SpamAssassin on Mac OS X Server

Mac OS X Server includes SpamAssassin that is called from amavisd and integrates with Postfix. SpamAssassin and ClamAV are called by amavisd in the standard setup.

Note: many of the commands used on this page require the installation of Developer Tools - install them before starting anything else.

Upgrading SpamAssassin

Mac OS X Server includes v.3.0.x, but 3.2.1 is the latest as of July 24, 2007. *Note 1: I personally do not know if 3.2 is compatible with Mac OS X Server's configuration, but hopefully it is; superficially at least, it appears to be working. Note 2: version 3.2.1 has a bug that prevents it from being installed via cpan or (perl -MCPAN -e shell) as root. Get the dependencies right with cpan/CPAN, then either do a force install, or download the tarball from cpan.org and do a normal install from source, making sure to run 'make test' as a non-root user.*

In general, to upgrade to the latest greatest version, simply issue the following command as an admin user from the command line:

```
sudo cpan -i Mail::SpamAssassin
```

There may be some configuration and updating to do with your CPAN install, but generally you can follow required modules and things turn out OK 😉.

Network Tests

Network tests (like URIBL) are disabled by default.

1. Turn on network tests in /etc/amavisd.conf Change the line `$sa_local_tests_only = 1;` to read `$sa_local_tests_only = 0;`
   2. Install Net::DNS. From the command line:

   ```
   sudo cpan -i Net::DNS
   ```

   Follow all the links for other required modules.

Note that this module does hit the network quite a bit for DNS lookups and it is highly recommended that you run a caching nameserver to speed things up and minimize network traffic. To do this, simply turn on DNS in Server Admin.app then point your DNS lookup to 127.0.0.1 in System Preferences -> Network.

3. Test with `spamassassin -D < path_to_some_test_message`. This will generate a lot of output which is quite useful in seeing what is working and what is not. For a test message, just paste a spam message (including all headers) into a text file and save it somewhere on the server where spamassassin can read it. In the debug output, check to see that a. DNS is available and b. a spam score is being applied to your test message. If URIBL tests are working and it recognizes a spam URL, there should be a nice table showing which blacklists the URL was found on and how many spam points your message has.

Bayesean Filtering

There is a switch in Server Admin -> Mail -> Settings -> Filters to update the virus and junkmail DBs but it may or may not work (at least for the spam DB, antivirus seems OK - YMMV). Here's what you can do to fix it:

1. Properly link the amavisd and clamav .spamassassin bayes DBs. Both ClamAV and amavis create DBs with bayes info (the statistical info that allows SpamAssassin to "learn" what is and isn't spam) - unfortunatley, they should be sharing that info. On the command line:

   ```
   sudo -s
   cd /var/clamav
   rm -rf .spamassassin
   ln -s /var/amavis/.spamassassin .spamassassin
   exit
   ```

   The rm command may or may not give you an error - it depends if you have been running your server for a while already and there is a .spamassassin directory to delete. If you do get the error that the directory does not exist, don't worry, just move on.
   2. Make sure that the cron script to learn Ham and Spam is in place, or replace it with another script (I use spamtrainer because I'm too lazy to fix Apple's install when spamtrainer will do it for me. As a bonus, spamtrainer will clean up messages that it has already analysed, so you don't have to.) Read the documentation that is included.
   3. Create system accounts with the short name junkmail and notjunkmail. These accounts do not exist by default, so you have to do it. In Workgroup Manager, make sure that you are in the local directory (not LDAP!!) and create the two users, (you don't need to give them home dirs) then enable mail for them. You can now redirect (not forward) any Spam to junkmail@yourdomain.com and Ham to notjunkmail@yourdomain.com and the system should learn from those messages. You can also create shared IMAP folders for these users so your IMAP users can just drag spam into the shared junkmail mailbox and it will automagically be learned. The standard Mac OS X Server install puts cyradm(1), the Cyrus administrative tool, in `/usr/bin/cyrus/admin/cyradm`, which is not in the default `PATH`. A very effective GUI for the Mac to set shared permissions on the inboxes for junkmail and notjunkmail is SirAdmin. If you run SirAdmin remotely, remember to turn on SSL; otherwise your password will be transmitted in cleartext.

4. Gather some spam (what, no spam??) and feed the bayes DB. SpamAssassin won't use bayes filtering until you have at least 200 messages. I have been told that it doesn't get really effective until your have about 2000 spam and 2000 ham messages in the DB. Note that the learning process is time and processor intensive! Schedule the DB updates to minimize impact on the server. When I build a server, I like to feed it big lumps of spam and ham before it goes on line. You can use spamtrainer to specify another mailbox (other than junkmail and notjunkmail) to use to learn spam and ham on a one time basis (handy!)
5. Check your Bayes DB. Run these commands to see what your install is using for a Bayes DB:

```
sudo su clamav
sa-learn --dump magic
exit
```

You should see some stats about how many Spam messages and ham messages you have in your Bayes DB. Note that this will return an error if the DB has never been updated. Feed the system some spam, then try again.

Lint (the nasty stuff that grows between your toes)

The standard Mac OS X Server install has a number of incorrect settings in /etc/mail/spamassassin/local.cf - run this command to see what's broken then fix it. Note that Mac OS X Server calls spamassassin from amavisd, and so ignores much of what is in local.cf - the config settings are found in amavisd.conf.

```
spamassassin --lint
```

Here's some typical output:

```
spamassassin --lint
[473] warn: config: failed to parse line, skipping: auto_learn 1
[473] warn: config: failed to parse line, skipping: safe_reporting 0
[473] warn: config: failed to parse line, skipping: use_terse_report 0
[473] warn: config: failed to parse line, skipping: subject_tag *** Warning: Junk Mail ***
[473] warn: config: failed to parse line, skipping: rewrite_subject 0
[473] warn: config: failed to parse, now a plugin, skipping: ok_languages en fr de ja
[473] warn: lint: 5 issues detected, please rerun with debug enabled for more information
```

In local.cf,

- auto_learn should be bayes_auto_learn
- safe_reporting should be report_safe (Thanks to Alex from topicdesk.com for pointing out this error!)
- use_terse_report can be safely commented out
- subject_tag can be commented out
- rewrite_subject can be commented out
- ok_languages can be commented out and this line uncommented in v310.pre `loadplugin Mail::SpamAssassin::Plugin::TextCat`. ok_languages is now a User setting - I haven't implemented this, so I'm not sure if the plugin is even installed or how best to configure it. YMMV.

Pyzor

The Pyzor package is not installed by default on Mac OS X Server, you'll have to install it manually.

- download pyzor from Sourceforge and unstuff the bzip archive. cd into the download directory and run these commands to build and install the package

```
python setup.py build
python setup.py install
```

Pyzor depends on py-gdbm which is most easily installed via DarwinPorts

- first install Darwinports, then `sudo port install py-gdbm` to install the py-gdbm package
- test the py-gdbm install with this command `python -c 'import gdbm' && echo 'gdbm found'`
- test with `spamassassin -D < path_to_spam_message` and see what Pyzor messages come up - you should be good to go.

Razor2

Not free for non-personal use. The Razor2 package is not installed by default on Mac OS X Server, you'll have to find it and install it manually. On the TODO list...

Other cf files

The cf files are stored in `/usr/local/share/spamassassin/`. You can create new files or drop in cf files that others have created like here. Note that the numbers at the beginning of the cf file names determines the order in which the filters are processed.

Links to other useful info

- http://developer.apple.com/server/fighting_spam.html Fighting Spam on Mac OS Server 10.3 - includes stuff about installing Razor and other useful tools
- http://www.afp548.com/article.php?story=20051127235810230 Greylisting - not exactly SpamAssassin related, but helpful in a global approach.
- http://www.afp548.com/ AFP548.com has gobs of good info about spam fighting
- http://en.wikipedia.org/wiki/Mac_OS_X/ Mac Os X Wiki
- Mac Os Articles
- http://osx.topicdesk.com/ the creators of the spamtrainer script and other software, as well as some Mac OS X tutorials (that were used in part for the writing of this article). Thanks to Alex who was kind enough to correct some of my errors on this page.