

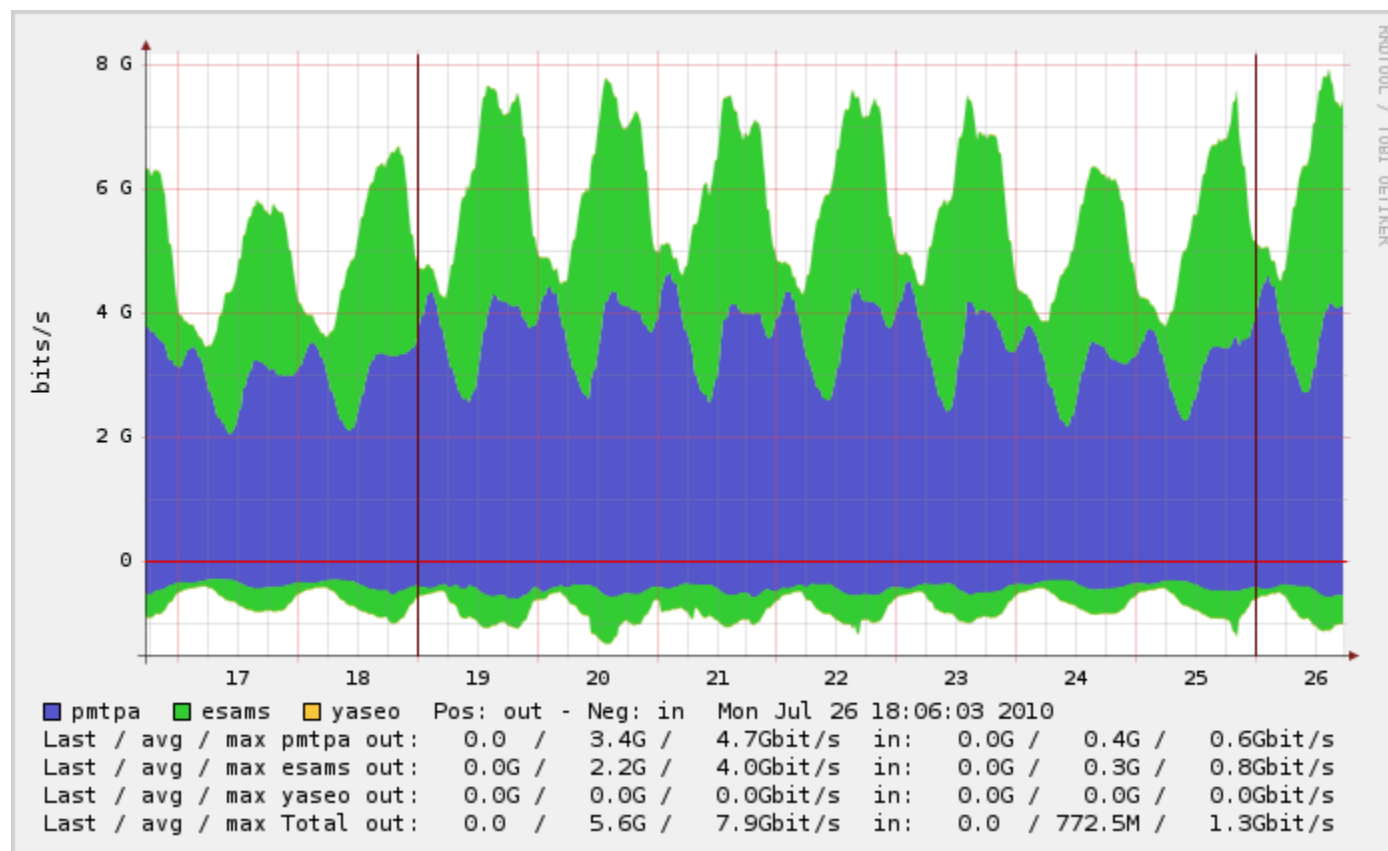
Here is an assessment of how much data got lost at server locke in last 3 months.  
 (we can't look further back, 1:1000 sampled log gets wiped after 3 months for privacy reasons).

Each squid separately numbers its messages. Per definition the average gap between these sequence numbers in the 1:1000 sampled should be 1000. This is an average for all squids, some squids might send more often than others.

The chart above shows how much the average gap really was, per week, per time of day.

As expected the loss varied per hour of day, depending on overall variation in traffic volume.

Compare chart below: <http://www.nedworks.org/~mark/reqstats//trafficstats-weekly.png>



July 22 server overload was fixed and around 14.00 hrs GMT message loss vanished almost entirely

In week 23 and 26/27 several unrelated abnormalities caused extra loss of data [EZ: details omitted for public version]

Ignoring these special weeks, for the remaining 10 'normal' weeks average gap between sequence numbers varied from 1247 to 1395, overall average for those 10 weeks was 1314.

So out of a 1,000,000 consecutive requests 761 (1,000,000/1314) got logged where it should have been 1000. So page views was actually 31% higher than the raw counts of the 1:1000 sampled log suggest.

The average monthly gap for days that were

- still available in the log, and before the problem was fixed (24/4 - mid 22/7)
- not on a day where an extra anomaly occurred (see previous mail)  
(these days were already ignored fully, and monthly totals extrapolated from remainder)

is as follows:

April avg gap is 1241 (+24.1% to be added to earlier April page views )

May avg gap is 1310 (+31.0% to be added..)

June avg gap is 1328 (+32.8% to be added..) July avg gap is 1470 (+47.0% to be added for first 21.5 days in July, and +0.03% for remaining 9.5 days) -> +29.5% for full month