Anti-Spam/Anti-Viral E-mail Gateway

This presentation will give an overview of a modular and extensible Anti-Spam/Anti-Viral filtering E-mail Gateway.

It is based upon an activity from Sam Hart's “Advanced Unix and Linux Administration” course *(presently taught through the U of A's Extended University)* which can be found online here:

www.geekcomix.com/cgi-bin/classnotes/wiki.pl?Setting_Up_An_Anti-SPAM_Gateway

Which is part of the larger collection of classnotes:

www.geekcomix.com/classnotes/

This system was originally based upon Scott Vintinner's excellent document *“Fairly-Secure Anti-SPAM Gateway Using OpenBSD, Postfix, Amavisd-new, SpamAssassin, Razor and DCC”*, which can be found here:

www.lawmonkey.org/anti-spam.html

Anti-Spam/Anti-Viral E-mail Gateway

Nomenclature:

- MTA – Mail Transport Agent
  The Mail Server
- MUA – Mail User Agent
  The E-mail client (Evolution, Pine, Outlook)
- LDA – The Local Delivery Agent
  Something which delivers mail locally
- UCE – Unsolicited Commercial E-mail
  Spam
- MP – Mail Processor
  Application that processes E-mail
- AV – Anti-Virus
  Typically some sort of Anti-Virus Software

Anti-Spam/Anti-Viral E-mail Gateway

- Debian GNU/Linux (OS)
  www.debian.org
- Postfix (MTA)
  www.postfix.org
- Amavisd-New
  www.ijs.si/software/amavisd/
- SpamAssassin
  www.spamassassin.org
- Vipul's Razor
  razor.sf.net
- DCC
  www.dcc-servers.net/dcc/
- Sophos Anti-Virus for Linux
  www.sophos.com
- Sophie
  www.vanja.com/tools/sophie/

Anti-Spam/Anti-Viral E-mail Gateway

Part 1:
Anti-Spam

Anti-Spam/Anti-Viral E-mail Gateway

Traditional Spam Filtering and Blocking:

- RBLs (Real-time Blackhole Lists, Relay Blocking Lists)

  Started out as System Administrators keeping lists of known Spammer IPs. Soon Administrators began sharing these IPs with eachother and Relay Blocking Lists were born.

  Now there are numerous RBL lists available for System Administrators to choose from.

- Insist on Valid Hostname

  Simple check inside MTA for valid source hostname

Anti-Spam/Anti-Viral E-mail Gateway

Limitations of RBLs: (see http://theory.whirlycott.com/~phil/antispam/rbl-bad/rbl-bad.html)

- Collateral Damage and Legitimate Users

  If you had a mail system that rejected all inbound emails, you would not have a very useful mail system. Likewise, as RBL lists grow in size and their use amongst mail system administrators increases, their value diminishes.

- Geopolitics and Blackholes

  A huge amount of SPAM is being sent through unsecured relays in Asia and South America. Using an RBL can effectively wall your mail system off from entire geographic regions. (See the Wired article, *"Not All Asian E-mail is Spam"*)

- Invisible Authorities

  Cases where RBL maintainers have intentionally blocked IPs of legitimate sources to censor or silence. Documented case where one popular RBL blocked the IPs of their business competitors.

- Appeals and Corrections

  A big problem with most RBLs is that it is very easy to get on them, but very hard to be removed. In essence, getting on to most RBLs means being tagged a spammer forever.

- Reliance on Unknown Third-Party

  Whomever is maintaining your RBL may be flaky, unreliable, or even a complete nut. Sometimes entire RBL sources go inactive, which may result in your server blocking the entire world.

Anti-Spam/Anti-Viral E-mail Gateway

Alternatives to RBLs:

- E-mail verification

  Each new contact must be verified either by the sender of the message or the message's recipient. (Not included in this gateway, but work is being done to do so in the future).

- Content Filtration
  - Filtration by message scanning

    Messages are scanned for patterns indicating “Spamminess”
  - Filtration by checksum

    Checksums are computed for each message and compared to a database of known Spam
  - Filtration by adaptive algorithms

    Messages are examined by algorithms that can be “taught” what is and what isn't Spam

Anti-Spam/Anti-Viral E-mail Gateway

Part 1.a:
Gateway Infrastructure

Anti-Spam/Anti-Viral E-mail Gateway

MTA: Postfix

Postfix is an easy to configure, yet very secure and efficient replacement for Sendmail. For our gateway, efficiency and modularity is important, so Postfix is a natural choice.

For security, it is best to have Postfix running as its own non-privileged user (such as “postfix” with groups “postfix” and “postdrop”).

Postfix can be configured to run in a chroot'ed environment easily for more security.

Anti-Spam/Anti-Viral E-mail Gateway

MP: Amavisd-New

Amavisd-New started as an anti-viral mail scanner, but has since evolved into more of a modular mail processor (MP).

Amavisd-New integrates very well with Postfix, but can also integrate with other MTAs (for example, there is a Sendmail Milter for it).

Amavisd-New is written largely in Perl and can be extended quit easily. New and additional components can be snapped in an integrated into existing infrastructure without needing to restart MTA.

Anti-Spam/Anti-Viral E-mail Gateway

MP: SpamAssassin

SpamAssassin is a mail processor which analyzes mail based upon a set of rules. The rules have scores associated with them. After message has been analyzed, an accumulative score is assigned to it.

SpamAssassin is also modular, and many people use it in lieu of Amavisd-New (however, you lose easy anti-viral integration).

SpamAssassin allows for user-level customization and rule score reassignments.

Anti-Spam/Anti-Viral E-mail Gateway

MP: Vipul's Razor

Innovative P2P Spam checksum distributive system. Almost guaranteed item caught is Spam

New Spam is identified by one client which then computes a unique checksum against that Spam and distributes that to the others on the Razor network.

Problems:

- Holes in network
- Requires user intervention
- Checksums often distribute slower than Spam
- Slight potential for abuse from Spammers

Anti-Spam/Anti-Viral E-mail Gateway

MP: DCC

Distributive Checksum Clearing house (DCC) is a large number of ISP mail servers (more than 120) which compares checksums for the mail that passes through them. If a single checksum is common among a large percentage of these ISPs (and client destinations) then it is very likely to be Spam.

Like Vipul's Razor, has very few false-positives. Unlike Vipul's Razor (which is P2P), DCC actually delivers checksums quickly enough for preventing Spam.

Anti-Spam/Anti-Viral E-mail Gateway

MP: Bayesian Filter

Ingenious technique by which an algorithm actually “learns” what is and what isn't Spam via training.

Can provide highly accurate Spam detection, if properly trained.

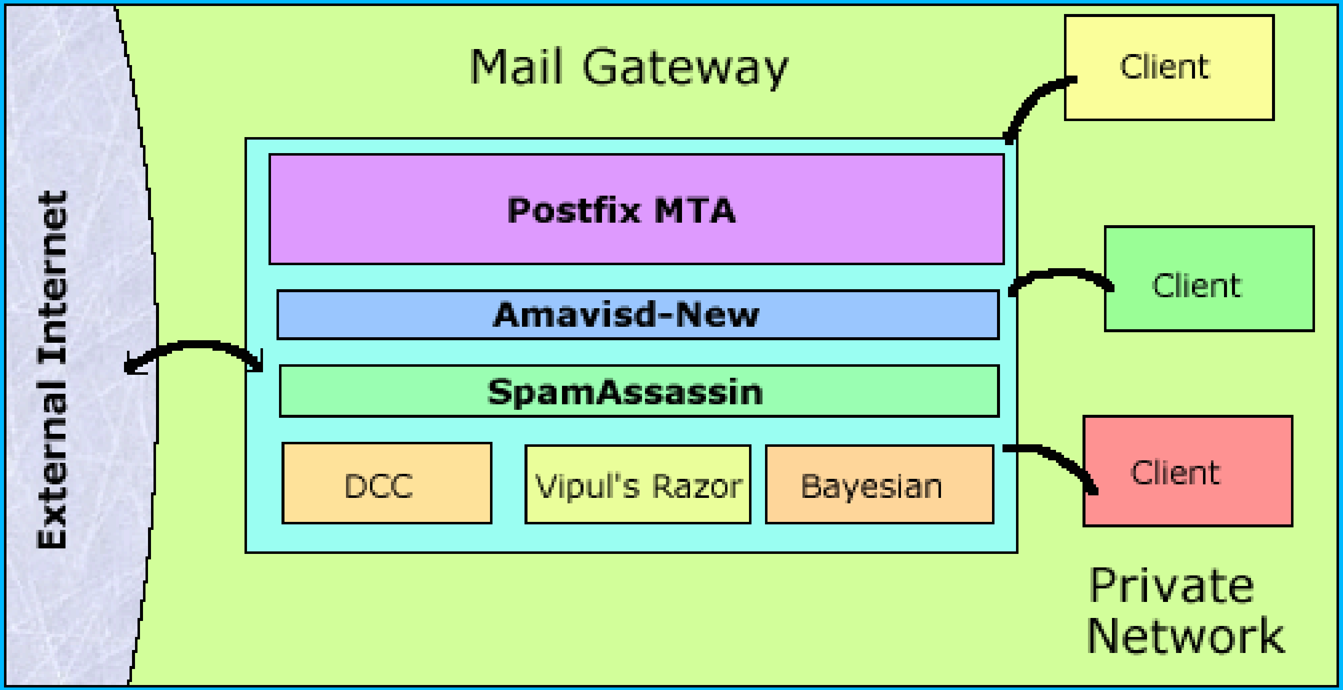
Problem:

Requires much user intervention- users must train the filter by sending items that are both “Spammy” and “not-Spammy”.

However, uses must send *full* message source, which many end-users do not know how to do (and those that do, may not have the patience).

Anti-Spam/Anti-Viral E-mail Gateway

Infrastructure Overview:



Anti-Spam/Anti-Viral E-mail Gateway

Part 1.b:
Gateway Configuration

Anti-Spam/Anti-Viral E-mail Gateway

MTA: Postfix

Setting up content-filter:

Postfix will communicate with Amavisd-New via an SMTP gateway on local machine. This does allow for the Amavisd-New filtering infrastructure (which includes SpamAssassin and all other MPs) to be moved off of the main mail delivery server if desired and onto a dedicated machine.

To enable Postfix to communicate with Amavisd-New, the following must be added to the main.cf configuration file (generally located in /etc/postfix):

```
content_filter = smtp-amavis:[127.0.0.1]:10024
```

Also, we want to define the “smtp-amavis” interface in master.cf:

```
smtp-amavis  unix -    -    y    -    2    smtp
    -o smtp_data_done_timeout=1200
    -o disable_dns_lookups=yes
 127.0.0.1:10025 inet  n    -    y    -    -    smtpd
    -o content_filter=
    -o local_recipient_maps=
    -o relay_recipient_maps=
    -o smtpd_restriction_classes=
    -o smtpd_client_restrictions=
    -o smtpd_helo_restrictions=
    -o smtpd_sender_restrictions=
    -o smtpd_recipient_restrictions=permit_mynetworks,reject
    -o mynetworks=127.0.0.0/8
    -o strict_rfc821_envelopes=yes
```

Anti-Spam/Anti-Viral E-mail Gateway

MTA: Postfix

Postfix Security:

We may wish at this point to configure Postfix to run in a protected chroot environment. Here, Postfix separates it's privileged and non-privileged processes by a chroot-jail so that we have an extra ring of security separating the outside world from our machine.

Setting up Postfix to run in chroot-mode is easy. We just edit master.cf (usually in /etc/postfix) such that everything has “chroot=y” except for “virtual” and “local” (these are for if we were running virtual servers, and for local mail delivery, which needs to deliver outside of the chroot jail).

MP: Amavisd-New

Configuring Amavisd-New:

The process for configuring Amavisd-New is very complicated, so we will only touch on the concepts here. For a more in-depth discussion, please see the following links:

- http://www.geekcomix.com/cgi-bin/classnotes/wiki.pl?UNIX03/Amavisd-New
- http://www.geekcomix.com/cgi-bin/classnotes/wiki.pl?UNIX03/Configure_Amavisd

Please note that these classnotes actually go through the rigmarole of setting up Amavisd-new in a chroot-jail. While this will increase security, it was largely included in the classnotes as an educational exercise. In practice, I usually do not enable the Amavisd-new chroot-jail, as it is not completely necessary and can shave off an hour or so from your setup.

Anti-Spam/Anti-Viral E-mail Gateway

MP: Amavisd-New

Configuring Amavisd-New, some topics to touch on:

Debugging: Initially you will want to set the Amavisd-new debugging levels rather high to help diagnose any problems you may have. However, you will want to lower them in a production environment as too much verbosity can really slow the system down.

Bouncing versus Passing: There are two basic options for what to do with mail tagged as Spam. You can bounce the mail back to sender, or pass the mail through to recipient (tagged as Spam, of course).

Spam Tag Levels: These options allow you to finely tune what is treated as Spam and what to do with it. The first tag level should be the lowest SpamAssassin score to flag a message as potential Spam (but still deliver). The second tag level should be the lowest SpamAssassin score to flag a message as *definite* Spam (then bounce or pass). The third level (kill level) is whatever score you want the message to be killed at.

MP: SpamAssassin

Configuring SpamAssassin:

Be sure to enable all of the external application checks that you will be using. For example, here we will want to enable Vipul's Razor, DCC, and Bayesian. If you are not using RBLs, then you will want to disable the RBL checks as well.

Inside of the main SpamAssassin configuration file (usually local.cf inside of /etc/spamassassin) you can also reconfigure rule scores depending upon your needs. I personally like the following tweaks:

```
score DCC_CHECK 4.000
score RAZOR2_CHECK 2.500
score BAYES_99 4.300
score BAYES_90 3.500
score BAYES_80 3.000
```

You can see all of the tests performed and their scores here:

http://www.spamassassin.org/tests.html

MP: Vipul's Razor

Configuring Vipul's Razor:

As Vipul's Razor is a P2P application, you will need to first register yourself on the Razor network. Your users can do this individually if they want, however, for our gateway we will want this to be system-wide. We start by running a script that verifies Vipul's Razor was installed correctly:

```
# razor-client
```

We then want to create the Razor files. To do this system-wide, we will run the following as root:

```
# razor-admin -create
```

Next, we want to create our user profile (which is only used to connect to the network and insert new Spam checksums). We could do this a number of ways, but the easiest would be to issue:

```
# razor-admin -register -user postmaster@domain.com
```

Finally, we move the .razor directory into Amavisd-new's home.

MP: Vipul's Razor

Configuring Vipul's Razor:

One final thing you may wish to do which will make your system all the more useful is to create a Spam-trap e-mail address or two for inserting new Spam into the Vipul's Razor network. This is completely optional (if you do not do it, you are effectively being a leech on the Razor community, but if you can deal with that, fine), but is typically considered good netiquette.

If you do make Spam-traps, you will probably want at least two:

- One for raw Spam from the internet (i.e., an e-mail address you intentionally sprinkle around the web and newsgroups to entice Spammers. Could also be integrated with a honeypot or tarpit).

- One for bouncing uncaught Spam into from a MUA. Note that you have to bounce the full message source in order for this to work.

Anti-Spam/Anti-Viral E-mail Gateway

MP: DCC

Configuring DCC:

There really isn't anything to configuring DCC. Once you have installed it and told SpamAssassin about it, it will just work.

However, if you are running Amavisd-new in chroot-jail, then you will have to make certain that everything DCC needs is inside this jail.

MP: Bayesian Filter

Configuring Bayesian Filter:

There are a variety of methods for implementing a Bayesian Filter. However, the one I used is the same as was mentioned by Scott Vintinner and comes as a Bash shell script which you can obtain here:

- http://www.geekcomix.com/cgi-bin/classnotes/wiki.pl?UNIX03/Bayesian_Learning_Script

However, if you do not plan on using the Bayesian Filter (because it requires so much more of your users than anything else here) then you can effectively skip this part (be sure to turn off the Bayesian Filter in SpamAssassin, though.)

Anti-Spam/Anti-Viral E-mail Gateway

Part 1.c:

Notes and Results for Anti-Spam Gateway

Anti-Spam/Anti-Viral E-mail Gateway

Items of Consideration:

I have been running something close to this setup on Geekcomix.com and Tux4Kids.net for 3 years now (DCC and Bayesian filters are new, and Amavisd-new replaced Procmail which was used previously). I have also run this exact setup on several of my client's networks (including one client which has 200+ employees). I also have several former students who have implemented this Gateway on their networks.

In this section, I will share with you what we have discovered about this setup.

Anti-Spam/Anti-Viral E-mail Gateway

Items of Consideration:

Resource Intensive:

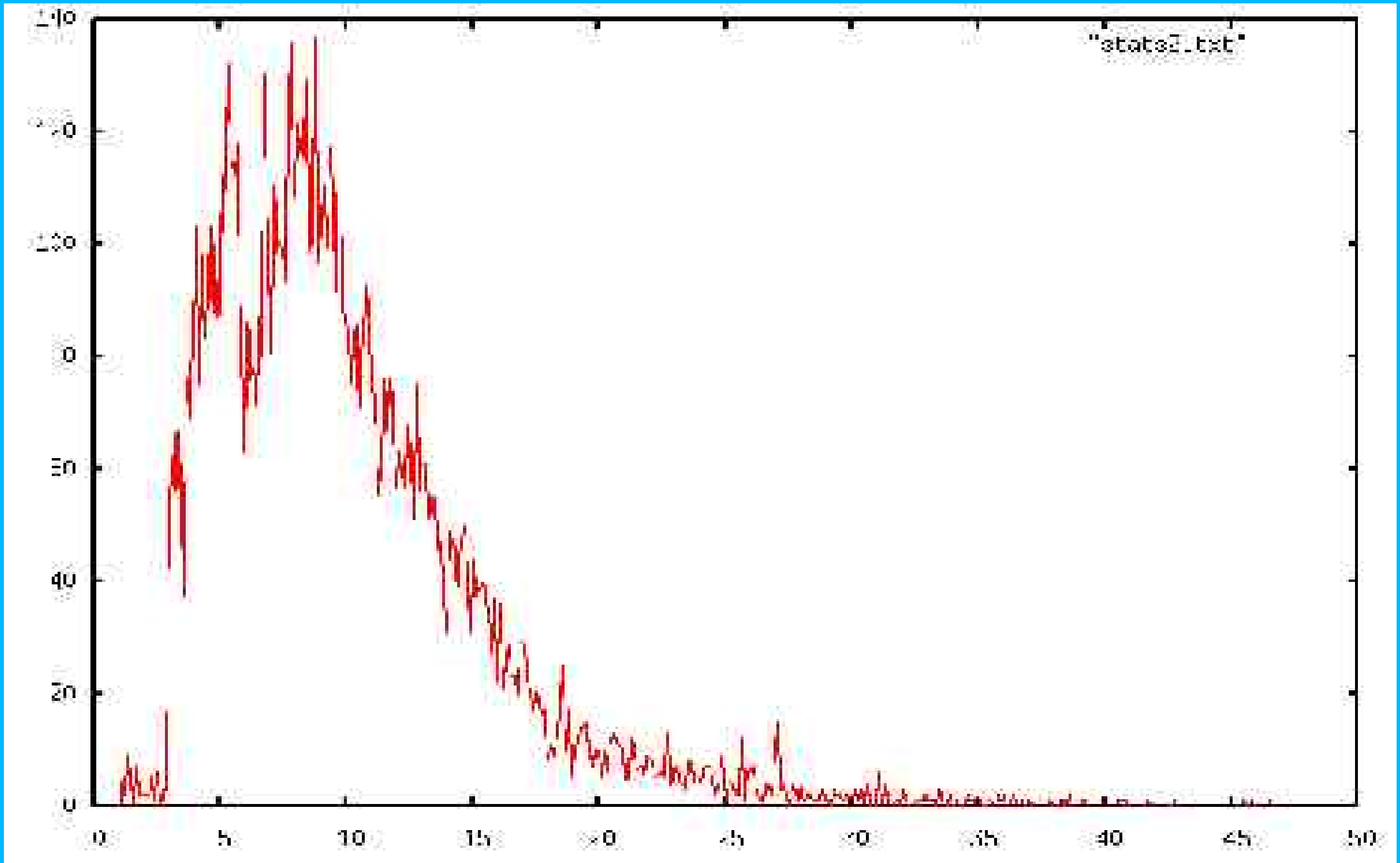
The Anti-Spam Gateway detailed here can be a rather resource intensive system for reasonably sized organizations (or organizations receiving a lot of e-mail). This is especially true once you include the Anti-Viral components (which we will see shortly).

As such, it is my recommendation that the Gateway be setup on its own dedicated machine. If you have some legitimate need to have the MTA be on the same machine as other processes (such as users logging in or a web-site) then the Amavisd-new infrastructure should probably be moved to another system.

Anti-Spam/Anti-Viral E-mail Gateway

Items of Consideration:

By far the hardest part of this setup is picking appropriate settings for your SpamAssassin tag levels.



Anti-Spam/Anti-Viral E-mail Gateway

Items of Consideration:

With appropriate tweaking of the SpamAssassin rules for your organization (and these are really quite minor tweaks as opposed to blocking and unblocking offending IPs, trust me) and of the three tag levels in Amavisd-new, you can obtain a highly efficient Spam-filtration system.

By way of a concrete example, at Geekcomix.com, between July 2002 and December 2003, this system has had the following results:

21,335	Items of Spam Caught and Tagged
961	False-negatives (uncaught Spam)*
1073	False-positives (un-Spam caught)

Some of my clients are reporting better results.

*: Okay, this includes a bunch of uncounted Sobig viri that I just dragged into my uncaught folder to get them out of my inbox.

Anti-Spam/Anti-Viral E-mail Gateway

Items of Consideration:

Top 10 Spam "Sources" (note that most, if not all, are forged)

```
+----+-----Author-----------------------------------+--Msg-+-Percent-+
|  1 | offers@permission-mail.com                   |   58 |  0.59 % |
|  2 | @flora.host4u.net                            |   35 |  0.36 % |
|  3 | support@oceanicspecials.com                  |   34 |  0.35 % |
|  4 | newsletter@thecareernews.com                 |   32 |  0.33 % |
|  5 | support@consumer-marketplace.com             |   27 |  0.28 % |
|  6 | bescotime@besco.com.cn                       |   25 |  0.25 % |
|  7 | t4k-art-admin@tux4kids.net                   |   20 |  0.20 % |
|  8 | TailWaggingOffer@e.ss01.net                  |   19 |  0.19 % |
|  9 | admin@geekcomix.com                          |   19 |  0.19 % |
| 10 | interkeysol@earthlink.net                    |   16 |  0.16 % |
+----+----------------------------------------------+------+---------+
|    | other                                        | 9531 | 97.10 % |
+----+----------------------------------------------+------+---------+
```

This list is from Febuary 2003, and does not reflect current Spam-box.

Anti-Spam/Anti-Viral E-mail Gateway

Items of Consideration:

Top 10 MUA Client Names Used

```
+----+----Mailer------------------------------------+--Msg-+-Percent-+
|  1 | (unknown)                                    | 3224 | 32.84 % |
|  2 | Microsoft Outlook Express 5.x                | 1406 | 14.32 % |
|  3 | Microsoft Outlook Express 6.x                | 1194 | 12.16 % |
|  4 | Microsoft Outlook                            |  583 |  5.94 % |
|  5 | Microsoft Outlook IMO                        |  415 |  4.23 % |
|  6 | AOL 7.0 for Windows US sub 118               |  354 |  3.61 % |
|  7 | QUALCOMM Windows Eudora                      |  333 |  3.39 % |
|  8 | The Bat!                                     |  298 |  3.04 % |
|  9 | MIME-tools 5.503 (Entity 5.501)              |  263 |  2.68 % |
| 10 | Internet Mail Service 5.x                    |  261 |  2.66 % |
+----+----------------------------------------------+------+---------+
|    | other                                        | 1485 | 15.13 % |
+----+----------------------------------------------+------+---------+
```

This list is from Febuary 2003, and does not reflect current Spam-box.

Anti-Spam/Anti-Viral E-mail Gateway

Items of Consideration:

Spam Delivery by Day of the Month

```
100% -----------------------------------------------------#---------
 90% -------#-#-------------------------------------#---#-#-#-------
 80% -------#-#-#---------#-#-#-------------#-------#-#-#-#-#-------
 70% -#---#-#-#-#-#-#-----#-#-#-#-----------#-#---#-#-#-#-#-#-#-#---
 60% -#-#-#-#-#-#-#-#---#-#-#-#-#-----#---#-#-#-#-#-#-#-#-#-#-#-#---
 50% -#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#---
 40% -#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#---
 30% -#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#---
 20% -#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#---
 10% -#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-#-
      * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
 day  1         6          12          18         24             31
```

This list is from Febuary 2003, and does not reflect current Spam-box.

Anti-Spam/Anti-Viral E-mail Gateway

Part 2:
Anti-Viral

Anti-Spam/Anti-Viral E-mail Gateway

FLOSS Anti-Viral Options Available Under Linux (and possibly for other Unixes):

- Amavis : Okay, so Amavis does still offer some Anti-Virus detection.
- ClamAV : Open-Source Anti-Virus software which can integrate with an MTA or act as a stand-alone server. Unique in Open-Source AVs in that it has an actively maintained viral database.
- OpenAntiVirus : Another Open-Source Anti-Virus software. OAV was the first from which others like ClamAV and Amavis we spawned.
- VirusHammer : Part of OAV, but with handy web-interface.
- MIME-Defang : Not specifically AV, but often works.

Anti-Spam/Anti-Viral E-mail Gateway

Proprietary Anti-Viral Solutions for Linux:

- Sophos AV : Barely passable utilities for Linux (and Windows for that matter), but with an excellent API for making Sophos everything it should have been.
- Norton AV : The Linux versions seem to be made as an afterthought. No useful API (that I know of) and limitations on how it can be integrated with your servers (e.g. Samba, MTA, FTP, etc).
- RAV : For the longest time the best Linux AV software hands down. Now owned by Microsoft.
- H+BDEV : Germany-based company that many Linux users swear by (primarily SuSE users, I wonder why?) I have had no experience with them.

Anti-Spam/Anti-Viral E-mail Gateway

AV: Sophos

Sophos has a barely passable Linux version of their software. They provide a “sweep” application that mirrors their Windows version, but which is unusable as a mail scanner (too resource intensive).

Sophos also provides a MailMonitor server application for integration with MTAs. However it is expensive, tricky to install, and limits what else you can filter with.

But the big draw for Sophos is their excellent API which allows other developers to extend (and fix) the Sophos functionality under Linux.

Anti-Spam/Anti-Viral E-mail Gateway

AV: Sophie

Sophie is an AV daemon which uses libsavi to scan any file passed to it for viri.

Whereas Sophos “sweep” is a resource hog and not something you'd want to scan each e-mail attachment with (it reloads the whole SAV library each time it's run), Sophie is lean and does things “the Unix way”. Sophie loads the SAV library when it first runs and caches it until needed.

Sophie can integrate with Amavisd-new and provide seemless integration of AV software into our existing gateway.

Anti-Spam/Anti-Viral E-mail Gateway

Part 2.b:
AV Configuration

Anti-Spam/Anti-Viral E-mail Gateway

AV: Sophos

Installation and Configuration:

Sophos requires there to be a non-privileged user and group for it to use when running. It also has very specific libc requirements.

Check the instructions included in your copy of Sophos for Unix/Linux for installation information.

You can create a script for updating Sophos with the latest IDEs. As an example, here is what I use:

http://raman.physics.arizona.edu/temp/sophupdate

Have this script run via cron daily.

Anti-Spam/Anti-Viral E-mail Gateway

AV: Sophie

Installation and Configuration:

Installing and configuring Sophie is rather involved, and I would point you to the following online resources from my classnotes:

http://www.geekcomix.com/cgi-bin/classnotes/wiki.pl?UNIX03/Integrating_Sophie_With_Amavisd

http://www.geekcomix.com/cgi-bin/classnotes/wiki.pl?UNIX03/Install_Sophie

http://www.geekcomix.com/cgi-bin/classnotes/wiki.pl?UNIX03/Configure_Amavisd_For_Sophie

Finally, you will want to set up Sophie so that it starts at init. If you installed Sophie by hand, then you will likely have to come up with your own script for this. I have a rudimentary one at this URL if you wish to take it and modify it for your needs:

http://www.geekcomix.com/cgi-bin/classnotes/wiki.pl?UNIX03/Add_Sophie_To_Init

Anti-Spam/Anti-Viral E-mail Gateway

Part 3:
Conclusion

Anti-Spam/Anti-Viral E-mail Gateway

There isn't much data for the AV section of this gateway. If you use Sophos and Sophie then the viruses are detected only if Sophos is kept up to date.

The Anti-Spam component of the Gateway will work only if you are willing to do some initial tweaking. Don't worry about it being too time consuming, it isn't. It really amounts to checking why a message was falsely tagged as Spam or not and seeing what rules you need to adjust the scoring for.

By far the most difficult part to figure out is how to set the three SpamAssassin tags in Amavisd-new. These three settings will make or break your gateway.