SpamAssassin
An Introduction

PacNOG I Workshop

June 20, 2005
Nadi, Fiji

Hervey Allen

What is it?

From http://spamassassin.apache.org/

SpamAssassin is an Apache Software Foundation project and is released under the Apache License.

...[it] is a mail filter to identify spam. It is an intelligent email filter which uses a diverse range of tests to identify unsolicited bulk email, more commonly known as Spam. These tests are applied to email headers and content to classify email using advanced statistical methods. In addition, SpamAssassin has a modular architecture that allows other technologies to be quickly wielded against spam and is designed for easy integration into virtually any email system.

What is it...Our description?

From Philip Hazel (Author of Exim):

SpamAssassin is a content-based filtering system. It is written in Perl, and is very CPU-intensive. SpamAssassin can perform a number of network-based lookup checks that can take a long time to complete. It may therefore not be suitable for high-volume mail systems. The configuration is complex, though a basic installation using the ports system is straighforward.

How can I use it?

- Integrated with DNS Blocklists
- Using checksum/digest based systems
- *Integrated with an MTA (exim, sendmail, qmail, postfix, etc.)*
- Single user installs under Unix
- With Procmail
- As a proxy server
- Against remote IMAP folders
- Integrated in commercial products

Where does it fit?

Generally speaking you should filter incoming mail to a server like this:

- Blacklists
- Whitelists
- Content-based solution (like SpamAssassin)

Then decide how you wish to integrate SpamAssassin in to your MTA.

Using Exim there are several ways...

Where does it fit into Exim?

From

http://wiki.apache.org/spamassassin/IntegratedInMta

1. As of Exim 4.50, by compiling Exim using WITH_CONTENT_SCAN=yes. As an Exim 4.4x Exiscan-extended ACL condition stack, Exim can reject spam after reading the body, but before Exim acknowledges acceptance of the email.
2. As an Exim 4.x loadable module or local_scan.c replacement. SA-Exim also allows SpamAssassin to reject spam before its accepted by your MTA.
3. As an Exim transport.

With Exim we won't use...

These integrated SpamAssassin features:

- DCC
- Pyzor
- Razor
- RBL checks
- Bayesian analysis

But, you can use them if you wish, or if they work for your situation.

Why not use...?

- DCC: Distruted Checksum Clearinghouses
  - http://www.rhyolite.com/anti-spam/dcc/
- Razor: Vipul's Razor Version 2
  - http://www.rhyolite.com/anti-spam/dcc/
- Pyzor
  - http://pyzor.sourceforge.net/

All are separate modules that must be installed to work with SpamAssassin. You should investigate and decide for yourself.

Why not use (2)...?

- Realtime Blacklist Checks (RBL)
- Bayesian logic checking system

Exim can already do RBL mail filtering for us as part of its basic configuration.

You can turn on Bayesian logic checking, but this is CPU-intensive. On a heavily loaded system this may cause problems. Test for yourself.

More resources

- SpamAssassin Configuration Generator (version 2.5):
  - http://www.yrex.com/spam/spamconfig.php
- Using SpamAssassin:
  - http://wiki.apache.org/spamassassin/UsingSpamAssassin
- SpamAssassin AutoWhitelist explained:
  - http://wiki.apache.org/spamassassin/AutoWhitelist