Mail filtering at full volume

Keeping e-mail safe using open source software

Des Keane
Sysadmin
Google, Inc.
June 25, 2005

Keeping e-mail safe and relevant

Challenges

- Spam - ~50%
- Viruses - ~14%
- Phishing attacks - ~1%
- Bounce messages - ~7%
- Useful mail - ~28%

Volume

- Load
- Downtime
- DoS
- Complexity

Google

Let's head to the armoury

Open source toolkit

- sendmail
  - http://www.sendmail.org/
- MIMEdefang
  - http://www.mimedefang.org/
- SpamAssassin
  - http://spamassassin.apache.org/
- ClamAV
  - http://www.clamav.net/

Google

sendmail

The original and the...

- milter API for mail filtering plugins
  - passes mail to plugins at SMTP time
  - we can reject mail with 550 response
  - we can modify mail before delivering it
  - we can inspect the message contents and metadata

Google™

MIMEdefang

Rules engine

- sendmail plugin using milter API (C code) - **mimedefang**
- “flock of daemons” approach - **mimedefang-multiplexor**
- preloads SpamAssassin into memory - **mimedefang.pl**
- rules written in perl, the sysadmin's friend - **mimedefang-filter**
- requires a flock of perl modules

- hooks for many virus scanners plus SpamAssassin
- methods to block/tag/manipulate mail, esp. attachments
- logging

Google

SpamAssassin

There is no substitute

- wide range of local and network tests
- test against headers and body
- easy creation of custom rules
- Bayesian learning
- hooks into popular RBLs
- whitelist and blacklist

Google

ClamAV

Open source virus scanner

- Very fast response times to outbreaks
- Good detection rates (~98%)
  - Better than some commercial scanners
- understands MIME and archive formats
- daemon scanner, **clamd**, as well as command line
- database updater - **freshclam**
- Create your own virus signatures

Google

Installation

Fedora Core 3 example

- Pre-installed: sendmail, SpamAssassin, [sendmail-devel]
- Get from dag.wieers.com:
  - **`yum install clamav clamd`**
  - **`yum install perl-IO-stringy perl-MIME-Base64 perl-MailTools perl-MIME-tools perl-Digest-SHA1 libnet perl-Mail-Audit perl-Time-HiRes perl-HTML-Tagset perl-HTML-Parser perl-Compress-Zlib perl-Archive-Zip`**
- Get from CPAN:
  - **`sudo cpan -i MIME::Base64 Mail::Audit`**
- Install MIMEdefang:
  - **`sudo /usr/sbin/useradd defang`**
  - **`./configure; make; sudo make install`**

Google

Getting started

Configuration files

- sendmail.mc, add:
  - INPUT_MAIL_FILTER(`mimedefang', `S=unix:/var/spool/MIMEDefang/mimedefang.sock, F=T, T=S:360s;R:360s;E:15m')
- clamd.conf, freshclam.conf
  - [if build from src] comment out line containing “Example”
- sa-mimedefang.cf
- mimedefang-filter
- init scripts
- update virus signatures: **`sudo freshclam`**

Google

MIMEdefang rules

Set policy

- Tag all spam messages:
  - `action_change_header("X-Spam-Score", "$hits ($score) $names");`
- Reject viruses:
  - `return action_bounce(“Rejecting because of virus $VirusName”;`
- Remove bad attachments and add header
  - `action_change_header("X-Bad-Attachment", "True ($fname)");`
- Notify administrator
  - `action_notify_administrator("An attachment of type $type, named $fname was dropped.\n");`

Google

Your own SpamAssassin rules

Configuration

- Enable network rules, e.g. URIDNSBL
  - Beware of RBLs bearing false promises
  - performance will drop due to DNS lookups
- Be careful with trusted_networks definition
  - **`clear_trusted_networks`**
    `trusted_networks 127/8 192.168/16`
- Sample local rule (**`sa-mimedefang.cf`**):
  - **`header PIGEON_WARNING Subject =~ /Muttley/`**
    `describe PIGEON_WARNING Untrustworthy carrier pigeon`
    `score PIGEON_WARNING 5.0 5.0 5.0 5.0`
- Beware of non-English language messages causing false positives

Google

Bayesian learning

The Curse of Bayes

- Turn off bayes auto_learning & expiry
- Turn on bayes journalling
- Very disk I/O intensive – consider tmpfs
- dbs are fragile and get corrupted
- backup them up regularly or start afresh
- To check bayesian database status:
  - **`sa-learn --dump magic`**
- Feeding bayes:
  - **`sa-learn --spam --mbox --showdots <mbox>`**
  - **`sa-learn --ham --mbox --showdots <mbox>`**

Google

Encrypted zips, rars and bad attachments

- looking inside zipfiles
  - Archive::Zip
- rar licensing
- bad attachment list
  - **ade|adp|app|asd|asf|asx|bas|bat|chm|cmd|com|cpl|crt|dll|exe|fxp|hlp|hta|hto|inf|ini|ins|isp|jse?|lib|lnk|mdb|mde|msc|msi|msp|mst|ocx|pcd|pif|prg|reg|scr|sct|sh|shb|shs|sys|url|vb|vbe|vbs|vcs|vxd|wmd|wms|wmz|wsc|wsf|wsh**

Google

Keeping up to date

Updates

- freshclam for signatures
- clamav engine upgrades
- SpamAssassin rules and engine
- MIME bugs
- false positives and bad versions
  - clamav 0.86, 0.85, 0.82, 0.81
  - *"POPULAR OPEN SOURCE virus scanner Clamav has been hastily updated this morning to remove a 'false positive': the scanner was detecting the GNU Public Licence as a virus."*
    http://www.theinquirer.net/?article=18919 (2004/10/06)

Google

Scaling issues

- cpu
- disk IO
  - more logging = more disk IO
  - tmpfs (bayes, /var/spool/MIMEDefang)
- rule size and number (memory)
- multiple virus scanners
- cap size of messages scanned for viruses or spam

Google™

Availability issues

- disk space (queues, quarantine, logs)
- sendmail load average checks
- bugs
- zip bombs (rare)
  - Oversized.Zip virus
  - false positives

- should be possible to build a complete new server within minutes

Google™

Monitoring

Keep on running...

- clamd
  - clamdwatch
- sendmail
- mimedefang multiplexor
- always tempfail
- system resources
- alerting without email!

Google™

Des Keane
Google, Inc.