Optimization and Design of Network Routing using Refined Asymptotic Approximations

Debasis Mitra, John A. Morrison*, and K. G. Ramakrishnan

Bell Laboratories, Lucent Technologies

Abstract

The problems of route optimization, and the sizing of virtual paths and explicit routes in wide-area multi-service broadband networks are considered. The problems are formulated at the call-level in the framework of multi-rate, circuit-switched, loss networks, with effective bandwidth encapsulating cell and packet-level behavior. Broadband networks are characterized by links with very large capacities in circuits, and are expected to support many services each having a characteristic bandwidth or rate. Various asymptotic results based on Uniform Asymptotic Approximations (UAA) have previously been obtained to reduce the complexity of the numerical calculations. This paper offers refinements (RUAA) to UAA to the loss probabilities for a single link, as well as their sensitivities to the offered traffic.

Network loss probabilities are obtained by solving fixed-point equations. Another system of equations determines the implied costs for services and links, which are used to guide the network optimization. Refined asymptotic approximations to the network's loss probabilities as well as to the implied costs are proposed based on the RUAA. The refinements are crucial for the accurate evaluations of the implied costs. The complexity of these calculations remains bounded as the link capacities and traffic intensities become increasingly large, and the complexity for the implied costs is independent of the number of services. A network design tool TALISMAN has been extended to implement the refined approximations. Numerical examples illustrate the accuracy of the RUAA.

1. Introduction

This paper considers the design and optimization, at the call level, of wide-area, broadband networks. Here the term call is meant to be synonymous with connections and flows. By viewing calls as circuits, which, for the duration of the call, requires constant bandwidth corresponding to the effective bandwidth that is characteristic of the service type of the call, these networks are treated as multirate, circuit-switched loss networks. Thus the effective bandwidth encapsulates traffic behavior and quality of service (QoS) issues at the cell and packet levels. This separation of levels, and the accompanying hiding of the cell and packet

*Consultant

level details, is essential for the tractability of call level design problems. The notion of effective bandwidth has been extensively researched, and reviews and original contributions are widely available, for instance in [9] and [28], and references therein. Broadband networks are characterized by links with very large capacities in circuits, and are expected to support many services with significant diversity in their effective bandwidths.

In an earlier paper [22], asymptotic results were used to reduce the complexity of the numerical calculations involved in multirate network design and optimization. More generally, a unified hybrid approach was given which allows asymptotic and nonasymptotic methods of calculation to be used cooperatively. In this earlier work a central element in the asymptotic approach is the uniform asymptotic approximation (UAA) developed by Mitra and Morrison [20] for the analyses of single links. The asymptotic regime for the UAA is large capacity C and offered traffic which is also large, $O(C)$. Prior work on the asymptotic analysis of circuit switched networks gave separate treatments for the underloaded, critically loaded and overloaded regimes (see for instance [6, 10, 12, 25] and other references in [20]), which the UAA renders unnecessary. In another context, Birman and Kogan [1] derived a uniform asymptotic approximation to the partition function for a single-chain closed product form queueing network. Subsequently, Kogan and Shenfild [18] did likewise for the partition function of the generalized multiclass Engset model. The complexity of calculating loss probabilities on a link carrying multirate traffic by the UAA is $O(1)$, i.e., bounded as $C \to \infty$. This contrasts with complexity $O(C)$ by the well-known, exact calculations due to Kaufman [13] and Roberts [26].

For calculating network loss probabilities we rely on the fixed point approach, which is based on the approximation of link independence. Excellent sources of information on this technique are Kelly [17] and Ross [27]. The nature of the errors introduced by the approximation has drawn considerable attention, both theoretical (see, for instance, [11], [23], and [27]) and empirical. The notion of implied costs is used crucially for calculating the gradient of the function to be maximized, called network revenue, during the optimization.

The equations for the implied costs were obtained [22] by extending the approach due to Kelly [7, 15, 16] to the multirate case.

Our computational experience has been that in computing link loss probabilities, the UAA is quite robust and accurate for typical realistic problems. However, numerical examples show that the UAA to the sensitivities of link loss probabilities is not very accurate in the important critically loaded regime. Moreover, as is pointed out in [22], this approximation is not useful in the overloaded regime. Consequently, refined uniform asymptotic approximations (RUAA) were derived for the loss probabilities, and their sensitivities to the offered traffic rates, for a single link [24]. Numerical results show that the RUAA to the loss probabilities is very accurate, and to their sensitivities is very good. Although the RUAA are more complicated than the UAA, the complexity of the calculations for a link remains $O(1)$ as $C \to \infty$.

In this paper we show how the UAA and RUAA to the loss probabilities, and their sensitivities, may be derived from those for the tail of the stationary distribution π of the utilized bandwidth in a link. The approximations to π for link capacity C and occupancy $C - j$, $j = O(1)$, are characterized by

$$\begin{aligned} \pi_{C-j}(C) &\sim \mathbb{B}(z^*)^j \qquad : \text{UAA} \\ &\sim \left[\mathbb{B}^{(1)} + \mathbb{B}^{(2)}(j-1) + \mathbb{B}^{(3)}(j-1)(j-2)\right](z^*)^j \qquad : \text{RUAA} \end{aligned}$$

The partial derivatives of the blocking probabilities with respect to the offered traffic rates of the different services are similarly approximated in the asymptotic regime by the same set of parameters, i.e., $(z^*; \mathbb{B})$ for UAA and $(z^*; \mathbb{B}^{(1)}, \mathbb{B}^{(2)}, \mathbb{B}^{(3)})$ for the RUAA.

In this paper we apply the RUAA for links to a network design problem in which the objective is to maximize network revenue, a weighted aggregate measure of carried calls over the network. The quantities which are subject to design are the proportion of offered traffic of each stream that is admitted to the network and, of the admitted portion, the proportions that are offered to the various routes from a set of predetermined admissible routes for each service and origin, destination pair. The design therefore is for joint admission control and routing. A hill-climbing solution is proposed [22], which relies on analytic computations of the gradient of the network revenue with respect to all the design variables in the problem. The gradient is obtained from the implied costs, $c_{s\ell}$ for service s on link ℓ. The analytic support for the algorithm is, first, in calculating the end-to-end loss probabilities for each service and route, and, second, in the calculations of the implied costs $\{c_{s\ell}\}$.

Use of the RUAA for the link analyses in the fixed-point equations for the network is straightforward, and this leads to improved accuracy of the solutions, exactly as expected from the prior results on solutions for links. The framework for the use of the RUAA in the calculation of the networks' implied costs is rather more complex than for the UAA, which is given in [22]. In this paper, first, a particular product-form structure to the partial derivatives of the link loss probabilities with respect to offered traffic rates is established, and subsequently this structure is extended to the networks' implied costs. We show that the latter are obtained from a system of $3L$ linear equations, while the number of equations are, respectively, L and SL for the UAA-based and non-asymptotic approaches [22], where S is the number of services. Thus, while the complexity of calculating the implied costs is somewhat more for the RUAA over the UAA, the important feature of independence from the number of services is maintained.

We have had considerable experience with the techniques described in the paper for solving the combined admission control and route optimization problems described above. The first observation is that for networks of current interest, where link capacities are 10^4 and higher, the method based on the RUAA provides solutions in time periods which are typically orders of magnitude shorter than for non-asymptotic techniques. In many cases where large capacity links are involved, the latter approach is beset with machine underflow and overflow problems, and no final design solutions are available. In such cases we have to rely on the stochastic and flow bounds described in [22] to establish plausibility of the solution obtained by the RUAA. A second observation concerns the relative performances of techniques based on the UAA and the RUAA. In our experience in some cases the hill-climbing procedure based on the UAA has not converged, while in every case convergence has been obtained with the RUAA. The explanation for such behavior lies in the difference in accuracy of the computed gradients of the network revenue. In such cases, the saving in computational and programming time stemming from the RUAA is large.

The routes in the admissible route sets considered in the formulation of the network design problem of this paper may be viewed as virtual paths, a standard concept in ATM networking, so that the problem considered here may alternately be viewed as one of optimally sizing the virtual path network [7], [29]. In the alternative Internet Protocol (IP) framework, the Internet Engineering Task Force through its Multi-Protocol Label Switching (MPLS) working group is considering the related concept of explicit routes [3]. The problem in this paper may also be viewed as one of optimally sizing the explicit routes. Also, link speeds of 155 Mbps and 620 Mbps are quite widely deployed in networks today. If there is a service for which the effective bandwidth of a call is 16 Kbps, then the link capacity C is about 9,687 and 38,750, respectively. Such large capacities provide some of the motivation for asymptotic techniques.

A tool for route optimization and logical network design called TALISMAN has been designed which incorporates our theoretical results, including all the refinements described in this paper. TALISMAN implements the hybrid approach described in this paper and it also incorporates rules for selecting the technique, either asymptotic or nonasymptotic, for each link.

In Section 2 we describe the model of the network, and then discuss the network design problem. In Section 3 we consider the analysis of the network. We first give the fixed point equations for determining the link loss probabilities, and then the equations for the implied costs. In Section 4 we consider a single link network. We first state the results for the exact analysis, and then summarize the UAA to the loss probabilities, and proceed to the RUAA to the loss probabilities and their partial derivatives with respect to the loads. In Section 5 we consider the network analysis and describe the fixed point equations, and the equations for the implied costs, obtained by RUAA. We also describe the hybrid systems of equations for the loss probabilities and the implied costs, which combine asymptotic and nonasymptotic analyses. Finally, in Section 6 we present some numerical examples which illustrate the accuracy of the RUAA.

2. Formulation

We first describe the model of the network, and then discuss a canonical network design problem.

2.1. Model

We consider a network with N nodes and L links which connect pairs of nodes, where link ℓ has capacity C_ℓ. There are S types of service, where service s requires bandwidth $d_{s\ell}$ on link ℓ. It is assumed that C_ℓ, $\ell = 1, 2, \ldots, L$ and $d_{s\ell}$, $s = 1, 2, \ldots, S$, $\ell = 1, 2, \ldots, L$ are positive integers. Let σ denote an origin-destination pair of nodes, and (s, σ) denote a stream of calls of service type s from the origin node to the destination node. The arrivals for stream (s, σ) are Poisson, with mean rate $\Lambda_{s\sigma}$. The set of admissible routes for stream (s, σ) is denoted by $\mathcal{R}(s, \sigma)$. Throughout this paper, the route sets $\{\mathcal{R}(s, \sigma)\}$ are assumed to be given and fixed. The outcome of an independent Bernoulli trial determines if an arriving call of stream (s, σ) is offered to the network, which is a form of admission control, and, if it is, to which route $r \in \mathcal{R}(s, \sigma)$ it is offered. The probabilities of the Bernoulli trials are determined by the design process. If an arriving call in stream (s, σ) is offered to a route $r \in \mathcal{R}(s, \sigma)$ it may be blocked and lost if there is insufficient bandwidth available on any link of the route r. If the call is accepted, the bandwidth on each link is simultaneously held for the duration of the call. The holding period of a call in stream (s, σ) is generally distributed with mean $1/\mu_{s\sigma}$, and is independent of earlier arrival and holding times.

The call arrivals of service type s on route $r \in \mathcal{R}(s, \sigma)$ are Poisson, with mean rate λ_{sr}, where

$$\sum_{r\in\mathcal{R}(s,\sigma)} \lambda_{sr} \leq \Lambda_{s\sigma} \ . \tag{2.1}$$

The corresponding traffic intensity is $\rho_{sr} = \lambda_{sr}/\mu_{s\sigma}$. It follows that

$$\sum_{r\in\mathcal{R}(s,\sigma)} \rho_{sr} \leq \bar{\rho}_{s\sigma} \ , \tag{2.2}$$

where $\bar{\rho}_{s\sigma} = \Lambda_{s\sigma}/\mu_{s\sigma}$. The probability

$$p_{s\sigma} = \frac{1}{\Lambda_{s\sigma}} \sum_{r\in\mathcal{R}(s,\sigma)} \lambda_{sr} \tag{2.3}$$

is a parameter of the admission control, i.e., exogenous calls of stream (s,σ) are offered to the network with Bernoulli probability $p_{s\sigma}$ and dropped with probability $(1-p_{s\sigma})$. Similarly, the probability that a call of service type s is offered to a route $r \in \mathcal{R}(s,\sigma)$ is $\lambda_{sr}/\Lambda_{s\sigma}$.

To recapitulate, the important elements of the model are fixed set of paths, admission control and state-independent routing. Also, the exogenous offered load of stream (s,σ) is $\bar{\rho}_{s\sigma}$ and, after admission control, the offered load of service type s on route r is ρ_{sr}, $r \in \mathcal{R}(s,\sigma)$.

2.2. Network design: offered traffic on routes which maximize revenue

Let e_{sr} be the revenue earned per carried call per unit time by calls of service type s on route r, and let L_{sr} be the equilibrium loss probability of service s on route $r \in \mathcal{R}(s,\sigma)$. Since $\rho_{sr}(1-L_{sr})$ is the carried load of service type s on route r, the long-run average revenue for the network is

$$W = \sum_{\sigma,s} \sum_{r\in\mathcal{R}(s,\sigma)} e_{sr}\rho_{sr}(1-L_{sr}) \ . \tag{2.4}$$

We let $\{\rho_{sr}\}$ be the design variables. The goal is to maximize the revenue subject to the constraint (2.2), i.e.,

$$\max_{\{\rho_{sr}\}:\ \sum_{r\in\mathcal{R}(s,\sigma)} \rho_{sr} \leq \bar{\rho}_{s\sigma},\ \ \forall s,\ \forall\sigma;\ \rho_{sr}\geq 0} W \ . \tag{2.5}$$

Note that the probabilities for admission control ($p_{s\sigma}$) and routing are recovered from the solution $\{\rho_{sr}\}$; in particular, for any (s, σ), if the constraint in (2.5) is binding then $p_{s\sigma} = 1$ and otherwise $p_{s\sigma} < 1$.

A simpler version of the network design problem is obtained by eliminating admission control, in which case the inequality in the constraint in (2.2) and (2.5) is replaced by equality. Except in heavy traffic, the solutions for the two problems are typically identical.

Observe that (2.4)–(2.5) is an optimization problem with linear constraints (2.5) and a nonlinear objective function (2.4). To evaluate W and $\partial W/\partial \rho_{sr}$ for given offered traffic rates $\{\rho_{sr}\}$, the techniques described in Sections 3–5 are used. Initially a steepest ascent direction is computed by projecting $\partial W/\partial \rho_{sr}$ on to the null space of (2.2). This gives a direction in which to move. A search is conducted along this direction vector for an improving solution, and the process is restarted, until a local maximum of (2.4) is obtained. The optimization algorithm is given in greater detail in [22]. Note that the procedure does not guarantee a global revenue maximum. In practice it is generally necessary to repeat the procedure for various initial conditions. Also, the multicommodity flow bound and the Erlang bound described in [22] are useful aids in deciding when to terminate the optimization procedure.

3. Network analysis

In this section we first give the fixed point equations for determining the loss probabilities, and then the equations for the implied costs.

3.1. Fixed point equations

The fixed point equations are derived on the basis of the well known link independence assumption. Each route which uses link ℓ contributes a load to link ℓ, which is Poisson with rate which is reduced by independent thinning by all other links in the route. Let $B_{s\ell}$ denote the loss probability of service s calls on link ℓ. Recall that link ℓ has capacity C_ℓ and service s requires bandwidth $d_{s\ell}$. Let $\nu_{s\ell}$ be the reduced load obtained after thinning of service type s calls offered to link ℓ, and let $\mathbf{d}_\ell = (d_{1\ell}, \ldots, d_{S\ell})$, $\boldsymbol{\nu}_\ell = (\nu_{1\ell}, \ldots, \nu_{S\ell})$. Then,

$$B_{s\ell} = \mathbb{L}_s(\mathbf{d}_\ell, \boldsymbol{\nu}_\ell, C_\ell) \ , \tag{3.1}$$

where the functions $\mathbb{L}_s$ may be calculated exactly by means of the recursion derived independently by Kaufman [13] and Roberts [26]. The complexity of calculating $B_{s\ell}$ by this procedure is $O(C_\ell)$ as $C_\ell \to \infty$.

With the link independence assumption, the reduced load is given by

$$\nu_{s\ell} = \sum_\sigma \sum_{r \in \mathcal{R}(s,\sigma): \ell \in r} \rho_{sr} \prod_{m \in r - \{\ell\}} (1 - B_{sm}) \ . \tag{3.2}$$

The fixed point equations for the network may be summarized as follows.

$$B_{s\ell} = \phi_{s\ell}(\boldsymbol{\nu}_\ell), \quad (s = 1, 2, \ldots, S; \ell = 1, 2, \ldots L) \ , \tag{3.3a}$$

$$\boldsymbol{\nu} = \boldsymbol{\Psi}(\mathbf{B}) \ , \tag{3.3b}$$

where $\boldsymbol{\nu} = \{\nu_{s\ell}\}_{s,\ell}$ and $\mathbf{B} = \{B_{s\ell}\}_{s,\ell}$. The function $\phi_{s\ell}$ is obtained by fixing $\mathbf{d}_\ell$ and C_ℓ in (3.1); $\boldsymbol{\Psi}$ is defined by (3.2).

The method of successive approximation is typically effective for solving (3.3). This method starts with initial values $\boldsymbol{\nu}_\ell^{(0)}$ and iterates, $i = 0, 1, 2, \ldots$

$$\begin{aligned} B_{s\ell}^{(i+1)} &= \phi_{s\ell}(\boldsymbol{\nu}_\ell^{(i)}), \quad (s = 1, 2, \ldots, S; \ \ell = 1, 2, \ldots, L) \ , \\ \boldsymbol{\nu}^{(i+1)} &= \boldsymbol{\Psi}(\mathbf{B}^{(i+1)}) \end{aligned} \tag{3.5}$$

until a convergence criterion is satisfied.

Under the link independence assumption, the loss probability L_{sr} of service s on route r is

$$L_{sr} = 1 - \prod_{\ell \in r} (1 - B_{s\ell}); \quad r \in \mathcal{R}(s, \sigma) \ . \tag{3.6}$$

3.2. Implied costs

From (3.1) and the explicit expression for $\mathbb{L}_s$ (see, for instance [13]), we obtained [22]

Proposition 3.1.

$$\frac{\partial B_{s\ell}}{\partial \nu_{t\ell}} = (1 - B_{t\ell})[\mathbb{L}_s(\mathbf{d}_\ell, \boldsymbol{\nu}_\ell, C_\ell - d_{t\ell}) - \mathbb{L}_s(\mathbf{d}_\ell, \boldsymbol{\nu}_\ell, C_\ell)] \ . \tag{3.7}$$

□

The revenue sensitivity to the offered loads, and equations for the implied costs, were determined [22] by extending the approach of Kelly [16] to the multirate case. This extension has been made independently by Faragó et al. [7]. A heuristic derivation has been given by Chung and Ross [4].

Proposition 3.2. *The sensitivity of the revenue W with respect to the offered load ρ_{sr}, when the revenue is calculated from the solution of the fixed point equations, is*

$$\frac{\partial W}{\partial \rho_{sr}} = (1 - L_{sr}) \left(e_{sr} - \sum_{\ell \in r} c_{s\ell} \right) \ , \tag{3.8}$$

where $c_{s\ell}$ $(s = 1, 2, \ldots, S; \ell = 1, 2, \ldots, L)$ are the implied costs. Moreover, if $\nu_{s\ell;r}$ denotes the thinned load of service type s on route r which is offered to link ℓ, i.e.,

$$\nu_{s\ell;r} = \rho_{sr} \frac{(1 - L_{sr})}{(1 - B_{s\ell})} = \rho_{sr} \prod_{m \in r - \{\ell\}} (1 - B_{sm}) \ , \tag{3.9}$$

then the implied costs satisfy the following system of SL linear equations,

$$c_{t\ell} = \frac{1}{(1 - B_{t\ell})} \sum_{\sigma, s} \sum_{r \in \mathcal{R}(s,\sigma) : \ell \in r} \frac{\partial B_{s\ell}}{\partial \nu_{t\ell}} \nu_{s\ell;r} \left(e_{sr} - \sum_{k \in r - \{\ell\}} c_{sk} \right) \ . \tag{3.10}$$

The expression for $\partial B_{s\ell} / \partial \nu_{t\ell}$ in the above equation is given in (3.7). □

Note, from (3.2), that the thinned load of service type s offered to link ℓ is

$$\nu_{s\ell} = \sum_{\sigma} \sum_{r \in \mathcal{R}(s,\sigma) : \ell \in r} \nu_{s\ell;r} \ . \tag{3.11}$$

Also note that the fixed point equations are independent of the equations for the implied costs, while the coefficients of the latter depend on the solution of the former. Specifically, $\{B_{t\ell}\}$, $\{\partial B_{s\ell} / \partial \nu_{t\ell}\}$ and $\{\nu_{s\ell;r}\}$ in (3.10) are obtained from the solution of the fixed point equations.

Note that the complexity of solving the system of equations in (3.10) is $O(S^3L^3)$. The equations are typically dense.

We are also interested in the sensitivity of loss on links with respect to the link capacity. Using linear interpolation, we define

$$\frac{\partial \mathbb{L}_s}{\partial_t C_\ell}(\mathbf{d}_\ell, \boldsymbol{\nu}_\ell, C_\ell) \triangleq \frac{1}{d_{t\ell}}[\mathbb{L}_s(\mathbf{d}_\ell, \boldsymbol{\nu}_\ell, C_\ell) - \mathbb{L}_s(\mathbf{d}_\ell, \boldsymbol{\nu}_\ell, C_\ell - d_{t\ell})] \; . \tag{3.12}$$

Then, by extending the approach of Kelly [16], we showed [22] that

$$c_{t\ell} = d_{t\ell}\frac{\partial W}{\partial_t C_\ell} \; , \tag{3.13}$$

which gives an alternate interpretation of the implied costs.

4. Single link network analysis

We first give the results from the exact analysis of a single link network, and then apply a uniform asymptotic approximation (UAA) to the loss probabilities, and a refined uniform asymptotic approximation (RUAA) to the loss probabilities and their partial derivatives with respect to the loads, to simplify the analysis. The RUAA is applied in Section 5 to networks with more than one link. The structural forms of the results obtained here are insightful and suggest the generalizations to networks given in the next section.

Since we are considering a single link network in this section, we suppress the link index ℓ, and the route index r throughout this section. Thus, in this section, the offered traffic to the single link network of service type s is ν_s.

4.1. Exact analysis

Let $n_s(s = 1, 2, \ldots, S)$ be the number of active calls of service s. Then [14, 19, 20] the stationary distribution is

$$\pi(n_1, \ldots, n_S) = \frac{1}{G(C)}\prod_{s=1}^{S} e^{-\nu_s}\frac{\nu_s^{n_s}}{n_s!}, \quad \sum_{s=1}^{S} d_s n_s \leq C \; , \tag{4.1}$$

where the normalization constant is

$$G(C) = \sum_{\sum_{s=1}^{S} d_s n_s \leq C} \prod_{s=1}^{S} e^{-\nu_s}\frac{\nu_s^{n_s}}{n_s!} \, . \tag{4.2}$$

The total bandwidth in use is $i = \sum_{s=1}^{S} d_s n_s$, and the distribution of i is

$$\pi_i(C) = [G(i) - G(i-1)]/G(C)\,. \tag{4.3}$$

Kaufman [13] and Roberts [26] showed that $\pi_i(C)$ satisfies the recursion

$$i\pi_i(C) = \sum_{s=1}^{S} d_s \nu_s \pi_{i-d_s}(C), \quad i = 1, \ldots, C\,, \tag{4.4}$$

where $\pi_i(C) = 0$ for $i < 0$, and $\sum_{i=0}^{C} \pi_i(C) = 1$. The complexity of solving the recursion (4.4) is $O(C)$.

The loss probability of calls of service s is

$$B_s = \sum_{j=0}^{d_s-1} \pi_{C-j}(C) \triangleq \mathbb{L}_s(\mathbf{d}, \boldsymbol{\nu}, C)\,. \tag{4.5}$$

From Propositions 3.1 and 3.2 we obtain

Proposition 4.1. *For a single link network, the revenue sensitivities are given by*

$$\frac{\partial W}{\partial \nu_s} = \left[1 - \sum_{j=0}^{d_s-1} \pi_{C-j}(C)\right](e_s - c_s)\,, \tag{4.6}$$

where the implied costs are

$$c_t = \sum_{s=1}^{S} e_s \nu_s \sum_{j=0}^{d_s-1} \left[\pi_{C-d_t-j}(C-d_t) - \pi_{C-j}(C)\right]. \tag{4.7}$$

□

From (4.2) it follows that

$$\sum_{C=0}^{\infty} G(C) z^C = (1-z)^{-1} \exp\left[\sum_{s=1}^{S} \nu_s (z^{d_s} - 1)\right], \tag{4.8}$$

for $|z| < 1$. If we apply Cauchy's formula [5], we obtain

$$G(C) = \frac{1}{2\pi i} \int_{|z|<1} \frac{\exp\left[\sum_{s=1}^{S} \nu_s (z^{d_s} - 1)\right]}{(1-z) z^{C+1}}\, dz\,, \tag{4.9}$$

where the integral is taken in a counter-clockwise direction around a circle of radius less than 1. Hence, from (4.3), we have the exact representation

$$\pi_{C-j}(C) = \frac{1}{2\pi i G(C)} \int_{|z|<1} \frac{\exp\left[\sum_{s=1}^{S} \nu_s (z^{d_s} - 1)\right]}{z^{C-j+1}}\, dz\,. \tag{4.10}$$

An exact representation for B_s follows from (4.5). Earlier we used that representation to obtain uniform asymptotic approximations to B_s, when the capacity C, and the traffic intensities ν_s $(s = 1, 2, \ldots, S)$, are commensurately large. We summarize our results in the next two subsections, and give the corresponding approximations to $\pi_{C-j}(C)$ for $j = O(1)$.

4.2. UAA

For completeness we first consider the UAA to the stationary distribution of the total bandwidth in use, and to the loss probabilities, for a single link. It is assumed that

$$C \gg 1;\ \nu_s = O(C),\ (s = 1, 2, \ldots, S), \tag{4.11}$$

where C is an integer. Also d_s $(s = 1, 2, \ldots, S)$ are positive integers, not large relative to C, and the greatest common divisor (g.c.d.) of $d_1, \ldots, d_S$ is assumed to be 1. We define

$$F(z) \triangleq \sum_{s=1}^{S} \nu_s (z^{d_s} - 1) - C \log z\,. \tag{4.12}$$

There is a unique positive solution z^* of $F'(z) = 0$, where the prime denotes derivative, so that

$$\sum_{s=1}^{S} \nu_s d_s (z^*)^{d_s} = C,\ z^* > 0\,. \tag{4.13}$$

Since $F'(z)$ is strictly monotonic increasing with z, any simple procedure, such as bisection, quickly yields z^*.

A uniform asymptotic approximation to the contour integral in (4.9) was derived by means of the technique given by Bleistein [2]. The approximation remains valid when the stationary point at $z = z^*$ is close to the pole at $z = 1$, and it is also valid when z^* is not close to 1. We have [20].

Proposition 4.2. *Let*

$$V = \sum_{s=1}^{S} \nu_s d_s^2 (z^*)^{d_s}\,, \tag{4.14}$$

and

$$M = \frac{1}{2} Erfc\left[sgn(1 - z^*) \sqrt{-F(z^*)} \right] + \frac{K e^{F(z^*)}}{\sqrt{2\pi V}}\,, \tag{4.15}$$

where

$$(4.16) \qquad K = \frac{1}{(1-z^*)} - \frac{\sqrt{V}\, sgn(1-z^*)}{\sqrt{-2F(z^*)}}, \; z^* \neq 1\,,$$

$$(4.17) \qquad K = \frac{1}{2} + \frac{1}{6V}\sum_{s=1}^{S} \nu_s d_s^3, \; z^* = 1\,,$$

and the complementary error function is given by

$$(4.18) \qquad Erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-\xi^2}\, d\xi\,.$$

Then, asymptotically,

$$(4.19) \qquad G(C) \sim M\,.$$

□

Since $F(1) = 0$, $F'(z^*) = 0$ and $(z^*)^2 F''(z^*) = V$, the expression for K in (4.16) remains finite as $z^* \to 1$, and its limiting value is given by (4.17). However, the numerical evaluation of (4.16) loses accuracy when z^* is close to 1. Consequently, we derived [24] a representation for K which gives numerically accurate results, whether or not z^* is close to 1, or even equal to 1. We also derived an alternate expression for $F(z^*)$, which leads to a more accurate numerical evaluation than (4.12) does when z^* is close to 1.

Since the integrand in (4.10) is analytic at $z = 1$, an asymptotic approximation to the contour integral may be obtained by the standard saddle-point method [8]. Then, from (4.19), we have

Proposition 4.3. *Let*

$$(4.20) \qquad \mathbb{B} = \frac{e^{F(z^*)}}{M\sqrt{2\pi V}}\,.$$

Then, asymptotically,

$$(4.21) \qquad \pi_{C-j}(C) \sim \mathbb{B}(z^*)^j, \quad j = O(1)\,.$$

□

The above is a noteworthy two-parameter, $(z^*;\mathbb{B})$, characterization of the geometric tail of the stationary link bandwidth utilization distribution, as obtained by the UAA. This characterization has not appeared before.

Proceeding to the calculation of the blocking probabilities, let $h^{(1)}(z^*;n)$ be the following polynomial in z^* of degree $(n-1)$,

$$h^{(1)}(z^*;n) = \sum_{j=0}^{n-1}(z^*)^j \,. \tag{4.22}$$

In the development below z^* will be fixed while various value of n will be considered. Hence, it is convenient to suppress the argument z^* in (4.22) and write $h^{(1)}(z^*;n) = h^{(1)}(n)$. From (4.5) and (4.21), we obtain

Proposition 4.4.

$$B_s \sim \mathbb{B}h^{(1)}(d_s) \quad (s = 1, 2, \ldots, S)\,. \tag{4.23}$$

□

The asymptotic approximations to the loss probabilities B_s given by Proposition 4.4 hold whether $0 < z^* < 1$, $1 - z^* = O(1/\sqrt{C})$ or $z^* > 1$, corresponding to the overloaded, critically loaded and underloaded regimes, respectively. Numerical results [20] illustrate that the approximation (4.23) is quite accurate in all three regimes. In [22] (4.23) was differentiated with respect to ν_t, to yield

$$\partial B_s/\partial \nu_t \sim \mathbb{B}(\mathbb{B} - 1 + z^*)h^{(1)}(d_s)h^{(1)}(d_t)\,. \tag{4.24}$$

Since $\mathbb{B} \sim 1 - z^*$ in the overloaded regime, corresponding to $0 < z^* < 1$, (4.24) is not useful there. Moreover, numerical examples [24] show that (4.24) is not very accurate in the important critically loaded regime, corresponding to $1 - z^* = O(1/\sqrt{C})$, since $\mathbb{B} - 1 + z^* = O(1/\sqrt{C})$ there. It is therefore necessary to refine the approximation (4.24). Refinements of (4.19), (4.21) and (4.23) are also obtained.

4.3. RUAA

We now give the RUAA to the stationary distribution of the link occupancy, i.e., the total bandwidth in use, and to the loss probabilities, and their partial derivatives with respect to the loads, for a single link. First, we have [24]

Proposition 4.5. *Let*

$$T = \sum_{s=1}^{S} \nu_s d_s^2 (d_s - 3)(z^*)^{d_s}\,, \quad Y = \sum_{s=1}^{S} \nu_s d_s^2 (d_s^2 - 6d_s + 11)(z^*)^{d_s}\,, \tag{4.25}$$

$$W = \sum_{s=1}^{S} \nu_s d_s^2 (d_s^3 - 10d_s^2 + 35d_s - 50)(z^*)^{d_s}\,, \tag{4.26}$$

and

$$\begin{aligned} E &= \frac{[1 - 3z^* + 3(z^*)^2]}{(1-z^*)^3} + \frac{T(1-2z^*)}{2V(1-z^*)^2} \\ &\quad + \frac{1}{8}\left(\frac{5T^2}{3V^2} - \frac{Y}{V}\right)\frac{1}{(1-z^*)} - \frac{V\sqrt{V}sgn(1-z^*)}{[-2F(z^*)]^{3/2}},\ z^* \neq 1 \end{aligned} \tag{4.27}$$

$$E = 1 + \frac{T}{2V} + \frac{5T^2}{24V^2} + \frac{35T^3}{432V^3} - \frac{Y}{8V}\left(1 + \frac{5T}{6V}\right) + \frac{W}{40V},\ z^* = 1\,. \tag{4.28}$$

Then, we have the refined uniform asymptotic approximation

$$G(C) \sim M\left(1 - \frac{\mathbb{B}E}{V}\right)\,. \tag{4.29}$$

□

The expression for E in (4.27) remains finite as $z^* \to 1$, and its limiting value is given by (4.28). However, the numerical evaluation of (4.27) loses accuracy significantly when z^* is close to 1. Consequently, a representation for E has been given in [24], which gives numerically accurate results, whether or not z^* is close to 1, or even equal to it.

We note that $V = O(C)$ and $E = O(1)$. A refined asymptotic approximation to contour integrals of the form in (4.10) is given in [24]. Combined with (4.29), and neglecting smaller order terms, we obtain

Proposition 4.6. *Let*

$$\mathbb{B}^{(1)} = \mathbb{B}\left\{1 + \frac{1}{V}\left[\frac{1}{8}\left(\frac{Y}{V} - \frac{5T^2}{3V^2}\right) + \mathbb{B}E\right]\right\}, \tag{4.30}$$

$$\mathbb{B}^{(2)} = \frac{\mathbb{B}T}{2V^2},\ \mathbb{B}^{(3)} = -\frac{\mathbb{B}}{2V}. \tag{4.31}$$

Then, we have the refined asymptotic approximation

$$\pi_{C-j}(C) \sim \left[\mathbb{B}^{(1)} + \mathbb{B}^{(2)}(j-1) + \mathbb{B}^{(3)}(j-1)(j-2)\right](z^*)^j, \quad j = O(1). \tag{4.32}$$

□

This characterization of the geometric tail of the stationary link bandwidth utilization distribution by the RUAA is new and important. It is a four-parameter characterization, $(z^*; \mathbb{B}^{(1)}, \mathbb{B}^{(2)}, \mathbb{B}^{(3)})$, which contrasts with the two-parameter characterization obtained earlier by the UAA. Minimal characterizations are important in our work on real-time network control, where the parameters are computed locally and disseminated throughout the network.

Proceeding to the calculation of the blocking probabilities, let $h^{(2)}(z^*; n)$ and $h^{(3)}(z^*; n)$ be two additional polynomials in z^* of degree $(n-1)$, which augment the polynomial in (4.22).

$$h^{(2)}(z^*; n) = \sum_{j=0}^{n-1}(j-1)(z^*)^j, \quad h^{(3)}(z^*; n) = \sum_{j=0}^{n-1}(j-1)(j-2)(z^*)^j. \tag{4.33}$$

As with the polynomial in (4.22), for convenience we will suppress the argument z^* here and write $h^{(i)}(z^*; n) = h^{(i)}(n)$, $i = 1, 2, 3$.

From (4.5) and (4.32), we obtain

$$\sum_{j=0}^{n-1} \pi_{C-j}(C) \sim \sum_{i=1}^{3} \mathbb{B}^{(i)} h^{(i)}(n), \quad n = O(1), \tag{4.34}$$

and, as a corollary,

Proposition 4.7. *The refined asymptotic approximation is*

$$B_s \sim \sum_{i=1}^{3} \mathbb{B}^{(i)} h^{(i)}(d_s) \quad (s = 1, 2, \ldots, S). \tag{4.35}$$

□

We now proceed to the calculations of the partial derivatives of the loss probabilities by the refined asymptotic approximation. First, note the expressions

$$\sum_{i=0}^{n-1} \pi_{C-i}(C) = [G(C) - G(C-n)]/G(C)\,, \tag{4.36}$$

which follows from (4.3), and

$$\frac{\partial B_s}{\partial \nu_t} = \frac{G(C-d_t)}{G(C)} \cdot \frac{G(C-d_s)}{G(C)} - \frac{G(C-d_s-d_t)}{G(C)}\,, \tag{4.37}$$

which corresponds to (3.7). Hence,

$$\begin{aligned} \frac{\partial B_s}{\partial \nu_t} \quad &= \quad \sum_{j=0}^{d_s+d_t-1} \pi_{C-j}(C) - \sum_{j=0}^{d_s-1} \pi_{C-j}(C) - \sum_{j=0}^{d_t-1} \pi_{C-j}(C) \\ &\quad + \sum_{j=0}^{d_s-1} \pi_{C-j}(C) \sum_{k=0}^{d_t-1} \pi_{C-k}(C)\,. \end{aligned} \tag{4.38}$$

Making use of (4.34), we obtain the refined asymptotic approximation,

$$\begin{aligned} \frac{\partial B_s}{\partial \nu_t} \quad &\sim \quad \sum_{i=1}^{3} \mathbb{B}^{(i)} \left[h^{(i)}(d_s+d_t) - h^{(i)}(d_s) - h^{(i)}(d_t) \right] \\ &\quad + \sum_{i=1}^{3} \sum_{j=1}^{3} \mathbb{B}^{(i)} \mathbb{B}^{(j)} h^{(i)}(d_s) h^{(j)}(d_t)\,. \end{aligned} \tag{4.39}$$

The above expression is not in the form of a sum of products of terms depending respectively on s and t, which is crucial for certain reductions in the subsequent network analysis. To obtain the requisite form we proceed as follows.

From (4.22) we obtain

$$h^{(1)}(d_s+d_t) - h^{(1)}(d_s) = (z^*)^{d_s} h^{(1)}(d_t)\,, \tag{4.40}$$

and

$$(z^*)^{d_s} - 1 = (z^* - 1) h^{(1)}(d_s)\,. \tag{4.41}$$

Hence,

$$h^{(1)}(d_s+d_t) - h^{(1)}(d_s) - h^{(1)}(d_t) = (z^* - 1) h^{(1)}(d_s) h^{(1)}(d_t)\,. \tag{4.42}$$

Similarly, from (4.33), we find that

$$h^{(2)}(d_s+d_t)-h^{(2)}(d_s)-h^{(2)}(d_t) = d_s(z^*)^{d_s}h^{(1)}(d_t)+(z^*-1)h^{(1)}(d_s)h^{(2)}(d_t)\,, \tag{4.43}$$

and

$$h^{(3)}(d_s+d_t)-h^{(3)}(d_s)-h^{(3)}(d_t) = d_s(d_s-1)(z^*)^{d_s}h^{(1)}(d_t)+2d_s(z^*)^{d_s}h^{(2)}(d_t)+(z^*-1)h^{(1)}(d_s)h^{(3)}(d_t)\,. \tag{4.44}$$

If we express $d_s(z^*)^{d_s}$ and $d_s(d_s-1)(z^*)^{d_s}$ in terms of $h^{(i)}(d_s)$, $i=1,2,3$, then we obtain expressions for the right-hand sides of (4.43) and (4.44) which are symmetric in s and t, but this is not necessary for our purposes.

If we use (4.42)–(4.44) in (4.39), then we obtain an expression for $\partial B_s/\partial \nu_t$ of the requisite form. However, it follows from (4.30) and (4.31), since $V=O(C)$, that

$$\mathbb{B}^{(1)} = \mathbb{B}\left[1+O\left(\frac{1}{C}\right)\right],\ \mathbb{B}^{(2)} = O\left(\frac{\mathbb{B}}{C}\right),\ \mathbb{B}^{(3)} = O\left(\frac{\mathbb{B}}{C}\right). \tag{4.45}$$

Hence, neglecting smaller order terms as we did in obtaining (4.32), we have

Proposition 4.8. *Let*

$$\begin{aligned} A_s^{(1)} &= [\mathbb{B}(\mathbb{B}-1+z^*)+(2\mathbb{B}-1+z^*)(\mathbb{B}^{(1)}-\mathbb{B})]h^{(1)}(d_s) \\ &\quad + \mathbb{B}^{(2)}[\mathbb{B}h^{(2)}(d_s)+d_s(z^*)^{d_s}]+\mathbb{B}^{(3)}[\mathbb{B}h^{(3)}(d_s)+d_s(d_s-1)(z^*)^{d_s}]\,, \end{aligned} \tag{4.46}$$

and

$$A_s^{(2)} = \mathbb{B}^{(2)}(\mathbb{B}-1+z^*)h^{(1)}(d_s)+2\mathbb{B}^{(3)}d_s(z^*)^{d_s},\ \ A_s^{(3)} = \mathbb{B}^{(3)}(\mathbb{B}-1+z^*)h^{(1)}(d_s)\,. \tag{4.47}$$

Then, we have

$$\frac{\partial B_s}{\partial \nu_t} \sim \sum_{i=1}^{3} A_s^{(i)}h^{(i)}(d_t)\,. \tag{4.48}$$

□

Note that the complexity of calculating the asymptotic approximations (4.35) and (4.48) to B_s and $\partial B_s/\partial \nu_t$ is $O(1)$, i.e., bounded as $C\to\infty$. Also, the forms in (4.23), (4.24), (4.35) and (4.48), which are specific to UAA and RUAA, are essential for the reduction in complexity of computing loss and implied costs in networks, as we shall see.

Corresponding to (3.7), we have

$$\mathbb{L}_s(\mathbf{d}, \boldsymbol{\nu}, C - d_t) - \mathbb{L}_s(\mathbf{d}, \boldsymbol{\nu}, C) = \frac{1}{(1 - B_t)} \frac{\partial B_s}{\partial \nu_t} . \tag{4.49}$$

Hence, from (4.5), (4.7) and (4.48), we obtain asymptotic approximations to the implied costs in a single link network,

$$c_t \sim \sum_{i=1}^{3} a_t^{(i)} \sum_{s=1}^{S} e_s \nu_s A_s^{(i)}; \; a_t^{(i)} = \frac{h^{(i)}(d_t)}{(1 - B_t)} . \tag{4.50}$$

We remark that the requirement that the g.c.d. of $d_1, \ldots, d_S$ be 1 is crucial. This must be taken into consideration if some of the elements of $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_S)$ are zero, see [21].

5. Network analysis based on RUAA

With the groundwork laid in Sections 3 and 4, it is straightforward to describe the fixed point equations obtained by UAA and RUAA, and the equations for the implied costs for the network obtained by RUAA.

5.1. Fixed point equations

The fixed point equations for the network are given in (3.3). The function $\boldsymbol{\Psi}$ is unchanged and defined by (3.2). The function $\boldsymbol{\phi}_\ell$ which maps $\boldsymbol{\nu}_\ell$ to $\mathbf{B}_\ell$ is, of course, quite different if UAA or RUAA is used for link analyses. For UAA the procedure is given in Proposition 4.2. Specifically (we re-introduce the link index), z_ℓ^* is obtained from (4.13), $F_\ell(z_\ell^*)$ is given by (4.12), V_ℓ, M_ℓ and K_ℓ are given by (4.14)–(4.17), and $h_\ell^{(1)}(d_{s\ell})$ and $\mathbb{B}_\ell$ by (4.22) and (4.20). Finally, for UAA,

$$B_{s\ell} = \mathbb{B}_\ell \, h_\ell^{(1)}(d_{s\ell}) . \tag{5.1}$$

The above procedure defines the functions

$$B_{s\ell} = U_{s\ell}(\boldsymbol{\nu}_\ell), \; (s = 1, 2, \ldots, S; \; \ell = 1, 2, \ldots, L) , \tag{5.2}$$

which contrasts with (3.3a).

For RUAA the procedure is given by Proposition 4.7. The quantities T_ℓ and Y_ℓ are given by (4.25), and E_ℓ, $h_\ell^{(2)}(d_{s\ell})$ and $h_\ell^{(3)}(d_{s\ell})$ by (4.27) and (4.33). Finally, (4.35) defines the functions

$$B_{s\ell} = R_{s\ell}(\boldsymbol{\nu}_\ell), \; (s = 1, 2, \ldots, S; \; \ell = 1, 2, \ldots, L)\,, \tag{5.3}$$

which contrasts with (3.3a).

In practical implementation we have found it to be efficient to base the optimization procedures on UAA at the early stages, and only later switch to RUAA, because of the somewhat greater computation burden associated with RUAA.

5.2. Implied costs

Equation (3.10) gives the system of linear equations satisfied by the implied costs. In (3.10) we substitute the expression obtained by RUAA, which is given by Proposition 4.8,

$$\frac{\partial B_{s\ell}}{\partial \nu_{t\ell}} = \sum_{i=1}^{3} A_{s\ell}^{(i)} h_\ell^{(i)}(d_{t\ell})\,, \tag{5.4}$$

and we let

$$a_{t\ell}^{(i)} = \frac{h_\ell^{(i)}(d_{t\ell})}{(1 - B_{t\ell})}\,. \tag{5.5}$$

This gives

$$c_{t\ell} = \sum_{i=1}^{3} a_{t\ell}^{(i)} \sum_{\sigma,s} \sum_{r\in\mathcal{R}(s,\sigma):\ell\in r} A_{s\ell}^{(i)} \nu_{s\ell;r} \left(e_{sr} - \sum_{k\in r-\{\ell\}} c_{sk} \right)\,. \tag{5.6}$$

Now define

$$\xi_\ell^{(i)} \triangleq \sum_{\sigma,s} \sum_{r\in\mathcal{R}(s,\sigma):\ell\in r} A_{s\ell}^{(i)} \nu_{s\ell;r} \left(e_{sr} - \sum_{k\in r-\{\ell\}} c_{sk} \right)\,. \tag{5.7}$$

Note that for all t $(1 \le t \le S)$ and ℓ $(1 \le \ell \le L)$,

$$c_{t\ell} = \sum_{i=1}^{3} a_{t\ell}^{(i)} \xi_\ell^{(i)}\,. \tag{5.8}$$

Now introduce (5.8) in (5.7) to obtain

$$\xi_\ell^{(i)} = \sum_{\sigma,s} \sum_{r\in\mathcal{R}(s,\sigma):\ell\in r} A_{s\ell}^{(i)} \nu_{s\ell;r} \left[e_{sr} - \sum_{k\in r-\{\ell\}} \sum_{j=1}^{3} a_{sk}^{(j)} \xi_k^{(j)} \right]\,. \tag{5.9}$$

Equation (5.9) is a complete system of equations in $\{\xi_\ell^{(i)}\}$. The parameters in this system of equations, namely, $\{\nu_{s\ell;r}\}$, $\{A_{s\ell}^{(i)}\}$ and $\{a_{sk}^{(j)}\}$ are all obtained directly from the solution of the fixed point equations in Section 5.1. Equation (5.9) has the property of being a system of only $3L$ linear equations. Once the solution of (5.9) has been obtained then all the implied costs $\{c_{t\ell}\}$ are obtained from (5.8). This procedure for obtaining the implied costs has complexity $O(27L^3)$, and its independence from the number of services constitutes one of the main contributions of this paper. The complexity of solving the linear system of equations (5.6) directly for the implied costs is $O(S^3L^3)$, which is greater if S is large. Central to the above reduction is the structure (5.4) implied by RUAA for links.

5.3. Hybrid system of network equations

We expect that not atypically networks will have some links which are not large and therefore not amendable to an asymptotic analysis, even though most links will be suitable for application of RUAA. We propose a procedure for such cases which, while based on a simple observation, considerably enhances the range of applications. The procedure, a hybrid of asymptotic and nonasymptotic analyses, gives closed systems of equations for the loss probabilities as well as the implied costs. Let

$$\mathcal{L}_E \triangleq \text{set of links which are subject to nonasymptotic analysis, and}$$

$$\mathcal{L}_A \triangleq \text{set of links which are subject to UAA and RUAA.}$$

Each link is an element of one of these sets.

Let us first consider the system of linear equations for the implied costs. From (5.7) and (5.8), for $\ell \in \mathcal{L}_A$,

$$\xi_\ell^{(i)} = \sum_{\sigma,s} \sum_{r\in\mathcal{R}(s,\sigma):\ell\in R} A_{s\ell}^{(i)} \nu_{s\ell;r} \times \left[e_{sr} - \sum_{k\in\mathcal{L}_E\cap r} c_{sk} - \sum_{k\in\mathcal{L}_A\cap(r-\{\ell\})} \sum_{j=1}^{3} a_{sk}^{(j)} \xi_k^{(j)} \right], \tag{5.10}$$

and $c_{t\ell}$ is given by (5.8). For $\ell \in \mathcal{L}_E$, from (3.7) and (3.10),

$$
\begin{aligned}
c_{t\ell} &= \sum_{\sigma,s} \sum_{r \in \mathcal{R}(s,\sigma):\ell \in r} \left[\mathbb{L}_s(\mathbf{d}_\ell, \boldsymbol{\nu}_\ell, C_\ell - d_{t\ell}) - \mathbb{L}_s(\mathbf{d}_\ell, \boldsymbol{\nu}_\ell, C_\ell)\right] \nu_{s\ell;r} \\
&\quad \times \left[e_{sr} - \sum_{k \in \mathcal{L}_E \cap (r - \{\ell\})} c_{sk} - \sum_{k \in \mathcal{L}_A \cap r} \sum_{j=1}^{3} a_{sk}^{(j)} \xi_k^{(j)} \right] .
\end{aligned}
\tag{5.11}
$$

Note that (5.10) and (5.11) form a complete set of equations in $\xi_\ell^{(i)}$ ($\ell \in \mathcal{L}_A;\ i = 1,2,3$) and $c_{s\ell}(\ell \in \mathcal{L}_E;\ s = 1,2,\ldots,S)$.

The above idea extends naturally to a hybrid system of nonlinear fixed point equations for the loss probabilities. In (3.3),

$$B_{s\ell} = U_{s\ell}(\boldsymbol{\nu}_\ell) \text{ or } R_{s\ell}(\boldsymbol{\nu}_\ell),\ \ell \in \mathcal{L}_A\,, \tag{5.12a}$$

$$= \phi_{s\ell}(\boldsymbol{\nu}_\ell),\ \ell \in \mathcal{L}_E\,, \tag{5.12b}$$

where $U_{s\ell}$ and $R_{s\ell}$ are as discussed in Section 5.1. As before $\boldsymbol{\nu} = \boldsymbol{\Psi}(\mathbf{B})$, and the iterative method of successive approximations applies in a straightforward manner.

6. Numerical results

Here the focus is on the comparative numerical accuracy of the UAA and the RUAA applied to network performance and network optimization, while in prior work [24] the focus has been on single links. In [24] comparison was made with the exact result (4.5) for two particular examples, and it was found that the UAA (see (4.23)) is quite accurate, and the RUAA (see (4.35)) is very accurate. The UAA and RUAA to the partial derivatives of the loss probabilities with respect to the loads, given by (4.24) and (4.48), respectively, were also compared with the exact result (4.37). It was found that overall the UAA is rather poor, but the RUAA is very accurate.

Here we compare the RUAA to the implied costs, given by (4.50), with the exact result (4.7), for the same two examples. The first example has capacity $C = 96$ and two services, with bandwidths $d_1 = 1$ and $d_2 = 6$, and corresponding earned revenues $e_1 = 1$ and $e_2 = 6$. The corresponding traffic intensities are $\nu_1 = 0.86\,\nu/1.7$ and $\nu_2 = 0.14\,\nu/1.7$, where

$\nu = \nu_1 d_1 + \nu_2 d_2$ is the total traffic intensity. In Table 1 we compare the RUAA to the implied costs, c_1 and c_2, with the exact results for $\nu = 96 + \beta\sqrt{96}$, where $\beta = -6, -5, \ldots, 3$, which includes the underloaded, critically loaded and overloaded regimes. It is seen that the RUAA is very accurate. We also compare the UAA to the implied costs (4.7), corresponding to the use of (4.24) in (3.10), and it is seen that overall the approximation is rather poor.

Table 1: Comparison of RUAA and UAA to the implied costs c_1 and c_2 with the exact results for two services with $C = 96$, $d_1 = 1 = e_1$, $d_2 = 6 = e_2$, $\nu_1 = 0.86\,\nu/1.7$, $\nu_2 = 0.14\,\nu/1.7$ and $\nu = 96 + \beta\sqrt{96}$.

β	c_1 (Exact)	c_1 (RUAA)	c_1 (UAA)	c_2 (Exact)	c_2 (RUAA)	c_2 (UAA)
-6	3.3225×10^{-4}	3.3226×10^{-4}	3.3415×10^{-4}	3.6205×10^{-3}	3.6206×10^{-3}	3.8238×10^{-3}
-5	5.9151×10^{-3}	5.9151×10^{-3}	5.9493×10^{-3}	5.4807×10^{-2}	5.4809×10^{-2}	5.8373×10^{-2}
-4	3.9940×10^{-2}	3.9941×10^{-2}	4.0174×10^{-2}	3.2409×10^{-1}	3.2411×10^{-1}	3.4960×10^{-1}
-3	1.3597×10^{-1}	1.3597×10^{-1}	1.3677×10^{-1}	9.9410×10^{-1}	9.9416×10^{-1}	1.0929
-2	2.8816×10^{-1}	2.8816×10^{-1}	2.8989×10^{-1}	1.9566	1.9567	2.2101
-1	4.5039×10^{-1}	4.5040×10^{-1}	4.5319×10^{-1}	2.9161	2.9165	3.4137
0	5.8681×10^{-1}	5.8683×10^{-1}		3.6918	3.6923	
1	6.8857×10^{-1}	6.8859×10^{-1}		4.2587	4.2595	
2	7.6095×10^{-1}	7.6098×10^{-1}		4.6585	4.6595	
3	8.1203×10^{-1}	8.1206×10^{-1}		4.9400	4.9412	

The second example has capacity $C = 130$ and three services, with bandwidths $d_1 = 1$, $d_2 = 3$ and $d_3 = 10$, and corresponding earned revenues $e_1 = 1$, $e_2 = 3$ and $e_3 = 10$. The corresponding traffic intensities are $\nu_1 = 2\nu/13 = \nu_2$ and $\nu_3 = \nu/26$, where $\nu = \nu_1 d_1 + \nu_2 d_2 + \nu_3 d_3$ is the total traffic intensity. In Table 2 we compare the RUAA to the implied costs, c_1, c_2 and c_3, with the exact results for $\nu = 46.8\,(20.8)\,171.6$, which includes the underloaded, critically loaded and overloaded regimes. Again the RUAA is very accurate.

We next consider the star network, depicted in Figure 1, which was studied by Chung and Ross [4]. They showed that the loss probabilities and revenue sensitivities, calculated from the fixed point equations for the network, agree quite well with simulated confidence

Table 2: Comparison of RUAA to the implied costs c_1, c_2 and c_3 with the exact results for three services with $C = 130$, $d_1 = 1 = e_1$, $d_2 = 3 = e_2$, $d_3 = 10 = e_3$, $\nu_1 = 2\nu/13 = \nu_2$ and $\nu_3 = \nu/26$.

ν	c_1 (Exact)	c_1 (RUAA)	c_2 (Exact)	c_2 (RUAA)	c_3 (Exact)	c_3 (RUAA)
46.8	5.3231×10^{-5}	5.3234×10^{-5}	1.8707×10^{-4}	1.8707×10^{-4}	1.1302×10^{-3}	1.1301×10^{-3}
67.6	2.9753×10^{-3}	2.9755×10^{-3}	9.8646×10^{-3}	9.8645×10^{-3}	4.7012×10^{-2}	4.7011×10^{-2}
88.4	2.4486×10^{-2}	2.4488×10^{-2}	7.7766×10^{-2}	7.7759×10^{-2}	3.1488×10^{-1}	3.1488×10^{-1}
109.2	6.9296×10^{-2}	6.9297×10^{-2}	2.1430×10^{-1}	2.1425×10^{-1}	7.9053×10^{-1}	7.9054×10^{-1}
130	1.1555×10^{-1}	1.1554×10^{-1}	3.5264×10^{-1}	3.5245×10^{-1}	1.2447	1.2448
150.8	1.5234×10^{-1}	1.5228×10^{-1}	4.6212×10^{-1}	4.6170×10^{-1}	1.5998	1.6001
171.6	1.8101×10^{-1}	1.8087×10^{-1}	5.4761×10^{-1}	5.4686×10^{-1}	1.8794	1.8802

intervals for the exact results obtained with the Monte Carlo summation method. The network has 5 nodes and 4 links, with capacities as shown in Figure 1.

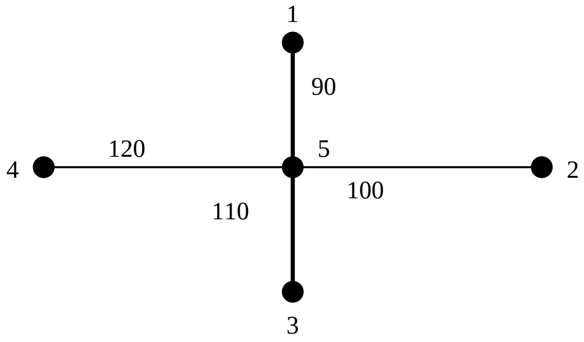

Figure 1: Star network, with link capacities as depicted, and bandwidths $d_{1\ell} = 1$ and $d_{2\ell} = 5$, $(\ell = 1, 2, 3, 4)$.

There are two services, with bandwidths 1 and 5, respectively, on all links. The earned revenue for each service and each route is shown in Table 3. Three cases, corresponding to light, moderate and heavy traffic, are considered. In each case the offered traffic depends only on the service, and not on the route. In light traffic $\rho_{1r} \equiv 9$ and $\rho_{2r} = 1.6$, in moderate traffic $\rho_{1r} \equiv 10$ and $\rho_{2r} \equiv 2$, and in heavy traffic $\rho_{1r} \equiv 15$ and $\rho_{2r} \equiv 3$.

Table 3: Routes r, and earned revenues e_{sr} for services $s = 1, 2$.

Route r	e_{1r}	e_{2r}
$1, 5, 2$	1.0	3.0
$1, 5, 3$	1.2	3.6
$1, 5, 4$	1.4	4.2
$2, 5, 3$	1.6	4.8
$2, 5, 4$	1.8	5.4
$3, 5, 4$	2.0	6.0

We first assume that the entire offered traffic is put on each route. In Table 4 we compare the RUAA to the implied costs $c_{1\ell}$ and $c_{2\ell}$, for links between node ℓ ($\ell = 1, 2, 3, 4$) and node

Table 4: Comparison of RUAA to the implied costs $c_{1\ell}$ and $c_{2\ell}$, $\ell = 1, 2, 3, 4$, with the results based on exact analysis for links, for the star network.

Traffic	$c_{1\ell}$ **(Exact)**	$c_{1\ell}$ **(RUAA)**	$c_{2\ell}$ **(Exact)**	$c_{2\ell}$ **(RUAA)**
Light	1.4161×10^{-2}	1.4161×10^{-2}	9.8503×10^{-2}	9.8506×10^{-2}
	2.9519×10^{-3}	2.9520×10^{-3}	2.2199×10^{-2}	2.2199×10^{-2}
	4.0917×10^{-4}	4.0918×10^{-4}	3.2972×10^{-3}	3.2973×10^{-3}
	4.3344×10^{-5}	4.3345×10^{-5}	3.7150×10^{-4}	3.7150×10^{-4}
Moderate	6.9558×10^{-2}	6.9560×10^{-2}	4.3041×10^{-1}	4.3043×10^{-1}
	2.3792×10^{-2}	2.3792×10^{-2}	1.5849×10^{-1}	1.5849×10^{-1}
	5.4634×10^{-3}	5.4634×10^{-3}	3.8981×10^{-2}	3.8981×10^{-2}
	9.5793×10^{-4}	9.5792×10^{-4}	7.2711×10^{-3}	7.2711×10^{-3}
Heavy	3.4639×10^{-1}	3.4641×10^{-1}	1.8253	1.8256
	2.9027×10^{-1}	2.9025×10^{-1}	1.5864	1.5864
	1.5889×10^{-1}	1.5887×10^{-1}	9.1345×10^{-1}	9.1333×10^{-1}
	6.8519×10^{-2}	6.8506×10^{-2}	4.1728×10^{-1}	4.1720×10^{-1}

5, with the results based on the exact analysis for links, both being based on the fixed point equations for the network. It is seen that the RUAA is very accurate, although slightly less so in the heavy traffic case. In Table 5 we compare the RUAA to the loss probabilities $B_{1\ell}$ and $B_{2\ell}$, $\ell = 1, 2, 3, 4$, with the results based on the exact analysis for links. Again the

RUAA is very accurate, although slightly less so in the heavy traffic case. We remark that the revenue and its gradients are given by (2.4) and (3.8), with the help of (3.6), in terms of $B_{s\ell}$ and $c_{s\ell}$.

Table 5: Comparison of RUAA to the loss probabilities $B_{1\ell}$ and $B_{2\ell}$, $\ell = 1, 2, 3, 4$, with the results based on exact analysis for links, for the star network.

Traffic	$B_{1\ell}$ **(Exact)**	$B_{1\ell}$ **(RUAA)**	$B_{2\ell}$ **(Exact)**	$B_{2\ell}$ **(RUAA)**
Light	4.6146×10^{-4}	4.6147×10^{-4}	3.4001×10^{-3}	3.4000×10^{-3}
	6.3401×10^{-5}	6.3401×10^{-5}	5.0052×10^{-4}	5.0051×10^{-4}
	6.7283×10^{-6}	6.7283×10^{-6}	5.6506×10^{-5}	5.6505×10^{-5}
	5.6466×10^{-7}	5.6466×10^{-7}	5.0164×10^{-6}	5.0163×10^{-6}
Moderate	2.9170×10^{-3}	2.9170×10^{-3}	1.9130×10^{-2}	1.9130×10^{-2}
	6.3498×10^{-4}	6.3498×10^{-4}	4.4637×10^{-3}	4.4636×10^{-3}
	1.0847×10^{-4}	1.0847×10^{-4}	8.1059×10^{-4}	8.1057×10^{-4}
	1.4701×10^{-5}	1.4701×10^{-5}	1.1610×10^{-4}	1.1610×10^{-4}
Heavy	3.9940×10^{-2}	3.9939×10^{-2}	2.0495×10^{-1}	2.0494×10^{-1}
	1.8151×10^{-2}	1.8148×10^{-2}	1.0148×10^{-1}	1.0146×10^{-1}
	6.8244×10^{-3}	6.8229×10^{-3}	4.1035×10^{-2}	4.1026×10^{-2}
	2.1181×10^{-3}	2.1175×10^{-3}	1.3551×10^{-2}	1.3547×10^{-2}

Next we consider the revenue optimization problem (2.5) for the star network. After optimization it was found that in both the light and moderate traffic cases equality holds in (2.2). However, in the heavy traffic case the optimization leads to dropping some of the service 2 traffic on route $1, 5, 2$, i.e., (2.2) holds with strict inequality, the optimal offered traffic after dropping being 1.94 rather than 3. This is consistent with the negative revenue gradient for service 2 on route $1, 5, 2$, which occurs when the entire offered traffic is put on each route, since, from Table 4, $c_{21} + c_{22} > 3$, the earned revenue. In Table 6 we compare the RUAA to the optimal revenue with the results based on the exact analysis for links. It is seen that the RUAA is very accurate.

Table 6: Comparison of RUAA to the optimal revenue with the results based on exact analysis for links, for the star network.

Traffic	**Exact**	**RUAA**
Light	124.1115	124.1115
Moderate	143.3038	143.3038
Heavy	199.5292	199.5306

Finally, we consider the Core NSF network described in an earlier paper [22]. This network has 8 nodes and 20 unidirectional links connecting 10 pairs of nodes in both directions. There are 6 services, with bandwidths $1, 3, 4, 6, 24$ and 40. Four different route sets are considered for each source-destination pair: route set 1 is the set of routes with the minimum number of hops (min-hop), route set 2 consists of all routes with number of hops ≤ 3, route set 3 consists of all routes with number of hops $\leq$ min-hop $+1$, and route set 4 consists of all routes with number of hops ≤ 4. The number of routes in these sets is 68, 102, 108 and 160, respectively. As stated in [22], for route set 1 the optimal revenue obtained by UAA is 20,387, whereas the "exact" result is 20,401. (To recapitulate, the "exact" method is different in that the link blocking probabilities are calculated by the exact Kaufman–Roberts recursions, and the implied costs are calculated by solving (3.10); both methods rely on the fixed point equations and on the common optimization algorithm.) For the remaining route sets UAA encounters convergence problems in the final iterations of the optimization procedure, because of errors in the gradients of the revenue function, so that the line search could not find an improving direction.

On the other hand, RUAA converges for all 4 route sets, using initial conditions $\rho_{sr} = 0$ for all s, r. To check the accuracy of the RUAA converged solution, we used the optimal traffic splits to the various admissible routes which it gave, and evaluated the corresponding revenue exactly. In Table 7 we compare the optimal revenue obtained by RUAA to that obtained by the exact analysis. We note that RUAA converges to the same optimal solution as the exact one, except for route set 4, where the RUAA result is quite close to the exact one.

Table 7: Comparison of RUAA to the optimal revenue with the results based on exact analysis for links, for the Core NSF network.

Route set	**Exact**	**RUAA**
1	20,401	20,401
2	21,012	21,012
3	21,020	21,020
4	21,029	21,022

References

[1] A. Birman and Y. Kogan. Asymptotic evaluation of closed queueing networks with many stations. *Commun. Statist. – Stochastic models*, 8:543–563, 1992.

[2] N. Bleistein. Uniform asymptotic expansions of integrals with stationary point near algebraic singularity. *Commun. Pure Appl. Math.*, 19:353–370, 1966.

[3] R. Callon, P. Doolan, N. Feldman, A. Fredette, G. Swallow, and A. Viswanathan. A framework for multiprotocol label switching. Work in progress. *Internet Draft* ⟨*draft-ietf-mpls-framework-01.txt*⟩, July 30 1997.

[4] S.-P. Chung and K. W. Ross. Reduced load approximations for multirate loss networks. *IEEE Trans. Commun.*, 41:1222–1231, 1993.

[5] E. T. Copson. *Theory of Functions of a Complex Variable.* Oxford University Press, London, 1948.

[6] S. P. Evans. Optimal bandwidth management and capacity provision in a broadband network using virtual paths. *Perform. Eval.*, 13:27–43, 1991.

[7] A. Faragó, S. Blaabjerg, L. Ast, G. Gordos, and T. Henk. A new degree of freedom in ATM network dimensioning: optimizing the logical configuration. *IEEE J. Select. Areas Commun.*, 13:1199–1206, 1995.

[8] M. V. Fedoryuk. Asymptotic methods in analysis. In R. V. Gamkrelidze, editor, *Analysis I.* Springer-Verlag, New York, 1989.

[9] R. Gallager, B. Hajek, F. P. Kelly, D. Mitra, and P. P. Varaiya (Guest Editors). Advances in the Fundamentals of Networking. *IEEE J. Select. Areas Commun.*, 13(6–7), August–September 1995.

[10] P. Gazdzicki, I. Lambadaris, and R. R. Mazumdar. Blocking probabilities for large multirate Erlang loss systems. *Adv. Appl. Prob.*, 25:997–1009, 1993.

[11] P. J. Hunt. Implied costs in loss networks. *Adv. Appl. Prob.*, 21:661–680, 1989.

[12] P. J. Hunt and F. P. Kelly. On critically loaded loss networks. *Adv. Appl. Prob.*, 21:831–841, 1989.

[13] J. S. Kaufman. Blocking in a shared resource environment. *IEEE Trans. Commun.*, 29:1474–1481, 1981.

[14] F. P. Kelly. *Reversibility and Stochastic Networks.* Wiley, New York, 1980.

[15] F. P. Kelly. Routing in circuit-switched networks: optimization, shadow prices and decentralization. *Adv. Appl. Prob.*, 20:112–144, 1988.

[16] F. P. Kelly. Fixed point models of loss networks. *J. Austr. Math. Soc. Ser. B*, 31:204–218, 1989.

[17] F. P. Kelly. Loss networks. *Ann. Appl. Prob.*, 1:319–378, 1991.

[18] Y. Kogan and M. Shenfild. Asymptotic solutions of generalized multiclass Engset model. In J. Labetoulle and J. W. Roberts, editors, *The Fundamental Role of Teletraffic in the Evolution of Telecommunications Networks, Proc. ITC 14*, volume 1b, pages 1239–1249, Elsevier, Amsterdam, 1994.

[19] S. S. Lam. Queueing networks with population size constraints. *IBM J. Res. Develop.*, 21:370–378, 1977.

[20] D. Mitra and J. A. Morrison. Erlang capacity and uniform approximations for shared unbuffered resources. *IEEE/ACM Trans. Networking*, 2:558–570, 1994.

[21] D. Mitra, J. A. Morrison, and K. G. Ramakrishnan. Unified approach to multirate ATM network design and optimization. In *Proc. 9th ITC Specialists Seminar, Teletraffic Modeling and Measurement in Broadband and Mobile Communication*, pages 77–94, Leidschendam, The Netherlands, 1995.

[22] D. Mitra, J. A. Morrison, and K. G. Ramakrishnan. ATM network design and optimization: a multirate loss network framework. *IEEE/ACM Trans. Networking*, 4:531–543, 1996.

[23] J. A. Morrison. Loss probabilities in a simple circuit-switched network. *Adv. Appl. Prob.*, 26:456–473, 1994.

[24] J. A. Morrison, K. G. Ramakrishnan, and D. Mitra. Refined asymptotic approximations to loss probabilities and their sensitivities in shared unbuffered resources. *SIAM J. Appl. Math.*, 59:494–513, published electronically November 18, 1998. Paper located in http://cm.bell-labs.com/who/mitra.

[25] M. I. Reiman. A critically loaded multiclass Erlang loss system. *Queueing Syst.*, 9:65–82, 1991.

[26] J. W. Roberts. Teletraffic models for the Telecom 1 integrated services network. In *Proc. 10th Internat. Teletraffic Congress*, Session 1.1, paper #2, Montreal, Canada, 1983.

[27] K. W. Ross. *Multirate Loss Models for Broadband Telecommunications Networks.* Springer, New York, 1995.

[28] M. Schwartz. *Broadband Integrated Networks.* Prentice Hall, New York, 1996.

[29] R. Siebenhaar. Optimized ATM virtual path bandwidth management under fairness constraints. In *Proc. IEEE GLOBECOM '94*, pages 924–928, 1994.