

Spamassassin at the Tech Computer Center



Bryan Hughes

2008-01-02 18:55

Abstract

User procedures for reducing unsolicited e-mail.

This publication is available in Web form¹ and also as a PDF document². Please forward any comments to tcc-doc@nmt.edu.

Table of Contents

1. What is <i>Spamassassin</i> ?	1
2. How <i>Spamassassin</i> works: your spam score limit	1
3. What you can do using the TCC mail filter	2
4. When the TCC mail filter is not enough: filtering with <i>procmail</i>	2
4.1. Writing your own spam test	4
4.2. Example of a <code>user_prefs</code> file	5
4.3. Example of a <code>.procmailrc</code> file	6

1. What is *Spamassassin*?

Spamassassin is a package that can help you filter spam (unsolicited commercial e-mail) out of your incoming mail.

2. How *Spamassassin* works: your spam score limit

The TCC's current implementation of *Spamassassin* is through a sendmail mail filter. The mail filter checks to see if the incoming mail is spam or not. Unlike other implementations of *Spamassassin*, the mail filter will not rewrite the mail's header.

When *Spamassassin* looks at your mail, it is assigned a score based on tests it runs on the mail. If the score exceeds a number called the *spam score limit*, the mail is considered spam and will not be delivered to your mailbox. The TCC's default spam score limit is 10.

The filtering that the TCC does may be sufficient for you. If not, see Section 4, "When the TCC mail filter is not enough: filtering with *procmail*" (p. 2).

¹ <http://www.nmt.edu/tcc/help/pubs/spamassassin/>

² <http://www.nmt.edu/tcc/help/pubs/spamassassin/spamassassin.pdf>

3. What you can do using the TCC mail filter

You can edit your `.spamassassin/user_prefs` file to adjust some spam settings.

- You can add whitelists and blacklists. Use a *whitelist* to identify the “good guys,” senders that you definitely want to see. Use a *blacklist* to identify sites you consider spam.

These will assign a -100 and +100 score respectively. Here are some examples:

```
whitelist_from *.nmt.edu
blacklist_from spammer@spam.comments
```

Use the `*` character as a “wild card” to match parts of the address.

- You can change the spam score limit by adding to your `.spamassassin/user_prefs` a line like this, where the number is the spam score limit value you want:

```
required_hits 7.5
```

Your `.spamassassin/user_prefs` file will have more examples of whitelists and blacklists. See also Section 4.2, “Example of a `user_prefs` file” (p. 5).

4. When the TCC mail filter is not enough: filtering with *procmail*

You can invoke `spamassassin` from *procmail* to have greater control over your spam settings. Using *procmail* to invoke `spamassassin` will tag your email and allow it to be filtered. The *procmail* step happens after the `spamassassin` mail filter is run, and mails dropped by the mail filter will not make it to *procmail*.

Note

To disable the mail filter for your account, create an empty file called `.disobeyspamassassin` in your home directory. On Linux systems, this command will do create that empty file:

```
touch ~/.disobeyspamassassin
```

There are now two options for invoking `spamassassin`:

1. If you do not want to write your own tests, but only have your mail tagged for filtering, then you will want to invoke `/usr/bin/spamc` in your `.procmailrc` file.
2. If you have the need to write your own spam tests (which are specified as a Perl regular expression), then you will want to invoke `/usr/bin/spamassassin` in your `.procmailrc` file. This method is much slower than the next one, and in some instances will cause the sender's mail client to time out.

If you want to write your own tests, edit your `.procmailrc` file, remove any content you didn't put there, and add this line:

```
:0fw: 1
|/usr/bin/spamassassin 2
```

- 1** • “:0” is the start of a new *procmail* recipe.

- “f” is a flag that tells *procmail* to filter the mail through the action (second) line.
- The “w” flag tells *procmail* to wait.
- The trailing “:” tells *procmail* to use a lock file.

2 The second line is the *action line* of the recipe. It tells *procmail* to pipe the mail into the program “/usr/bin/spamassassin”

If you don't want to write your own rules, just have your mail tagged by using this action line instead:

```
|/usr/bin/spamc
```

Procmail can also be used to sort your mail. Examples can be found in the manual pages for `procmailrc`³ and `procmailrc`⁴ on TCC computers running Linux.

More information about procmail can be found on the `procmail.org` homepage⁵.

The invocation of spamassassin in your `.procmailrc` file opens up several new ways to handle spam. The whitelist, blacklist and score changes from the mail filter section apply here also.

1. The header of your mail will now be tagged with spam information. In SquirrelMail, clicking on *view full header* when a message is open will show the mail header with *Spamassassin* tags.
2. If the mail is considered spam, the body will be rewritten with *Spamassassin* information and the original message will be included as an attachment.
3. You can change the `required_hits` variable in `.spamassassin/user_prefs`.
4. You can write your own spam tests.

When tagging your header, *Spamassassin* will add three or four lines that look like this:

```
X-Spam-Flag: YES
```

will be added if the spam score is greater than your `required_hits`.

```
X-Spam-Checker-Version: SpamAssassin 2.64 (2004-01-11) on mailhost.nmt.edu
```

This shows the version and host information.

```
X-Spam-Level: *****
```

A string of asterisks (*) will be added to show the integer part of the spam score. For example, a spam score of 7.3 will have seven asterisks.

```
X-Spam-Status: Yes, hits=7.7 required=5.0 tests=FORGED_MUA_OUTLOOK,
HTML_30_40,HTML_FONTCOLOR_RED,HTML_FONT_INVISIBLE,HTML_MESSAGE,
HTML_TAG_BALANCE_A,MIME_BASE64_LATIN,MIME_BASE64_TEXT,MIME_HTML_ONLY,
MSGID_FROM_MTA_HEADER,RCVD_IN_NJABL_DUL,RCVD_IN_SORBS_DUL
autolearn=no version=2.64
```

This tag shows which tests scored on this mail, showing the total spam score and your effective `required_hits` value.

³ <http://www.nmt.edu/bin/man?procmailrc>

⁴ <http://www.nmt.edu/bin/man?procmailex>

⁵ <http://www.procmail.org/>

Mail that has a score greater than your `required_hits` will be marked as spam and the body will be rewritten similar to this example:

```
----- Start SpamAssassin results -----
This mail is probably spam. The original message has been altered
so you can recognize or block similar unwanted mail in future.
If this message is NOT spam please forward it to postmaster@nmt.edu
and we will correct our filters not to block this any more. If you
would like to know more about SpamAssassin please see:
http://spamassassin.org/tag/ for details.

Content analysis details: (17.3 hits, 5.0 required)
0.3 NO_REAL_NAME From: does not include a real name
2.8 FORGED_MX_HOTMAIL Forged hotmail.com Received 'from mx' header
1.4 RCVD_FAKE_HELO_DOTCOM Received contains a faked HELO hostname
1.2 DEAR_SOMETHING BODY: Contains 'Dear (something)'
3.0 ROUND_THE_WORLD_LOCAL Received: says mail sent around the world (HELO)

2.0 SUBJ_ALL_CAPS Subject is all capitals
0.0 FORGED_HOTMAIL_RCVD Forged hotmail.com 'Received:' header found
0.1 RCVD_IN_SORBS_DUL RBL: SORBS: sent directly from dynamic IP address
[165.165.20.168 listed in dnsbl.sorbs.net]
1.6 RCVD_IN_NJABL_DUL RBL: NJABL: dialup sender did non-local SMTP
[165.165.20.168 listed in combined.njabl.org]
4.3 CONFIRMED_FORGED Received headers are forged
0.8 MSGID_FROM_MTA_HEADER Message-Id was added by a relay
----- End of SpamAssassin results -----
```

4.1. Writing your own spam test

The *Spamassassin* documentation⁶ goes deeper into the types of tests you can write. However, one type of test will be covered here.

This test will check line by line of the entire raw body of the mail for what we describe in the regular expression.

Here's an example. Spam will usually forge the `From:` line in mail. Usually a spammer will change the `From:` line each time they send you spam. Suppose you notice that the mail always has a hyperlink pointing to “`http://www.blatantspam.com`”. Here is a raw body test looking for this url.

```
rawbody BLATANT_SPAM /http:\\\\www\\.blatantspam\\.com 1
describe BLATANT_SPAM Spam from blatantspam.com, added 01 January 2005 2
score BLATANT_SPAM 10 3
```

- 1** The `rawbody` line describes the type of test we are running: a test that searches the body of the mail for a given pattern. The new test will be called `BLATANT_SPAM`. The rest of the line is a Perl-style regular expression specifying the pattern to be matched.
- 2** The `describe` line documents what the test is for, including the date this rule was added.
- 3** The `score` line specifies the score adjustment to be added when this pattern is found in the message body.

⁶ <http://spamassassin.apache.org/doc.html>

4.2. Example of a user_prefs file

Here is an example of the `.spamassassin/user_prefs` file in one spam-filtering solution. This example shows the user-defined rule described in the previous section.

```
# SpamAssassin user preferences file. See
# 'perldoc Mail::SpamAssassin::Conf' for details of what can be
# tweaked.
#
# Lines starting with a # are comments and will not be read
# by spamassassin
#####

required_hits 10

# Whitelist and blacklist addresses are now file-glob-style patterns, so
# "friend@somewhere.com", "*@isp.com", or "*.domain.net" will all work.
# whitelist_from someone@somewhere.com

whitelist_from *nmt.edu
blacklist_from spammer@spam.com

# Add your own customized scores for some tests below. The default scores
# are read from the installed spamassassin rules files, but you can
# override them here. To see the list of tests and their default scores,
# go to http://spamassassin.org/tests.html.
#
# score SYMBOLIC_TEST_NAME n.nn

score DATE_MISSING 10
score TO_EMPTY 5
score NIGERIAN_SUBJECT6 7.4

# Speakers of Asian languages, like Chinese, Japanese and Korean,
# will almost definitely want to uncomment the following lines.
# They will switch off some rules that detect 8-bit characters,
# which commonly trigger on mails using CJK character sets,
# or that assume a western-style charset is in use.
#

score HTML_COMMENT_8BITS 0
score UPPERCASE_25_50 0
score UPPERCASE_50_75 0
score UPPERCASE_75_100 0

# User defined tests here
#

rawbody BLATANT_SPAM /http:\\\\www\\.blatantspam\\.com/i
describe BLATANT_SPAM Spam from blatantspam.com, added 01 January 2005
score BLATANT_SPAM 10
```

4.3. Example of a .procmailrc file

This is an example of a .procmailrc file that goes with the preceding .spamassassin/user_prefs file.

```
:Ofw:
|/usr/bin/spamassassin

# Catch SPAM
:0:
* ^X-Spam-Flag: YES
$HOME/spam
```

This will create a file in your home directory called **spam**. Any mail flagged with a “X-Spam-Flag: YES” header line will be sent to this file.

In SquirrelMail, you can add this to your folders by clicking on *Folders* (next to *addresses* and *options* at the top). In the *Unsubscribe/Subscribe* section find your **spam** file, select it, and click on *subscribe*. After you refresh the page, **spam** will be listed as a mail folder on the left.