Analysis of Circuit Queues in Tor

Karsten Loesing

August 25, 2009

1 Background on queues and buffers in Tor

Whenever a relay receives a cell, this cell is appended to the outbound queue of the corresponding circuit. As soon as there is room in the outgoing buffer of the connection to the next relay in the circuit, the cell is removed from the queue and the cell body is written to that buffer. The bytes reside in the outgoing connection buffers until they can be sent over the TLS socket to the previous or next relay in the client's circuit. Multiple circuits can share the same connection, so that the cells of multiple queues can be waiting to be written to a single outgoing buffer.

This analysis focuses on circuits and the cells that are kept in their queues, rather than on connections to other relays and bytes waiting in their buffers. Possible changes to circuit queues include reducing queue lengths to reduce the latency of sending cells and changing the scheduling of processing cells competing for the same connection. Future analyses might also focus more on connections.

The graphs in this report are based on statistics gathered by a few relays starting on July 20, 2009. Table 1 lists the relays including their exit configuration, authority status, and operator name.

2 Format of measured data

The data format of the graphs in this report is as follows:

```
written 2009-07-25 00:24:27 (86400 s)
processed-cells 5978,223,56,22,9,5,5,4,3,1
queued-cells 5.40,0.60,0.03,0.01,0.00,0.00,0.00,0.00,0.00,0.01
time-in-queue 1919,1838,127,116,124,99,47,58,111,447
number-of-circuits-per-share 15351
```

Every 24 hours, these five lines are appended to a local file. The `written` line contains the end time of the measurement interval as well as the interval length in seconds. The `processed-cells` line contains the number of processed cells by deciles in descending order from loudest to quietest circuits. The `queued-cells`

Table 1: Relays measuring cell statistics

Nickname	Exit	Authority	Operator
echelon1	yes	no	Karsten Loesing
echelon2	yes	no	Karsten Loesing
ephemer2	no	no	Steven J. Murdoch
gabelmoo	no	yes	Karsten Loesing
gonzales	no	no	Karsten Loesing
hamsterrad	no	no	Karsten Loesing
ides	no	yes	Mike Perry
moria1	no	yes	Roger Dingledine
moria2	no	yes	Roger Dingledine
nottheNSA	yes	no	Andrew Lewman
trusted	no	no	Jacob Appelbaum
vallenator	no	no	Hans Schnehl

line denotes the mean number of cells contained in queues of the circuit deciles. The `time-in-queue` line contains the mean time that cells spend in circuit queues in milliseconds. Finally, `number-of-circuits-per-share` states how many circuits are included in every decile.

3 Total number of circuits

The first data set that can be visualized from the statistics is the number of circuits that can be derived from the `number-of-circuits-per-share` line. Figure 1 shows the number of circuits over time periods of 24 hours. The graph shows that the number of circuits varies from relay to relay, mostly depending on the relays' advertised bandwidth.

4 Classification of circuits by processed cells

The total number of circuits does not reflect how loud or quiet these circuits are. Therefore, circuits are classified by the number of cells that they have processed as written in the `processed-cells` line. Figure 2 shows the classification result: the loudest 10% of all circuits have pushed more than 10,000 cells in the mean, the quietest 10% only a single cell. The monotonic decrease in this graph can be explained from the definition of louder circuits processing more cells than quieter ones.

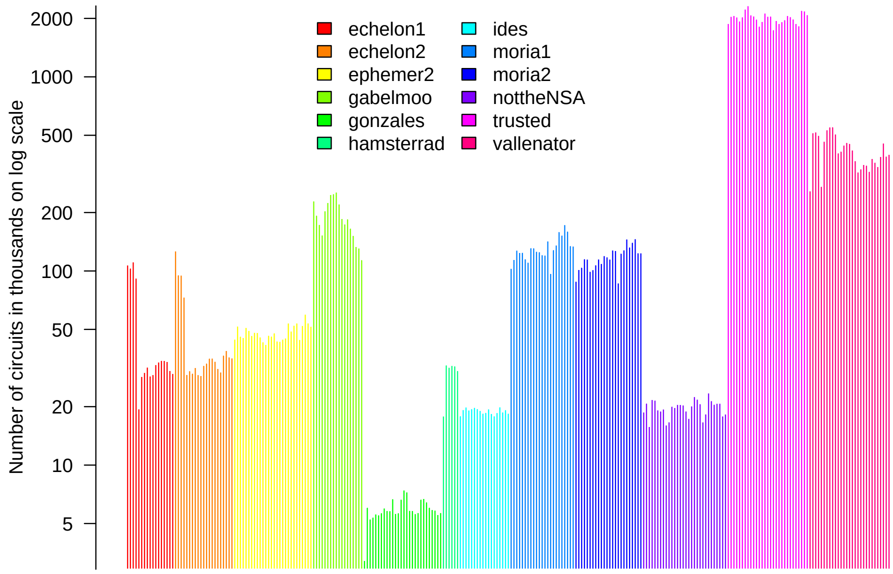

Figure 1: Total number of circuits per day

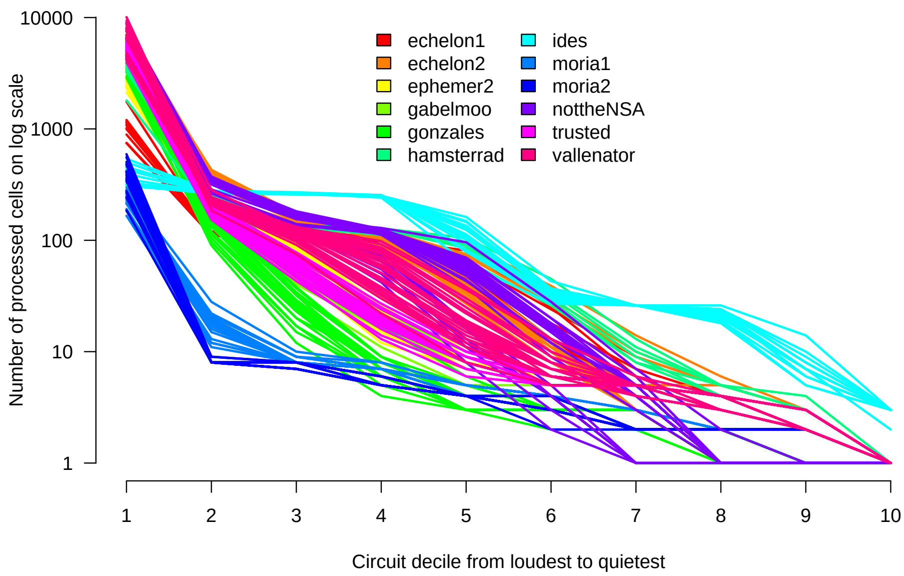

Figure 2: Mean number of processed cells per circuit

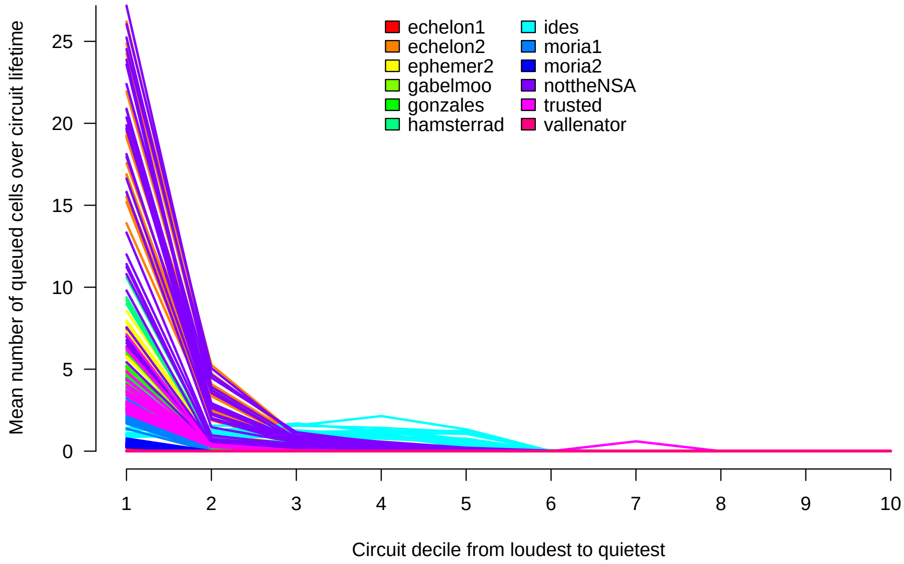

Figure 3: Mean number of cells in queues

5 Queued cells by circuit loudness

Even though it is interesting to learn about the classification of circuits by their loudness, there is not much to be changed or improved. Whether or not a circuit is loud depends only on the usage behavior. It is questionable if this behavior can be changed (for example by discouraging high-volume applications), and if so, only in the long term.

The mean number of queued cells as a function of circuit loudness is a more interesting metric here. These numbers are contained in the `queued-cells` line. Figure 3 shows these numbers over the measurement interval of 24 hours. The loudest 10% of all circuits have up to 27 cells in their queues in the mean, the 10–20% loudest only up to 5, and so on. The intuition is that louder circuits have (and should have) their cells queued for a bit while quieter circuits have their cells delivered quickly rather than waiting in queues.

6 Mean time cells spend in queue

The mean time that cells spend in circuit queues is probably the most important metric here. This waiting time directly contributes to the latency that users experience. It would be good to have a short latency for the quieter circuits. Figure 4 shows the mean time of cells by circuit deciles. The general trend is that the quieter the circuits are, the less time cells spend in queues, although there are exceptions to this trend.

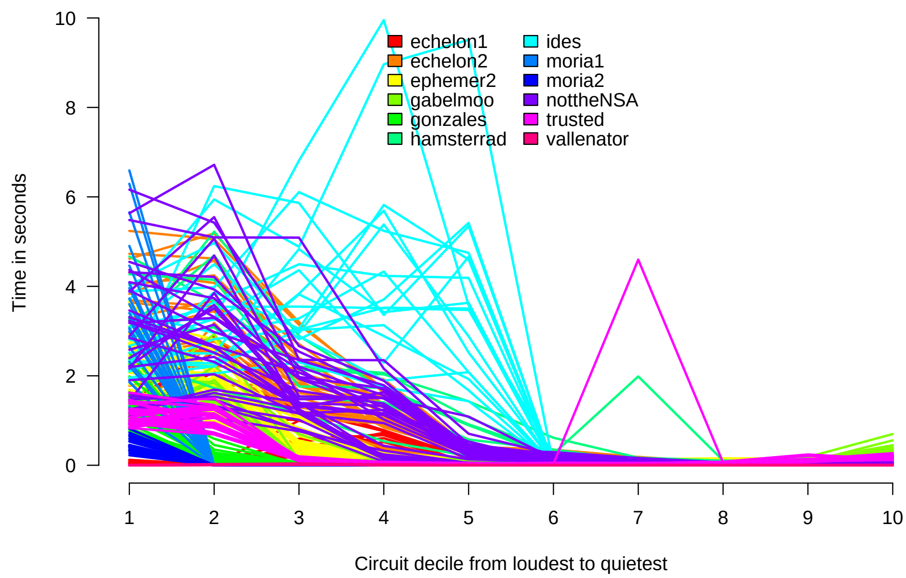

Figure 4: Mean time that cells spend in circuit queues

7 Conclusion

This analysis has shown a few characteristics of circuit queues in Tor. Results include total numbers of processed cells, a classification of circuits by the number of processed cells, the mean number of cells in queues, and the mean time that cells spend in circuit queues. If possible, the waiting times for the quieter circuits should be reduced, even at the cost of increasing the waiting time for the loudest circuits. The rationale is that the loudest circuits do not use Tor for low-latency applications anyway, but for high-volume applications. The result would be that Tor becomes more attractive for interactive applications. This analysis permits measurement of the effectiveness of design changes in the future.