A reduced load approximation accounting for link interactions in a loss network

M.R. Thompson P.K. Pollett

Department of Mathematics, The University of Queensland, Brisbane, Australia.

Abstract

This paper is concerned with evaluating the performance of loss networks. Accurate determination of loss network performance can assist in the design and dimensioning of telecommunications networks. However, exact determination can be difficult and generally cannot be done in reasonable time. For these reasons there is much interest in developing fast and accurate approximations. We develop a reduced load approximation that improves on the famous Erlang fixed point approximation (EFPA) in a variety of circumstances. We illustrate our results with reference to a range of networks for which the EFPA may be expected to perform badly.

1 Introduction

We shall use the standard model for a circuit-switched teletraffic network. The network consists of a finite set of links J and the j-th link comprises a co-operative group of C_j circuits. Upon connection of a call an end-to-end route is established such that a call initiated on route r seizes a_{jr} circuits from one or more of the links in J. For simplicity, we will assume that $a_{jr} = 1$ if link j is part of route r; otherwise $a_{jr} = 0$. More general models may allow $a_{jr} \in \{0, 1, 2, \ldots, C_j\}$. The $(a_{jr}; j \in J)$ circuits remain exclusively dedicated to the connection as long as it is maintained, even when no information is being transferred. When the call is terminated, all of the circuits are released simultaneously and are then available to be used by future calls. Denote the set of all routes by R, the routing matrix $(a_{jr}; j \in J, r \in R)$ by A, and write $j \in r$ as an abbreviation for $j \in \{i \in J : a_{ir} > 0\}$. Rather than identifying a call by its origin and destination points, a call is identified by its route, and we assume that arriving calls are requesting to be connected along a particular route. There are no waiting arrangements for calls that cannot be connected immediately; a call that arrives to find insufficient capacity on one or more of the links along its route is blocked from service and is then lost. The proportions $(L_r; r \in R)$ of calls that are expected to be lost on the various routes form a natural measure of network efficiency.

The usual state description tracks the number of calls in progress on each of the routes. Let $\boldsymbol{Y} = (Y_r; r \in R)$, where Y_r is the number of route-r calls in progress. Due to the capacity constraints, $\boldsymbol{Y}$ takes values in the subset $S = S(\boldsymbol{C})$ of $\mathbb{N}^R$ given by

$$S(\boldsymbol{C}) = \left\{ \boldsymbol{n} \in \mathbb{N}^R : \sum_{r \in R} a_{jr} n_r \leq C_j, \ j \in J \right\}. \tag{1}$$

We will suppose that calls for each route arrive in independent Poisson streams, with route-r calls arriving at rate ψ_r. Further, we will suppose that calls have an exponentially distributed duration after being connected. Under these assumptions, $\boldsymbol{Y}$ is a reversible Markov process and its equilibrium distribution has a product form. Let the mean holding time of calls on route r be ϕ_r^{-1}. Define P to be the probability measure under which $(Y_r; r \in \mathcal{R})$ are independent Poisson random variables with means $\nu_r = \psi_r/\phi_r$, $r \in \mathcal{R}$. This would be the equilibrium measure for the usage on each of the routes were the system not to have any capacity constraints. The restriction $\boldsymbol{Y}$ to S is a truncation of a reversible Markov process and its equilibrium probability measure is thus given by

$$\pi(\mathcal{A}) = P(\mathcal{A} | \boldsymbol{Y} \in S), \qquad \text{for all } P\text{-measurable } \mathcal{A}. \tag{2}$$

Under π, $\boldsymbol{Y}$ is still reversible (Kelly (1979), Corollary 1.10), and thus the form of π can be easily obtained from the detailed balance equations,

$$\psi_r \pi(\boldsymbol{Y} = \boldsymbol{n}) = (n_r + 1)\phi_r \pi(\boldsymbol{Y} = \boldsymbol{n} + \boldsymbol{e}_r), \qquad \boldsymbol{n}, \boldsymbol{n} + \boldsymbol{e}_r \in S. \tag{3}$$

(Here $\boldsymbol{e}_r$ represents the unit vector with a 1 in the r-th position.) In the present context, there is no need to distinguish between the traffic load ν_r and the arrival rate ψ_r, for replacing ψ_r by ν_r and ϕ_r by 1 does not alter (3). Thus, without loss of generality, it may be assumed that the mean holding time for all call types is 1.

The form of π can also be derived directly from definition (2). For instance, if $K \subseteq J$ and $R_K = \{r \in R : \sum_{j \in K} a_{jr} > 0\}$ is the set of routes that use at least one link in K, then the marginal distribution of the numbers of calls on routes in R_K is

$$\pi\left(\boldsymbol{Y}_{R_K} = \boldsymbol{n}_{R_K}\right) = \frac{P\left(\boldsymbol{Y} \in S | \boldsymbol{Y}_{R_K} = \boldsymbol{n}_{R_K}\right) P\left(\boldsymbol{Y}_{R_K} = \boldsymbol{n}_{R_K}\right)}{P\left(\boldsymbol{Y} \in S\right)},$$

where $\boldsymbol{Y}_{R_K} = \boldsymbol{n}_{R_K}$ is shorthand for the event $\cap_{r \in R_K}\{Y_r = n_r\}$. Noticing that

$$P\left(\boldsymbol{Y} \in S | \boldsymbol{Y}_{R_K} = \boldsymbol{n}_{R_K}\right) = P\left(\sum_{r \notin R_K} a_{jr} Y_r \leq C_j - \sum_{r \in R_K} a_{jr} n_r, \ j \in J \right)$$

is a function (call it θ_K) only of $\boldsymbol{n}_{\partial R_K} = (n_r : r \in R_K \cap R_{J \setminus K})$, we are lead to

$$\pi\left(\boldsymbol{Y}_{R_K} = \boldsymbol{n}_{R_K}\right) = \frac{\theta_K\left(\boldsymbol{n}_{\partial R_K}\right)}{G(\boldsymbol{C})} \prod_{r \in R_K} \frac{\nu_r^{n_r}}{n_r!}, \tag{4}$$

where $G(\boldsymbol{C})$ is a normalising constant chosen so that the distribution π sums to unity. Expression (4) is due to Zachary and Ziedins (1999). It implies that the equilibrium distribution for the loss network is a Markov random field.

The loss probability on a route can be calculated from π: for any route $r \in R_K$,

$$\begin{aligned}\pi\left(\boldsymbol{Y}+\boldsymbol{e}_r \in S\right) &= P(\boldsymbol{Y}\in S)^{-1} \sum_{\boldsymbol{n}_{R_K}\in S_{R_K}} P\left(\boldsymbol{Y}+\boldsymbol{e}_r \in S \middle| \boldsymbol{Y}_{R_K} = \boldsymbol{n}_{R_K}\right) P\left(\boldsymbol{Y}_{R_K} = \boldsymbol{n}_{R_K}\right) \\ &= P(\boldsymbol{Y}\in S)^{-1} \sum_{\boldsymbol{n}_{R_K}\in S_{R_K}} P\left(\boldsymbol{Y}\in S \middle| \boldsymbol{Y}_{R_K} = \boldsymbol{n}_{R_K}+\boldsymbol{e}_r\right) P\left(\boldsymbol{Y}_{R_K} = \boldsymbol{n}_{R_K}\right) \\ &= P(\boldsymbol{Y}\in S)^{-1} \sum_{\boldsymbol{n}_{R_K}\in S_{R_K}} P\left(\boldsymbol{Y}\in S \middle| \boldsymbol{Y}_{R_K} = \boldsymbol{n}_{R_K}\right) P\left(\boldsymbol{Y}_{R_K} = \boldsymbol{n}_{R_K}-\boldsymbol{e}_r\right),\end{aligned}$$

where, for any $\mathcal{R} \subseteq R$, $S_{\mathcal{R}} = S_{\mathcal{R}}(\boldsymbol{C}) = \left\{\boldsymbol{n}_{\mathcal{R}} \in \mathbb{N}^{\mathcal{R}} : \sum_{r\in\mathcal{R}} a_{jr} n_r \le C_j,\ j \in J\right\}$ is the projection of S onto $\mathbb{N}^{\mathcal{R}}$. Thus, the probability a call requesting route r arrives to find one or more of the links in r full is

$$L_r = 1 - \frac{\sum_{\boldsymbol{n}_{R_K}\in S_{R_K}} \theta_K(\boldsymbol{n}_{\partial R_K}) P\left(\boldsymbol{Y}_{R_K} = \boldsymbol{n}_{R_K} - \boldsymbol{e}_r\right)}{\sum_{\boldsymbol{n}_{R_K}\in S_{R_K}} \theta_K(\boldsymbol{n}_{\partial R_K}) P\left(\boldsymbol{Y}_{R_K} = \boldsymbol{n}_{R_K}\right)}. \tag{5}$$

In the case $K = J$, equation (5) has the concise form $L_r = 1 - G(\boldsymbol{C} - A\boldsymbol{e}_r)/G(\boldsymbol{C})$.

Unfortunately, calculating the loss probabilities using $G(\boldsymbol{C})$ is often intractable. Direct normalisation of the distribution π in (2) entails summing over the space S, and, even for moderately sized networks, it is apparent from (1) that the number of distinct states in S is large and grows rapidly with the number of routes, and also with the link capacities. In fact, the problem of evaluating π in this way is $\#P$-complete (Louth, Mitzenmacher and Kelly (1994)). Thus, there is strong evidence to suggest that an algorithm for finding the loss probabilities in polynomial time using G does not exist.

An alternative to evaluating G is to find θ_K and then use (5) to calculate L_r. Choose a collection of links H disjoint from K. Then,

$$P\left(\boldsymbol{Y} \in S \middle| \boldsymbol{Y}_{R_K} = \boldsymbol{n}_{R_K}\right) = \sum_{\substack{\boldsymbol{m}_{R_{K\cup H}} \in S_{R_{K\cup H}}: \\ \boldsymbol{m}_{R_K} = \boldsymbol{n}_{R_K}}} P\left(\boldsymbol{Y} \in S \middle| \boldsymbol{Y}_{R_{K\cup H}} = \boldsymbol{m}_{R_{K\cup H}}\right) P\left(\boldsymbol{Y}_{R_H} = \boldsymbol{m}_{R_H} \middle| \boldsymbol{Y}_{R_K} = \boldsymbol{n}_{R_K}\right),$$

(recall that $\boldsymbol{m}_{R_K} = \boldsymbol{n}_{R_K}$ is shorthand for $\cap_{r\in R_K} m_r = n_r$). Thus, θ_K satisfies the recurrence

$$\theta_K\left(\boldsymbol{n}_{\partial R_K}\right) = \sum_{\substack{\boldsymbol{m}_{R_H} \in S_{R_H}: \\ \boldsymbol{m}_{R_K\cap R_H} = \boldsymbol{n}_{R_K\cap R_H}}} \theta_{K\cup H}\left(\boldsymbol{m}_{\partial R_{K\cup H}}\right) \prod_{r\in R_H\setminus R_K} \frac{\nu_r^{m_r}}{m_r!} e^{-\nu_r}, \tag{6}$$

where $\boldsymbol{m}_{\partial R_{K\cup H}}$, the argument of $\theta_{K\cup H}$, is actually the vector $(m_r : r \in \partial R_H \setminus R_K)$ joined with $(n_r : r \in \partial R_K \setminus R_H)$. The functions θ_K often have a natural factorisation, which refines (6) and sometimes reduces the complexity of the problem to the point where exact calculation of the loss probabilities is tractable. If the problem is still too large, Zachary and Ziedins (1999) suggest imposing a product form on θ_K and then using (6) as the basis of an approximation scheme. This approach, which is very accurate for a wide range of networks, is reviewed in Section 3.

2 Link interactions

Instead of $\boldsymbol{Y}$, it is frequently convenient to work in terms of the link utilisations. Overlapping routes are, in some sense, competing for usage of the circuits on the links that they share. It is this competitiveness between routes that cause the volumes of carried traffic on the links to affect one another and makes analysing the process so interesting and challenging.

For each link j, let $U_j = \sum_{r \in R_j} Y_r$ be the capacity used on link j. Only in exceptional circumstances is there sufficient information encapsulated in a state description that lists only the links' utilisation for the process $\boldsymbol{U} = (U_j; j \in J)$ to be Markovian. While transitions $\boldsymbol{U} \to \boldsymbol{U} + A\boldsymbol{e}_r$ are made at a rate $\sum_{r \in R} \nu_r \, 1_{\{\boldsymbol{U}+A\boldsymbol{e}_r \leq \boldsymbol{C}\}}$ that only depends on the state $\boldsymbol{U}$, the rate at which transitions of the form $\boldsymbol{U} \to \boldsymbol{U} - A\boldsymbol{e}_r$ occur is Y_r, a quantity that cannot always be determined from $\boldsymbol{U}$.

Under P, U_j is the superposition of independent Poisson streams and is therefore marginally distributed as a Poisson random variable with mean $\rho_j = \sum_{r \in R_j} \nu_r$. The joint probability generating function of U_i and U_j is

$$\mathbb{E}_P \left(s^{U_i} t^{U_j} \right) = \exp \left[\sum_{r \in R} \nu_r \left(s^{a_{ir}} t^{a_{jr}} - 1 \right) \right].$$

Clearly, if there is at least one route that uses both links i and j then, even under P, the links will not operate independently. With capacity constraints the link interactions become complicated, and, the effects of link blocking not only influence the behaviour of neighbouring links but tend to propagate throughout the network. Owing to this complexity a useful explicit expression for $\pi(\boldsymbol{U} = \boldsymbol{u})$ does not usually exist. However, these probabilities do form the unique solution to the equations

$$u_j \, \pi(\boldsymbol{U} = A\boldsymbol{n}) = \sum_{r \in R_j} \nu_r \, \pi(\boldsymbol{U} = A\boldsymbol{n} - A\boldsymbol{e}_r), \qquad j \in J, \, \boldsymbol{n} \in S, \tag{7}$$

where $u_j = \sum_{r \in R_j} n_r$. Recurrence (7) was established by Dziong and Roberts (1987); a neat derivation is given in Zachary (1991).

We have described the classical loss network model, similar to that of Kelly (1986). It also arises in variety of different contexts. Appropriate choices of A and $\boldsymbol{C}$ for the linear constraints will lead to simple models for fixed-line networks (Ross and Tsang 1990, Girard 1990, Kelly 1991), cellular mobile networks (Everitt and Macfadyen 1983, Boucherie and Mandjes 1998), computer database access problems (Mitra and Weinberger 1984), and other kinds of telecommunications networks (Whitt 1985, Ross 1995). Part of the model's appeal is that it can easily be extended to include call acceptance criteria that cannot necessarily be expressed using a linear constraint $A\boldsymbol{Y} \leq \boldsymbol{C}$. Provided those controls preserve the reversibility of the process $\boldsymbol{Y}$, even the product-form distribution π in (4) applies. Unfortunately, this is not the case for admission policies such as trunk reservation (Key 1990, Hunt and Laws 1997) or virtual partitioning (Borst and Mitra 1998, Mitra and Ziedins 1996). Nor does the product-form result hold for networks allowing alternative routing.

3 The Markov field method

Sometimes the network has special structure that allows quicker determination of the loss probabilities. Again fix some $K \subseteq J$. For each link $j \in J \setminus K$, let $X_j = \sum_{r \notin R_K} a_{jr} Y_r$ denote the capacity used on link j due to calls on routes not in R_K. Under P, the X_j and X_k corresponding to two links j and k for which there does not exist a common route ($\notin R_K$) are independent random variables. That is to say, if $a_{jr} > 0$ only for the routes in the set $\{r \notin R_K : a_{kr} = 0\}$, then X_j and X_k are independent. This means that θ_K can be factorised as

$$\theta_K \left(\boldsymbol{n}_{\partial R_K}\right) = \prod_i \theta_{K,H_i}(\boldsymbol{n}_{R_K \cap R_{H_i}}), \tag{8}$$

where $\cup_i H_i = J \setminus K$ and each H_i is a group of links satisfying $\sum_{r \notin R_K} a_{jr} a_{kr} > 0$ if and only if j and k belong to the same group. Furthermore,

$$\theta_{K,H_i}(\boldsymbol{n}_{R_K \cap R_{H_i}}) = P\left(X_j \leq C_j - \sum_{r \in R_K} a_{jr} n_r,\ j \in H_i \right). \tag{9}$$

Expression (8) separates the calculation of θ_K into the smaller calculations of θ_{K,H_i}, which may then be conducted in parallel.

The factorisation of θ_K substantially simplifies the recurrence (6). Let $K_1, \ldots, K_d$ be collections of links that form a complete covering of J and let $\sim$ be the relation on pairs of link groups α and β in $\mathcal{K} = \{K_1, \ldots, K_d\}$ defined by $\alpha \sim \beta$ if and only if $R_\alpha \cap R_\beta \neq \emptyset$. Now suppose that the collections of links $\mathcal{K}$ are chosen so that

$$\theta_\alpha \left(\boldsymbol{n}_{\partial R_\alpha}\right) = \prod_{\beta \sim \alpha} \theta_{\alpha\beta} \left(\boldsymbol{n}_{R_\alpha \cap R_\beta}\right), \qquad \text{for each } \alpha \in \mathcal{K}. \tag{10}$$

This choice is always possible: at worst, $d = 1$ and $K_1 = J$. In a network that has been decomposed this way, recurrence (6) implies that, for each $\alpha \in \mathcal{K}$ and $\beta \sim \alpha$,

$$\theta_{\alpha\beta} \left(\boldsymbol{n}_{R_\alpha \cap R_\beta}\right) = \sum_{\substack{\boldsymbol{m}_{R_\beta} \in S_{R_\beta}: \\ \boldsymbol{m}_{R_\alpha \cap R_\beta} = \boldsymbol{n}_{R_\alpha \cap R_\beta}}} \prod_{\substack{\gamma \sim \beta \\ \gamma \neq \alpha}} \theta_{\beta\gamma} \left(\boldsymbol{m}_{R_\beta \cap R_\gamma}\right) \prod_{r \in R_\beta \setminus R_\alpha} \frac{\nu_r^{m_r}}{m_r!} e^{-\nu_r}. \tag{11}$$

This finer recursion relates $\theta_{\alpha\beta}$ to only those $\theta_{\beta\gamma}$ for which $\alpha \sim \beta$ and $\beta \sim \gamma$ and suggests that it may be solved efficiently using block iterative methods. Then, using (8) and (5), the loss probabilities can be determined from

$$L_r = 1 - \frac{\sum_{\boldsymbol{n}_{R_\alpha \cap R_\beta} \in S_{R_\alpha \cap R_\beta}} \theta_{\alpha\beta}^{(\boldsymbol{C} - A\boldsymbol{e}_r)}\left(\boldsymbol{n}_{R_\alpha \cap R_\beta}\right) \theta_{\beta\alpha}^{(\boldsymbol{C} - A\boldsymbol{e}_r)}\left(\boldsymbol{n}_{R_\alpha \cap R_\beta}\right) P\left(\boldsymbol{Y}_{R_\alpha \cap R_\beta} = \boldsymbol{n}_{R_\alpha \cap R_\beta}\right)}{\sum_{\boldsymbol{n}_{R_\alpha \cap R_\beta} \in S_{R_\alpha \cap R_\beta}} \theta_{\alpha\beta}\left(\boldsymbol{n}_{R_\alpha \cap R_\beta}\right) \theta_{\beta\alpha}\left(\boldsymbol{n}_{R_\alpha \cap R_\beta}\right) P\left(\boldsymbol{Y}_{R_\alpha \cap R_\beta} = \boldsymbol{n}_{R_\alpha \cap R_\beta}\right)}, \tag{12}$$

where

$$\theta_{\alpha\beta}^{(\boldsymbol{C} - A\boldsymbol{e}_r)} \left(\boldsymbol{n}_{R_\alpha \cap R_\beta}\right) = P\left(\sum_{\rho \notin R_\alpha} a_{j\rho} Y_\rho \leq C_j - a_{jr} - \sum_{\rho \in R_\alpha} a_{j\rho} n_\rho,\ j \in \beta \right).$$

When $r \in R_\alpha \setminus R_\beta$,

$$\theta^{(\boldsymbol{C}-A\boldsymbol{e}_r)}_{\alpha\beta}(\boldsymbol{n}_{R_\alpha \cap R_\beta}) = \theta_{\alpha\beta}(\boldsymbol{n}_{R_\alpha \cap R_\beta}), \qquad \text{for all } \boldsymbol{n}_{R_\alpha \cap R_\beta} \in S_{R_\alpha \cap R_\beta},$$

and if $r \in R_\alpha \cap R_\beta$,

$$\theta^{(\boldsymbol{C}-A\boldsymbol{e}_r)}_{\alpha\beta}(\boldsymbol{n}_{R_\alpha \cap R_\beta}) = \theta_{\alpha\beta}(\boldsymbol{n}_{R_\alpha \cap R_\beta} + \boldsymbol{e}_r), \qquad \text{for all } \boldsymbol{n}_{R_\alpha \cap R_\beta} \in S_{R_\alpha \cap R_\beta},$$

in either case, and otherwise $\theta^{(\boldsymbol{C}-A\boldsymbol{e}_r)}_{\alpha\beta}$ satisfies relations (8) and (11) with S replaced by $S(\boldsymbol{C} - A\boldsymbol{e}_r)$.

The factorisation (10) holds if the graph $(\mathcal{K}, \sim)$ contains no 3-cycles. Thus, the extent to which J may be separated into the subsets $\mathcal{K}$ is limited. This might present a problem. If R_β contains more than a few routes, the space S_{R_β} might be too large to sum over, and evaluation of $\theta_{\alpha\beta}$ in (11) might still prove to be a formidable task. In this case, imposing a product form on θ_K may lead to a good approximation. This is the approach that Zachary and Ziedins (1999) take in developing their Markov field approximation (MFA) method. It is a general approximation scheme of which the Erlang fixed point approximation (EFPA) is a special case.

4 The Erlang fixed point approximation

In the EFPA the loss probability for route r is estimated to be

$$L_r = 1 - \prod_{i \in r}(1 - B_i), \tag{13}$$

with $B_1, B_2, \ldots, B_J$ a solution to the system of equations

$$\begin{aligned} B_j &= E(\rho_j, C_j), \qquad j \in J, & (14) \\ \rho_j &= \sum_{r \in R_j} \nu_r \prod_{i \in r \setminus \{j\}} (1 - B_i), \qquad j \in J, & (15) \end{aligned}$$

where

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

is Erlang's formula for the blocking probability on a single isolated link with Poisson traffic offered at rate ν. The EFPA has the effect of replacing the true probability measure π by a more amenable measure $\mathcal{P}$. Under $\mathcal{P}$, each link j is assumed to be offered a stream of traffic at a constant rate ρ_j. If indeed this were the case, the equilibrium probability distribution for $\boldsymbol{U}$ would be $\mathcal{P}(\boldsymbol{U} = \boldsymbol{u}) = \prod_{j \in J} \mathcal{P}(U_j = u_j)$, where

$$\mathcal{P}(U_j = u) = \frac{\rho_j^u}{u!} \left(\sum_{n=0}^{C} \frac{\rho_j^n}{n!} \right)^{-1}.$$

This amounts to the assumption that the links operate independently. Under $\mathcal{P}$, the probability that link j is full is B_j in equation (14).

Except in the most trivial of circumstances, calls on routes that use link j do not arrive at a constant rate. In state $\boldsymbol{Y}$ with $\boldsymbol{U} = A\boldsymbol{Y}$, calls requesting link j as part of their route actually arrive at rate $a_j(\boldsymbol{U}) = \sum_{r\in R_j} \nu_r \, 1_{\{\boldsymbol{U}+A\boldsymbol{e}_r \le \boldsymbol{C}\}}$. Whenever $\sum_{r\in R_i} Y_r < C_i$, for all i, the arrivals seen by link j form a Poisson stream with rate $\sum_{r\in R_j} \nu_r$. In states $\boldsymbol{Y}$ for which one or more links i are full (that is, $\sum_{r\in R_i} Y_r = C_i$), the arrival stream for link j only includes calls that can be accepted without violating the capacity constraints. When j is full, $a_j(\boldsymbol{U}) = 0$.

The quantity ρ_j given in expression (15) can be interpreted as an expected arrival rate under the distribution $\mathcal{P}$:

$$\begin{aligned} \rho_j(u_j) &= \mathbb{E}_{\mathcal{P}}(a_j(\boldsymbol{U})|U_j = u_j) \\ &= \sum_{r\in R_j} \nu_r \mathcal{P}\left(\boldsymbol{U} + A\boldsymbol{e}_r \le \boldsymbol{C} | U_j = u_j\right) \\ &= \sum_{r\in R_j} \nu_r \mathcal{P}\left(\cap_{i\in r}\{U_i + 1 \le C_i\} | U_j = u_j\right) \\ &= \sum_{r\in R_j} \nu_r \prod_{i\in r\setminus\{j\}} (1 - B_i)\, 1_{\{u_j < C_j\}}. \end{aligned} \tag{16}$$

(For a more general description of the idea of expected rates, see Pollett and Thompson (2001).) The system comprising (14) and (15) is simply stating that, for each link $j \in J$, the likelihood of congestion and the intensity of offered traffic should be consistent. Kelly (1986) proved that, for the model under consideration, there is a unique fixed point $(B_1, \ldots, B_J) \in [0,1]^J$ of the system.

The EFPA fits the MFA framework described in Section 3. Specifically, the EFPA can be realised by assuming

$$\mathcal{P}\left(\boldsymbol{Y}_{R_j} = \boldsymbol{n}_{R_j}\right) \propto \prod_{i\sim j} \theta_{ji}\left(\boldsymbol{n}_{R_j\cap R_i}\right) \prod_{r\in R_j} \frac{\nu_r^{n_r}}{n_r!},$$

along with

$$\theta_{ji}\left(\boldsymbol{n}_{R_j\cap R_i}\right) = \prod_{r\in R_i\cap R_j} (1 - B_i)^{n_r},$$

for individual links i and j in J.

The EFPA is known to be effective under a variety of limiting regimes. Kelly (1991) proved that the estimates for a network with fixed routing and no controls tend towards the exact probabilities when (i) the link capacities and arrival rates are increased at the same rate, keeping the network topology fixed (Kelly limiting regime), and (ii) (Ziedins and Kelly 1989) the number of links and routes are increased while the link loads are held constant (diverse routing limit). The EFPA performs least well in highly linear networks and in circumstances where the offered traffic loads are roughly equal to the capacities (critically loaded).

The relationships between $\mathcal{P}$ and the probability measures P and π are interesting enough to mention. If there were no capacity constraints, then all three would imply that U_j is a Poisson random variable with mean $\sum_{r\in R_j} \nu_r$, but only under $\mathcal{P}$ do the links operate independently. When the constraints $A\boldsymbol{Y} \le \boldsymbol{C}$ are added, π and $\mathcal{P}$ bear little resemblance. They may not even be equivalent measures.

The true distribution π restricts $\boldsymbol{U}$ to the set $\{\boldsymbol{u} \in \mathbb{N}^J : \exists \boldsymbol{n} \in S : A\boldsymbol{n} = \boldsymbol{u}\}$, whereas the approximate distribution $\mathcal{P}$ assigns non-zero probability mass to all of the states $\{\boldsymbol{u} \in \mathbb{N}^J : \boldsymbol{u} \leq \boldsymbol{C}\}$. Only if the routing matrix A has rank J will the two sets coincide. Some extra care should be taken when applying the EFPA to networks with $rank(A) < J$ (Kelly 1991). One thing that the two measures π and $\mathcal{P}$ do share, is a common expression for the expected utilisation of link j:

$$\begin{aligned}\mathbb{E}_\pi(U_j) &= \sum_{k=0}^{C_j} \sum_{\boldsymbol{u}:u_j=k} k\, \pi(\boldsymbol{U} = \boldsymbol{u}) \\ &= \sum_{k=0}^{C_j} \sum_{\boldsymbol{u}:u_j=k} \sum_{r \in R_j} \nu_r\, \pi(\boldsymbol{U} = \boldsymbol{u} - A\boldsymbol{e}_r) \qquad \text{(from (7))} \\ &= \sum_{r \in R_j} \nu_r (1 - L_r). \end{aligned} \tag{17}$$

The construction of $\mathcal{P}$ using reduced load rates ρ_j, as given by (16), ensures that $\mathbb{E}_{\mathcal{P}}(U_j)$ is also equal to $\sum_{r\in R_j} \nu_r(1-L_r)$, but this time L_r is only an estimate of the loss probability of calls on route r as calculated from (13). The marginal distribution of U_j under both P and $\mathcal{P}$ is $(\rho_j^u/u!,\ u = 0, 1, \ldots, C_j)$ appropriately normalised. It appears as though $\mathcal{P}(U_j = u_j)$ has adopted the exact form of $P(U_j = u_j)$ with the transition rate $\sum_{r\in R_j} \nu_r$ of $U_j \to U_j + 1$ replaced by the reduced rates (15), so that $\mathbb{E}_{\mathcal{P}}(U_j)$ is consistent with (17). This observation is the motivation for our two-link approximation.

5 A two-link approximation

An estimate of the route loss probabilities, which is more accurate than those in (13), can be obtained by taking into account the link interdependencies. This *two-link approximation* is achieved by approximating the joint distribution of the usage on *pairs* of links (the EFPA effectively estimates this distribution on single links). The approximation is as follows. For each pair of links i, j, let

$$h_{ij}(u_{i|j}, u_{ij}, u_{j|i}) = \frac{\prod_{m=0}^{u_{i|j}-1} \rho_{i|j}(m)}{u_{i|j}!} \frac{\prod_{m=0}^{u_{ij}-1} \rho_{ij}(m)}{u_{ij}!} \frac{\prod_{m=0}^{u_{j|i}-1} \rho_{j|i}(m)}{u_{j|i}!}, \tag{18}$$

for $(u_{i|j}, u_{ij}, u_{j|i}) \in \mathbb{N}^3 : u_{i|j} + u_{ij} \leq C_i, u_{j|i} + u_{ij} \leq C_j$, where

$$\rho_{i|j}(u) = \sum_{r \in R_i \setminus R_j} \nu_r \sum_{u_{ij}=0}^{\min(C_i-u, C_j)} \prod_{k \in r} \left(1 - B_{k|i}(u + u_{ij})\right) \frac{\sum_{v=0}^{C_j-u_{ij}} h_{ij}(u, u_{ij}, v)}{\sum_{w=0}^{C_i-u-1} \sum_{v=0}^{C_j-w} h_{ij}(u, w, v)}, \tag{19}$$

$$\rho_{ij}(u) = \sum_{r \in R_i \cap R_j} \nu_r \sum_{u_{i|j}=0}^{C_i-u-1} \prod_{k \in r} \left(1 - B_{k|i}(u_{i|j} + u)\right) \frac{\sum_{v=0}^{C_j-u-1} h_{ij}(u_{i|j}, u, v)}{\sum_{w=0}^{C_i-u-1} \sum_{v=0}^{C_j-u-1} h_{ij}(w, u, v)}, \tag{20}$$

and

$$B_{k|i}(u_i) = \begin{cases} \frac{\sum_{u_{ik}=0}^{\min(C_k,u_i)} h_{ki}(C_k - u_{ik}, u_{ik}, u_i - u_{ik})}{\sum_{u_k=0}^{C_k} \sum_{u_{ik}=0}^{\min(u_k,u_i)} h_{ki}(u_k - u_{ik}, u_{ik}, u_i - u_{ik})}\,, & \text{if } k \neq i, \\ 1_{\{u_i = C_i\}}, & \text{if } k = i. \end{cases} \tag{21}$$

These equations will be derived in Section 7. They form a set of equations in the unknowns $\boldsymbol{B} = (\boldsymbol{B}_{k|i}; i, k \in J)$, where $\boldsymbol{B}_{k|i} = (B_{k|i}(m); m \leq C_i) \in \mathbb{R}^{C_i}$. Existence of a fixed point is guaranteed by Brouwer's Fixed Point Theorem. To see this, let $\Omega_k = \{\boldsymbol{x}_k \in \prod_{i \in J} \mathbb{R}^{C_i} : \boldsymbol{0} \leq \boldsymbol{x}_k \leq \boldsymbol{1}\}$, and observe that

$$f_{k|i}^{u_i}(\boldsymbol{B}) = \begin{cases} \frac{\sum_{u_{ik}=0}^{\min(C_k,u_i)} h_{ki}(C_k - u_{ik}, u_{ik}, u_i - u_{ik})}{\sum_{u_k=0}^{C_k} \sum_{u_{ik}=0}^{\min(u_k,u_i)} h_{ki}(u_k - u_{ik}, u_{ik}, u_i - u_{ik})}\,, & \text{if } k \neq i, \\ 1_{\{u_i = C_i\}}, & \text{if } k = i, \end{cases}$$

is a continuous mapping from $\Omega = \prod_{k \in J} \Omega_k$ into $[0, 1]$. Thus with $\boldsymbol{f} = (f_{k|i}^{u_i}; u_i = 0, \ldots, C_i, k, i \in J)$, we have $\boldsymbol{f}(\Omega) \subseteq \Omega$, and therefore $\boldsymbol{f}$ has at least one fixed point in Ω.

The loss probabilities can be estimated using $\boldsymbol{h} = (h_{ij}; i, j \in J)$. Losses on two-link routes, for example, have

$$L_r = 1 - \frac{\Phi_{ij}(C_i - 1, C_j - 1)}{\Phi_{ij}(C_i, C_j)}\,, \qquad \text{if } r = \{i, j\}, \tag{22}$$

where

$$\Phi_{ij}(C_i, C_j) = \sum_{u_i=0}^{C_i} \sum_{u_j=0}^{C_j} \sum_{k=0}^{\min(u_i,u_j)} h_{ij}(u_i - k, k, u_j - k)\,.$$

Calls that use the single link $r = \{i\}$ are lost with probability

$$B_i = 1 - \frac{\Phi_{ij}(C_i - 1, C_j)}{\Phi_{ij}(C_i, C_j)}\,, \tag{23}$$

where j is any link with a route common to i.

The rationale for the approximation is as follows. The traffic offered to a subsystem consisting of two arbitrary links, i and j, can be classified as either (i) link i only, (ii) link j only, or (iii) common to both links. Correspondingly, let $U_{i|j} = \sum_{r \in R_i \setminus R_j} Y_r$, $U_{j|i} = \sum_{r \in R_j \setminus R_i} Y_r$ and $U_{ij} = \sum_{r \in R_i \cap R_j} Y_r$ be, respectively, the number of calls using link i, the number using link j, and the number on routes using both i and j. This is a natural way to classify the traffic offered to the subsystem. Without capacity constraints, the joint distribution of the link utilisations $U_i = U_{i|j} + U_{ij}$ and $U_j = U_{j|i} + U_{ij}$ is

$$P(U_i = u_i, U_j = u_j) = \sum_{k=0}^{\min(u_i,u_j)} P(U_{i|j} = u_i - k, U_{ij} = k, U_{j|i} = u_j - k)\,,$$

where

$$P(U_{i|j} = u_{i|j}, U_{ij} = u_{ij}, U_{j|i} = u_{j|i}) = \frac{\rho_{i|j}^{u_{i|j}}}{u_{i|j}!}\frac{\rho_{ij}^{u_{ij}}}{u_{ij}!}\frac{\rho_{j|i}^{u_{j|i}}}{u_{j|i}!}e^{-(\rho_{i|j}+\rho_{ij}+\rho_{j|i})}, \qquad (24)$$

with $\rho_{ij} = \sum_{r\in R_i\cap R_j} \nu_r$, $\rho_{i|j} = \sum_{r\in R_i\setminus R_j} \nu_r$ and $\rho_{j|i} = \sum_{r\in R_j\setminus R_i} \nu_r$. To construct a reduced load approximation we shall replace the aggregate rates ρ_{ij}, $\rho_{i|j}$ and $\rho_{j|i}$ in (24) with reduced load rates, and we isolate the subsystem composed of traffic offered to links i and j. Motivated by the form of (24), let us suppose for the moment that $\pi(U_{i|j} = u_{i|j}, U_{ij} = u_{ij}, U_{j|i} = u_{j|i})$ has the form $h_{ij}(u_{i|j}, u_{ij}, u_{j|i})/\Phi_{ij}(C_i, C_j)$. If this were the case then questions concerning call blocking could be answered easily. For instance, the probability that link i is full would be B_i in expression (23), the probability that either link i or link j are full would be L_r in expression (22), and the conditional probability that link k is full given link i carries u_i calls would be $B_{k|i}(u_i)$ in expression (21). To ensure that the traffic offered to the subsystem is consistent with blocking in other parts of the network, the rates ρ_{ij}, $\rho_{i|j}$ and $\rho_{j|i}$ are replaced by state-dependent reduced load rates. For example, expression (19) for $\rho_{i|j}(u_{i|j})$ is just $\rho_{i|j} = \sum_{r\in R_i\setminus R_j} \nu_r$, but reduced by an estimate of the expected blocking on the other links $k \in r$ such that $r \in R_i \setminus R_j$ when link i is carrying $u_{i|j}$ calls that are not also carried by link j.

6 Examples

In this section we examine the performance of the two-link reduced load approximation when applied to a suite of simple networks. To compare its accuracy with that of other approximations, we have used relative errors: specifically, the difference between the approximate value and the exact loss probability, expressed as a proportion of the exact value. These exact values were calculated using expression (5). In Section 6.5 we compare the computation times of the EFPA, the MFA, and our two-link approximation with the time it takes to compute the exact loss probabilities.

6.1 A star network

Consider a private computing network consisting of a number of workstations linked to a central mainframe in a star configuration. Each workstation is linked directly to the central processor. Any exchange of information between workstations must be via the central mainframe. This structure is quite common and in the past it was a popular design for computing environments. As such, the backbone of many networks in existence today is a number of star configurations with a few additional links to improve resilience (Lloyd-Evans 1996).

In a star network, each link carries a single-link traffic as well as sharing two-link traffic with each of the other links. For simplicity, we will assume that the network is completely symmetric: the link capacities are the same ($C_j = C$ for all $j \in J = \{1, 2, \ldots, l\}$), each link is offered single-link traffic at the same rate ν_1 and the $l-1$ streams of two-link traffic are each offered at rate ν_2.

6.1.1 The Erlang fixed point approximation

When considered in isolation, the arrivals at any given link consist of $l-1$ streams at rate ν_2, each thinned by a factor $(1-B)$, and, one traffic stream at rate ν_1. Thus, the EFPA for the loss of single-link and two-link calls, respectively, are given by

$$L_1 = B \qquad \text{and} \qquad L_2 = 1-(1-B)^2,$$

where B is the solution to

$$B = E\big(\nu_1 + (l-1)\nu_2(1-B), C\big).$$

6.1.2 The two-link approximation

The two-link reduced load approximation is obtained by solving the system comprising (25) and (26) below. By the symmetry of the network, $B_{k|i}(u) = B(u)$ and $\rho_{i|j}(u) = \rho(u)$ are independent of i and j. Since the longest route consists of only two links, $\rho_{ij}(u) = \nu_2$. The parameters $B(u)$ and $\rho(u)$ satisfy

$$\rho(u) = \nu_1 + (J-2)\nu_2 \sum_{w=0}^{C-u-1} \big(1 - B(w+u)\big) \frac{\sum_{v=0}^{C-w} \frac{\prod_{m=0}^{u-1}\rho(m)}{u!} \frac{\nu_2^w}{w!} \frac{\prod_{m=0}^{v-1}\rho(m)}{v!}}{\sum_{k=0}^{C-u-1} \sum_{v=0}^{C-k} \frac{\prod_{m=0}^{u-1}\rho(m)}{u!} \frac{\nu_2^k}{k!} \frac{\prod_{m=0}^{v-1}\rho(m)}{v!}}, \tag{25}$$

and

$$B(u) = \frac{\sum_{w=0}^{\min(C,u)} \frac{\prod_{m=0}^{C-w-1}\rho(m)}{(C-w)!} \frac{\nu_2^w}{w!} \frac{\prod_{m=0}^{u-w-1}\rho(m)}{(u-w)!}}{\sum_{v=0}^{C} \sum_{w=0}^{\min(v,u)} \frac{\prod_{m=0}^{v-w-1}\rho(m)}{(C-w)!} \frac{\nu_2^w}{w!} \frac{\prod_{m=0}^{u-w-1}\rho(m)}{(u-w)!}}, \qquad \text{for } u = 0,\dots,C-1. \tag{26}$$

Under this scheme, the loss probabilities are estimated to be

$$L_1 = 1 - \frac{\Phi(C-1,C)}{\Phi(C,C)} \qquad \text{and} \qquad L_2 = 1 - \frac{\Phi(C-1,C-1)}{\Phi(C,C)}, \tag{27}$$

with

$$\Phi(u_i,u_j) = \sum_{x=0}^{u_i} \sum_{y=0}^{u_j} \sum_{k=0}^{\min(x,y)} \frac{\prod_{m=0}^{x-k-1}\rho(m)}{(x-k)!} \frac{\nu_2^k}{k!} \frac{\prod_{m=0}^{y-k-1}\rho(m)}{(y-k)!}.$$

6.1.3 Zachary and Ziedins' method

In Section 4 of their paper, Zachary and Ziedins (1999) describe a generic approximation for networks that exhibit a certain degree of symmetry. For the star model, the approximation is achieved by replacing the existing probability measure π under which

$$\pi\left(\boldsymbol{Y}_{R_j} = \boldsymbol{n}_{R_j}\right) = \frac{\theta\left(\boldsymbol{n}_{\partial R_j}\right)}{G(\boldsymbol{C})} \prod_{r\in R_j} \frac{\nu_r^{n_r}}{n_r!}, \qquad \text{for all } j \in J,$$

by $\mathcal{P}$ with

$$\mathcal{P}\left(\boldsymbol{Y}_{R_j}=\boldsymbol{n}_{R_j}\right) \propto \prod_{k=1}^{l-1} \lambda\left(\boldsymbol{n}_{R_j \cap R_k}\right) \prod_{r \in R_j} \frac{\nu_r^{n_r}}{n_r!}, \qquad \text{for all } j \in J,$$

where λ is given by

$$\lambda\left(\boldsymbol{n}_{R_j \cap R_k}\right) \propto \sum_{\substack{\boldsymbol{m}_{R_k} \in S_{R_k}: \\ \boldsymbol{m}_{R_j \cap R_k}=\boldsymbol{n}_{R_j \cap R_k}}} \prod_{i=1}^{l-2} \lambda\left(\boldsymbol{m}_{R_k \cap R_i}\right) \prod_{r \in R_k \backslash R_j} \frac{\nu_r^{m_r}}{m_r!}.$$

Under $\mathcal{P}$, instances of blocking of single-link and two-link routes have the respective likelihoods

$$L_1=\frac{\sum_{k=0}^{C-1} \lambda(k) \lambda(k+1) \frac{\nu_2^k}{k!}}{\sum_{k=0}^{C} \lambda(k) \lambda(k) \frac{\nu_2^k}{k!}} \qquad \text{and} \qquad L_2=\frac{\sum_{k=0}^{C-1} \lambda(k+1) \lambda(k+1) \frac{\nu_2^k}{k!}}{\sum_{k=0}^{C} \lambda(k) \lambda(k) \frac{\nu_2^k}{k!}}.$$

This scheme is labelled MFA.

Figure 1 compares the relative errors in the MFA, EFPA, and two-link reduced load approximation schemes. The network considered had five links and five circuits per link. The x-axes have the single-link arrival rate ν_1 varying over $[0,10]$. We have chosen $\nu_2=\nu_1/4$, so that each link is offered roughly equal proportions of single-link and two-link traffic. It is apparent that the two-link approximation compares favourably with the EFPA over most of the region tested. The accuracy of the two-link scheme is only marginally worse than the MFA.

6.2 A ring network

Reduced load approximations such as the EFPA tend to perform least well in networks of linear structure, with the links joined end-to-end or in a cycle. A popular test case is the ring network, where the links are arranged in a loop with adjacent pairs of links sharing routes.

As with the star network, we assume a high degree of symmetry in the model. Suppose that all links have the same capacity C and that there are only two types of traffic. Single-link traffic is offered to each link, $1,2,\ldots,l$, at a common rate ν_1 and two-link traffic is offered to each pair of adjacent links, $\{1,2\},\{2,3\},\ldots,\{l,1\}$, at rate ν_2.

The MFA is applicable to the star network. Indeed, successive applications of recurrence (6) provides a means of *exact* analysis in reasonable time; see Zachary and Ziedins (1999) for details.

6.2.1 The Erlang fixed point approximation

Arguing that every link sees one traffic stream at rate ν_1 and two streams at ν_2 thinned by a factor $(1-B)$, representing the proportion of calls accepted on neighbouring links, the reduced load rate for the EFPA is $\nu_1+2\nu_2(1-B)$. The EFPAs for the loss of single-link and two-link calls, respectively, are given by

$$L_1=B \qquad \text{and} \qquad L_2=1-(1-B)^2,$$

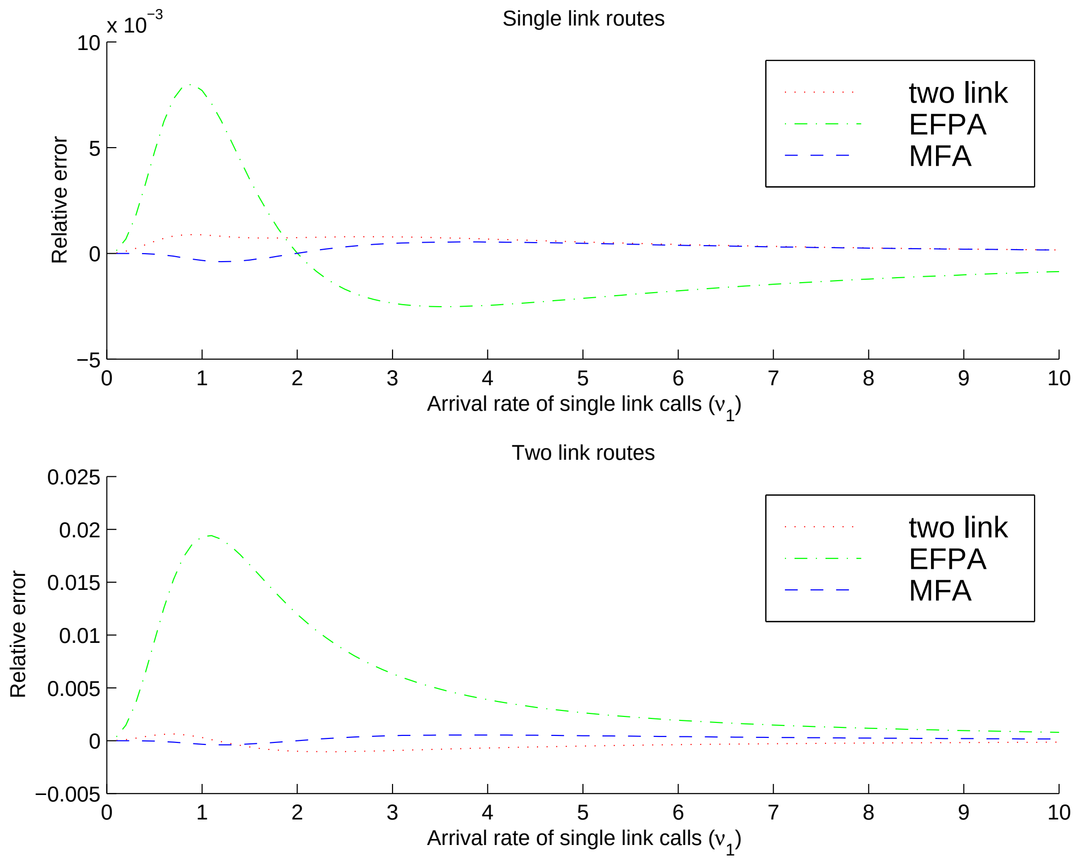

Figure 1: Accuracy for a star network ($J = 5$, $C = 5$, $\nu_2 = \nu_1/2$)

where B is the solution to

$$B = E\big(\nu_1 + 2\nu_2(1 - B), C\big).$$

6.2.2 The two-link approximation

The EFPA is accurate when links are blocked almost independently of one another. Unfortunately, the link utilisations are sometimes significantly dependent. This is particularly true of linear and cyclic networks, such as the ring. The two-link approximation is an attempt to account for the link interactions. The approximation used for the star network requires only minor modification for the ring network. In fact, the only change is that

$$\rho(u) = \nu_1 + \nu_2 \sum_{w=0}^{C-u-1} \big(1 - B(w+u)\big) \frac{\sum_{v=0}^{C-w} \frac{\prod_{m=0}^{u-1}\rho(m)}{u!} \frac{\nu_2^w}{w!} \frac{\prod_{m=0}^{v-1}\rho(m)}{v!}}{\sum_{k=0}^{C-u-1} \sum_{v=0}^{C-k} \frac{\prod_{m=0}^{u-1}\rho(m)}{u!} \frac{\nu_2^k}{k!} \frac{\prod_{m=0}^{v-1}\rho(m)}{v!}},$$

instead of (25) (in the ring network each link i carries a single two-link route $\{i, i+1\}$, not shared with an adjacent link $i-1$). Expression (26) for $B(u)$ and expressions (27) for the loss probabilities remain unaltered.

6.2.3 The method of Bebbington, Pollett and Ziedins

A similar approximation for the ring network was previously devised by Bebbington, Pollett and Ziedins (1997) (here labelled BPZ). In both their *Approximation II* and our two-link approximation, the rates are reduced by a usage-dependent factor $(1 - B(m))$. Link i is offered three streams of traffic. Let Y_i, $Y_{i,i+1}$ and $Y_{i-1,i}$ be the numbers currently carried on the respective streams. Taking into account the cyclic structure of the network, we write $i = 1$ for $i = l + 1$. For $m = 0, \ldots, C-1$, they define

$$B(m) = \mathcal{P}\left(Y_i + Y_{i,i+1} + Y_{i-1,i} = C \,|\, Y_{i-1} + Y_{i-1,i} = m\right),$$

whereas our approximation requires

$$B(m) = \mathcal{P}\left(Y_i + Y_{i,i+1} + Y_{i-1,i} = C \,|\, Y_{i-1} + Y_{i-1,i} + Y_{i-2,i-1} = m\right).$$

Aside from this, the schemes are the same. The event $\{Y_{i-1} + Y_{i-1,i} = m\}$ yields more information than does $\{Y_{i-1} + Y_{i-1,i} + Y_{i-2,i-1} = m\}$ in determining the likelihood of $\{Y_i + Y_{i,i+1} + Y_{i-1,i} = C\}$.

Figure 2 shows that the relative errors in the estimates from the BPZ scheme are negligible when compared with our two-link approximation and the EFPA. Both two-link approximations improve on the EFPA.

6.3 A linear network

In this and the following section the accuracy of the two-link approximation is compared with the EFPA on a network in which the links are joined end-to-end. Typically, reduced load approximations perform poorly in linear networks.

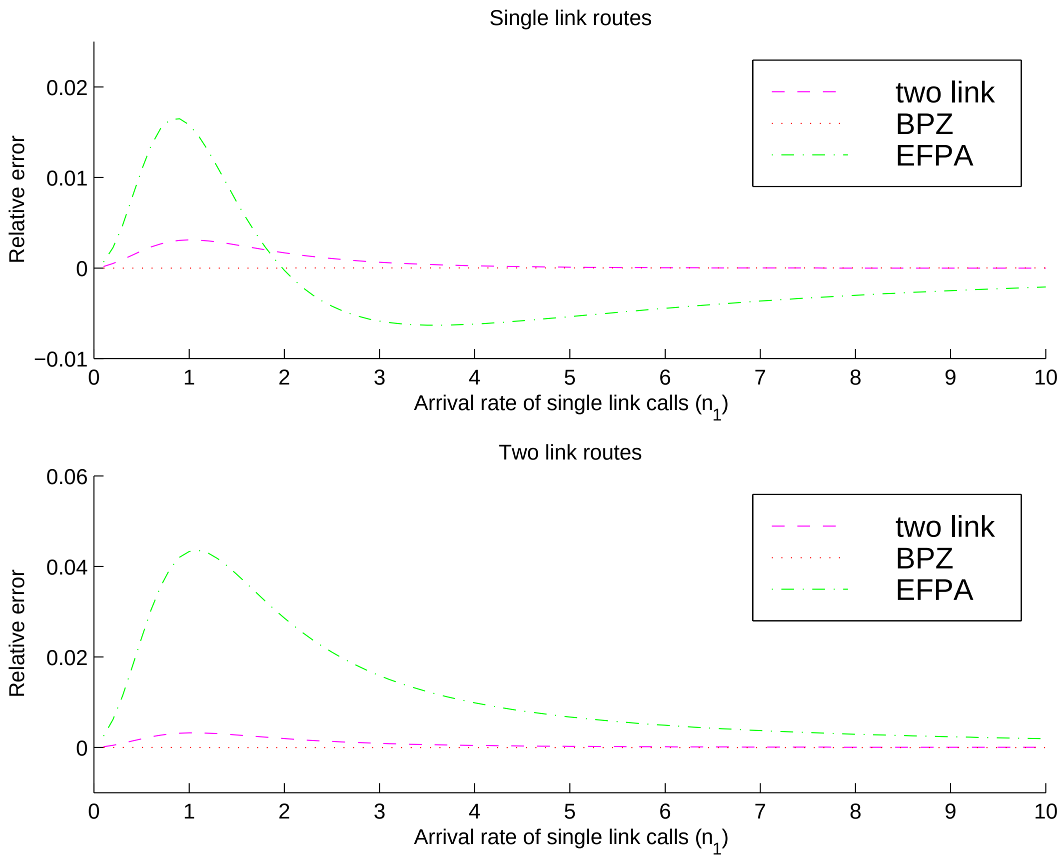

Figure 2: Accuracy for a ring network ($J = 5$, $C = 5$, $\nu_2 = \nu_1/2$)

Consider a network of l links labelled $1, \ldots, l$. Suppose that each link is offered a stream of single-link traffic, and that each of the links $i \in \{2, \ldots, l-1\}$ share two-link routes $\{i-1, i\}$ and $\{i, i+1\}$, with each of their neighbouring links. Thus, there are l single-link routes and $l-1$ two-link routes. For simplicity, assume that calls on single-link routes arrive at a common rate ν_1, and that calls on each two-link route arrive at rate ν_2.

6.3.1 The Erlang fixed point approximation

The EFPA for the route loss probabilities is

$$\begin{aligned} L_i &= B_i, \qquad i = 1, \ldots, l, \\ L_{i,i+1} &= 1 - (1 - B_i)(1 - B_{i+1}), \qquad i = 1, \ldots, l-1, \end{aligned}$$

where $(B_i; i = 1, \ldots, l)$ is the solution to

$$\begin{aligned} B_1 &= \nu_1 + \nu_2 (1 - B_2), \\ B_i &= \nu_1 + \nu_2 (1 - B_{i-1}) + \nu_2 (1 - B_{i+1}), \qquad i = 2, \ldots, l-1, \\ B_l &= \nu_1 + \nu_2 (1 - B_{l-1}). \end{aligned}$$

6.3.2 The two-link approximation

The two-link reduced load approximation for this network is as follows. For $u = 0, \ldots, C-1$, set $\rho_{ij}(u) = \nu_2$ for all $i, j = 1, \ldots, l$ such that $j = i-1$ or $j = i+1$, $\rho_{1|2}(u) = \nu_1$, $\rho_{l|l-1}(u) = \nu_1$, and let $(B_{j|i}(u); i, j = 1, \ldots, l, j = i-1 \,\text{or}\, j = i+1)$ be a solution to the system of equations

$$B_{j|i}(u) = \frac{\sum_{u_{ij}=0}^{\min(C,u)} h_{ji}(C - u_{ij}, u_{ij}, u - u_{ij})}{\sum_{u_j=0}^{C} \sum_{u_{ij}=0}^{\min(u_j,u)} h_{ji}(u_j - u_{ij}, u_{ij}, u - u_{ij})}, \qquad i, j = 1, \ldots, l,\ j = i-1 \text{ or } j = i+1,$$

$$\rho_{i|i-1}(u) = \nu_1 + \nu_2 \sum_{k=0}^{C-u-1} \big(1 - B_{i+1|i}(u+k)\big) \frac{\sum_{w=0}^{C-k} h_{i-1,i}(w, k, u)}{\sum_{v=0}^{C-u-1} \sum_{w=0}^{C-v} h_{i-1,i}(w, v, u)}, \qquad i = 2, \ldots, l-1,$$

$$\rho_{i|i+1}(u) = \nu_1 + \nu_2 \sum_{k=0}^{C-u-1} \big(1 - B_{i-1|i}(u+k)\big) \frac{\sum_{w=0}^{C-k} h_{i,i+1}(u, k, w)}{\sum_{v=0}^{C-u-1} \sum_{w=0}^{C-v} h_{i,i+1}(u, v, w)}, \qquad i = 2, \ldots, l-1,$$

where

$$h_{ij}(u_{i|j}, u_{ij}, u_{j|i}) = \frac{\prod_{m=0}^{u_{i|j}-1} \rho_{i|j}(m)}{u_{i|j}!} \frac{\nu_2^{u_{ij}}}{u_{ij}!} \frac{\prod_{m=0}^{u_{j|i}-1} \rho_{j|i}(m)}{u_{j|i}!},$$

for $i, j = 1, \ldots, l$, $j = i - 1$ or $j = i + 1$. Estimates for the loss probabilities on single-link routes are

$$\begin{aligned} L_i &= 1 - \frac{\Phi_{i,i+1}(C-1,C)}{\Phi_{i,i+1}(C,C)}, \quad \text{for } i = 1, \ldots, l-1, \quad \text{and} \\ L_i &= 1 - \frac{\Phi_{i,i-1}(C-1,C)}{\Phi_{i,i-1}(C,C)}, \quad \text{for } i = 2, \ldots, l, \end{aligned}$$

where

$$\Phi_{ij}(C_i, C_j) = \sum_{u_i=0}^{C_i} \sum_{u_j=0}^{C_j} \sum_{k=0}^{\min(u_i,u_j)} \frac{\prod_{m=0}^{u_i-k-1} \rho_{i|j}(m)}{(u_i-k)!} \frac{\nu_2^k}{k!} \frac{\prod_{m=0}^{u_j-k-1} \rho_{j|i}(m)}{(u_j-k)!},$$

for $i = 1, \ldots, l$ and j a link adjacent to i. For certain links there may be more than one possible estimate for the loss probability. For example, this scheme produces two estimates for the loss probability on link 2:

$$L_2 = 1 - \frac{\Phi_{12}(C, C-1)}{\Phi_{12}(C,C)} \qquad \text{and} \qquad L_2 = 1 - \frac{\Phi_{23}(C-1, C)}{\Phi_{23}(C,C)}.$$

In practice there is no way of knowing which estimate will be the most accurate. Both estimates achieved greater precision than the EFPA for the network tested here. There is no ambiguity in estimating the loss probabilities on two-link routes. For $i = 1, \ldots, l-1$,

$$L_{i,i+1} = 1 - \frac{\Phi_{i,i+1}(C-1, C-1)}{\Phi_{i,i+1}(C,C)}.$$

In Figures 3, 4 and 5, the relative errors in the loss probability estimates for the EFPA and the two-link approximation are compared. The network tested had 5 links, each with a carrying capacity of 5 calls. The single-link route arrival rate ν_1 was varied over $[0, 10]$ and ν_2 was set at $\nu_1/2$. By symmetry, there are only three single-link routes and two two-link routes to distinguish. In this test case, the two-link approximation provided a significant improvement in accuracy over the EFPA for each of the two-link routes (Figure 5), the single-link route using an end link (top pane of Figure 3) and the single-link route that uses the centre link (bottom pane of Figure 3). The single-link route that uses a link second from the end was the only one with multiple loss estimates. In Figure 4 the relative errors of the estimates of L_2, using Φ_{12} and Φ_{23}, are compared with the EFPA. Both two-link estimates show a significant improvement over the EFPA.

6.4 A linear network with three-link routes

As a final example, we will analyse a linear network, which is the same as the one in the previous example, except that there are additional traffic streams spanning groups of three adjacent links. The presence of these three-link routes increases the difficulty of accurately approximating the loss probabilities, because of the need to account for an increase in the amount interaction between links. Furthermore, their presence destroys the simple structure needed for the Zachary and Ziedins (1999) recursion to work.

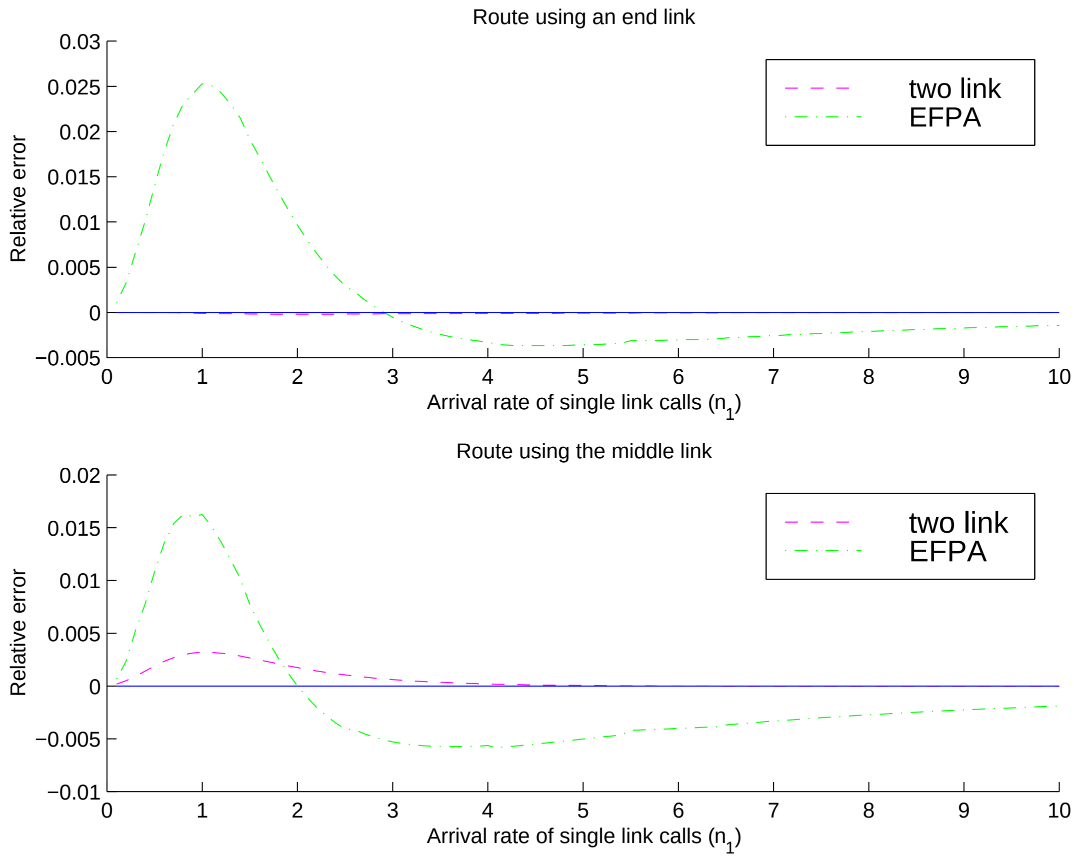

Figure 3: Accuracy for a line network (5 links, $C = 5$, $\nu_2 = \nu_1/2$)

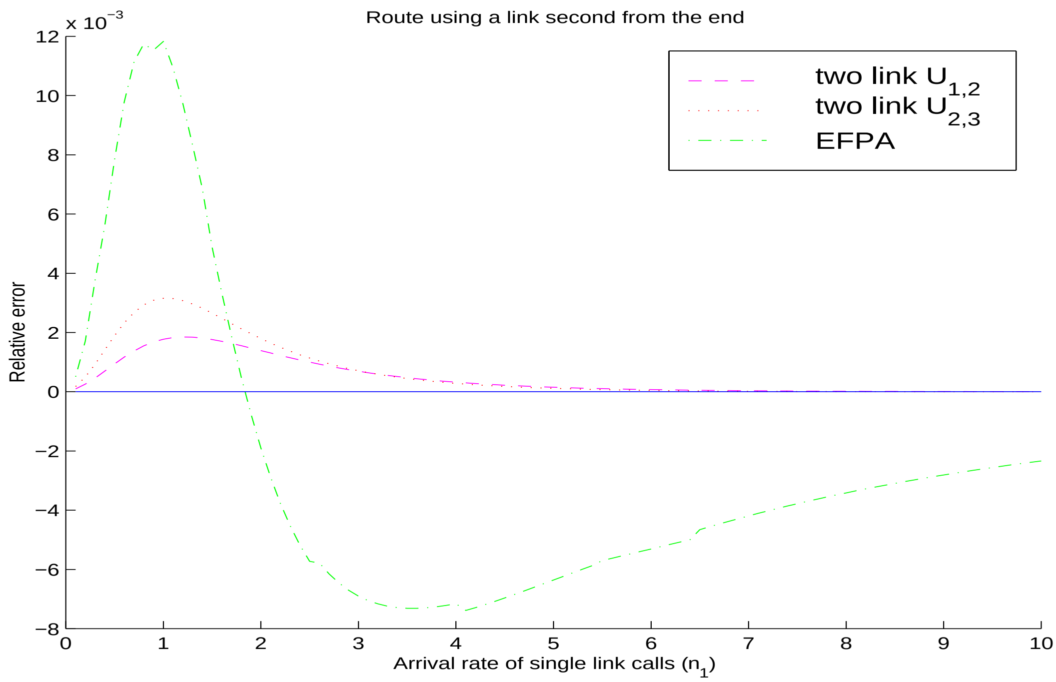

Figure 4: Accuracy for a line network (5 links, $C = 5$, $\nu_2 = \nu_1/2$)

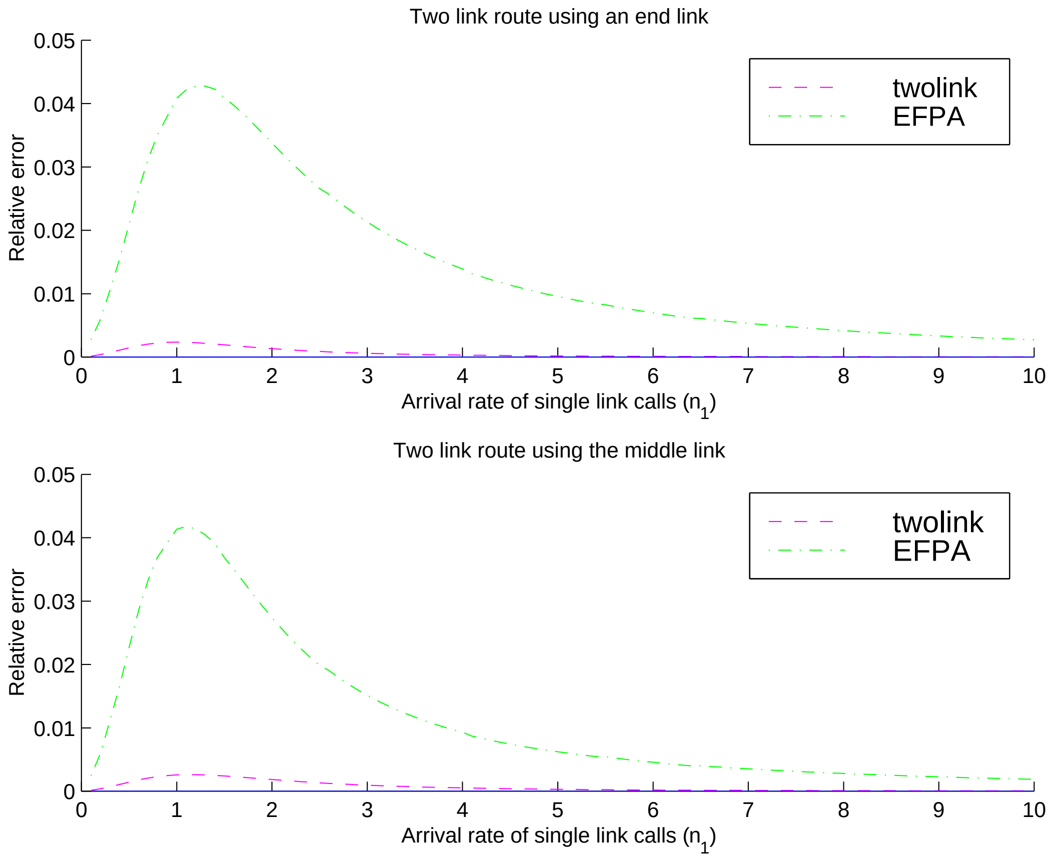

Figure 5: Accuracy for a line network (5 links, $C = 5$, $\nu_2 = \nu_1/2$)

6.4.1 The Erlang fixed point

Assume that traffic for each of the three-link routes arrives at common rate ν_3. The EFPA for the route loss probabilities is

$$\begin{aligned} L_i &= B_i, \quad i = 1, \ldots, l, \\ L_{i,i+1} &= 1 - (1 - B_i)(1 - B_{i+1}), \quad i = 1, \ldots, l-1, \\ L_{i,i+1,i+2} &= 1 - (1 - B_i)(1 - B_{i+1})(1 - B_{i+2}), \quad i = 1, \ldots, l-2, \end{aligned}$$

where $(B_i; i = 1, \ldots, l)$ is the solution to

$$\begin{aligned} B_1 &= E\left(\nu_1 + \nu_2(1 - B_2) + \nu_3(1 - B_2)(1 - B_3), C\right), \\ B_2 &= E\big(\nu_1 + \nu_2(1 - B_1) + \nu_2(1 - B_3) + \nu_3(1 - B_1)(1 - B_3) \\ &\quad + \nu_3(1 - B_3)(1 - B_4), C\big), \\ B_i &= E\big(\nu_1 + \nu_2(1 - B_{i-1}) + \nu_2(1 - B_{i+1}) + \\ &\quad \nu_3(1 - B_{i-2})(1 - B_{i-1}) + \nu_3(1 - B_{i-1})(1 - B_{i+1}) \\ &\quad + \nu_3(1 - B_{i+1})(1 - B_{i+2}), C\big), \quad i = 3, \ldots, l-2, \\ B_{l-1} &= E\big(\nu_1 + \nu_2(1 - B_l) + \nu_2(1 - B_{l-2}) \\ &\quad + \nu_3(1 - B_l)(1 - B_{l-2}) + \nu_3(1 - B_{l-2})(1 - B_{l-3}), C\big), \\ B_l &= E\left(\nu_1 + \nu_2(1 - B_{l-1}) + \nu_3(1 - B_{l-1})(1 - B_{l-2}), C\right). \end{aligned}$$

6.4.2 The two-link approximation

For $i, j = 1, \ldots, l$, let

$$h_{i,j}(u_{i|j}, u_{i,j}, u_{j|i}) = \frac{\prod_{m=0}^{u_{i|j}-1} \rho_{i|j}(m)}{u_{i|j}!} \frac{\prod_{m=0}^{u_{i,j}-1} \rho_{i,j}(m)}{u_{i,j}!} \frac{\prod_{m=0}^{u_{j|i}-1} \rho_{j|i}(m)}{u_{j|i}!},$$

and $\Phi_{i,j}(C, C) = \sum_{u_i=0}^{C} \sum_{u_j=0}^{C} \sum_{u_{i,j}=0}^{\min(u_i,u_j)} h_{i,j}(u_i - u_{i,j}, u_{i,j}, u_j - u_{i,j})$. We propose to estimate the loss probabilities on single and two-link routes as

$$\begin{aligned} L_i &= 1 - \frac{\Phi_{i,i+1}(C-1, C)}{\Phi_{i,i+1}(C, C)}, \text{ for } i = 1, \ldots, l-1, \quad \text{or} \\ L_i &= 1 - \frac{\Phi_{i,i-1}(C-1, C)}{\Phi_{i,i-1}(C, C)}, \text{ for } i = 2, \ldots, l, \end{aligned}$$

and $L_{i,i+1} = 1 - \Phi_{i,i+1}(C-1, C-1)/\Phi_{i,i+1}(C, C)$, for $i = 1, \ldots, l-1$. Loss probabilities on three-link routes $\{i, i+1, i+2\}$ are then estimated as $L_{i,i+1,i+2} = 1 - (1 - L_{i,i+1})(1 - L_{i+2})$.

Applying our technique to this network requires us to estimate $(B_{i|j}(u), u = 0, \ldots, C)$ for each ordered pair of links (i, j) such that $|i - j| \leq 2$. Although there is no difficulty implementing the procedure for this network, it exposes a potential problem with the procedure: that, for large networks with routes spanning many links, the number of parameters needing to be estimated may be large and this may lead to excessive demands on memory. One possible solution is to have the analyst identify links i for which $B_{i|j}(u)$ is expected be approximately constant with respect to u. An algorithmic approach might then treat as constant all those $B_{i|j}(u)$'s for which the correlation between blocking events on the two links was relatively weak.

In the present context, let us make the simplifying assumption that $B_{i|j}(u) = B_i$ whenever $|i-j| \geq 2$. Under this assumption, estimates of the marginal reduced load rates are

$$\begin{aligned}
\rho_{1|2}(u) &= \nu_1, \\
\rho_{2|1}(u) &= \nu_1 + (\nu_2 + \nu_3(1-B_4)) \sum_{k=0}^{C-u-1} (1 - B_{3|2}(u+k)) H^{(1)}_{2,1}(k,u), \\
\rho_{2|3}(u) &= \nu_1 + \nu_2 \sum_{k=0}^{C-u-1} (1 - B_{1|2}(u+k)) H^{(1)}_{2,3}(k,u), \\
\rho_{i|i-1}(u) &= \nu_1 + (\nu_2 + \nu_3(1-B_{i+2}) \sum_{k=0}^{C-u-1} (1 - B_{i+1|i}(u+k)) H^{(1)}_{i,i-1}(k,u), \\
&\qquad \text{for } i = 3, \ldots, l-3,
\end{aligned}$$

$$\begin{aligned}
\rho_{i|i+1}(u) &= \nu_1 + (\nu_2 + \nu_3(1-B_{i-2})) \sum_{k=0}^{C-u-1} (1 - B_{i-1|i}(u+k)) H^{(1)}_{i,i+1}(k,u), \\
&\qquad \text{for } i = 3, \ldots, l-3,
\end{aligned}$$

$$\rho_{l-1|l-2}(u) = \nu_1 + \nu_2 \sum_{k=0}^{C-u-1} (1 - B_{l|l-1}(u+k)) H^{(1)}_{l-1,l-2}(k,u),$$

$$\begin{aligned}
\rho_{l-1|l}(u) &= \nu_1 + (\nu_2 + \nu_3(1-B_{l-3})) \sum_{k=0}^{C-u-1} (1 - B_{l-2|l-1}(u+k)) H^{(1)}_{l-1,l}(k,u), \\
\rho_{l|l-1}(u) &= \nu_1,
\end{aligned}$$

where $H^{(1)}_{i,j}(k,u) = \sum_{w=0}^{C-k} h_{i,j}(u,k,w) / \sum_{v=0}^{C-u-1} \sum_{w=0}^{C-v} h_{i,j}(u,v,w)$. And, the joint reduced load rates are

$$\begin{aligned}
\rho_{1,2}(u) &= \nu_2 + \nu_3 \sum_{k=0}^{C-u-1} (1 - B_{3|2}(k+u)) H^{(2)}_{2,1}(k,u), \\
\rho_{i,i+1}(u) &= \nu_2 + \nu_3 \sum_{k=0}^{C-u-1} (1 - B_{i-1|i}(k+u)) H^{(2)}_{i,i+1}(k,u) \\
&\quad + \nu_3 \sum_{k=0}^{C-u-1} (1 - B_{i+2|i+1}(k+u)) H^{(2)}_{i+1,i}(k,u), \\
&\qquad \text{for } i = 2, \ldots, l-2,
\end{aligned}$$

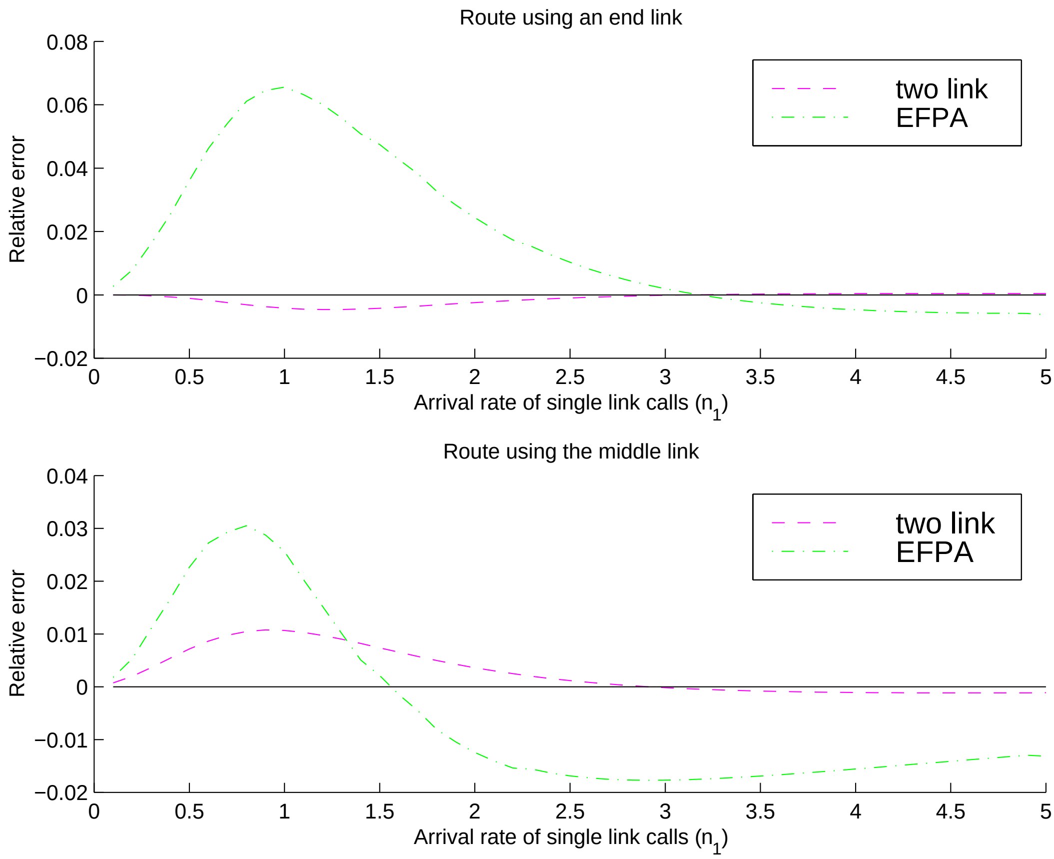

Figure 6: Accuracy for a line network (5 links, $C = 5$, $\nu_2 = \nu_1/2$)

$$\begin{aligned}
\rho_{i,i-1}(u) &= \nu_2 + \nu_3 \sum_{k=0}^{C-u-1} (1 - B_{i+1|i}(k+u)) H^{(2)}_{i,i-1}(k,u) \\
&\quad + \nu_3 \sum_{k=0}^{C-u-1} (1 - B_{i-2|i-1}(k+u)) H^{(2)}_{i-1,i}(k,u), \\
&\qquad\qquad\qquad\qquad\qquad \text{for } i = 3, \ldots, l-1, \\
\rho_{l,l-1}(u) &= \nu_2 + \nu_3 \sum_{k=0}^{C-u-1} (1 - B_{l-2|l-1}(k+u)) H^{(2)}_{l-1,l}(k,u),
\end{aligned}$$

where $H^{(2)}_{i,j}(k,u) = \sum_{w=0}^{C-u-1} h_{i,j}(k,u,w) / \sum_{v=0}^{C-u-1} \sum_{w=0}^{C-u-1} h_{i,j}(v,u,w)$.

We compare the relative errors in the proposed two-link approximation with those of the Erlang fixed point approximation in Figures 6, 7, 8, and 9. Our approximation shows an improvement for all of the single-link routes. On the routes where multiple approximations are possible, it may be beneficial to take an average of the approximations. Since we cannot be sure which approximation will be the more accurate beforehand, this would make the results more robust. Interestingly, neither of the two-link approximations are consistently better than the other (see Figure 7). Significant improvements over the EFPA are also observed in Figure 8 for the two-link routes. On the three-link routes, our proposed approximation again improves on the EFPA (see Figure 9).

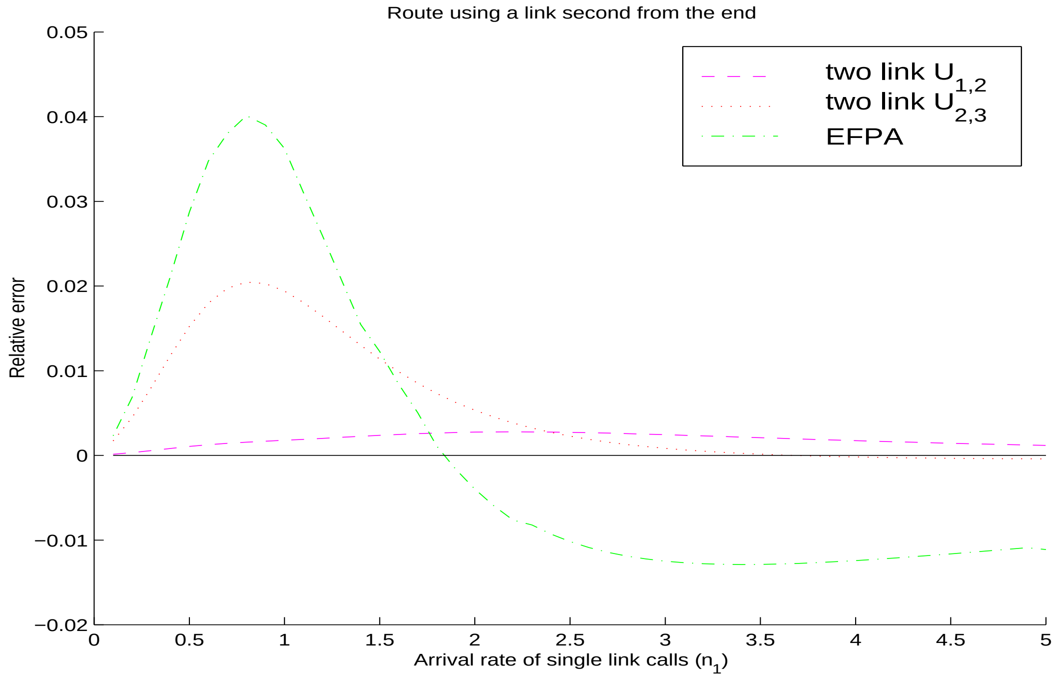

Figure 7: Accuracy for a line network (5 links, $C = 7$, $\nu_2 = \nu_1/2, \nu_3 = \nu_1/3$)

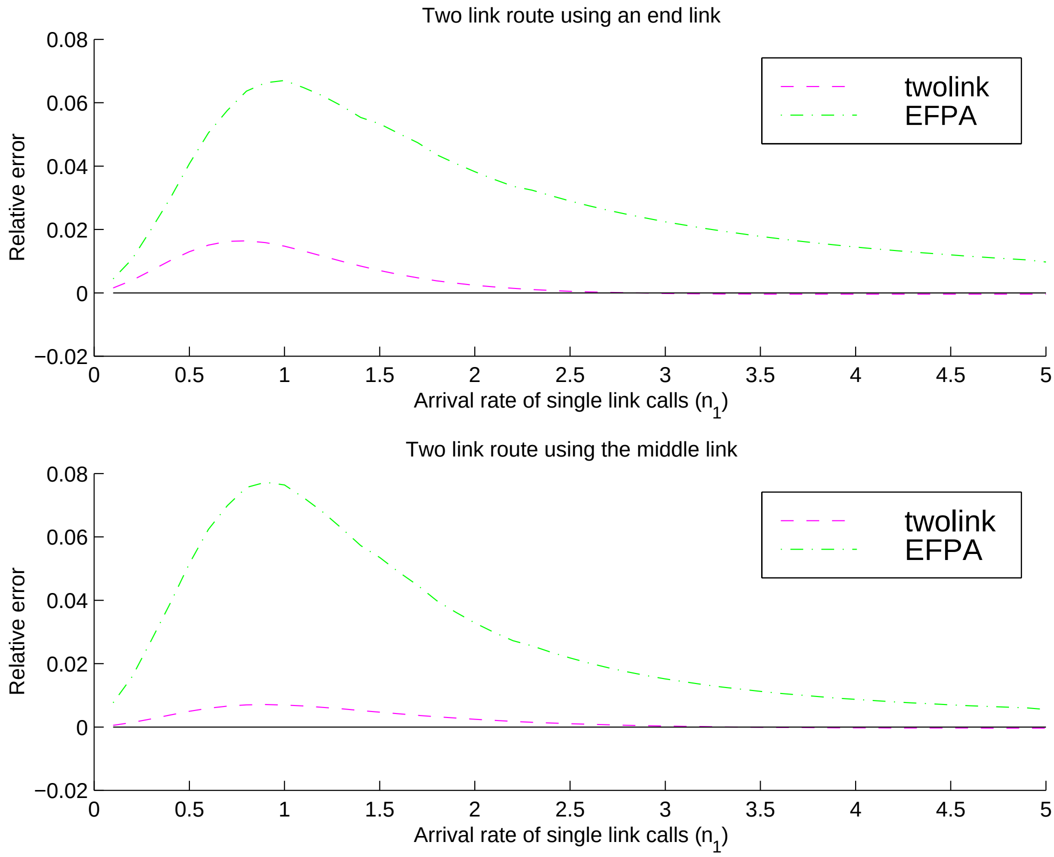

Figure 8: Accuracy for a line network (5 links, $C = 7$, $\nu_2 = \nu_1/2, \nu_3 = \nu_1/3$)

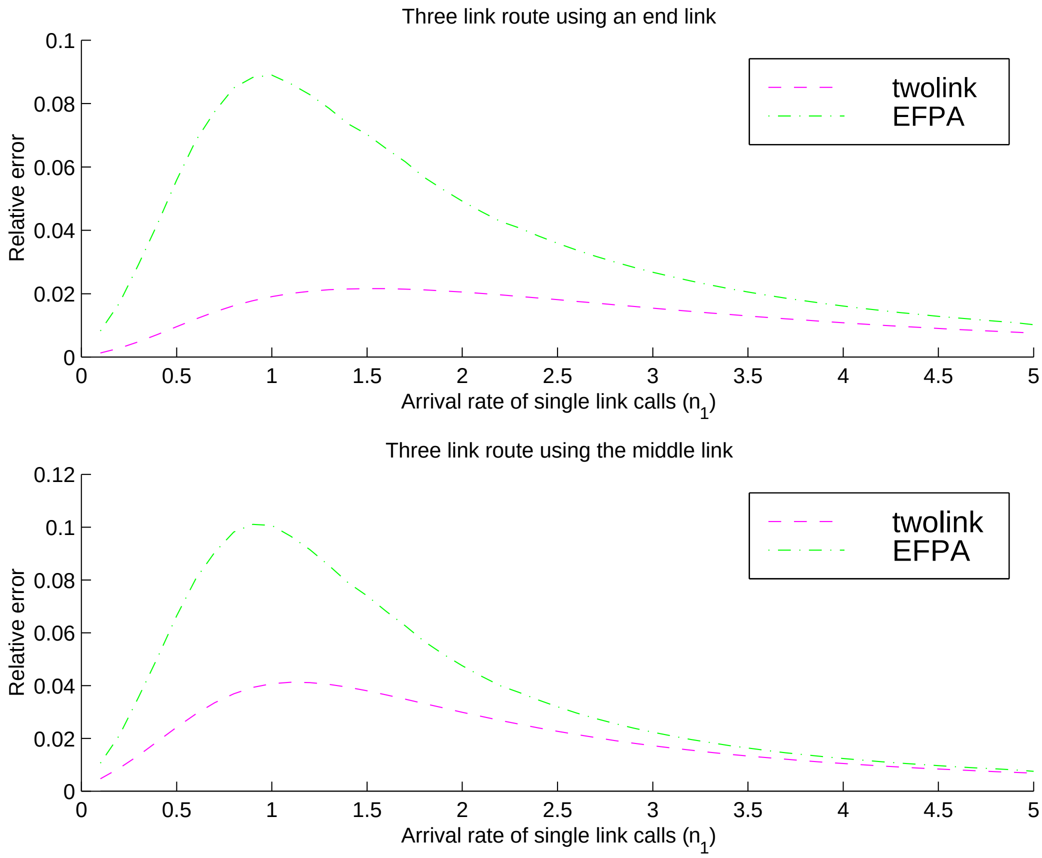

Figure 9: Accuracy for a line network (5 links, $C = 7$, $\nu_2 = \nu_1/2, \nu_3 = \nu_1/3$)

6.5 Computation times

There is a clear trade off between the effort required to calculate an approximation and the accuracy of the approximation. In general, the calculation of the exact loss probabilities for teletraffic networks of realistic size is too computationally expensive. On the other hand, the Erlang fixed point approximation provides a means by which one may estimate losses very quickly. The use of this method on large capacity networks with heavy traffic or with diverse routing is justified by the limit theorems of Kelly (1991) and of Ziedins and Kelly (1989). Considering its simplicity, the EFPA is impressively accurate, even when these limiting regimes are not in force. In terms of computational effort, our two-link approximation is situated somewhere between the EFPA and the exact calculations. In return for this extra effort we would expect the two-link approximation to provide estimates that are closer to the exact values than the EFPA estimates. Interestingly, though, this is not necessarily the case (see Figure 4).

To numerically solve the fixed point equations of the EFPA, MFA, and the two-link approximation, typically one would make an initial guess and then, using the relevant equations, compute successive refinements until a desired rate of change criterion is met. Therefore, the computation times of these fixed point approximations depend on the initial guess, the speed at which the successive estimates converge to a fixed point, and, on the time to compute each refinement. The last component would be the most affected by increases in network size.

The computation times for each of the methods are compared in Figure 10. The test network used was the star network described in Section 6.1. For each level of capacity we have run our programs several times over a range of rates and recorded the execution times. Plotted are the averages of these trials. What we can gauge from this graph are the rates at which the computation times increase with the addition of extra capacity. Clearly, the time taken to compute the exact values increases most rapidly with network size. As expected, the Erlang fixed point approximation has the shallowest slope and is therefore the least taxing of all the methods. The EFPA's variability in observed clock times is probably due to the sharing of processor time with unrelated background processes, but may also be due to slight variations in clock speed over time. The MFA and the two-link approximation have comparable slopes. In this example, the MFA starts with a better absolute computation time but this increases at a faster rate than that of the two-link approximation. This effect may be implementation specific or network specific.

The two-link approximation is the result of an attempt to achieve greater accuracy than the Erlang fixed point approximation. The trade off, as seen clearly in Figure 10, is an increase in computational effort. The two-link approximation is comparable to the Zachary and Ziedins' Markov Field method in the sense that both have computation times which are greater than the Erlang fixed point methods, while begin significantly faster than algorithms for calculating loss probabilities exactly.

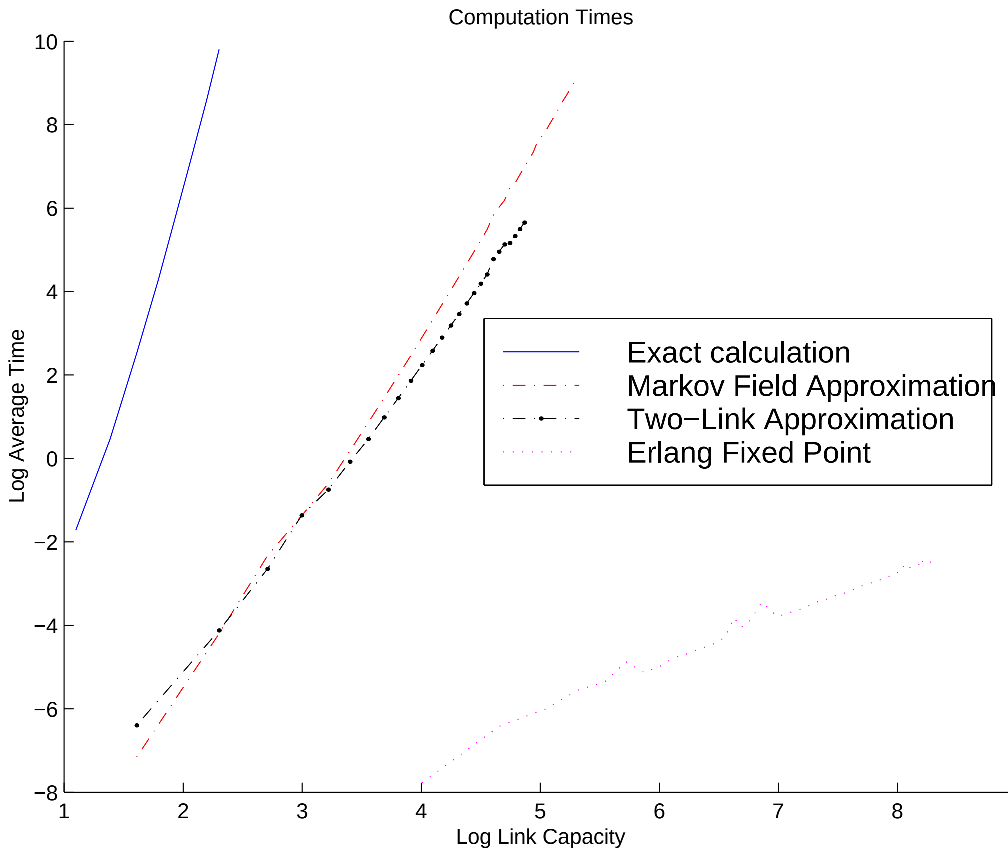

Figure 10: Change in computation time with increasing capacity C.

7 Derivation of the two-link approximation

In this section we derive the fixed-point equations for the two-link reduced load approximation of Section 5. Recall the way that we classified traffic offered to links i and j. We had introduced $U_{i|j} = \sum_{r\in R_i\setminus R_j} Y_r$, $U_{j|i} = \sum_{r\in R_j\setminus R_i} Y_r$ and $U_{ij} = \sum_{r\in R_i\cap R_j} Y_r$. When capacity constraints are present, questions concerning $\boldsymbol{U}_{ij} = (U_{i|j}, U_{ij}, U_{j|i})$ are generally not easily answered. Let us now introduce new, independent processes $\tilde{\boldsymbol{U}}_{ij} = (\tilde{U}_{i|j}, \tilde{U}_{ij}, \tilde{U}_{j|i})$, for each pair of links $i, j \in J$. We shall suppose $\tilde{\boldsymbol{U}}_{ij}$ is a continuous-time Markov chain that approximates the π-behaviour of $\boldsymbol{U}_{ij}$ in the space $S_{ij} = S_{ij}(C_i, C_j) = \{(u_{i|j}, u_{ij}, u_{j|i}) : u_{i|j}+u_{ij} \le C_i, u_{j|i}+u_{ij} \le C_j\}$. Suppose that $\tilde{\boldsymbol{U}}_{ij}$ makes transitions

$$
\begin{array}{lcll}
(u_{i|j}, u_{ij}, u_{j|i}) & \to & (u_{i|j}-1, u_{ij}, u_{j|i}), & \text{at rate } u_{i|j},\\
(u_{i|j}, u_{ij}, u_{j|i}) & \to & (u_{i|j}, u_{ij}-1, u_{j|i}), & \text{at rate } u_{ij},\\
(u_{i|j}, u_{ij}, u_{j|i}) & \to & (u_{i|j}, u_{ij}, u_{j|i}-1), & \text{at rate } u_{j|i},\\
(u_{i|j}, u_{ij}, u_{j|i}) & \to & (u_{i|j}+1, u_{ij}, u_{j|i}), & \text{at rate } \rho_{i|j}(u_{i|j})1_{\{u_{i|j}+u_{ij}\le C_i\}},\\
(u_{i|j}, u_{ij}, u_{j|i}) & \to & (u_{i|j}, u_{ij}+1, u_{j|i}), & \text{at rate } \rho_{ij}(u_{ij})1_{\{u_{i|j}+u_{ij}\le C_i, u_{j|i}+u_{ij}\le C_j\}},\\
(u_{i|j}, u_{ij}, u_{j|i}) & \to & (u_{i|j}, u_{ij}, u_{j|i}+1), & \text{at rate } \rho_{j|i}(u_{j|i})1_{\{u_{j|i}+u_{ij}\le C_j\}},
\end{array}
$$

and no other transitions are possible. Then, the stationary distribution for $\tilde{\boldsymbol{U}}_{ij}$ is

$$
\mathcal{P}\left(\tilde{\boldsymbol{U}}_{ij} = (u_{i|j}, u_{ij}, u_{j|i})\right) = \Phi_{ij}(C_i, C_j)^{-1} \frac{\prod_{m=0}^{u_{i|j}-1} \rho_{i|j}(m)}{u_{i|j}!} \frac{\prod_{m=0}^{u_{ij}-1} \rho_{ij}(m)}{u_{ij}!} \frac{\prod_{m=0}^{u_{j|i}-1} \rho_{j|i}(m)}{u_{j|i}!}. \quad (28)
$$

The partition function $\Phi_{ij}(C_i, C_j)$ is chosen so that $\mathcal{P}$ sums to 1 over the set S_{ij}:

$$
\Phi_{ij}(C_i, C_j) = \sum_{u_i=0}^{C_i} \sum_{u_j=0}^{C_j} \sum_{k=0}^{\min(u_i,u_j)} \frac{\prod_{m=0}^{u_i-k-1} \rho_{i|j}(m)}{(u_i-k)!} \frac{\prod_{m=0}^{k-1} \rho_{ij}(m)}{k!} \frac{\prod_{m=0}^{u_j-k-1} \rho_{j|i}(m)}{(u_j-k)!}.
$$

Our aim is to choose $\rho_{i|j}(\cdot)$, $\rho_{ij}(\cdot)$ and $\rho_{j|i}(\cdot)$ such that the behaviour of $\tilde{\boldsymbol{U}}_{ij}$, with its assumed transition structure, best approximates that of $\boldsymbol{U}_{ij}$. We assign these quantities expected rates.

Let $\tilde{S} = \prod_{i,j\in J} S_{ij}$ and $\boldsymbol{\Lambda}_{i|j}(u) = \{(\boldsymbol{u}, \boldsymbol{v}) \in \tilde{S} \times \tilde{S} : u_{i|j} = u, v_{i|j} = u+1\}$, for $u = 0, 1, \ldots, C_i - 1$. Then $\rho_{i|j}(u)$ defined as $r(\boldsymbol{\Lambda}_{i|j}(u))$:

$$
\rho_{i|j}(u) = \mathbb{E}_{\mathcal{P}}\left(q(\tilde{\boldsymbol{U}}, \boldsymbol{\Lambda}_{i|j}(u, \tilde{\boldsymbol{U}})) \,\middle|\, \tilde{U}_{i|j} = u, \tilde{U}_{i|j} + \tilde{U}_{ij} < C_i\right), \quad (29)
$$

where

$$
q(\boldsymbol{u}, \boldsymbol{\Lambda}_{i|j}(u, \boldsymbol{u})) = \sum_{r\in R_i\setminus R_j} \nu_r \prod_{k\in r\setminus\{i\}} 1_{\{u_{k|i}+u_{ki}<C_k\}} 1_{\{u+u_{ij}<C_i\}}.
$$

Expression (29) can be evaluated partially as follows:

$$
\mathbb{E}\left(q(\tilde{\boldsymbol{U}}, \boldsymbol{\Lambda}_{i|j}(u, \tilde{\boldsymbol{U}})) \,\middle|\, \tilde{U}_{i|j} = u, \tilde{U}_{i|j} + \tilde{U}_{ij} < C_i\right) = \mathbb{E}\left(\alpha_{i|j}(\tilde{U}_{i|j} + \tilde{U}_{ij}, \tilde{U}_{j|i} + \tilde{U}_{ij}) \,\middle|\, \tilde{U}_{i|j} = u, \tilde{U}_{i|j} + \tilde{U}_{ij} < C_i\right),
$$

where $\alpha_{i|j}(u_i, u_j) = \mathbb{E}(q(\tilde{\boldsymbol{U}}, \boldsymbol{\Lambda}_{i|j}(u, \tilde{\boldsymbol{U}}))|E(u_i, u_j))$, and

$$E(u_i, u_j) = \left\{\tilde{U}_{i|k} + \tilde{U}_{ik} = u_i,\ k \in J \setminus \{i\}\right\} \cap \left\{\tilde{U}_{j|k} + \tilde{U}_{jk} = u_j,\ k \in J \setminus \{j\}\right\}$$

is the event that links i and j have utilisations u_i and u_j respectively. The function $\alpha_{i|j}(u_i, u_j)$ is the expected rate of transitions in the set $\{(\boldsymbol{u}, \boldsymbol{v}) \in \tilde{S} \times \tilde{S} : u_{i|j} + u_{ij} = u_i, u_{j|i} + u_{ij} = u_j, v_{i|j} = u_{i|j} + 1\}$. It simplifies to

$$\alpha_{i|j}(u_i, u_j) = \begin{cases} 0\,, & \text{if } u_i = C_i; \\ \displaystyle\sum_{r \in R_i \setminus R_j} \nu_r \mathcal{P}\left(\tilde{U}_{k|i} + \tilde{U}_{ik} < C_k,\ k \in r \setminus \{i\} \middle| E(u_i, u_j)\right), & \text{otherwise.} \end{cases}$$

Extending the rationale of independent blocking, characteristic of the EFPA, we now assume that pairs of links $\{i, j\} \in J$ index *independent* random processes $\tilde{\boldsymbol{U}}_{ij}$. Under this assumption,

$$\begin{aligned} \alpha_{i|j}(u_i, u_j) &= \sum_{r \in R_i \setminus R_j} \nu_r \prod_{k \in r \setminus \{i\}} \mathcal{P}\big(\tilde{U}_{k|i} + \tilde{U}_{ik} < C_k \big| \tilde{U}_{i|k} + \tilde{U}_{ik} = u_i\big) 1_{\{u_i < C_i\}} \\ &= \sum_{r \in R_i \setminus R_j} \nu_r \prod_{k \in r} \big(1 - B_{k|i}(u_i)\big)\,, \end{aligned}$$

where $B_{k|i}(u_i)$ is the likelihood that link k is full when link i is known to have u_i circuits busy. This quantity is estimated to be

$$\begin{aligned} B_{k|i}(u_i) &= \frac{\sum_{u_{ik}=0}^{\min(C_k, u_i)} \mathcal{P}\big(\tilde{U}_{k|i} = C_k - u_{ik}, \tilde{U}_{ik} = u_{ik}, \tilde{U}_{i|k} = u_i - u_{ik}\big)}{\sum_{u_k=0}^{C_k} \sum_{u_{ik}=0}^{\min(u_k, u_i)} \mathcal{P}\big(\tilde{U}_{k|i} = u_k - u_{ik}, \tilde{U}_{ik} = u_{ik}, \tilde{U}_{i|k} = u_i - u_{ik}\big)} \\ &= \begin{cases} \frac{\sum_{u_{ik}=0}^{\min(C_k, u_i)} h_{ki}(C_k - u_{ik}, u_{ik}, u_i - u_{ik})}{\sum_{u_k=0}^{C_k} \sum_{u_{ik}=0}^{\min(u_k, u_i)} h_{ki}(u_k - u_{ik}, u_{ik}, u_i - u_{ik})}\,, & \text{if } k \neq i, \\ 1_{\{u_i = C_i\}}, & \text{if } k = i, \end{cases} \end{aligned}$$

with $h_{ki}(u_{k|i}, u_{ki}, u_{k|i}) \propto \mathcal{P}\big(\tilde{\boldsymbol{U}}_{ki} = (u_{k|i}, u_{ki}, u_{k|i})\big)$ in S_{ki}. Thus, we have an expression for the reduced load marginal rate of arrivals to link i that do not use link j:

$$\rho_{i|j}(u) = \sum_{r \in R_i \setminus R_j} \nu_r \sum_{v=0}^{\min(C_i - u, C_j)} \prod_{k \in r} \big(1 - B_{k|i}(u + v)\big)\, \mathcal{P}\big(\tilde{U}_{ij} = v \big| \tilde{U}_{i|j} = u, \tilde{U}_{i|j} + \tilde{U}_{ij} < C_i\big).$$

Expression (19) results when $\mathcal{P}\big(\tilde{U}_{ij} = u_{ij} \mid \tilde{U}_{i|j} = u, \tilde{U}_{i|j} + \tilde{U}_{ij} < C_i\big)$ is estimated by

$$\frac{\sum_{v=0}^{C_j - u_{ij}} h_{ij}(u, u_{ij}, v)}{\sum_{w=0}^{C_i - u - 1} \sum_{v=0}^{C_j - w} h_{ij}(u, w, v)}.$$

Expression (20) for the reduced load rate $\rho_{ij}(u)$ of arrivals corresponding to transitions in $\boldsymbol{\Lambda}_{ij}(u) = \{(\boldsymbol{u}, \boldsymbol{v}) \in \tilde{S} \times \tilde{S} : u_{ij} = u, v_{ij} = u + 1\}$, $u = 0, 1, \ldots, \min(C_i - 1, C_j - 1)$, is derived in a similar way. The quantity $\alpha_{ij}(u_i, u_j)$ representing the expected rate at which calls that cause an increase in the utilisation of both resource i and j are arriving when $U_i = u_i$ and $U_j = u_j$, is

$$\alpha_{ij}(u_i, u_j) = \mathbb{E}\left(\sum_{r\in R_i\cap R_j} \nu_r \prod_{k\in r\setminus\{i,j\}} 1_{\{\tilde{U}_{k|i}+\tilde{U}_{ki}<C_k\}} \,\middle|\, E(u_i, u_j)\right) 1_{\{u_i<C_i, u_j<C_j\}}\,,$$

which leads to

$$\alpha_{ij}(u_i, u_j) = \begin{cases} 0, & \text{if } u_j = C_j; \\ \sum_{r\in R_i\cap R_j} \nu_r \prod_{k\in r}\big(1 - B_{k|i}(u_i)\big)\,, & \text{otherwise.}\end{cases}$$

Setting $\rho_{ij}(u) = r(\mathbf{\Lambda}_{ij}(u))$, we get

$$\begin{aligned}\rho_{ij}(u) &= \mathbb{E}\left(\alpha_{ij}(\tilde{U}_{i|j} + \tilde{U}_{ij}, \tilde{U}_{j|i} + \tilde{U}_{ij})\,\middle|\, \tilde{U}_{ij} = u, \tilde{U}_{i|j} + \tilde{U}_{ij} < C_i, \tilde{U}_{j|i} + \tilde{U}_{ij} < C_j\right) \\ &= \sum_{r\in R_i\cap R_j} \nu_r \sum_{u_{i|j}=0}^{C_i-u-1} \prod_{k\in r\setminus\{j\}} \big(1 - B_{k|i}(u_{i|j} + u)\big) \\ &\qquad \mathcal{P}\left(\tilde{U}_{i|j} = u_{i|j} \,\middle|\, \tilde{U}_{ij} = u, \tilde{U}_{i|j} + \tilde{U}_{ij} < C_i, \tilde{U}_{j|i} + \tilde{U}_{ij} < C_j\right).\end{aligned}$$

Expression (20) follows on using

$$\frac{\sum_{v=0}^{C_j-u_{ij}-1} h_{ij}(u_{i|j}, u, v)}{\sum_{w=0}^{C_i-u-1} \sum_{v=0}^{C_j-u-1} h_{ij}(w, u, v)}$$

to estimate the latter conditional probability. The loss probabilities may be estimated using Φ_{ij}. Losses on two-link routes, $r = \{i, j\}$, have

$$L_r = 1 - \pi(U_i < C_i, U_j < C_j) \approx 1 - \frac{\Phi_{ij}(C_i - 1, C_j - 1)}{\Phi_{ij}(C_i, C_j)}\,.$$

Calls that use the single link i are lost with probability

$$B_i = 1 - \pi(U_i < C_i) \approx 1 - \frac{\Phi_{ij}(C_i - 1, C_j)}{\Phi_{ij}(C_i, C_j)}\,.$$

The approximation for B_i depends on j because the distribution of $\tilde{U}_{i|j} + \tilde{U}_{ij}$ is different from that of $\tilde{U}_{i|k} + \tilde{U}_{ik}$. As a result, the loss estimated using Φ_{ij} may be different from the estimate using Φ_{ik}.

Acknowledgements

We would like to thank the Associate Editor and the referee for valuable suggestions. We would also like to thank Ilze Ziedins for valuable discussions on this work and Nick Denman for comments on an earlier draft of the paper. The support of the Centre for Mathematics and its Applications at the Australian National University, for an award to attend the National Symposium "Non-linear Time Series, Stochastic Networks and Allied Modern Statistical Techniques" (Australian National University, 28 June–2 July, 2000), was of great assistance with Mark Thompson's work relating to this paper. The support of the Australian Research Council is also gratefully acknowledged.

References

Bebbington, M., Pollett, P. and Ziedins, I. (1997) Improved fixed point methods for loss networks with linear structure. In *Proc. 4th Int. Conf. Telecom.*, Vol. 3, Melbourne, Australia, pp. 1411–1416.

Borst, S. and Mitra, D. (1998) Virtual partitioning for robust resource sharing: computational techniques for heterogeneous traffic. *IEEE J. Selected Areas in Communications* **16**, 668–678.

Boucherie, R. and Mandjes, M. (1998) Estimation of performance measures for product form cellular mobile communications networks. *Telecom. Sys.* **10**, 321–354.

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

Everitt, D. and Macfadyen, N. (1983) Analysis of multi-cellular mobile radiotelephone systems with loss. *British Telecom. Technol. J.*, 37–45.

Girard, A. (1990)*Routing and Dimensioning in Circuit-Switched Networks*, Addison-Wesley.

Hunt, P. and Laws, C. (1997) Optimization via trunk reservation in single resource loss systems under heavy traffic. *Ann. Appl. Probab.* **7**, 1058–1079.

Kelly, F. (1979) *Reversibility and Stochastic Networks*, Wiley series in Probability and Mathematical Statistics, Wiley, Chichester.

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

Kelly, F. (1991) Loss networks. *Ann. Appl. Probab.* **1**, 319–378.

Key, P. (1990) Optimal control and trunk reservation in loss networks. *Probab. Eng. Informat. Sci.* **4**, 203–242.

Lloyd-Evans, R. (1996) *Wide Area Network Performance and Optimization: Practical Strategies for Success*, Data Communications and Networks Series, Addison-Wesley, Harlow.

Louth, G., Mitzenmacher, M. and Kelly, F. (1994) Computational complexity of loss networks. *Theoret. Comput. Sci.* **125**, 45–59.

Mitra, D. and Weinberger, P. (1984) Probabilistic models of database locking: solutions, computational algorithms and asymptotics. *J. Assoc. Comput. Mach.* **31**, 855–878.

Mitra, D. and Ziedins, I. (1996) Virtual partitioning by dynamic priorities: fair and efficient resource-sharing by several services. In (B. Plattner ed.) *1996 International Zurich Seminar on Digital Communications*, Lecture Notes in Computer Science, Broadband Communications, Springer, pp. 173–185.

Ross, K. (1995) *Multiservice Loss Models for Broadband Telecommunication Networks*, Telecommunication Networks and Computer Systems, Springer-Verlag, New York.

Ross, K. and Tsang, D. (1990) Teletraffic engineering for product-form circuit-switched networks. *Adv. Appl. Probab.* **22**, 657–675.

Pollett, P.K. and Thompson, M.R. (2002) A new method for analysing the equilibrium and time-dependent behaviour of Markovian models. *Math. Computer Modelling* (to appear).

Whitt, W. (1985) Blocking when service is required from several facilities simultaneously. *AT&T Tech. J.* **64**, 1807–1856.

Zachary, S. (1991) On blocking in loss networks. *Adv. Appl. Probab.* **23**, 355–372.

Zachary, S. and Ziedins, I. (1999) Loss networks and Markov random fields. *J. Appl. Probab.* **36**, 403–414.

Ziedins, I. and Kelly, F. (1989) Limit theorems for loss networks with diverse routing. *Adv. Appl. Probab.* **21**, 804–830.