filter must then reproduce this decision in the best possible way. However, a spam filter must also produce good results for the user even in its basic configuration, otherwise the effort required to 'train' the filter would be shifted to individual users, which defeats the point of an automatic system.

As spam is constantly changing, ongoing filter updates are also very important; the best filter is useless if it is supplied with outdated information. Spam often comes in waves. These spam waves typically last between a few minutes and a couple of hours. A filter update after the wave has finished would have no effect, and a large volume of unwanted emails would pass through and reach users' inboxes.

A fast and tiered response is therefore important in the event of a spam wave. First, automatic measures take effect, and then in a second step, new rules that are specially designed for the new

type of spam are created and activated. As spam messages also change constantly during a wave, often this process must be repeated multiple times in the course of a wave.

Infomail

The majority of spam emails are very clearly identifiable as spam, but there is a grey area where it is not easy to determine whether the recipient would want to receive an email or not. This grey area covers many newsletters, for instance. Newsletters sent by email are legitimate provided that legal requirements are met. In these cases, it cannot be assumed that the emails are unsolicited. The same is true for initial contact by email; although in such cases the email received is unsolicited, it cannot automatically be assumed that it is undesirable.

Manage Protect defines "infomail" as emails that users do not consider undesirable but which disrupt the day-to-day workflow and distract from more important emails. Separate info mail handling is therefore a useful and cost saving feature of the Manage Protect spam filter.

Cloud Architecture

Manage Protect uses the principle of the public cloud to protect its customers' systems and networks against attacks from the Internet. To do this, MPmail systems are set up outside the customer's internal network, like the outworks of a fortress. MPmail forms the gateway for conveying emails from the Internet into the customer's internal network. This offers significant benefits:

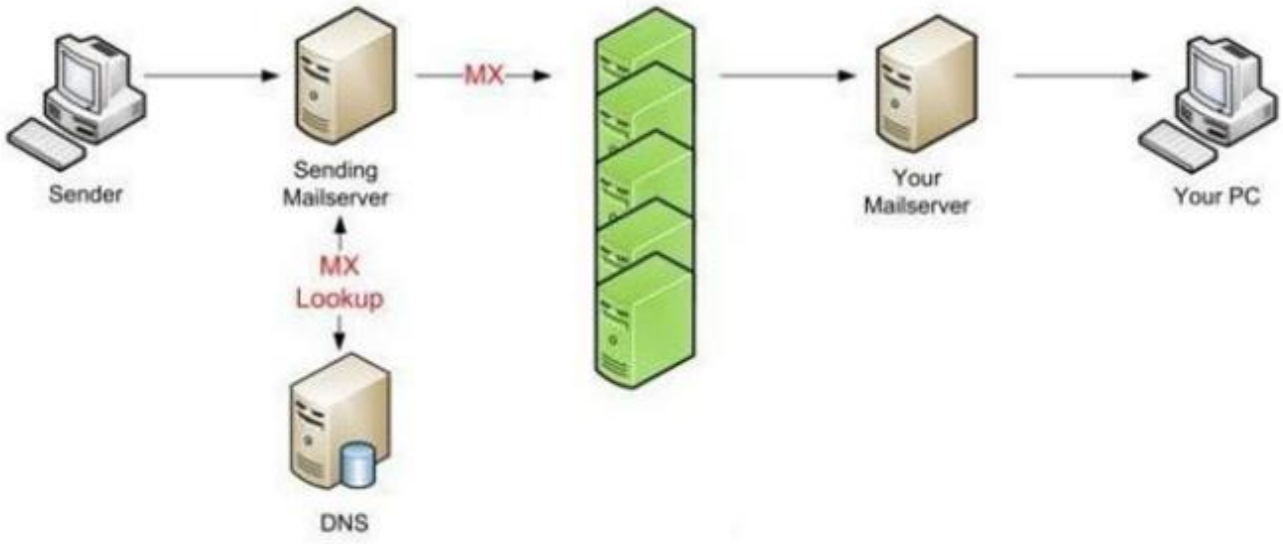
- Extremely redundant and distributed overall system layout, making it more powerful and more resilient against attacks than individual installations
- Defense against attacks malware, etc. outside the customer's infrastructure, reducing strain and threats to the infrastructure
- Simplification of protection systems at the perimeter of the customer's network, making them more resilient and cost effective, and significantly reducing the administrative load
- 24/7 system monitoring by security experts, ensuring security even outside the working hours of the customer's own IT staff.

Manage Protect solutions offer full multi-client capability, and have tiered user permissions. Data is stored hierarchically, ensuring that users are only able to access the relevant hierarchy level at all times. Manage Protect systems are run from several distributed and secure data centres.

Incoming Mail

Manage Protect systems are set up as a firewall between the Internet and the customer's IT infrastructure. For incoming email traffic, it is therefore important to ensure that emails sent to the customer from the Internet are routed via Manage Protect systems and not directly to the customer's infrastructure.

On the internet, the mail exchanger record (MX record) in the domain name service (DNS) specifies which email server is responsible for receiving a domain's emails. To reroute the incoming email traffic, the MX record for the customer's domain is changed so that it points to the Manage Protect mail gateway. Emails to this domain are then automatically transmitted without further intervention from the sender's mail server to the Manage Protect infrastructure, where they are checked and forwarded to the customer's email server.



The customer's firewall is also set up so that only emails from Manage Protect networks pass through the firewall. This is necessary as malicious emails could otherwise be sent directly to the mail server, bypassing the MX record and therefore bypassing Manage Protect protection. After changing the firewall, any direct delivery attempt will fail, and the customer's systems are definitively safe from spam.

The new setting also ensures that denial of service (DoS) attacks on the customer's mail infrastructure are considerably more difficult. The aim of these attacks is to completely prevent network traffic by flooding lines and systems with extremely high numbers of data packets. If these attacks do not target the Manage Protect infrastructure but the customer's infrastructure by bypassing the MX record, they are now easily blocked by the firewall and not allowed through to the mail server.

Processing Steps

By altering the MX record, wanted and unwanted emails to the customer's domain are sent to the Manage Protect gateway systems. There, emails are processed in two stages: Preliminary gateway analysis (blocking) and active analysis with final virus check.

Preliminary Gateway Analysis

The primary goal in the preliminary analysis stage is to detect known senders of unwanted emails (spam) and malicious content before the message is finally transmitted. Detection should take place in this very early stage if possible so as not to use unnecessary bandwidth or processing resources.

While the connection is established by the forwarding mail server, information about the sending server, such as its unique Internet IP address, is received and analysed by the Manage Protect systems. Using this information and extensive, carefully maintained databases, Manage Protect can detect definitely known spammers and prevent transmission. This process is frequently referred to as IP black listing. Similar mechanisms are used to monitor the number of connection attempts by the sending server and to limit them if necessary, thus providing protection against denial of service or email flooding.

If analysis during connection does not produce a negative result, further examination of the data will begin. Checks during transmission of header information establish whether this is an attempt to deliver unwanted emails or even to transmit computer viruses, worms and links to hijacked and infected websites.

The following detection mechanisms are used during Preliminary Gateway Analysis:

- Monitoring the number of connection attempts
- Blocking specific IP addresses that are definitely known to be spammers
- Checking the existence of the recipient (SMTP pass-through check)
- Checking header and subject information
- Compliance with email traffic transmission standards
- Detecting clear spam patterns

The sender does not receive positive acknowledgement that the email has been received unless all stages of the preliminary analysis in the Manage Protect spam filter are successfully completed without detecting known spam, computer viruses or other unwanted or potentially harmful content. If the email is classified as spam during the preliminary analysis, the connection is terminated and an error message returned instead of an acknowledgement.

A special feature is the SMTP pass-through check. This feature can be adjusted for a domain by the administrator, and ensures that the destination server is checked to determine whether the email can be delivered before it is accepted by the Manage Protect gateway. This ensures at an early stage that emails to non-existent addresses or recipients with full inboxes for example are not transmitted. In the event of an error message from the destination server, this exact error message is forwarded to the sending server and the email is rejected.

Statistically, 95% of spam delivery attempts are currently detected and rejected during preliminary gateway analysis (as of October 2012). Nonetheless, the preliminary analysis is deliberately set up to allow emails through in case of doubt. The primary goal here is to avoid false positives, i.e. incorrect classification - and therefore rejection - of legitimate emails as spam. If an email is incorrectly rejected during preliminary analysis, which is statistically extremely rare, the content of the error message is used by the sending mail server to inform the sender of the rejection. The sender can then contact the Manage Protect support team via a neutral website, http://www.cloud-security.net/ and inform them of the incorrect rejection. The Manage Protect support team will immediately check the message and take the necessary steps to prevent this from being repeated. The sender will be informed of the measures taken.

Administrators and users also have a complete overview of all accepted and rejected emails at all times in the Manage Protect control panel. Details such as the sender and recipient email addresses, date and time, subject, IP address of the sending server and the status message of the receiving mail server are displayed and used as search criteria. Because the message ID is also displayed, this can also help when searching for a particular message on the mail server.

Active Analysis

If an email has passed the preliminary analysis processes, the sending mail server will receive confirmation of successful transmission. By contrast, in the event of a negative preliminary analysis, acknowledgement of receipt is denied and an appropriate error message is transmitted to the sending server instead.

Users of the Manage Protect spam filter service can control the next processes. Selective white listing allows email receipt to be specifically and individually permitted for particular domains and recipients, while blacklisting similarly allows emails from particular senders to be marked as spam.

Very powerful further options for handling messages are available by entering rules in the Manage Protect spam filter compliance module.

Various aspects of the received email are checked during active analysis. By using highly specialised algorithms developed by Manage Protect, email content is systematically analysed depending on the content. Content is evaluated structurally, syntactically, semantically and heuristically. As well as text content, image information and text in images are also analysed. To do this, real time optical character recognition (OCR) is also used if necessary to detect spam messages sent as images. External content or links to websites are compared with known spam patterns.

In addition to this, all emails undergo a final computer virus check before delivery to the recipient. Known and detected viruses are completely blocked and not delivered to the user. New and previously unknown viruses are detected by the early warning system using outbreak detection. Emails are also checked for phishing links and URLs to malicious code or websites that pose a security risk.

The following detection mechanisms are used in this stage:

- Heuristic filter
- Bayesian filter
- Content analysis
- Optical Character Recognition (OCR)
- Domain and user-specific black/white lists
- Virus check

The Manage Protect content filter works donwnstream of spam detection. During this stage, emails with undesirable attachments are blocked or have their attachments removed according to adjustable criteria.

The email is not delivered to the recipient until it has passed the checks in all of the processes. If an email is identified as unwanted during one of these checks, the email is logged and moved to quarantine, and the recipient is informed via a status email. The status email frequency is configurable. Access to emails in quarantine is also possible via the Manage Protect control panel.

Using comprehensive, specialist content-dependent algorithms, Manage Protect can contractually guarantee a spam detection rate of 99.9% for the spam filter service. A rate of over 99.99% is consistently achieved in practice. The contractually guaranteed false positive rate is below 0.0004%.

Infomail Handling

Emails that users do not generally consider undesirable but which disrupt the day-to-day workflow and distract from more important emails, e.g. user-requested newsletters, are classed as “info mail” by Manage Protect. Because these messages are not spam in the strict sense, they are handled separately.

Administrators can activate the infomail filter for their entire domain in the control panel. It is also possible to configure whether users can activate or deactivate info mail handling themselves, regardless of the domain settings. When the infomail filter is activated, infomails and newsletters are filtered out and placed in quarantine like spam emails. These emails are separately marked as

"infomail" in both the spam report and control panel. This provides users with an overview of the infomails they receive, and allows them to deliver individual infomails with a click.

Email Buffers and Caches

If the SMTP pass-through check is activated and the email server is temporarily not accepting incoming emails, Manage Protect will forward this information to the sending mail server in incoming email traffic. Normally, the sending mail server will then take over buffering. If this is not the case, or if the buffer time in the sending mail server is too short, the email will normally be reported to the sender as undeliverable by the sending mail server.

If an SMTP pass-through check is not activated, Manage Protect will buffer incoming emails for seven days as standard if the recipient email server is unreachable or if email delivery is temporarily impossible for other reasons.

For technical reasons and to improve filter effectiveness, Manage Protect does not recommend deactivating the SMTP pass-through check without good reason. However, in the event of long downtime for a mail server with SMTP pass-through check activated, Manage Protect recommends that customers contact Manage Protect support. They can temporarily deactivate the SMTP pass-through and activate the buffer.

The contents of cached emails are deleted from Manage Protect systems after complete transmission, unless the customer uses the continuity service or archive service.

Virus Protection

For incoming email traffic, all emails undergo a final computer virus check before delivery to the recipient. The check consists of a series of mechanisms. The following processes are used to protect against viruses:

- Modified ClamAV Scanner
  - The signature-based ClamAV is used for known viruses. Manage Protect modifies and expands the available virus signatures to achieve their required standards for detection capability, speed, and false positive prevention. Signatures are updated at least every half hour
- Hornet Security Virus Scanner
  - The Hornet Security virus scanner uses signatures that are specially developed and optimised for viruses spread by email. Signature updates every 60 seconds allow very fast, more flexible and more comprehensive detection than with a generic scanner such as ClamAV
- Hornet Security Phishing Filter
  - The phishing filter analyses links/URLs in emails to detect downloadable malicious code. For this purpose, the filter detects downloadable malicious script commands, etc. This allows phishing emails and malicious drive-by downloads to be detected
- Hornet Security Outbreak Engine
  - MPmail constantly analyses incoming emails for unusual attachments, links, senders or content using honeypot accounts (email addresses with only one purpose, receiving spam). Signatures are derived from this analysis within an extremely short time (typically < 5 minutes)
- G Data Signature Exchange Data

- Manage Protect works with G DATA to detect and successfully filter out new viruses and phishing variants at an early stage. To do this, data with potentially malicious code are continuously exchanged, analysed and the results of the analysis used to detect subsequent transmission of the malicious code

Archives are scanned for viruses up to the eights level, this means that viruses are also found in archives that are in turn stored in archives. Archives with more than eight levels are treated as virus infected emails. Password protected (encrypted) archives are the exception to this. These are not scanned for viruses.

All emails that are identified as malicious by the mechanisms above are marked and treated as virus infected emails in quarantine. Only administrators, and not users, can deliver emails after they are classified as virus infected emails in quarantine.

Manage Protect guarantees an email traffic virus detection rate of 99.99% in relation to the total number of incoming transmission attempts.

Outbound Email Security

The Manage Protect relay service is included in the Manage Protect spam filter service. The relay service increases security in outgoing email traffic. To use the relay service, the Manage Protect mail relay is entered as a smart host or relay on the customer's mail server. Outgoing emails are then sent by the customer's mail server to the Manage Protect mail relay, where they are checked and forwarded to the destination address on the Internet.

The following checks are carried out in the relay:

- Check for known viruses
- Check for compliance with email policies regarding attachments
- Basic check for spam distribution
- Check for unusual email traffic such as active bots in the customer network

It is also sensible to close the firewall for outgoing email traffic, so that emails from the internal network can only be sent to Manage Protect relays. This will make it considerably more difficult for any potentially active bots in the internal network to send out spam.

Sending outgoing emails via the Manage Protect relays also activates a further feature, Manage Protect bounce management. Bounce emails are messages generated automatically by mail servers to provide information about a user's absence or an email delivery delay, for example. Genuine bounces are desirable; the problem is bounces that are received by a user due to fake sender addresses.

Spammers frequently use fake sender addresses to send their messages. For example, if spam is addressed to non-existent email addresses, this frequently creates bounce emails that are sent to the supposed sender's address. In extreme cases, bounce attacks generated in this way can have a serious adverse effect on the innocent supposed sender's mail systems.

Manage Protect bounce management ensures that only genuine bounces in incoming email traffic are sent to the recipient, and bounces in response to spam with fake sender addresses are reliably filtered out. All that is required for this is for email to be sent via the Manage Protect relay.

When sending via the Manage Protect relay, a group-based configurable email footer/disclaimer can also be attached, automatically, to outgoing emails. This can be used for marketing purposes, for instance, or to ensure that legally required details are included on business correspondence.

Use of the Manage Protect relay is also a requirement for revision-proof archiving of outgoing email if the Manage Protect archive service is used.

Content Filter (Attachment Filter)

Many companies have guidelines for sending or receiving email attachments. This is intended to limit the volume of data sent by email, and to prevent users from sending undesirable formats (e.g. only PDFs transmitted, but not Office files) or receiving data that poses a potential risk to system security (e.g. executable files).

The Manage Protect spam filter service allows guidelines for sending and receiving email attachments to be set specifically by user, group or domain. Emails with forbidden attachments in outgoing traffic are rejected. Emails with forbidden attachments in incoming traffic are either rejected or delivered without the forbidden attachment, depending on settings. In the latter case, the complete email is cached in quarantine. Emails that have been quarantined due to a forbidden attachment can be delivered in full to the recipient's inbox by administrators at the click of a button.