www.photocase.de

A Home-Grown Spam Filter

SPAM HUNTER

Even if you do not have a mail server of your own, SpamAssassin can help you filter out unsolicited junk mail. This article describes how SpamAssassin collaborates with Evolution and KMail. Thunderbird, on the other hand, has its own simple spam-detection feature. **BY OLIVER FROMMEL**

Spammers invent more perfidious tricks every day to confuse spam detection software. For example, they modify the text and subject line to leave just a minimum of legible information. They also add punctuation between the individual letters within a word to prevent simple tests for text strings: for example, *V.a.l.i.u.m* is not the same as *Valium*.

SpamAssassin gives you the weapons you need to fight spam. SpamAssassin checks for gappy brand names and detects a whole range of suspicious products. Spammers who use tricks to slip brand names into your inbox are actually shooting themselves in the foot, because brand names of the types typically used by spammers are unlikely to occur in normal email messages. This reduces the number of false positives, that is, the number of legitimate messages classified as spam.

Hunt for Spam

Our spam detective also has a few tricks up its sleeve to improve its success rate: SpamAssassin queries servers that register the spam known to be in the wild.

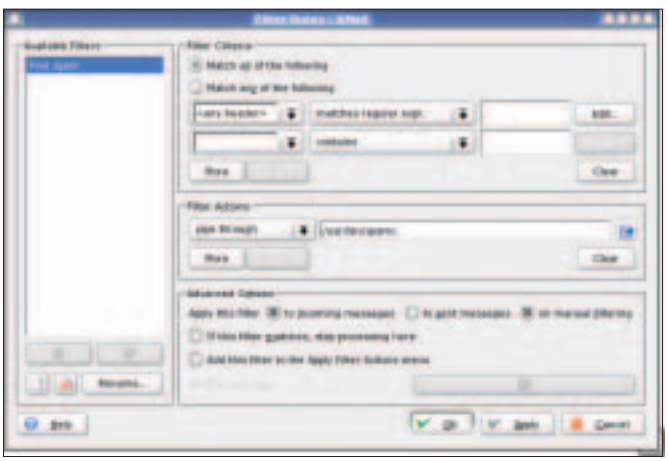

Figure 1: The Filter window in KMail. The button for adding a new filtering rule is on the lower left.

This is a similar to the approach used by a blacklist, such as the Spamhaus list [1], which lists spam servers. SpamAssassin performs several hundred checks for each message. If you are interested, there is a page with a list of tests on the website at [2].

When SpamAssassin discovers a spam message, it tags the message by adding additional header lines, for example *X-Spam-Flag: YES*. Other filters can then read the header and handle the message as defined by the user. This could mean binning the messsage, or moving the message to a special mail folder. SpamAssassin adds a few header lines with additional information about the tests:

```
X-Spam-Status: Yes, ➥
hits=7.1 tagged_above=➥
-99.0 required=5.0➥
 tests=BAYES_99,➥
FORGED_YAHOO_RCVD, ➥
MORTGAGE_PITCH
```

Besides the spam tag (*Yes*), the line indicates the number of *hits*, the spam threshold (*required*), and the tests that have been performed.

Home-Grown Filter

If you do not run your own mail server, you obviously cannot run SpamAssassin on the mail server. As a workaround, you can run the tool on the client, that is, on the machine that fetches your mail from the server. There is a downside to the workaround, though: you need to download your mail before

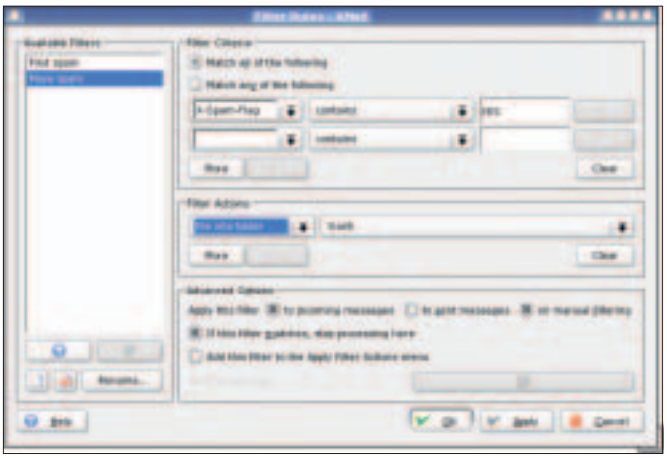

Figure 2: The second filtering rule, "Move spam", tells KMail to move any messages it has designated as spam to a mail folder.

Figure 3: Adding filters in the Evolution filter dialog.

Figure 4: A filtering rule that checks all messages that have the *X-Spam-Flag* set to *YES* in the header.

you can check it. This means that junk mail can still use up your valuable resources, but at least it will not show up in the mail program.

KMail Integration

If you have a new KMail version (1.7 or better), you can simply enable the integrated SpamAssassin support (see the Progress box). For older versions of KMail, you will need to take a detour via the normal mail filter, that is, via the *Settings* | *Configure Filters* menu item (Figure 1).

In the window that then follows, click the button shown in Figure 1 to create a new rule, which you could call *Find Spam*. In the selection list, click on *< any header >*, and then click on *matches regular expression* in the field to the right. In the text box on the right, enter a dot as a regular expression matching any character that might occur in an email. This completes setting up a filtering rule that will match any incoming message.

Select *pipe through* below *Filter Actions* and type *spamc* in the text box to the right. After closing the dialog, KMail will run *spamc* for each message that you download from your POP account. The command line program contacts the SpamAssassin server running in the background and passes the email message to the server for spam testing. You might like to perform a manual test with the filter just to make sure. To do so, select *Apply filter* in the *Message* menu, or press [Ctrl-J].

Fedora sets up SpamAssassin to tag unsolicited mail with a *[SPAM]* label in the subject line. Of course, SpamAssassin will still add new header lines to the email message to document the spam analysis. To view these additional headers, select *View* | *Header* | *All Headers*, and then look for the lines that start with *X-Spam*.

A second filtering rule will remove junk mail that is labeled as spam. To add the rule, open the first menu item and select *X-Spam-Flag*. If the next field already says *contains*, you just need to add *YES* in the third field. The filter action we need here is *move to folder* with *Trash* as the target. It is important to deselect *If this filter matches, stop processing here ...* for the first rule we defined (*Find Spam*) to allow KMail to apply the second filter.

Whenever you fetch mail, the mail is now processed by the two filters, which will hopefully send any junk to the trashcan. On the downside, this means fetching mail will take a bit longer because KMail needs to launch *spamc* for each message. Current KMail versions need fewer resources and are slightly quicker.

Filtering with Evolution

Setting up a filter in Evolution (Version 1.4.6) follows an approach similar to the one we just followed with KMail, although the menus have slightly different names and are organized slightly differently. Clicking on *Tools* | *Filter* opens a filter dialog where you can click *Add* to pop up a new window (Figure 2).

First assign a name to the rule, for example *Find Spam*. In the list at the top, which will originally read *Sender*,

Installation

SpamAssassin is a popular tool, so ready-to-run packages are available for many distributions. Most distros actually put SpamAssassin on the installation media, allowing a simple package manager install. Both Fedora Core 2 and Suse 9.1 include Version 2.6.3.

There is an RPM package with the current SpamAssassin 3.0.1 at *http://atrpms.net/*. However, installing the RPM does require advanced skills, as standard distributions typically lack the required current libraries. Although it is easy enough to google for the libraries, installing them on your own distribution may be difficult due to the number of dependencies you may need to resolve.

This is why we will be focusing on SpamAssassin 2.6 in this article. The Antispam software requires the following Perl package: perl-Time-HiRes, perl-Digest-SHA1, perl-Digest-HMAC and perl-Net-DNS. If you use a GUI-based package manager for the install, the package manager will automatically handle the dependencies. If not, your mileage may vary.

Debian users can run *apt-get install spamassassin* to install SpamAssassin, however, you will have to make do with Version 2.64.

SpamAssassin needs to be running in the background to allow your mail program to look for spam. Fedora has a *chkconfig spamassassin on* parameter that sets up the program persistently and launches the program at the same time. On Suse, you need to type *insserv spamd* instead. Debian uses the *update-rc.d* script for this purpose. The script expects the service name (*spamassassin*), the *start* command, the script number, and finally the runlevel.

Progress

A more recent version of KMail (for example 1.7.1) or Evolution (2.0 or later) can save you a lot of work. Modern mail programs have many anti-spam mechanisms. You are actually required to install SpamAssassin before you can install Evolution 2.0 on Fedora Core 3.

KMail allows gives users a choice between Bogofilter, Annoyance, and SpamAssassin. When the program launches, it automatically searches your computer for anti-spam tools.

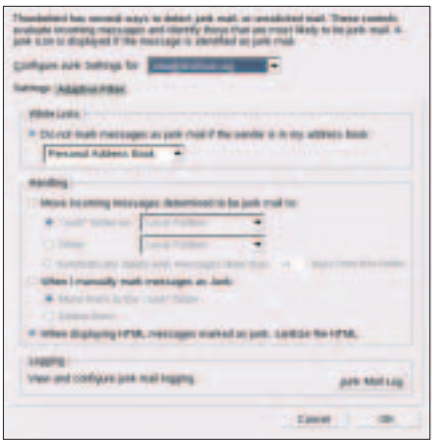

Figure 5: Setting up the integrated junk filter in the Thunderbird mail client.

select *Regex Match*, and type a dot in the text field. This is a rule to handle any incoming mail, just like the one we created for Kmail. In the bottom part of the window, select the *Pipe Message to Shell Command* item. Then type */usr/bin/spamc* in the box to the right.

Using the same steps, now set up a second command called *Move Spam*; this will use the *Specific Header* criterion, *X-Spam-Flag*, and *YES* as the content (Figure 3).

Low Budget

If you prefer to avoid installing an extra program and setting up filters in the mail client, the Mozilla offshoot Thunderbird [3] has a simpler approach. Thunderbird has a mechanism that gradually learns what a user considers spam.

The configuration starts with *Tools | Junk Mail Controls | Settings* (Figure 4). In this tab, you need to disable the option for *Move incoming messages determined to be junk mail to:* to allow you to train the integrated filter.

Now close the dialog and click to download your mail from the server. Thunderbird will tag all your email as junk by placing a trashcan icon next to the date. Now go through the list carefully and click to remove the icon for any legitimate messages. The next time you test Thunderbird, the match rate should be far better. When you are happy with the training results, change the setting in the junk filter dialog to tell Thunderbird to move spam messages to the trash folder.

Lifetime Assignment

A client-based spam filter requires you to download spam email messages before filtering, but if you do not have direct access to the server, this is still your best option.

SpamAssassin references updated lists on the Internet, and its detection rate is quite high due to the number of tests it applies. SpamAssassin has a lot of potential, allowing you to add new rules and feed junk mail to the program to train it. The SpamAssassin manpage *sa-learn* gives you more information. ■

INFO

[1] *http://www.spamhaus.org/sbl/*

[2] Integrated tests in SpamAssassin 2.6 *http://spamassassin.apache.org/tests_2_6x.html*

[3] Thunderbird email client : *http://www.mozilla.org/products/thunderbird/*