ON RECENT ADVANCES IN LOSS NETWORK PERFORMANCE ANALYSIS

Mark Thompson,
Dept. of Mathematics,
The University of Queensland,
QLD 4072, AUSTRALIA.
email: mrt@maths.uq.edu.au

Abstract

Loss networks are a class of resource allocation models which have proved useful in the study and design of communication networks, including cellular mobile networks, integrated services digital networks, database structures and multiprocessor architectures. This paper is concerned with estimating performance measures of loss networks such as congestion probabilities. Although many of these measures have explicit formulae, it often requires too much processor time to perform the calculations, making it necessary to use approximation techniques. Specifically we compare the Erlang fixed point approximation with the recently-proposed Markov random field method. The principles behind both will be presented. Simple examples are chosen to demonstrate each method's application, and comparison is made in terms of relative accuracy, computational time and complexity.

LOSS NETWORKS; STOCHASTIC MODELS; TELECOMMUNICATIONS

1 Introduction

The most rudimentary loss network is the Erlang loss model. Proposed by A. K. Erlang, circa 1917, the model is that of a single line in a telephone network with a fixed carrying capacity. Calls arrive as a Poisson stream with a constant rate. If a call arrives to find at least one circuit free, it is connected, and holds one circuit for the duration of that call; otherwise, if all circuits are in use, the call is blocked and lost. The model can be regarded as a multi-server queue with no waiting room.

Multidimensional generalisations of the Erlang loss model have been studied extensively but not exhaustively. For instance, it is common to include multiple links and a variety of call types where each call type has its own arrival rate and could require to be connected across multiple links, multiple circuits per link, or within a common bandwidth on all links. Further refinements considered include state dependent arrival rates, and controls on the acceptance criteria such as trunk reservation, call repacking, and dynamic routing (Kelly 1991, Kelly 1986, Choudhury, Leung & Whitt 1995*a*).

We shall be concerned with networks with multiple links and call types without controls. The following notation is chosen to suit our purpose: let the links of a loss network be labelled $1, \ldots, J$ and suppose each link j comprises C_j circuits; the examples we examine have $C_j = C$ for all j. Each member r of the routes set, R, corresponds to a vector $a_{\cdot,r}$ where the non-negative integer $a_{j,r}$ represents the number of circuits required from link j for a call requesting route r. We shall frequently use R_K to denote the set of routes which use a link in $K \subseteq J$, and ∂R_K as the set of routes requiring links both within and outside of K. Calls requesting route r arrive as a Poisson stream of rate ν_r independent of every other route and independent of the network's state. A newly arriving call is blocked and lost if there are fewer than $a_{j,r}$ circuits free on any of the links j; otherwise it is connected and simultaneously holds $a_{j,r}$ circuits for the call duration. Call holding times are independent and identically distributed exponential random variables and, without loss of generality, we take the mean holding time to be 1.

Let $\mathbf{n} = (n_r, r \in R)$, where n_r represents the number of calls in progress using route r, and taking values in the state space $S = S(\mathbf{C}) = \{\mathbf{n} \in \mathbb{N}^{|R|} : \sum_{r \in R} a_{j,r} n_r \leq C_j, \forall j\}$. The process $(\mathbf{n}(t), t \geq 0)$ is a continuous-time Markov chain. Since we will be dealing with subsets of links, we extend this notation slightly by introducing $\mathbf{n}_{R_K}$ to symbolise the number of calls on each of the routes in R_K. Formally, $\mathbf{n}_{R_K} = (n_r, r \in R_K)$ and is in $S_{R_K} = S_{R_K}(\mathbf{C}) = \{\mathbf{n}_{R_K} \in \mathbb{N}^{|R_K|} : \sum_{r \in R_K} a_{j,r} n_r \leq C_j, \forall j\}$. Note that the vector $\mathbf{n}_{R_J}$ is simply $\mathbf{n}$.

The unique equilibrium distribution for $\mathbf{n}(t)$ is $\pi = (\pi(\mathbf{n}); \mathbf{n} \in S)$, where

$$\pi(\mathbf{n}) = G(\mathbf{C})^{-1} \prod_{r \in R} \frac{{\nu_r}^{n_r}}{n_r!}, \tag{1}$$

and $G(\mathbf{C})$ is a normalising constant chosen so that the distribution sums to unity. It is often possible to write explicit expressions for the network's performance measures in terms of this function G. For example, the probability that a new call of type r is blocked (**L**oss) and the long-term rate at which class r calls are admitted to the system (**Th**roughput) are respectively

$$\begin{aligned} L_r &= 1 - \frac{G(\mathbf{C} - A.e_r)}{G(\mathbf{C})}, \qquad \text{and} \\ Th_r &= \nu_r \frac{G(\mathbf{C} - A.e_r)}{G(\mathbf{C})}, \end{aligned}$$

where e_r is the unit vector with a 1 in the rth position. Ideally the normalising constant would be calculated by summing the distribution over every allowable state in S. However the state space for this process grows rapidly in the numbers of links and circuits, thereby rendering exact determination G a formidable task. In fact, this problem is known to be NP-complete in the number of distinct routes; evidence that an always fast and exact sequential algorithm probably doesn't exist. For this reason there is motivation to explore methods of accurately estimating the relevant quantities.

2 The Erlang fixed point approximation

There is a fairly broad class of approximation techniques often referred to as reduced load approximations. The general approach is to diminish the arrival rates of offered traffic to a subnetwork (which may be a single link) by a factor equal to the probability that a new call on that route would not be blocked on the other links of its path. Typically this method leads to a set of fixed point equations for which there exists a (not necessarily unique) solution. The underlying assumption is that of independent blocking between the individual subnetworks and, although invalid in most non-trivial situations, does yield particularly good results when traffic correlations are small.

The famous Erlang fixed point approximation (EFPA) is a member of the reduced load class, one which analyses each link as a separate subnetwork. The EFPA performs well asymptotically. Kelly (1991) proved that the estimates for a network with fixed routing and no controls tend towards the exact probabilities (i) when the link capacities and arrival rates are increased simultaneously keeping the network topology fixed (Kelly limiting regime), and (ii) (Ziedins & Kelly 1989) when the number of links and routes are increased while the link loads are held constant (diverse routing limit).

The EFPA is a solution to the set of fixed point equations

$$B_j = E(\rho_j, C_j), \quad j = 1, \ldots, J, \tag{2}$$

$$\rho_j = \sum_{r \in R} a_{j,r} \nu_r \prod_{i \in r \setminus \{j\}} (1 - B_i), \quad j = 1, \ldots, J, \tag{3}$$

where Erlang's formula, $E(\nu, C)$, gives the probability that the Erlang loss model is fully utilised and is given by

$$E(\nu, C) = \frac{\nu^C}{C!} \left(\sum_{n=0}^{C} \frac{\nu^n}{n!} \right)^{-1} .$$

The interpretation is that B_j is the probability that link j is full given its offered traffic load is ρ_j. ρ_j is an approximation obtained by considering the carried traffic on link j: the throughput of link j is $(1 - B_j)\rho_j$ and $\sum_{r \in R} a_{j,r} \nu_r (1 - L_r)$ is the sum of the contributions made by each route r to j's carried load. Applying the independent blocking assumption yields

$$\begin{aligned} L_r &= 1 - \Pr(\text{call accepted on each link } i \in r), \\ &= 1 - \prod_{i \in r} \Pr(\text{call accepted on link } i), \\ &= 1 - \prod_{i \in r} (1 - B_i), \end{aligned}$$

and the expression (3) follows by equating the two expressions for carried traffic on j and substituting for L_r. Kelly (1986) proved that there is a unique vector $(B_1, \ldots, B_J) \in [0, 1]^J$ satisfying (2) and (3).

3 The Markov random field method

It is natural to think of the process $\mathbf{n}$ as having spatial properties since it is the interactions between resources caused by competing calls on overlapping routes that makes the process difficult and interesting to analyse. In a recent paper, Zachary

& Ziedins (1999) describe a new approach, highlighting the application of Markov random field theory to the problem. They consider the equilibrium distribution stated in (1) as a finite random field on the set of resources (alternatively routes) and note that the process is Markovian with respect to various neighbour relations. For instance the relation, $\sim$, defined on the set of links as $(i \sim j) \Leftrightarrow (\exists r : i \in r \wedge j \in r)$ induces a graph in which the links are the nodes and two resources, i and j, are connected if they are used by at least one common route. The graph $(J, \sim)$ is a Markov random field. This gives some useful preliminary results.

Theorem 1. *The normalising constant θ_K for the marginal distribution of $\mathbf{n}_{R_K}$ depends only on the numbers of calls active on routes in the boundary set of K. Formally,*

$$\pi_{R_K}(\mathbf{n}_{R_K}) = \theta_K(\mathbf{n}_{\partial R_K}) \prod_{r \in R_K} \frac{{\nu_r}^{n_r}}{n_r!}. \tag{4}$$

Theorem 1 provides the basis for the Markov random field method. It is a useful starting point because it specifies exactly which dependencies are apparent between subsets of resources. In addition, comparisons of equation (4) for subnetworks of the form K and $K \setminus \{j\}$ often yield recurrence relations for the constant θ_K.

The next result is important for two reasons. It's application can dramatically decrease the computational effort required. Conversely, when the premise fails, this result justifies and suggests product form approximations for the normalising constant.

Theorem 2. *When a subnetwork, K, partitions the associated graph's external nodes (those not in K) into distinct groups, $G_1, G_2, \ldots, G_d$, the normalising constant has a product form with each group contributing a factor,*

$$\theta_K(\mathbf{n}_{\partial R_K}) = \prod_{i=1}^{d} \theta_i(\mathbf{n}_{\partial R_K \cap R_{G_i}}).$$

In the special case that the Markov random field is a tree, that is, a connected graph which becomes disconnected upon the removal of any edge, Theorem 2 along with equation (4) may be used to derive a recurrence relation to perform an exact analysis. For convenience, let

$$\psi_{R_{K'}}(\mathbf{n}_{R_K}) \;=\; \left\{ \mathbf{n}_{R_{K'}} \in S_{R_{K'}} : (\mathbf{n}_{R_{K'}})_r = (\mathbf{n}_{R_K})_r, \forall r \in R_K \right\},$$

then by saying $\mathbf{m}_{R_{K'}} \in \psi_{R_{K'}}(\mathbf{n}_{R_K})$ we mean the components of $\mathbf{m}_{R_{K'}}$ and $\mathbf{n}_{R_K}$ agree for all the routes they have in common. Consider the subsets of J, $K' = \{j, k\}$, and $K = \{j\}$, for any two resources such that $j \sim k$. We can write down two distinct expressions for the marginal distribution of $\mathbf{n}_{R_K}$:

$$\begin{aligned}
\pi_{R_K}(\mathbf{n}_{R_K}) &= \prod_{l:l\sim j} \theta_{jl}(\mathbf{n}_{R_j\cap R_l}) \prod_{r\in R_j} \frac{{\nu_r}^{n_r}}{n_r!}, \qquad \text{and} \\
\pi_{R_K}(\mathbf{n}_{R_K}) &= \sum_{\mathbf{m}_{R_{K'}}\in\psi_{R_{K'}}(\mathbf{n}_{R_K})} \pi_{R_{K'}}(\mathbf{m}_{R_{K'}}), \\
&= \sum_{\mathbf{m}_{R_{K'}}\in\psi_{R_{K'}}(\mathbf{n}_{R_K})} \left(\prod_{\substack{l:l\sim j \\ l\neq k}} \theta_{jl}(\mathbf{m}_{R_j\cap R_l}) \right) \left(\prod_{\substack{l:l\sim k \\ l\neq j}} \theta_{kl}(\mathbf{m}_{R_k\cap R_l}) \right) \prod_{r\in R_{K'}} \frac{{\nu_r}^{m_r}}{m_r!}, \\
&= \prod_{\substack{l:l\sim j \\ l\neq k}} \theta_{jl}(\mathbf{n}_{R_j\cap R_l}) \prod_{r\in R_K} \frac{{\nu_r}^{n_r}}{n_r!} \sum_{\mathbf{m}_{R_{K'}}\in\psi_{R_{K'}}(\mathbf{n}_{R_K})} \prod_{\substack{l:l\sim k \\ l\neq j}} \theta_{kl}(\mathbf{m}_{R_k\cap R_l}) \prod_{r\in R_{K'}\setminus R_K} \frac{{\nu_r}^{m_r}}{m_r!}.
\end{aligned}$$

Comparing the right hand sides we get

$$\theta_{jk}(\mathbf{n}_{R_j\cap R_k}) = \sum_{\mathbf{m}_{R_{K'}}\in\psi_{R_{K'}}(\mathbf{n}_{R_K})} \prod_{\substack{l:l\sim k \\ l\neq j}} \theta_{kl}(\mathbf{m}_{R_k\cap R_l}) \prod_{r\in R_{K'}\setminus R_K} \frac{{\nu_r}^{m_r}}{m_r!}. \tag{5}$$

Relation (5) provides a tractable means to calculate quantities of interest exactly for any loss network whose associated graph is a tree. Furthermore this relation makes a useful approximation for some common structures which don't fit the tree specification but do become tree-like asymptotically. For this to be the case there must be no routes offered which have more than two links. In the next section we analyse two networks which are not tree-like but have only single and two-link traffic. For these examples we find equation (5) provides a good approximation.

4 Symmetric networks

We now examine two loss networks. The first, termed a fully connected network, can be pictured as a collection of nodes with every pair of nodes joined by a single link. Also under consideration is a star network in which links join each of a collection of outer nodes to a single central one. Two-link calls may request connection on any two links which share a common node. In the star network, for example, any pair of links is valid for a two-link call. In contrast, for each link in every fully connected network of four nodes or more there is at least one other link which doesn't share a common node. For simplicity suppose both networks to have $C_j = C$ circuits per link and two types of traffic: single-resource traffic arriving at each link at rate ν_1 and two-link traffic arriving at rate ν_2.

These seemingly dissimilar networks are related through their EFP approximations in that for every fully connected network there exists a star network with equivalent EFPA. Each link in the four node fully connected network is offered one stream of single-link traffic and four streams of two-link traffic. By symmetry the Erlang fixed point equations reduce to

$$\begin{aligned}
E_1 &= E(\nu_1 + 4\nu_2(1-E_1), C), \\
E_2 &= 1-(1-E_1)^2,
\end{aligned}$$

where E_1 and E_2 are the approximate loss probabilities for single and two-link calls respectively. The five point star network has an identical EFPA.

Expressions for the exact loss probabilities may be expressed in terms of the marginal distributions of calls using a common link or pair of links:

$$
\begin{aligned}
L_1 &= 1 - \left(\sum_{S_{R_j}(\mathbf{C}-e_j)} \pi_{R_j}(\mathbf{n}_{R_j}) \right) \left(\sum_{S_{R_j}(\mathbf{C})} \pi_{R_j}(\mathbf{n}_{R_j}) \right)^{-1}, \\
L_2 &= 1 - \left(\sum_{S_{R_j}(\mathbf{C}-e_j-e_k)} \pi_{R_j}(\mathbf{n}_{R_j}) \right) \left(\sum_{S_{R_j}(\mathbf{C})} \pi_{R_j}(\mathbf{n}_{R_j}) \right)^{-1},
\end{aligned}
$$

for any links $j \sim k$. Using the network's symmetry, equation (5) reduces to

$$
\theta(n_{j4}) = \sum_{m_j+m_{j1}+m_{j2}+m_{j3} \leq C-n_{j4}} \frac{{\nu_1}^{m_j}}{m_j!} \prod_{i=1}^{3} \theta(m_{ji}) \frac{{\nu_2}^{m_{ji}}}{m_{ji}!}, \tag{6}
$$

where we've labelled link j's neighbours 1 through 4. Now, approximating the marginal π_{R_j} using Theorem 2 and equation (6) we have

$$
\begin{aligned}
\sum_{S_{R_j}(\mathbf{C}-e_j)} \pi_{R_j}(\mathbf{n}_{R_j}) &\approx \sum_{n_j+n_{j1}+n_{j2}+n_{j3}+n_{j4} \leq C-1} \frac{{\nu_1}^{n_j}}{n_j!} \prod_{i=1}^{4} \theta(n_{ji}) \frac{{\nu_2}^{n_{ji}}}{n_{ji}!}, \\
&\approx \sum_{n_{j4} \leq C-1} \theta(n_{j4}) \theta(n_{j4}+1) \frac{{\nu_2}^{m_{j4}}}{m_{j4}!}.
\end{aligned}
$$

Similar computations imply that L_1 and L_2 may be approximated by

$$
\begin{aligned}
M_1 &= 1 - \frac{\sum_{n=0}^{C-1} \theta(n)\theta(n+1) \frac{{\nu_2}^n}{n!}}{\sum_{n=0}^{C} \theta(n)^2 \frac{{\nu_2}^n}{n!}}, \qquad \text{and} \\
M_2 &= 1 - \frac{\sum_{n=0}^{C-1} \theta(n+1)^2 \frac{{\nu_2}^n}{n!}}{\sum_{n=0}^{C} \theta(n)^2 \frac{{\nu_2}^n}{n!}}, \qquad \text{respectively,}
\end{aligned}
$$

where θ is a solution to (6). Note that although the symmetric star network and fully connected networks have different equilibrium distributions with the exception of the three link case, their approximations under this scheme are identical.

We now compare the relative accuracy of the Markov field method to the EFPA. In the bottom two panes of figure 1 there are qualitative plots of the relative error in the estimates of the probability a two-link call is lost in the fully connected four node symmetric network with $C = 5$. The region plotted is ν_1 and ν_2 varying from 0 to 10. These graphs indicate that both approximations perform well asymptotically with the most sensitivity observed under light loads.

Both approximation schemes also provide good single-link blocking estimates (top panes) with the exception of the region where ν_1 is small. When there is a large amount of two-link traffic both methods tend to underestimate the probability that a single link is full. This observation is attributed to packing constraints between the links. For example suppose all neighbours of a particular link are full, then even

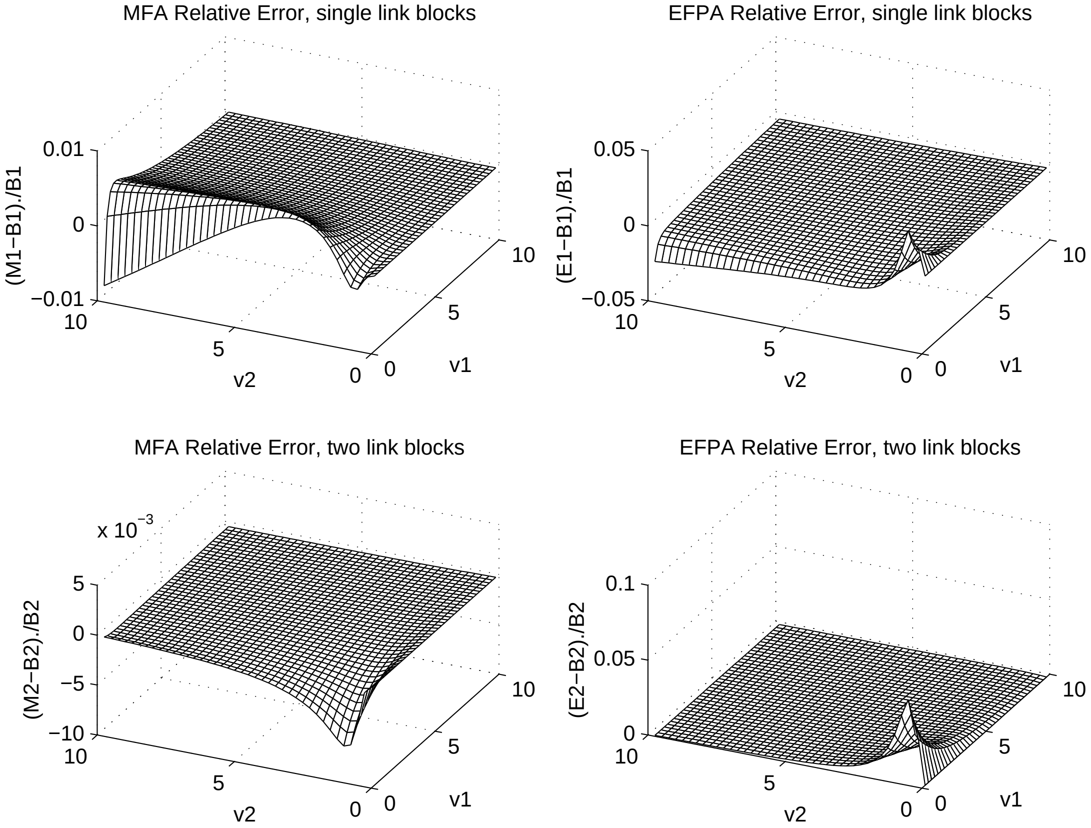

Figure 1: Relative error in link loss probabilities, four node fully connected network, $C = 5$.

if this link has free capacity any new two-link calls cannot be accepted. When the vast majority of arriving traffic is of the two-link variety this situation arises often enough to be significant.

Besides the edge ν_1-small, the largest relative error in the Markov random field approximation appears to be along the line $\nu_1 = \frac{\nu_2}{10}$. Figures 2 and 3 are plots of relative errors in estimating single-link and two-link blocking respectively. In this region the Markov field method is comparable to the EFPA. The EFPA records the largest relative error when the network is close to being critically loaded (offered traffic $\approx$ C); the Markov field method offers a marked improvement in this vicinity.

In terms of computational times and complexity, the Markov random field method yields a dramatic improvement over calculating the exact quantities, but it is not as quick as solving the Erlang fixed point equations. Since neither of the approximation schemes depend on the number of links in the network, we compare the computation times as the link capacities C are varied in figure 4. Of interest on these graphs are the rates of increase in computational time, rather than a direct time comparison, since the former decides the algorithmic complexity. The solid line in the top pane shows the time taken to perform the exact calculations increasing rapidly with the number of circuits per link. The computation time of the Markov random field method also increases non-linearly but has a much slower rate, implying that this approximation may not be feasible for large networks. The Erlang fixed point equations, on the other hand, take almost linear time to solve numerically.

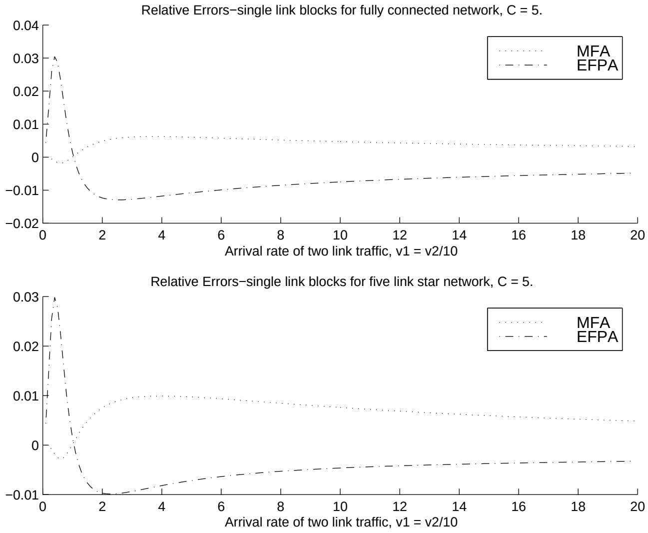

Figure 2:

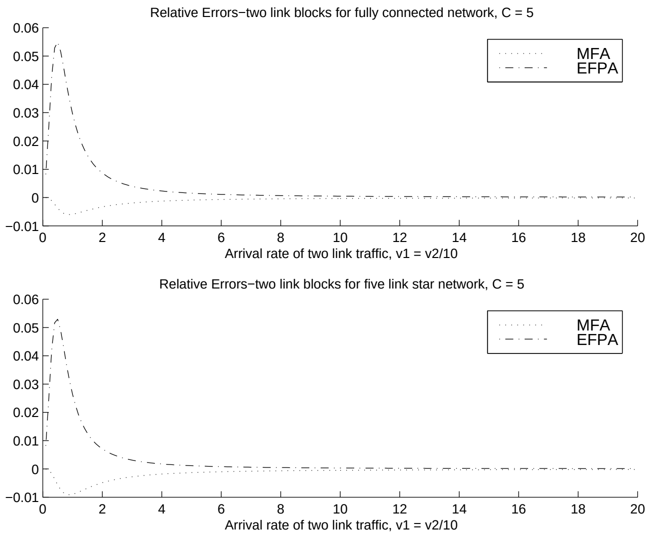

Figure 3:

5 Summary

There is a clear trade off between time and accuracy associated with these techniques. Calculating the partition function, G, is feasible for the small networks presented here. However the processor time required for this exact method increases rapidly with the network's size, making it unsuitable for most real-world models. The Erlang fixed point method, in contrast, is typically very quick to execute and is relatively simple to code. The EFPA provides estimates which are asymptotically correct and is therefore good when either diverse routing or the Kelly limiting regime is operating. The Markov random field method can be regarded as a compromise between calculating G exactly and the EFPA; requiring slightly extra effort to provide a refined approximation.

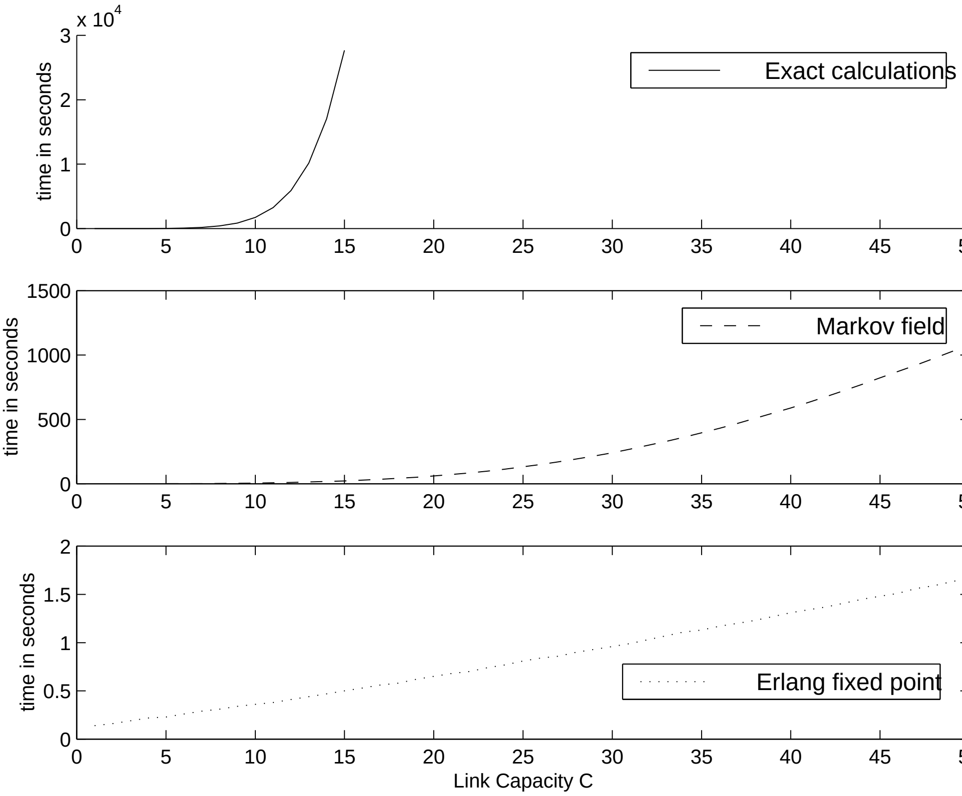

Figure 4: Comparison of computation times taken to calculate loss for 200 values of ν_1 and ν_2.

References

Bean, N. & Stewart, M. (1998), 'A new perspective on the normalisation of invariant measures for loss networks and other product form systems', *Mathematical and Computer Modelling* p. to appear.

Bebbington, M., Pollett, P. & Ziedins, I. (1997), Improved fixed point methods for loss networks with linear structure, *in* 'Proceedings of the 4th International Conference on Telecommunications', Vol. **3**, Melbourne, Australia, pp. 1411–1416.

Choudhury, G., Leung, K. & Whitt, W. (1995*a*), 'An algorithm to compute blocking probabilities in multi-rate multi-class multi-resource loss models', *Adv. Appl. Prob.* **27**, 1104–1143.

Choudhury, G., Leung, K. & Whitt, W. (1995*b*), 'Calculating normalization constants of closed queuing networks by numerically inverting their generating functions', *Journal of the Association for Computing Machinery* **42**(5), 935–970.

Dziong, Z. & Roberts, J. (1987), 'Congestion probabilities in a circuit-switched integrated services network', *Performance Evaluation* **7**, 267–284.

Hoover, H. & Ruzzo, W. (1995), *Limits to parallel computation: P-completeness theory*, Oxford University Press, Oxford.

Kelly, F. (1986), 'Blocking in large circuit-switched networks', *Adv. Appl. Prob.* **18**, 473–505.

Kelly, F. (1987), 'One-dimensional circuit-switched networks', *The Annals of Probability* **15**(3), 1166–1179.

Kelly, F. (1991), 'Loss networks', *The Annals of Applied Probability* **1**(3), 319–378.

Pallant, D. (1992), 'A reduced load approximation for cellular mobile networks including handovers', *A.T.R.* **26**(2), 21–29.

Pallant, D. & Taylor, P. (1994), 'Approximation of performance measures in cellular mobile networks with dynamic channel allocation', *Telecommunication Systems* **3**, 129–163.

Ross, K. (1995), *Multiservice loss models for broadband telecommunication networks*, Springer-Verlag, New York.

Zachary, S. & Ziedins, I. (1999), 'Loss networks and markov random fields', *J. Appl. Probab.* **36**, to appear.

Ziedins, I. & Kelly, F. (1989), 'Limiting theorems for loss networks with diverse routing', *Adv. Appl. Prob.* **21**, 804–830.