



positive in the region  $\mu_i > K_i$ . In addition

$$H''_{i,j}(a, b) = \frac{(\mu_i - K_i)^2 + K_i(\mu_i - K_i)}{\Lambda \mu_i (\mu_i - K_i)^3}, \quad a = b = j \quad (15)$$

$$H''_{i,j}(a, b) = \frac{0.5(\mu_i - K_i)^2 + 0.5K_i(\mu_i - K_i)}{\Lambda \mu_i (\mu_i - K_i)^3}, \quad a \neq b = j || b \neq a = j \quad (16)$$

$$H''_{i,j}(a, b) = 0, \quad a, b \neq j \quad (17)$$

The matrix  $H'''_i$  is defined by summing  $H''_{i,j}$  over all  $j$

$$H'''_i = \sum_{j=0}^{N-1} H''_{i,j} \quad (18)$$

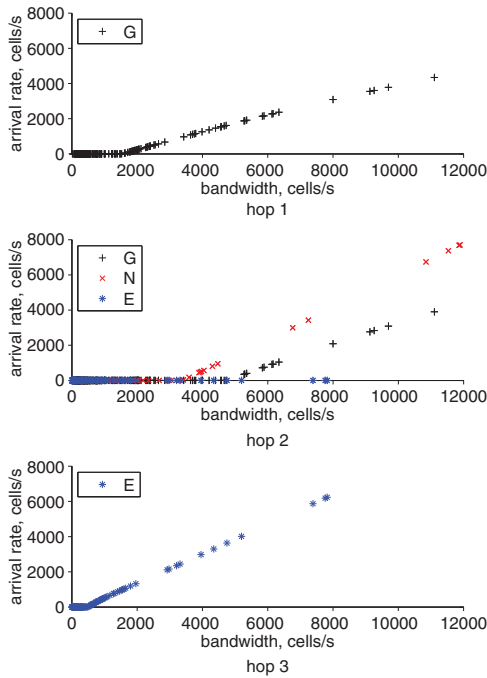
Noting that each term on the leading diagonal will equal (15) for exactly one of the matrices  $H''_{i,j}$  and 0 for all the others, and that all terms not on the leading diagonal will equal (16) for two of the matrices  $H''_{i,j}$  and 0 for all the others

$$H'''_i(a, b) = \frac{(\mu_i - K_i)^2 + K_i(\mu_i - K_i)}{\Lambda \mu_i (\mu_i - K_i)^3}, \quad \text{all } a, b \quad (19)$$

which is positive semi-definite, because each element is the same and positive in the region  $\mu_i > K_i$ . Finally, expressing the matrix  $H_i$  as the sum of the matrix  $H'''_{i,j}$  and the matrices  $H''_{(i,j)}$ , i.e. by using (9), (14) and (19)

$$\begin{aligned} H'_i &= \sum_{j=0}^{N-1} H'_{i,j} \\ &= \sum_{j=0}^{N-1} H''_{i,j} + \sum_{j=0}^{N-1} H'''_{i,j} \\ &= \sum_{j=0}^{N-1} H''_{i,j} + H'''_i \end{aligned} \quad (20)$$

which is positive semi-definite, because it is the sum of positive semi-definite matrices, as shown in (14) and (19). Note that the inclusion of the linear equality constraints (2) and (4) does not affect this property.



**Fig. 1** Optimal arrival rates for snapshot of Tor nodes at 50% usage

*Numerical example for Tor:* The theory is applied to a snapshot of the Tor network. Usage is assumed to be 50%, as is the typical loading of Tor [3]. Fig. 1 shows the optimal node arrival rates for each node type: guard (G), normal (N) and exit (E) for each of the three hops. This leads to an expected cell latency of 1.3 ms compared to 11.7 ms using the original node probability weightings. Note that many of the lower bandwidth nodes have zero arrival rate (i.e. zero probability of being chosen). Intuitively, this can be understood by considering that the minimum waiting time at these nodes (i.e. the service time) is greater than the expected waiting time for the remainder of the node population. Note that the plots in Fig. 1 have a similar shape to that of the one-hop network [3, Fig. 1], providing further evidence that this is likely to be the global minimum.

As well as demonstrating that the proposed method for assigning node selection probabilities significantly outperforms the current method, it is also necessary to consider how robust the proposed method is. This can be achieved by evaluating the expected cell latency for varying usages, given that the selection assignment probabilities have been optimised for 50% usage (i.e. to mimic the situation where the network status varies and the assignment probabilities have not been updated accordingly). This leads to the result that the proposed method optimised for 50% usage outperforms the current method between 0 and 60% usage; however, for usages  $> \sim 62\%$  the arrival rate exceeds the service rate for some nodes, and thus the expected cell latency tends to infinity.

*Conclusion:* A general solution to minimising the expected cell latency in multi-hop M/D/1 queuing networks has been derived, and it has been shown that the optimisation surface is convex. This has been applied to the Tor anonymity network at 50% usage, and it has been shown that the expected cell latency can be reduced from 11.7 ms with the original node selection probability method (i.e. the node selection probability is proportional to its bandwidth) to 1.3 ms with the proposed method. Furthermore, it has been shown that the proposed method leads to a reduced latency even if the usage is not known exactly.

The derivation has assumed that nodes have infinite queue capacity, which is not the case in reality. Therefore, it would be interesting to run a network simulation with nodes with finite buffer queues to verify that the proposed method would actually lead to improved results in an actual network.

© The Institution of Engineering and Technology 2014

12 June 2014

doi: 10.1049/el.2014.2136

One or more of the Figures in this Letter are available in colour online.

S.J. Herbert, S.J. Murdoch and E. Punskaya (*University of Cambridge, Cambridge, United Kingdom*)

E-mail: Steven.Herbert@cl.cam.ac.uk

## References

- 1 Dingleline, R., Mathewson, N., and Syverson, P.: 'Tor: the second-generation onion router'. USENIX Security Symp., 2004, p. 21
- 2 Murdoch, S., and Watson, R.: 'Metrics for security and performance in low-latency anonymity systems'. Int. Symp. on Privacy Enhancing Technologies, 2008, pp. 115–132
- 3 Dingleline, R., and Murdoch, S.: 'Performance improvements on Tor, or why Tor is slow and what we're going to do about it'. Technical Report, 2009
- 4 Kendall, D.: 'Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain', *Ann. Math. Stat.*, 1953, pp. 338–354
- 5 Khinchin, A.: 'Mathematical theory of a stationary queue', 1932