A Reduced Load Approximation Accounting for Link Interactions in a Loss Network

M. R. THOMPSON[†]   m. thompson@qic.com.au
Department of Mathematics, The University of Queensland, Brisbane, Australia.

P. K. POLLETT   pkp@maths.uq.edu.au
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.

Keywords: Loss networks, Blocking probabilities, Erlang fixed point

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

---

[†]   Requests for reprints should be sent to Department of Mathematics, The University of Queensland, Brisbane, Australia.

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 ν_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. Without loss of generality, let the mean holding time of calls be 1. Define P to be the probability measure under which $(Y_r; r \in \mathcal{R})$ are independent Poisson random variables with means ν_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 (Corollary 1.10 of [8]), 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.) The solution of (3) is the equilibrium distribution

$$\pi(\boldsymbol{Y} = \boldsymbol{n}) = G(\boldsymbol{C})^{-1} \prod_{r \in R} \frac{\nu_r^{n_r}}{n_r!}, \tag{4}$$

where $G(\boldsymbol{C})$ is a normalising constant chosen so that the distribution π sums to unity. The probability a call requesting route r arrives to find one or more of the links in r full is $L_r = 1 - G(\boldsymbol{C} - A\boldsymbol{e}_r)/G(\boldsymbol{C})$.

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 [13]. Thus, there is strong evidence to suggest that an algorithm for finding the loss probabilities in polynomial time using G does not exist.

We have described the classical loss network model similar to that of Kelly [9]. 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 [17], [6], [10], cellular mobile networks [5], [3], computer database access problems [14], and other kinds of telecommunications networks [19], [16]. 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 [11], [7] or virtual partitioning [2], [15]. Nor does the product-form result hold for networks allowing alternative routing.

2. The Erlang Fixed Point Approximation

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

$$L_r = 1 - \prod_{i \in r}(1 - B_i), \tag{5}$$

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

$$B_j = E(\rho_j, C_j), \qquad j \in J, \tag{6}$$

$$\rho_j = \sum_{r \in R_j} \nu_r \prod_{i \in r \setminus \{j\}} (1 - B_i), \qquad j \in J, \tag{7}$$

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 $\mathcal{P}$. For each link j, let $U_j = \sum_{r \in R_j} Y_r$ be the capacity used on link j. 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 $\boldsymbol{U} = (U_j\,;\,j \in J)$ 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 (6). Kelly [9] 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 is known to be effective under a variety of limiting regimes. Kelly [10] 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) [22] 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).

3.   A Two-Link Approximation

An estimate of the route loss probabilities, which is more accurate than those in (5), 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

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

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

$$\rho_{i|j}(u) =$$
$$\sum_{r \in R_i \backslash R_j} \nu_r \sum_{l=0}^{\min(C_i - u, C_j)} \prod_{k \in r} \left(1 - B_{k|i}(u+l) \right) \frac{\sum_{v=0}^{C_j - l} h_{ij}(u, l, v)}{\sum_{w=0}^{C_i - u - 1} \sum_{v=0}^{C_j - w} h_{ij}(u, w, v)}, \quad (8)$$

$$\rho_{ij}(u) =$$

$$\sum_{r \in R_i \cap R_j} \nu_r \sum_{l=0}^{C_i-u-1} \prod_{k \in r} \left(1 - B_{k|i}(l+u)\right) \frac{\sum_{v=0}^{C_j-u-1} h_{ij}(l,u,v)}{\sum_{w=0}^{C_i-u-1} \sum_{v=0}^{C_j-u-1} h_{ij}(w,u,v)}, \quad (9)$$

and

$$B_{k|i}(u_i) = \begin{cases} \dfrac{\sum_{l=0}^{\min(C_k,u_i)} h_{ki}(C_k-l,l,u_i-l)}{\sum_{m=0}^{C_k} \sum_{l=0}^{\min(m,u_i)} h_{ki}(m-l,l,u_i-l)}, & \text{if } k \neq i, \\ 1_{\{u_i=C_i\}}, & \text{if } k = i. \end{cases} \quad (10)$$

These equations will be derived in Section 5. 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.

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

$$L_r = 1 - \frac{\Phi_{ij}(C_i-1, C_j-1)}{\Phi_{ij}(C_i, C_j)}, \qquad \text{if } r = \{i, j\}, \quad (11)$$

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

$$B_i = 1 - \frac{\Phi_{ij}(C_i-1, C_j)}{\Phi_{ij}(C_i, C_j)}, \quad (12)$$

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 but not j, the number using link j and not i, 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

$$P(U_{i|j}=u_{i|j}, U_{ij}=u_{ij}, U_{j|i}=u_{j|i}) = \frac{\rho_{i|j}^{u_{i|j}}}{u_{i|j}!}\frac{\rho_{ij}^{u_{ij}}}{u_{ij}!}\frac{\rho_{j|i}^{u_{j|i}}}{u_{j|i}!}e^{-(\rho_{i|j}+\rho_{ij}+\rho_{j|i})}, \quad (13)$$

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 (13) with reduced load rates, and we isolate the subsystem composed of traffic offered to links i and j. Motivated by the form of (13), 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 (12), the probability that either link i or link j are full would be L_r in expression (11), 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 (10). 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 (8) for $\rho_{i|j}(u_{i|j})$ is just $\rho_{i|j} = \sum_{r\in R_i\setminus R_j}\nu_r$ 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.

4.   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 directly from $G(\boldsymbol{C})$.

4.1.   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 [12].

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.

4.1.1.   The two-link approximation

The two-link reduced load approximation is obtained by solving the system comprising (14) and (15) 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

$$\rho(u) = \nu_1 + (J - 2)\nu_2$$
$$\times \sum_{w=0}^{C-u-1} \left(1 - B(w+u)\right) \frac{\sum_{v=0}^{C-w} \frac{\prod_{m=0}^{u-1} \rho(m)}{u!} \frac{\nu_2^w}{w!} \frac{\prod_{m=0}^{v-1} \rho(m)}{v!}}{\sum_{k=0}^{C-u-1} \sum_{v=0}^{C-k} \frac{\prod_{m=0}^{u-1} \rho(m)}{u!} \frac{\nu_2^k}{k!} \frac{\prod_{m=0}^{v-1} \rho(m)}{v!}}, \quad (14)$$

and

$$B(u) = \frac{\sum_{w=0}^{\min(C,u)} \frac{\prod_{m=0}^{C-w-1} \rho(m)}{(C-w)!} \frac{\nu_2^w}{w!} \frac{\prod_{m=0}^{u-w-1} \rho(m)}{(u-w)!}}{\sum_{v=0}^{C} \sum_{w=0}^{\min(v,u)} \frac{\prod_{m=0}^{v-w-1} \rho(m)}{(C-w)!} \frac{\nu_2^w}{w!} \frac{\prod_{m=0}^{u-w-1} \rho(m)}{(u-w)!}} \quad (15)$$

for $u = 0, \ldots, C - 1$. Under this scheme, the loss probabilities are estimated to be

$$L_1 = 1 - \frac{\Phi(C-1, C)}{\Phi(C, C)} \qquad \text{and} \qquad L_2 = 1 - \frac{\Phi(C-1, C-1)}{\Phi(C, C)}, \quad (16)$$

with

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

4.1.2.   Zachary and Ziedins' method

In Section 4 of their paper, Zachary and Ziedins [21] 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

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

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

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

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.

4.2. 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
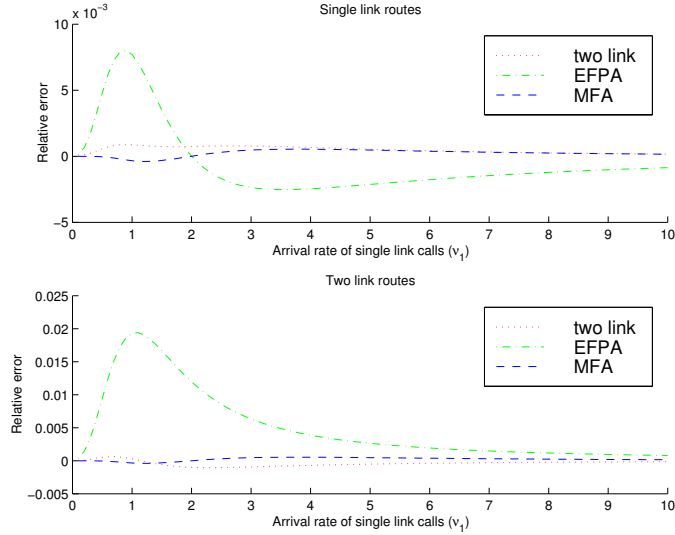
Figure 1. Accuracy for a star network ($J = 5$, $C = 5$, $\nu_2 = \nu_1/2$)

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.

4.2.1.   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

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

instead of (14) (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 (15) for $B(u)$ and expressions (16) for the loss probabilities remain unaltered.

4.2.2.   The method of Bebbington, Pollett and Ziedins

A similar approximation for the ring network was devised by Bebbington, Pollett and Ziedins [1] (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 $\{Y_i + Y_{i,i+1} + Y_{i-1,i} = C\}$.
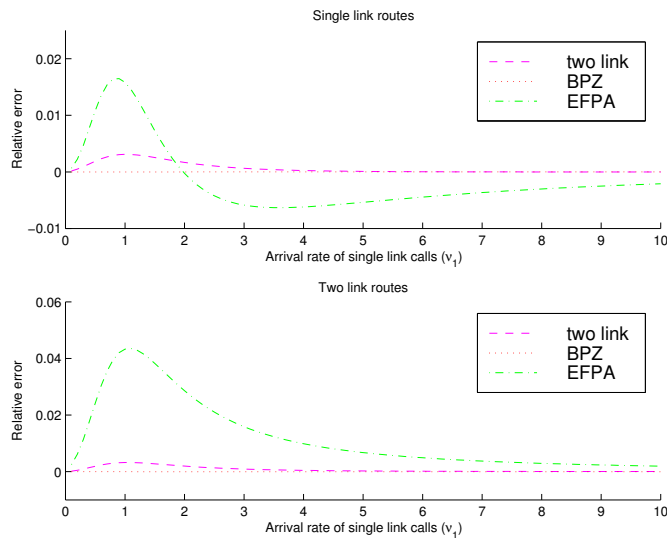

Figure 2. Accuracy for a ring network ($J = 5$, $C = 5$, $\nu_2 = \nu_1/2$)

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.

4.3. A linear network with three-link routes

As a final example, we will analyse a linear network in which there are 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 [21] recursion to work.

4.3.1. 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

$$L_i = 1 - \frac{\Phi_{i,i+1}(C-1,C)}{\Phi_{i,i+1}(C,C)}, \text{ for } i = 1, \ldots, l-1, \text{ or}$$

$$L_i = 1 - \frac{\Phi_{i,i-1}(C-1,C)}{\Phi_{i,i-1}(C,C)}, \text{ for } i = 2, \ldots, l,$$

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 here 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, 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 $\rho_{1|2}(u) = \nu_1$,

$$\rho_{2|1}(u) = \nu_1 + (\nu_2 + \nu_3(1 - B_4)) \sum_{k=0}^{C-u-1} (1 - B_{3|2}(u + k))H_{2,1}^{(1)}(k, u),$$

$$\rho_{2|3}(u) = \nu_1 + \nu_2 \sum_{k=0}^{C-u-1} (1 - B_{1|2}(u + k))H_{2,3}^{(1)}(k, u),$$

$$\rho_{i|i-1}(u) = \nu_1 + (\nu_2 + \nu_3(1 - B_{i+2}) \sum_{k=0}^{C-u-1} (1 - B_{i+1|i}(u + k))H_{i,i-1}^{(1)}(k, u),$$

$$\rho_{i|i+1}(u) = \nu_1 + (\nu_2 + \nu_3(1 - B_{i-2})) \sum_{k=0}^{C-u-1} (1 - B_{i-1|i}(u + k))H_{i,i+1}^{(1)}(k, u),$$

for $i = 3, \ldots, l - 3$, $\rho_{l|l-1}(u) = \nu_1$,

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

$$\rho_{l-1|l}(u) = \nu_1 + (\nu_2 + \nu_3(1 - B_{l-3})) \sum_{k=0}^{C-u-1} (1 - B_{l-2|l-1}(u + k))H_{l-1,l}^{(1)}(k, u),$$

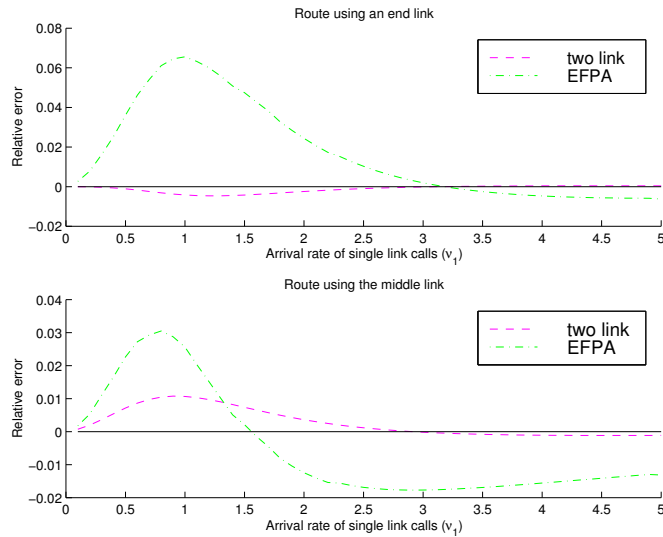where $H_{i,j}^{(1)}(k, u) = \sum_{w=0}^{C-k} h_{i,j}(u, k, w) / \sum_{v=0}^{C-u-1} \sum_{w=0}^{C-v} h_{i,j}(u, v, w)$. And, the joint reduced load rates are

$$\rho_{1,2}(u) = \nu_2 + \nu_3 \sum_{k=0}^{C-u-1} (1 - B_{3|2}(k + u))H_{2,1}^{(2)}(k, u),$$

$$\rho_{i,i+1}(u) = \nu_2 + \nu_3 \sum_{k=0}^{C-u-1} (1 - B_{i-1|i}(k + u))H_{i,i+1}^{(2)}(k, u)$$
$$+ \nu_3 \sum_{k=0}^{C-u-1} (1 - B_{i+2|i+1}(k + u))H_{i+1,i}^{(2)}(k, u),$$

for $i = 2, \ldots, l - 2$,

$$\rho_{i,i-1}(u) = \nu_2 + \nu_3 \sum_{k=0}^{C-u-1} (1 - B_{i+1|i}(k+u)) H_{i,i-1}^{(2)}(k, u)$$
$$+ \nu_3 \sum_{k=0}^{C-u-1} (1 - B_{i-2|i-1}(k+u)) H_{i-1,i}^{(2)}(k, u),$$

for $i = 3, \ldots, l - 1$,

$$\rho_{l,l-1}(u) = \nu_2 + \nu_3 \sum_{k=0}^{C-u-1} (1 - B_{l-2|l-1}(k+u)) H_{l-1,l}^{(2)}(k, u),$$

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

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

We compare the relative errors in the proposed two-link approximation with those of the Erlang fixed point approximation in Figures 3, 4, 5, and 6. 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
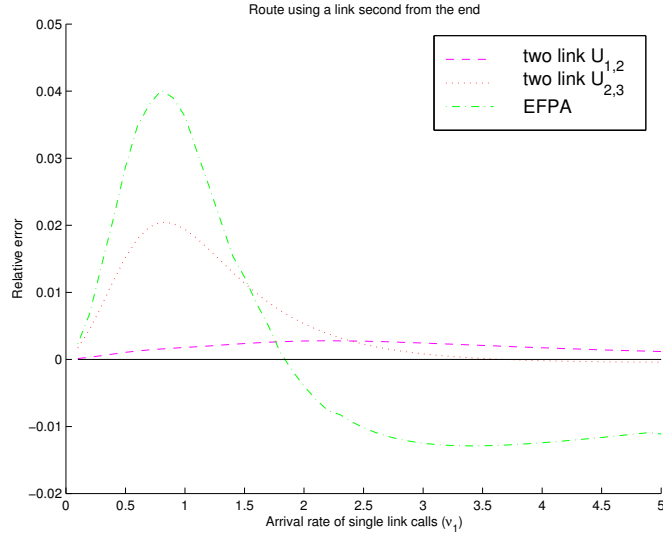
Figure 4. Accuracy for a line network (5 links, $C = 7$, $\nu_2 = \nu_1/2, \nu_3 = \nu_1/3$)

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 4). Significant improvements over the EFPA are also observed in Figure 5 for the two-link routes. On the three-link routes, our proposed approximation again improves on the EFPA (see Figure 6).

5.   Derivation of the Two-Link Approximation

In this section we derive the fixed-point equations for the two-link reduced load approximation of Section 3. 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 $S_{ij} = S_{ij}(C_i, C_j) = \{(u_{i|j}, u_{ij}, u_{j|i}) : u_{i|j} + u_{ij} \leq C_i, u_{j|i} + u_{ij} \leq$
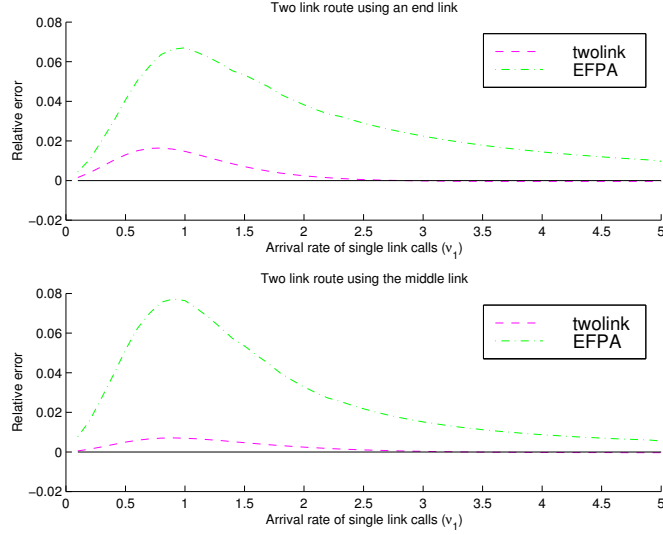
Figure 5. Accuracy for a line network (5 links, $C = 7$, $\nu_2 = \nu_1/2, \nu_3 = \nu_1/3$)

$C_j\}$. Suppose that $\tilde{\boldsymbol{U}}_{ij}$ makes transitions from $(u_{i|j}, u_{ij}, u_{j|i})$ to

$$
\begin{aligned}
(u_{i|j} - 1, u_{ij}, u_{j|i}), &\qquad \text{at rate } u_{i|j}, \\
(u_{i|j}, u_{ij} - 1, u_{j|i}), &\qquad \text{at rate } u_{ij}, \\
(u_{i|j}, u_{ij}, u_{j|i} - 1), &\qquad \text{at rate } u_{j|i}, \\
(u_{i|j} + 1, u_{ij}, u_{j|i}), &\qquad \text{at rate } \rho_{i|j}(u_{i|j}) 1_{\{u_{i|j} + u_{ij} \leq C_i\}}, \\
(u_{i|j}, u_{ij} + 1, u_{j|i}), &\qquad \text{at rate } \rho_{ij}(u_{ij}) 1_{\{u_{i|j} + u_{ij} \leq C_i, u_{j|i} + u_{ij} \leq C_j\}}, \\
(u_{i|j}, u_{ij}, u_{j|i} + 1), &\qquad \text{at rate } \rho_{j|i}(u_{j|i}) 1_{\{u_{j|i} + u_{ij} \leq C_j\}},
\end{aligned}
$$

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

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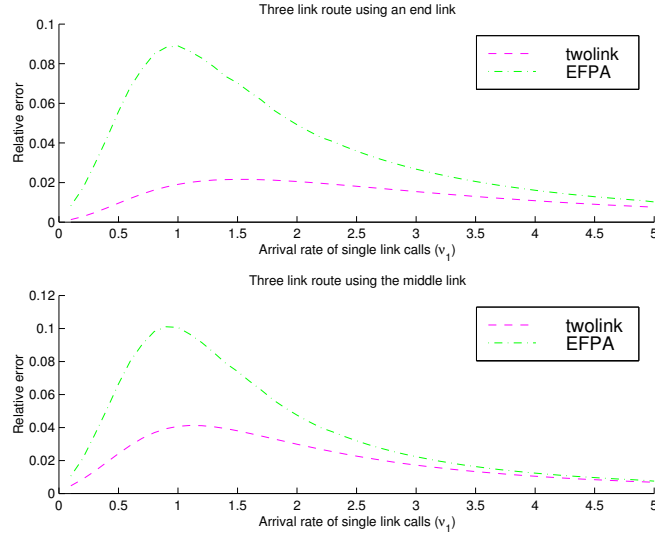$$

Figure 6. Accuracy for a line network (5 links, $C = 7$, $\nu_2 = \nu_1/2, \nu_3 = \nu_1/3$)

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 $r(\boldsymbol{\Lambda}_{i|j}(u))$:

$$\rho_{i|j}(u) = \mathbb{E}_{\mathcal{P}} \left(q(\tilde{\boldsymbol{U}}, \boldsymbol{\Lambda}_{i|j}(u, \tilde{\boldsymbol{U}})) \,\Big|\, \tilde{U}_{i|j} = u, \tilde{U}_{i|j} + \tilde{U}_{ij} < C_i \right), \qquad (17)$$

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 (17) can be evaluated partially as follows:

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

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 $\alpha_{i|j}(u_i, u_j) = 0$ if $u_i = C_i$ and

$$\alpha_{i|j}(u_i, u_j) = \sum_{r \in R_i \setminus R_j} \nu_r \mathcal{P}\left(\tilde{U}_{k|i} + \tilde{U}_{ik} < C_k, \ k \in r \setminus \{i\}\Big|E(u_i, u_j)\right),$$

otherwise. 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,

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

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

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

with $h_{ki}(u_{k|i}, u_{ki}, u_{k|i}) \propto \mathcal{P}(\tilde{\boldsymbol{U}}_{ki} = (u_{k|i}, u_{ki}, u_{k|i}))$ 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:

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

(8) results when $\mathcal{P}\left(\tilde{U}_{ij} = u_{ij} \mid \tilde{U}_{i|j} = u, \tilde{U}_{i|j} + \tilde{U}_{ij} < C_i\right)$ 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 (9) for the reduced load rate $\rho_{ij}(u)$ of arrivals corresponding to transitions in $\mathbf{\Lambda}_{ij}(u) = \{(\boldsymbol{u}, \boldsymbol{v}) \in \tilde{S} \times \tilde{S} : u_{ij} = u, v_{ij} = u+1\}$, $u = 0, 1, \ldots, \min(C_i - 1, C_j - 1)$, 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

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

which leads to

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

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

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

Expression (9) 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}.

Acknowledgments

References

1. M. Bebbington, P. Pollett and I. Ziedins. Improved fixed point methods for loss networks with linear structure. In W. Lavery, editor, *Proceedings of the 4th International Conference on Telecommunications*, Vol. 3, Melbourne, Australia, pages 1411–1416. 1997.
2. S. Borst and D. Mitra. Virtual partitioning for robust resource sharing: computational techniques for heterogeneous traffic. *IEEE Journal on Selected Areas in Communications*, 16:668–678, 1998.
3. R. Boucherie and M. Mandjes. Estimation of performance measures for product form cellular mobile communications networks. *Telecommunication Systems* 10:321–354, 1998.
4. Z. Dziong and J. Roberts. Congestion probabilities in a circuit-switched integrated services network. *Performance Evaluation*, 7:267–284, 1987.
5. D. Everitt and N. Macfadyen. Analysis of multi-cellular mobile radiotelephone systems with loss. *British Telecom Technology Journal*, 1:37–45, 1983.

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

7. P. Hunt and C. Laws. Optimization via trunk reservation in single resource loss systems under heavy traffic. *The Annals of Applied Probability*, 7:1058–1079, 1997.

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

9. F. Kelly. Blocking probabilities in large circuit-switched networks. *Advances in Applied Probability*, 18:473–505, 1986.

10. F. Kelly. Loss networks. *The Annals of Applied Probability*, 1:319–378, 1991.

11. P. Key. Optimal control and trunk reservation in loss networks. *Probability in the Engineering and Informational Sciences*, 4:203–242, 1990.

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

13. G. Louth, M. Mitzenmacher and F. Kelly. Computational complexity of loss networks. *Theoretical Computer Science*, 125:45–59, 1994.

14. D. Mitra and P. Weinberger. Probabilistic models of database locking: solutions, computational algorithms and asymptotics. *Journal of the Association for Computer Machinery* 31:855–878, 1984.

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

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

17. K. Ross and D. Tsang. Teletraffic engineering for product-form circuit-switched networks. *Advances in Applied Probability*, 22:657–675, 1990.

18. P. Pollett and M. Thompson. A new method for analysing the equilibrium and time-dependent behaviour of Markovian models. *Mathematical and Computer Modelling* (to appear).

19. W. Whitt. Blocking when service is required from several facilities simultaneously. *AT&T Technical Journal* 64:1807–1856, 1985.

20. S. Zachary. On blocking in loss networks. *Advances in Applied Probability*, 23:355–372, 1991.

21. S. Zachary and I. Ziedins. Loss networks and Markov random fields. *Journal of Applied Probability*, 36:403–414, 1999.

22. I. Ziedins and F. Kelly. Limit theorems for loss networks with diverse routing. *Advances in Applied Probability*, 21:804–830, 1989.