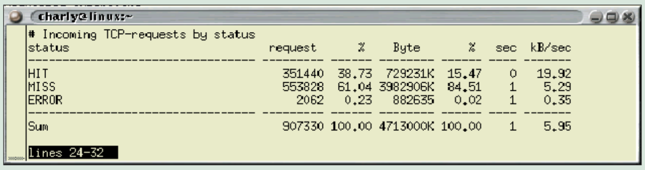
The Sysadmin's Daily Grind: Calamaris

Fishing for Squid

The Squid proxy logs its activities in a fairly illegible logfile. The Calamaris analysis tool uses its tentacles to dig up relevant information from this file and serves the statistics up to the admin – in HTML format, if required.

BY CHARLY KÜHNAST

Anyone needing to analyze the logfiles produced by the popular Squid [1] proxy server, will really appreciate a tool such as Calamaris [2] by Cord Beermann. The squid tool is written in Perl and not only tackles Squid logfiles, but also those produced by NetCache, Inktomi Traffic Server, Cisco Content Engine and a few others. Calamaris extracts cache efficiency, performance and load statistics for the proxy from the logfile.



```
# Incoming TCP-requests by status
status                                request      %      Byte      %   sec  kB/sec
----------------------------------- --------- ------ --------- ------ ---- -------
HIT                                    351440  38.73   729231K  15.47    0   19.92
MISS                                   553828  61.04  3982906K  84.51    1    5.29
ERROR                                    2062   0.23    882635   0.02    1    0.35
----------------------------------- --------- ------ --------- ------ ---- -------
Sum                                    907330 100.00  4713000K 100.00    1    5.95

lines 24-32
```

Figure 1: Calamaris provides a compact report if you do not specify any parameters

To try this out, I simply fed Calamaris a logfile without specifying any additional parameters:

```
/usr/local/bin/calamaris < /var↵
/squid/logs/proxy1.access.log
```

After quite a while (during which Calamaris was mainly pre-occupied with asking the nameserver a lot of questions), this friendly denizen of the deep provided me with a fairly compact report. The report told me that Squid had served up about 4.6 GBytes to its clients today, answering 38 percent of all requests from its cache, although this made up only 15 percent of data volume (see Figure 1). The reason for this is that one tends to cache small objects – my Squids will not cache objects of more than 1 Mbyte, for example.

Top Ten Sites

Now I would like to know which URLs are the most popular with my users. For this, I append the *-d 10* parameter:

```
/usr/local/bin/calamaris -d 10 ↵
-n -U M < /var/squid/logs/↵
proxy1.access.log
```

I stipulated *-n* to stop Calamaris sending all those lengthy lookups to the nameserver. This reduces the time required to parse the logfile, which by now contains about a million lines, to three and a half minutes. *-U M* tells the tool to display the volume of data transferred in MBytes. The final report tells me that *www.google.com* is the most popular URL. Now that was a surprise.

Let's take another look at the performance: I want Calamaris to tell me the proxy throughput at 30 minute intervals – the parameter that does this is *-P 30*:

```
/usr/local/bin/calamaris -d 10 ↵
-n -U M -P60 < /var/squid/logs/↵
proxy1.access.log
```

The results were to be expected: The throughput is extremely good before 8:00 a.m. and after 5:00 p.m. and at its worst around 11:00 a.m. and 2:30 p.m. These values are not particularly alarming, but I will keep an eye on them in future.

Now just to spice things up a bit, I would like to publish these reports on my Web server.

Fortunately, Calamaris can produce HTML, either complete HTML documents or using *<html>*, *<head>*, *<body>* and other tags that allow me to add the tool's output to existing HTML documents, as PHP or server side includes. I opted for the latter variant:

```
/usr/local/bin/calamaris -d 10 ↵
-n -U M -F html-embed < /var/↵
squid/logs/proxy1.access.log > ↵
report`date +%Y%m%d`.html
```

The Calamaris manpage lists a variety of additional report functions that you may find useful. Enjoy. ■

INFO

[1] Squid: *http://www.squid-cache.org*

[2] Calamaris: *http://cord.de/tools/squid/calamaris/*

THE AUTHOR

Charly Kühnast is a Unix System Manager at the data-center in Moers, near Germany's famous River Rhine. His tasks include ensuring firewall security and availability and taking care of the DMZ (demilitarized zone).

SYSADMIN