E-mail Content Scanning with Exim 4

Tim Jackson (tim@timj.co.uk)

Overview

- Introduction to content scanning
- Content scanning at the MTA – issues
- Methods of implementing content scanning
  - Accept-and-scan
  - SMTP-time
- Software required
- Brief overview of common software
- A look at Exiscan
- Other considerations
- Conclusions

Introduction

Basic Rationale & Considerations

- Defend against viruses, spam or other unwanted messages
- A final check for messages that have made it through other checks
- *Not* a substitute for DNSBLs and other non-content checks! False positives based on content typically occur more frequently than with checks not based on content (e.g. DNSBL lookups)
- Need to consider a variety of policy and "good practice" issues, not just technical ones

Inbuilt methods for content scanning

- SMTP time: Rules making decisions based on content (e.g. $message_body) in the DATA ACL
- Post-SMTP: Exim/Sieve filters (per-user or system filters)

Some common external content scanning software

- SpamAssassin – Free software spam checker
- Clam Antivirus – Free software virus scanner
- Sophos/sophie – commercial virus scanner/Free software daemon

Content scanning at the MTA

Advantages

- Centralised – easy to apply consistently across large user populations
- Transparent – no need for end users to install or configure software, or to learn about content scanning
- Easy to maintain – all in one place

Disadvantages

- Confuses the role of an MTA, which normally only *transports* mail rather than taking action based on its content
- Can be a blunt instrument, takes control away from users (although this cuts both ways – can be a good thing)
- Centralises burden on mail servers – resource usage needs to be considered.

Two primary ways to implement content scanning

- Accept-and-scan
- SMTP-time scanning

Method: Accept-and-scan

- Accept the message as normal, and process content later
- Internal – filters make decisions on message content
- External software - routers used to pass the message to external software - typically either fails the message and generates a bounce, or re-injects into Exim

Advantages

- Easy, does not require any extensions to Exim
- Easy to allow complete per-user configuration

Disadvantages

- Having accepted the message, what to do if it's detected as virus/spam?
  - Drop it silently – makes for an unreliable mail system. Not recommended.
  - Tag and/or deliver to a separate destination – OK, but means that the recipient still ultimately gets the spam/virus
  - Create a bounce – bad practice in the current climate where most senders are faked
    - You will end up “collateral spamming” innocent third parties – adding to the problem!
    - Even if the (faked) sender doesn't exist, you will add load to some innocent party's systems, and your queues will fill up with frozen bounces
- Typically involves scanning a message multiple times for multiple recipients

Method: SMTP-time scanning

Scan during the SMTP DATA phase

Advantages

- Elegant: if you're not going to accept a message, better to reject outright
- Reduces collateral spam (major consideration – best practice)
- No more queues filled with bounces

Disadvantages

- Requires enough resources to scan quickly & return back to SMTP session – risk of duplicates
- Stretches strict RFC compliance slightly – though shouldn't cause interoperability problems
- Per-user configuration options limited – content scanning only takes place once per *message*, not per *recipient*.

We focus on SMTP-time scanning.

Software Required

Need:

- Exim :-)
- Content-scanning patch to Exim – “glue” to pass the mail from Exim to the external software and return a result
- Scanning software (virus/spam scanners etc.)
  - Should be daemonised if possible for performance

Before diving in, consider the *policies* to be implemented as well as the *tools* to do it.

Content-scanning patches

- Exiscan – “Swiss army knife” - support for lots of external anti-spam/anti-virus tools including SpamAssassin, Sophos/sophie, Kaspersky, ClamAV, Brightmail, generic command line etc. Also has useful in-built MIME-based tools. Operates in the ACL system.
  http://duncanthrax.net/exiscan-acl/

- SA-Exim – single-purpose spam-scanning patch for SpamAssassin. Extensive & detailed functionality though increasingly most can be done with Exiscan. Includes 'greylisting', 'tarpitting' and more. Operates using the local_scan system and separate configuration file.
  http://marc.merlins.org/linux/exim/sa.html

- FFPA – current status unknown, extension to Exiscan which allows detection of attachments based on their actual file type, not just their file extensions.
  [No known website at present – contact Tony Sheen <tony.sheen@uk.mci.com>]

Scanning Software

Anti-virus

- ClamAV
  - Free software; in practice very good. Regular signature updates.
  - Daemonised; includes separate daemon to monitor for signature updates.
  - Some people have reported scalability and stability issues, though many use it successfully.
  - Variation called 'nclamd', supposedly more stable, will probably be merged eventually
- Sophos
  - Commercial software, with support.
  - Doesn't include a daemon, but free software 'sophie' daemon stable and works well
- Others
  - Kaspersky, ScannerDaemon etc.

Anti-spam

- SpamAssassin
  - Written in Perl
  - Primarily pattern-matching, but includes some other checks such as DNSBL lookups, Razor etc.
  - Works on a “points” system – each pattern (or rule) matched scores points (positive or negative)
  - The points scores allocated are normally used to flag or reject mails which exceed certain threshold scores (e.g flag at score=5, reject at score=10)

- Can be used to modify message content for detected spam, marking it as such and making the original mail an attachment. Works with SA-Exim, not with Exiscan.
- Includes Bayesian learning and analysis system
- Now widely used, so “good” spammers tailor their messages to avoid SA hits => maintaining your own custom rulesets and/or using “add ons” (many at http://www.rulesemporium.com/) is a good idea to maintain effectiveness.

- Others
  - Spamprobe, bogofilter, CRM114 etc. Not currently supported directly by any Exim patches mentioned

Exiscan

- Source code patch to Exim, maintained by Tom Kistner, normally released together with or shortly after Exim releases.
- Due to popularity, many Exim binary distributions/packages are pre-patched with Exiscan
- Elegant integration. Hooks into the Exim ACL system
- Provides several additions:
  - New options in DATA ACL to call external scanning software
  - Inbuilt MIME decoder
  - MIME checking to detect serious MIME errors (often indicative of malware)
  - File extension matching (e.g. to block all .pif files)
  - Regular expression matching of decoded or raw MIME parts
  - New ACL: acl_smtp_mime – called once per MIME part

Some brief Exiscan examples

- Too many possibilities to cover everything
- Comprehensive documentation & examples on Exiscan site

Reject spam

```
deny message   = This message was classed as spam
     condition = ${if <{$message_size}{80k}{1}{0}}
     spam      = nobody
```

Reject viruses

```
deny message   = Message contains a virus ($malware_name)
     malware   = *
```

MIME checking

```
deny  message = Serious MIME defect detected ($demime_reason)
      demime = *
      condition = ${if >{$demime_errorlevel}{2}{1}{0}}
```

Other considerations

The MX problem

- Consider whether you really need multiple MXes
- If you have more than one server, all need identical protection, to:
  - Avoid spam 'backdoors'
  - Avoid collateral spam

The multiple-recipient problem

- Affects scenarios where not all users have the same scanning preferences
- No easy way round it due to limitations of SMTP, but reasonable workarounds available
- If you have to offer options, try to keep the choices simple. Even if “one size doesn't fit all”, maybe “two sizes” do?
- Consider SMTP defers (multiple-scan-profile method)
  http://www.exim.org/pipermail/exim-users/Week-of-Mon-20031006/061151.html

A common usage scenario: Exim as a transparent front end

- Drop in front of existing mail system
- Route messages to the “real” MX using a manualroute router

Conclusions

- Content scanning is a useful tool as part of a wider policy framework
- Needs responsible planning and implementation to avoid amplifying the problem or moving the burden to someone else
- The Exiscan patch is widely used, stable and powerful, allows scanning at SMTP time for:
  - Anti-virus
  - Anti-spam
  - File extension blocking
  - Regular expression blocking
- With so many easy, powerful and free solutions, at the very least *some* basic content scanning (e.g. file extension blocking) is highly recommended and can be achieved with modest resources.

This summary, detailed notes from this talk, and Content Scanning HOWTO available at:

http://www.timj.co.uk/computing/software/exim/