WAVELENGTH CONTINUOUS WDM NETWORKS WITH NON-UNIFORM SERVICE RATES

Lachlan L. H. Andrew
ARC Special Research Centre for Ultra-Broadband Information Networks (CUBIN),
Department of Electrical and Electronic Engineering, University of Melbourne, Vic 3010, Australia.

Abstract

Many circuit switched networks have stationary state probabilities which depend on the arrival and service rates only through their ratio. This paper shows that this is not the case for WDM networks without wavelength conversion, and for cellular mobile networks, and investigates the implications for fast simulation techniques.

1 Introduction

It is commonly assumed that the stationary state distribution of a circuit switched network with Markov traffic (Poisson arrivals, exponential holding times) is uniquely determined by the load in Erlangs, ρ_i, offered to each route i, independent of the specific arrival and service rates. If the service rate is equal for all routes then the insensitivity follows by simply scaling in time. Moreover networks with Poisson arrivals and "reversible" channel assignment [1] have a closed form solution in terms of ρ_i. However, there are many networks which are not of these types, where the insensitivity does not apply.

Today's time division multiplex (TDM) networks with time-slot interchange are reversible, as are random channel assignment (which performs poorly [2]) and maximum packing [3] (which is unimplementable). However, with real channel assignment, neither cellular networks (future access networks) nor wavelength continuous wavelength division multiplexing (WDM) networks without wavelength conversion (future backbones) are reversible.

This paper quantifies the effect on the blocking probabilities of increasing the assumed arrival and service rates of a "focus" subset of the routes. If this change causes little change in the probabilities, simulating the former can reduce the computation required to find the blocking probability on a given route; increasing the rates of the focus set decreases the proportion of processing spent on other routes. This is particularly important for simulation schemes such as that of [4], in which each simulation is limited to evaluating the blocking on a single route.

In [5], hybrid packed/circuit networks operating on multiple timescales were studied, and it was found that the results only differed by a few percent when the ratio of timescales was changed. The present work differs from this by showing that the difference is significantly smaller in certain circumstances.

2 Motivation

In [4], importance sampling [6] was applied to circuit switched networks with fixed routing and non-reversible channel allocation. The algorithm is based on the concept of quasi-regenerative cycles, and used the fact that

$$B_i = \frac{\mathsf{E}[X_i(T^{(i)})]}{\mathsf{E}[T^{(i)}]}, \tag{1}$$

where B_i is the probability of blocking on route i, $T^{(i)}$ is the length of a quasi-regenerative cycle for route i, and $X_i(T^{(i)})$ is the amount of time within a quasi-regenerative cycle that calls to route i would be blocked. In order to estimate $X_i(T^{(i)})$, the network was simulated is simulated with "twisted" arrival and service rates for all routes which intersect route i. Call those routes which intersect route i a "cluster". Let μ be the service rate, common to all routes, let λ be the *sum* of the arrival rates to all routes within the cluster, and let

$$\lambda^*(s) = \lambda + s(\mu - \mu^*(s)), \tag{2}$$

$$\mu^*(s+1) = \frac{\lambda\mu}{\lambda^*(s)} \tag{3}$$

for $s \geq 1$, starting with $\mu^*(1) = 0$. At the start of a quasi-regenerative cycle, when there is exactly one call in the cluster, the algorithm "accelerates" the blocking of route i, until a blocking state is reached, after which it allows simulation to continue as normal until the end of the quasi-regenerative cycle. During acceleration, when the total number of calls in the cluster is s, the arrival and departure rates of routes in the cluster are scaled by factors of $\lambda^*(s)/\lambda$ and $\mu^*(s)/\mu$, respectively.

The changes of arrival and departure rate introduced by the acceleration are compensated exactly by multiplying by the appropriate likelihood ratio. (See [4] for details.) However, an improvement in simulation efficiency may be possible by additionally altering the arrival and departure rates of routes outside the cluster. To understand this, notice that the above changes to arrival and departure rates depend on the particular route which is being accelerated. Although a single simulation, without importance sampling, can be used to estimate all of the $\mathsf{E}[T^{(i)}]$s, a separate simulation is required to evaluate $\mathsf{E}[X_i(T^{(i)})]$ for each different route, i.

Because each separate simulation is only interested in a small section of the network, it seems inefficient to simulate the dynamics of the entire network. Let C be the number of routes in the cluster, and N be the total number of routes in the network. If the arrival rate is equal for all routes, then the number of events simulated could be reduced by a factor of

$$\frac{C + (N - C)/f}{N}$$

if both the arrival and departure rates of the $N - C$ routes outside the cluster were scaled down by a factor of f. The question then arises, what impact would this have on the value of $X_i(T^{(i)})$? The offered load in Erlangs would be unchanged, but since the network is not reversible, $X_i(T^{(i)})$ is changed slightly.

With a re-scaling of the time variable, reducing the rate of the events for routes outside a cluster is equivalent to increasing the rate of events inside the cluster by the same factor. This paper investigates the change of blocking probability due to focusing the simulation on one cluster or one single route. The impact of the following is investigated: the factor by which the service rates differ (focus factor, f), the choice of the focus set, the number of wavelengths per link, Λ, and the load, ρ_i.

3 Focus Factor

The impact of f was measured by simulating a 3×3 mesh torus with $\Lambda = 16$ and using first fit wavelength assignment [2], with $\rho_i = 2.5$ for all routes. Fixed routing was used, with all two link paths being an "upper corner" (Figure 1). When the blocking on a single route, r, is of interest, it is natural to increase the arrival and service rates of route r only. Focus was applied to one horizontal single link route. The impact was quantified as the percentage increase, I, i.e., the percentage by which the blocking probability on r with focusing exceeds that in an identical network without focusing:

$$I = \left(\frac{p(B_{r,\text{focus}})}{p(B_{r,\text{uniform}})} - 1 \right) \times 100\%.$$

This is plotted against f in Figure 2 as the curve "route focus". Rather than increasing steadily with f, I asymptotes to a constant value. This is because of a separation of time-scales, as follows.

Let the state of the network be the row vector $n \in \{0,1\}^{N\Lambda}$, with the N groups of Λ components denoting the occupancy of the Λ wavelengths in use on each of the N routes. Let the "focus" routes be the first $n\Lambda$ components, and let $M = 2^{(N-n)\Lambda}$. Write the $2^{N\Lambda} \times 2^{N\Lambda}$ transition

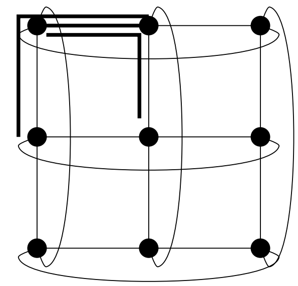

Figure 1. 3×3 mesh torus, with focus routes indicated.

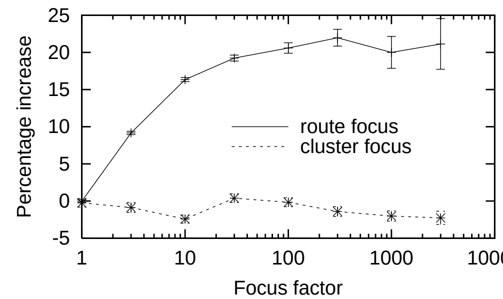

Figure 2. Percentage increase in blocking of one-link route for various focus factors, when the focus is on one route and when it is on the whole cluster. $\Lambda = 16$.

rate matrix of the original network as the block matrix

$$B = \begin{pmatrix} A_1 + D_{11}^1 & D_{12} & \dots & D_{1M} \\ D_{21} & A_2 + D_{22}^1 & \dots & D_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ D_{M1} & D_{M2} & \dots & A_M + D_{MM}^1 \end{pmatrix},$$

where $A_i = (a_{ijk})$ with $a_{ijj} = 0$, and D_{ij} are diagonal matrices, since only one component of n can change at a time. The transition rate matrix of the focussed network is then

$$A^f = \begin{pmatrix} A_1 + D_{11}^f & D_{12}/f & \dots & D_{1M}/f \\ D_{21}/f & A_2 + D_{22}^f & \dots & D_{2M}/f \\ \vdots & \vdots & \ddots & \vdots \\ D_{M1}/f & D_{M2}/f & \dots & A_M + D_{MM}^f \end{pmatrix},$$

where D_{ii}^f is such that the row sums equal to zero.

As f increases, the stationary distribution tends rapidly to $\pi^*(n) = (x_1, x_2, \dots, x_N)$, where x_i is a left eigenvector of $A_i + \text{diag}(d_{i1}, \dots, d_{i,N\Lambda})$, and $d_{ij} = \sum_{k=1}^{N\Lambda} a_{ijk}$. If the network is reversible then $\pi^*(n) = \pi(n)$, the stationary distribution of the original network, but that is not be the case in general. This corresponds to the first $N\Lambda$ components of the state settling into a quasi-equilibrium in between consecutive transitions of the remaining components.

For a fixed number of events, the uncertainty in the blocking of r increases with f for these simulations. This is because the transitions between quasi-equilibria become too slow, resulting in a poor estimate of the probability of each occurring.

4 Choice of Focus Set

When focusing is a means to accelerate the simulation of a network in which the service rates are in fact equal, the change in blocking probability observed above is highly undesirable. Since the blocking on a route is entirely determined by the state of its cluster, it may be hoped that focusing on r's *whole* cluster will reflect more accurately the blocking probability of r in the true network. Note that this is what would naturally be done if focusing were applied to the algorithm of [4]. The curve "cluster focus" in Figure 2 shows the results with the focus placed on the cluster of r. In this particular case, the error introduced by the focusing is very much less than that introduced by focusing on a single route. This is due, in part, to the fact that for each of the M configurations of the non-focus routes, the route of interest can be in any of the 2^Λ possible configurations. However, more work is required to determine the true degree of the approximation, and how generally it may be applied.

The problem of increased variance as f increases is much less than for the route-focused simulations. This is because the blocking probability on r is more similar in different quasi-equilibria, since now most quasi-equilibria include many possible states for the routes intersecting r.

5 Capacity Per Link

To determine the impact of the number of wavelengths on of each link, a two link network was simulated with link capacities from 2 to 45 wavelengths. Focus of $f = 100$ was applied to one single link route and the two link route (i.e., the cluster of one of the single link routes). The arrival rate was increased with the size of the links to maintain a constant utilisation of 0.5 Erlang per wavelength. Note that in this case, the "unfocused" link is the one operating on a different timescale from the rest of the network. Figure 3 shows the percentage decrease in the blocking probability of the focused and unfocused single link routes. As for the 3×3 network, the impact which focusing has on the blocking on a particular route is considerably smaller when all of the routes which intersect with that route operate on the same time scale as it. The impact on the unfocused link increases with Λ, but again the increase is not steady. Rather than approaching a plateau, the variation with Λ is quite complex. There is a degree of symmetry between the impact on the "fast" route and the "slow" route, with a peak in the latter for $15 < \Lambda < 30$ being mirrored in a dip in the former. The size of the jagged peaks for $\Lambda \approx 20$ is approximately one standard deviation of the statistical uncertainty in the measurements. This indicates that they may be more than simply statistical errors, but it is not clear what their origin may be.

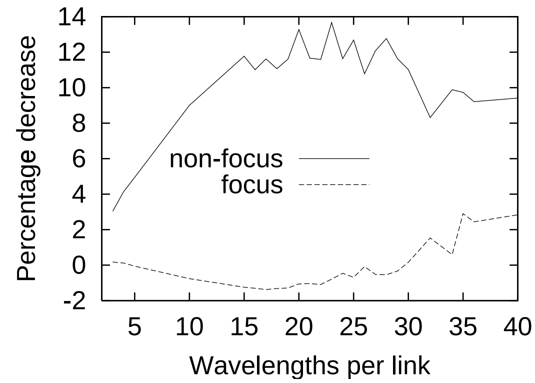

Figure 3. Percentage decrease in estimated blocking of focused and unfocused one-link routes for various link sizes.

6 Load

By keeping the load per wavelength constant while increasing the number of wavelengths per link, the simulations of the previous section produced decreasing blocking probabilities due to the improved trunking efficiency. In order to isolate the effects of reduced blocking probability, further simulations were performed in which the number of wavelengths per link was held fixed while the load per link was reduced. Figure 4 sows the relative error introduced as a function of the blocking probability for a 3×3 network with $\Lambda = 16$ wavelengths per link using a focus factor of $f = 1000$. This clearly shows that the blocking probability affects the error. The estimated blocking probability decreases with decreasing blocking, both using cluster focus and route focus. However, this corresponds to an increase in error for cluster focus and a decrease in error for route focus. For this network, it appears that route focus may actually produce less error than cluster focus when the blocking rate is very low, in contrast to the results found in previous sections. It is not clear why this is the case, or how these results would generalise to other networks. Note that the apparent "elbow" at around 0.0002 blocking in the route-focus curve is not statistically significant, given the accuracy of the simulations.

7 Conclusion

A WDM network with first fit wavelength assignment can have markedly different blocking probabilities, depending on the time scales of different routes, even with purely Markov traffic. The difference increases with link capacity, and with the disparity in time scales, but asymptotes to a constant. The variation for a given route is small if all routes which intersect it also operate on its time scale. This suggests that it may be possible to improve on the performance of the fast simulation algorithm of [4] by slowing the timescale of events in selected parts of the network.

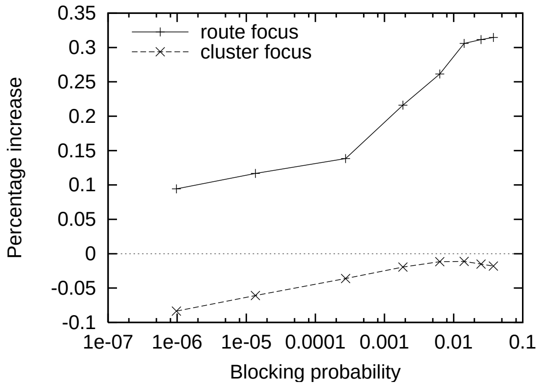

Figure 4. Percentage decrease in estimated blocking of focused and unfocused one-link routes against blocking probability for loads from $\rho_i = 2$ to $\rho_i = 4.5$.

References

[1] Frank P. Kelly. *Reversibility and Stochastic Networks*, (New York, NY: John Wiley and Sons, 1979).

[2] H. Harai, M. Murata, and H. Miyahara. Performance analysis of wavelength assignment policies in all-optical networks. *IEEE J. Select. Areas Commun.*, 16(7), 1998, 1051–1060.

[3] D. Everitt and N. W. Macfadyen. Analysis of multicellular mobile radiotelephone systems with loss. *Br. Telecom Technol. J.*, 1(2), 1983, 37–45.

[4] F. J. Vázquez-Abad, L. Andrew, and D. Everitt. Estimation of blocking probabilities in cellular networks with dynamic channel assignment. *ACM Trans. Model. Comput. Simul.*, 12, 2002. To appear.

[5] Ke-Qiang Liao and Lorne G. Mason. An approximate performance model for a multislot integrated services system. *IEEE Trans. Commun.*, 37(3), 1989, 211–221.

[6] Philip Heidelberger. Fast simulation of rare events in queueing and reliability models. *ACM Trans. Model. Comput. Simul.*, 5(1), 1995, 43–85.