The Sysadmin's Daily Grind: Spamstats

Dirt Tracker

The popular Spamassassin tool reliably filters genuine messages from the usual digital garbage. The Spamstats program monitors the spam killer and reports on its activities, enabling you to actively tailor the filters and monitor spam trends. BY CHARLY KÜHNAST

Although there is legislation to prevent spam, this has had no noticeable effect on the tons of unsolicited mail that hit most mailboxes. Most admins use Spamassassin [1] to defend their networks, and the tool performs very well. Unfortunately, that means you need a tool to monitor the spam killer's activities: enter Spamstats.

Spamstats [2] creates precise statistics for email handling and calculates the spam quota. The compressed package weighs in at a mere 13.5 Kbytes, and provides a very readable *README* that tells you all about the way Spamstats works. The script itself is written in Perl, and requires the Perl modules Getopt::Long and Compress::Zlib. If you do not have these, you should surf to CPAN to obtain them:

```
perl -MCPAN -e 'shell'
cpan> install Compress::Zlib
cpan> install Getopt::Long
```

Following this, a call to

```
/usr/local/bin/spamstats➥
0.4b5.pl -help
```

outputs a short overview of the parameters that Spamstats accepts. Only one of these parameters is really necessary, and that is *-f* /path/logfile.

In the current version, Spamstats reads logfiles created by Exim, Postfix, and Sendmail, in combination with Spamassassin's *spamd*. Qmail support is planned. Spamstats can handle multiple logs at the same time; they can even be compressed. That means the following syntax

```
/usr/local/bin/spamstats➥
0.4b5.pl -f /var/log/mail /var/➥
archive/altmail.gz
```

is perfectly okay. I tried this out with my own Postfix/Spamassassin – but only for the current log that contains new entries created after 0:00 hours – this produced the output shown in Figure 1.

Good Morning

Hey, 23 percent spam quota – that's less than my normal average! But then again it is Monday morning, so the credit card vendors and "cheapest Vi*gr* suppliers" will not be online yet.

If you want to publish your Spamstats on the Web, you can enable HTML output with the *-html* flag. And there is another gimmick: If you ask, Spamstats will tell you which email accounts have been hit hardest by spammers. For example, if I want to know which three accounts have been hit hardest by spammers, I can simply type

```
/usr/local/bin/spamstats➥
0.4b5.pl -f /var/log/mail➥
 -number 3
```

In my case, the winner is the account I use for Usenet posting – and that is no big surprise. Maybe I should feed the data I collected to RRDTool to visualize my spam trends over a longer period of time? But then again, the results might be depressing. ■

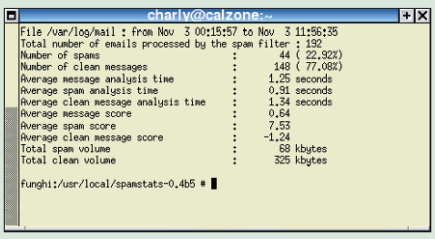

Figure 1: 23 percent of all received messages are spam

INFO

[1] Spamassassin: *http://eu3.spamassassin.org*

[2] Spamstats: *http://www.gryzor.com/tools/#spamstats*

THE AUTHOR

Charly Kühnast is a Unix System Manager at the datacenter in Moers, near Germany's famous River Rhine. His tasks include ensuring firewall security and availability and taking care of the DMZ (demilitarized zone). Although Charly started out on IBM mainframes, he has been working almost exclusively with Linux since 1995. To stay in shape he tries to get in some karate training on his leisure time.