ATM NETWORK DESIGN AND OPTIMIZATION: A MULTIRATE LOSS NETWORK FRAMEWORK*

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

Bell Labs, Lucent Technologies
Murray Hill, NJ 07974

Abstract

ATM network design and optimization at the call-level may be formulated in the framework of multirate, circuit-switched, loss networks with effective bandwidth encapsulating cell-level behavior. Each service supported on the ATM network is characterized by a rate or bandwidth requirement. Future networks will be characterized by links with very large capacities in circuits and by many rates. Various asymptotic results are given to reduce the attendant complexity of numerical calculations. A central element is a uniform asymptotic approximation (UAA) for link analyses. Moreover, a unified hybrid approach is given which allows asymptotic and nonasymptotic methods of calculations to be used cooperatively. Network loss probabilities are obtained by solving fixed point equations. A canonical problem of route and logical network design is considered. An optimization procedure is proposed, which is guided by gradients obtained by solving a system of equations for implied costs. A novel application of the EM algorithm gives an efficient technique for calculating implied costs with changing traffic conditions. Finally, we report numerical results obtained by the software package TALISMAN, which incorporates the theoretical results. The network considered has 8 nodes, 20 links, 6 services and as many as 160 routes.

Key words: effective bandwidth, asymptotic approximations, fixed point equations, revenue maximization, implied costs, steepest ascent, routing, EM algorithm, flow bounds, Erlang bounds.

*An earlier version of this paper was presented at the IEEE Infocom '96 Conference. The paper was selected as one of its top papers and referred to the Transactions for possible publication after the Transactions' own independent review.

ATM NETWORK DESIGN AND OPTIMIZATION: A MULTIRATE LOSS NETWORK FRAMEWORK

Debasis Mitra, John A. Morrison, and K. G. Ramakrishnan
Bell Labs, Lucent Technologies
Murray Hill, NJ 07974

1. INTRODUCTION

This paper addresses some basic questions of design and optimization of wide area, broadband ATM networks. All the questions addressed are at the connection, or call, layer. In fact, ATM networks are viewed in this paper as multirate circuit-switched, loss networks. The framework that supports this viewpoint relies on the concept of "effective bandwidth" for encapsulating all cell-level behavior, including multiplexing and related quality of service issues. This separation of levels, which hides details of lower levels, is essential for the tractability of high level design problems. Fortuitously, during this decade the notion of effective bandwidth has been extensively researched. See [1] for reviews and original contributions describing recent work for various traffic and buffer types. Future networks will integrate a multitude of services with diverse traffic characteristics, ranging from constant bit rate to highly bursty, variable bit rate. In the view of this paper, the description of each service includes a characteristic bandwidth or rate for each link which carries the service. Notably, there already exists a substantial body of work on single-rate, circuit-switched network design [2, 3]. However, further examination indicates that there are various features which make the new network design problems substantially different and more formidable. This paper addresses some of these questions.

Two features deserve particular note. The first concerns size. Consider a link with bandwidth 150 Mbps, which carries traffic of various services with the smallest service rate being 8 Kbps. In the circuit switched model such a link is viewed as having 18,750 circuits. Since high speed optical transmission links can carry 2.5 Gbps, the link capacity in circuits, which we denote generically by C, is expected to be very large in future networks. The second noteworthy feature is that the number of services, S, can be expected to be large, say, in the range 10–25 in the near future, and higher fairly soon. For instance, each leaky bucket regulator may be associated with a particular effective bandwidth [4]. This paper gives results which specifically address and help network design problems in which C is large and S may not be small.

The main contribution of this paper is the systematic incorporation of asymptotics in various stages of the computations in multirate network design and optimization. More generally, we give a unified, hybrid framework which allows asymptotic and nonasymptotic techniques to cooperate and complement each other. A central element in the asymptotic approach is the uniform asymptotic approximation (UAA) developed by Mitra and Morrison [5] for the analysis of single links. The asymptotic regime for the UAA is large C and offered traffic which is also large, $O(C)$. Prior work on the asymptotic analysis of circuit switched networks required separate treatments for the underloaded, critical and overloaded regimes (see for instance [6, 7, 8, 9]), which the UAA renders unnecessary. The blocking probabilities obtained by specializing the UAA to each of these regimes gives results which had been known earlier, such as the results due to Reiman [7] and Evans [8] for the critical regime, and Gazdzicki, Lambadaris and Mazumdar [9] for the overloaded and underloaded regimes and the critical point. Other references to related works are in [5]. The complexity of calculating blocking probabilities on a link carrying multirate traffic by the UAA is $O(1)$, i.e., bounded, fixed as $C \to \infty$. This contrasts with complexity $O(C)$ by the well-known, exact calculations due to Kaufman [10] and Roberts [11].

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 [12], Whitt [13], Girard [3] and Ross [14]. The nature of the errors introduced by the approximation has drawn considerable attention, both theoretical (see, for instance, [15], [16] and [14]) and empirical. The notion of implied costs is used crucially for calculating network revenue sensitivity during the optimization. The equations for the implied costs are obtained by extending the approach due to Kelly [17, 18, 19, 20] to the multirate case. The exact equations for the implied costs in the multirate network are linear with dimension SL, where L is the number of links in the network, and typically not sparse. Hence the complexity of solving these equations is $O(S^3L^3)$. The memory requirements are commensurately high. We show that the UAA gives a particular product-form structure to the sensitivity of links, which is used subsequently to reduce the number of linear equations for the network's implied costs to only L. The reduction of complexity to $O(L^3)$ is a key contribution of this paper.

Girard and Ho [21] give a general technique for network optimization, which relies on the fixed point approach and an approximation due to Labourdette and Hart [22] for link analyses. Girard and Ho emphasize the importance of speed in the link analyses. This is not hard to appreciate since each network design and optimization problem requires many network solutions, each network solution requires many iterations and each iteration requires L link analyses. A similar argument shows the leveraged advantage of reduced complexity of solving the equations for the implied costs.

It should be noted that the asymptotic techniques work well only in their appropriate regimes. (This of course explains the attraction of the hybrid approach, which combines asymptotic and nonasymptotic techniques.) Generally our experience has been that in computing link blocking probabilities, the UAA is quite robust and accurate for typical realistic problems. For computing sensitivities of link probabilities the UAA is sometimes less accurate, even on the important critical and underloaded regimes. Accurate refinements based on higher order terms in the uniform asymptotic expansion have been obtained but cannot be reported here for lack of space. These refinements retain the important feature of having numerical complexity which is independent of C and S, the link capacity and the number of services, respectively.

The network revenue function is generally nonconcave, as our computational experience confirms. Hence we have to resort to repeated applications of the steepest ascent optimization procedure from various initial conditions. To calibrate the quality of the best solution obtained, it is very helpful to have bounds on the optimal revenue and on the lost bandwidth. In Section 7 we give two classes of bounds, one of which, called flow bounds, is obtained by maximizing revenue in a corresponding deterministic, multirate multicommodity flow problem. We also give Erlang bounds, which take into account traffic stochasticity and is based on the notions of cuts and aggregation of the links in each cut into a single link.

This paper focuses on a single canonical problem in ATM network design, namely, the calculation of rates of traffic offered to various routes connecting origin-destination node pairs, which maximize network revenue. Network revenue is a proxy for network performance. The routes may be considered to be virtual paths, a concept which is supported by ATM standards. Hence, the problem considered here may alternately be viewed as one of optimally sizing the virtual path network [19, 23].

A natural extension of the design problem considered in this paper is the real-time adaptation of the sizes of routes or virtual paths to changing traffic loads. In the framework of this paper the main burden of resizing is the computation of the new implied costs to reflect the altered traffic loads. Section 8, which is on "Adaptation," gives an iterative algorithm for solving the equations for the implied costs, which uses the prior implied costs as starting values. This is a novel application of the EM algorithm [24, 25, 26, 27], which is a well-known technique for maximum likelihood estimation.

Another canonical problem is the design of least cost networks. This is a higher level problem in that it subsumes the optimum sizing of routes or virtual paths. Other canonical problems have been identified.

We have designed an initial prototype of a software package called TALISMAN (**T**ool for **A**na**L**ys**I**s and **S**ynthesis of **M**ultirate **A**TM **N**etworks), which incorporates our theoretical results for solving multi-layered network design problems. 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. TALISMAN is written in C++ and is modularized to reflect the layered problem and analytic structure. The software structure of TALISMAN will be described in a future publication. Extensive numerical results obtained by exercising TALISMAN on a network with 8 nodes, 20 links, 6 services and route sets containing up to 160 routes are reported in Section 9. Also given there are comparative flow bounds obtained by solving LPs, Erlang bounds and results on the EM algorithm applied to the computation of implied costs. In [28] we report numerical results obtained by TALISMAN by purely asymptotic and hybrid techniques for loss probabilities and implied costs in a simple network, which was introduced and extensively studied by Chung and Ross [20]. The numbers, which are for light, moderate and heavy traffic, are in fairly close agreement.

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. The links are unidirectional, i.e., carry traffic in one direction only. There are S types of service, where service s requires bandwidth $d_{s\ell}$ on link ℓ. The script letters $\mathcal{L}$, $\mathcal{N}$ and $\mathcal{S}$ will denote the corresponding sets. 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 to. 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.,

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

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 for given offered traffic rates $\{\rho_{sr}\}$, it is necessary to analyze the network for the loss probabilities $\{L_{sr}\}$. The dependence of the loss probabilities on the offered traffic rates is complicated. Our optimization algorithm is iterative. The complete process needs to be repeated for various initial conditions. An example of initial conditions is $\rho_{sr} = 0$, $\forall\ s \in \mathcal{S}$ and $r \in \mathcal{R}(s, \sigma)$, $\forall \sigma$, and another is uniform splitting of traffic among all routes with the minimum number of hops.

We analyze the network using the techniques described in Sections 3–6, to obtain W and $\partial W/\partial \rho_{sr}$. We then compute a steepest ascent direction by projecting $\partial W/\partial \rho_{sr}$ on to the null space of (2.2). This gives us a direction in which to move. We then search along this direction vector for an improving solution, and restart the iterative process, until a local maximum of (2.4) is obtained. The optimization algorithm is given below as Algorithm A.

Algorithm A: Offered Traffic on Routes which Maximize Revenue

input: N; L; S; C_ℓ, $1 \le \ell \le L$; $d_{s\ell}$, $1 \le s \le S$, $1 \le \ell \le L$; $\mathcal{R}(s, \sigma)$, $1 \le s \le S$, $\sigma \in \mathcal{N} \times \mathcal{N}$; $\Lambda_{s\sigma}$, $\mu_{s\sigma}$, $1 \le s \le S$, $\sigma \in \mathcal{N} \times \mathcal{N}$; e_{sr}, $\rho_{sr}^{(0)}$, $1 \le s \le S$, $r \in \mathcal{R}(s, \sigma)$, $\sigma \in \mathcal{N} \times \mathcal{N}$.

output: W^*, the (locally) optimum revenue and ρ^*_{sr}, the (locally) optimum rates of traffic offered to routes.

1. Initialize: $k \leftarrow 0$

   evaluate $W^{(0)}$

   not_converged $\leftarrow$ true

2. **While** (not_converged) **do**

2.1. $L_{sr}, \partial W/\partial \rho_{sr} \leftarrow$ Network_solve $(N, L, S, C_\ell, d_{s\ell}, \mathcal{R}(s,\sigma), \rho_{sr}^{(k)}, e_{sr})$

2.2. $\mathbf{d} \leftarrow$ Null_space_project $\left(\partial W/\partial \rho_{sr}^{(k)}\right)$

2.3. $\rho_{sr}^{(k+1)} \leftarrow$ Line_search $(\rho_{sr}^{(k)}, \mathbf{d})$

2.4. evaluate $W^{(k+1)}$

2.5. **if** $(W^{(k+1)} - W^{(k)})/W^{(k)} \leq \epsilon$, not_converged $\leftarrow$ false

2.6. $k \leftarrow k+1$

end_while

The procedure Network_solve in the above algorithm performs two tasks: first, it solves the hybrid system of nonlinear fixed point equations (see Sec. 6) by successive approximations; second, it solves the system of linear equations for computing the implied costs by either direct factorization or the EM algorithm. The procedure Line_search in line 2.3 performs a one dimensional optimization to compute the maximum of the revenue function along the direction $\mathbf{d}$.

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 [10] and Roberts [11]. The complexity of calculating $B_{s\ell}$ by this procedure is $O(C)$ as $C \to \infty$.

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.

$$B_{s\ell} = \phi_{s\ell}(\boldsymbol{\nu}_\ell), \quad (s = 1, 2, \ldots, S; \ell = 1, 2, \ldots L) \ , \tag{3.4a}$$

$$\boldsymbol{\nu} = \boldsymbol{\Psi}(\mathbf{B}) \ , \tag{3.4b}$$

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 (3.4). This method starts with initial values $\boldsymbol{\nu}_\ell^{(0)}$ and iterates, $i = 0, 1, 2, \ldots$

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

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 [10]), we obtain in Appendix A

Proposition 3.1.

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

The revenue sensitivity to the offered loads, and equations for the implied costs, are determined in Appendix B by extending the approach of Kelly [18] to the multirate case. This extension has also been made independently by Faragó et al. [19]. A heuristic derivation has been given by Chung and Ross [20].

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 $c_{s\ell}$ ($s = 1, 2, \dots, S; \ell = 1, 2, \dots, L$) 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 $\partial B_{s\ell}/\partial \nu_{t\ell}$ given in (3.7) should be substituted in (3.10). □

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 $O(S^3 L^3)$. The equations are not typically expected to be sparse.

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 [18], we show in Appendix B 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. While implied costs are positive quantities typically, and almost certainly for engineered networks, there are instances of hierarchical networks and certain networks associated with paradoxes for which the costs are negative.

4. SINGLE LINK NETWORK ANALYSIS

We first state the results from the exact analysis of a single link network, and then apply a uniform asymptotic approximation (UAA) to simplify the analysis. The UAA 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

We summarize the main, exact results.

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

$$\frac{\partial W}{\partial \nu_s} = (1 - B_s)(e_s - c_s) \ , \tag{4.1}$$

where the loss probability of type s calls is

$$B_s = \mathbb{L}_s(\mathbf{d}, \boldsymbol{\nu}, C), \quad (s = 1, 2, \ldots, S) \ , \tag{4.2}$$

and the implied costs are

$$c_t = \sum_{s=1}^{S} e_s \nu_s [\mathbb{L}_s(\mathbf{d}, \boldsymbol{\nu}, C - d_t) - \mathbb{L}_s(\mathbf{d}, \boldsymbol{\nu}, C)], \quad (t = 1, 2, \ldots, S) \ . \qquad \Box \tag{4.3}$$

4.2. UAA

We now summarize the UAA to the loss probabilities in a single link, derived by Mitra and Morrison [5]. It is assumed that

$$C \gg 1; \quad \nu_s = O(C), \quad (s = 1, 2, \ldots, S) \ , \tag{4.4}$$

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

$$F(z) \triangleq \sum_{s=1}^{S} \nu_s(z^{d_s}-1) - C\log z; \quad V(z) \triangleq \sum_{s=1}^{S} d_s^2 \nu_s z^{d_s} \ . \tag{4.5}$$

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

$$\sum_{s=1}^{S} d_s \nu_s (z^*)^{d_s} = C, \quad z^* > 0 \ . \tag{4.6}$$

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

Proposition 4.2. *Let*

$$\begin{aligned} b_s &= \frac{[1-(z^*)^{d_s}]}{(1-z^*)}, \quad z^* \neq 1 \ ; \\ &= d_s, \qquad\qquad\quad z^* = 1 \ , \end{aligned} \tag{4.7}$$

and

$$B = \frac{e^{F(z^*)}}{M\sqrt{2\pi V(z^*)}} \ , \tag{4.8}$$

where

$$\begin{aligned} M &= \frac{1}{2}\,\mathrm{Erfc}\left[\mathrm{sgn}(1-z^*)\sqrt{-F(z^*)}\right] \\ &\quad + \frac{e^{F(z^*)}}{\sqrt{2\pi}}\left[\frac{1}{\sqrt{V(z^*)}(1-z^*)} - \frac{\mathrm{sgn}(1-z^*)}{\sqrt{-2F(z^*)}}\right] \ , \quad z^* \neq 1 \ , \end{aligned} \tag{4.9}$$

$$M = \frac{1}{2}\left\{1 + \frac{1}{\sqrt{2\pi V(1)}}\left[1 + \frac{1}{3V(1)}\sum_{s=1}^{S} d_s^3 \nu_s\right]\right\}, \quad z^* = 1 \ , \tag{4.10}$$

and the complementary error function is given by

$$\mathrm{Erfc}\,(y) = \frac{2}{\sqrt{\pi}}\int_y^{\infty} e^{-x^2}\,dx \ . \tag{4.11}$$

Then, asymptotically,

$$B_s \sim b_s B, \quad (s = 1,2,\ldots,S) \ . \qquad \Box \tag{4.12}$$

These approximations hold whether $0 < z^* < 1$, $z^* \approx 1$ or $z^* > 1$, corresponding to the overloaded, critical and underloaded regimes, respectively. In Appendix C we prove the following result which gives an analogous approximation to the derivative of B_s with respect to ν_t.

Proposition 4.3. *Let*

$$a_t = \frac{(z^*)^{d_t}}{(1 - B_t)} - 1 \ . \tag{4.13}$$

Then,

$$\mathbb{L}_s(\mathbf{d}, \boldsymbol{\nu}, C - d_t) - \mathbb{L}_s(\mathbf{d}, \boldsymbol{\nu}, C) = \frac{1}{(1 - B_t)} \frac{\partial B_s}{\partial \nu_t} \sim a_t B_s \ . \qquad \square \tag{4.14}$$

Note that the complexity of calculating B_s and $\partial B_s / \partial \nu_t$ is $O(1)$, i.e., bounded, fixed as $C \to \infty$. Moreover, the product forms in (4.12) and (4.14), which are specific to UAA, are critical for the reduction in complexity of computing loss and implied costs in networks.

If we use (4.12) and (4.14) in (4.3), we obtain asymptotic approximations to the implied costs in a single link network,

$$\begin{aligned} & c_t \sim a_t \xi, \\ \text{where} \quad & \xi = \textstyle\sum_{s=1}^{S} e_s \nu_s B_s \sim B \sum_{s=1}^{S} e_s \nu_s b_s \ . \end{aligned} \tag{4.15}$$

If $0 < z^* < 1$, i.e., in the overloaded regime, $B_t \sim 1 - (z^*)^{d_t}$, so that $a_t \sim 0$, and it is necessary to refine the approximation. These refined approximations have been obtained but are not given here (partly to save space and partly because the overloaded regime is not of great interest).

We emphasize 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 [28]. We have

Proposition 4.4. *Suppose that $\nu_s > 0$, $1 \le s \le K$ and $\nu_s = 0$, $K + 1 \le s \le S$, and that the g.c.d. of $d_1, \ldots, d_K$ is g. Let*

$$\widehat{C} = \lfloor C/g \rfloor; \quad \widehat{d}_s = \lfloor C/g \rfloor - \lfloor (C - d_s)/g \rfloor, \quad 1 \le s \le S \ . \tag{4.16}$$

Then, if C and $\mathbf{d}$ are replaced by $\widehat{C}$ and $\widehat{\mathbf{d}}$, the UAA to B_s holds for $1 \le s \le S$, and the analogous approximations to $\partial B_s / \partial \nu_t$ hold if $1 \le s \le K$, **or** *$1 \le t \le K$.* $\square$

We note, from (4.16), that $\widehat{d}_s = d_s / g$, $1 \le s \le K$. We remark that in general the correct asymptotic approximations are not obtained for $\partial B_s / \partial \nu_t$, $K + 1 \le s \le S$, $K + 1 \le t \le S$. However, since $\nu_s = 0$, $K + 1 \le s \le S$, these quantities do not occur in the equations for the implied costs, so this does not matter.

5. NETWORK ANALYSIS BASED ON UAA

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

5.1. Fixed Point Equations

The fixed point equations for the network are given in (3.4). 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 is used for link analyses. The procedure is given in Proposition 4.2. Specifically (we re-introduce the link index), the functions $F_\ell(z)$ and $V_\ell(z)$ are defined in (4.5), the point z_ℓ^* where the minimum $F_\ell(z_\ell^*)$ occurs is calculated, $b_{s\ell}$ and B_ℓ calculated as in (4.7) and (4.8), and finally

$$B_{s\ell} = b_{s\ell} B_\ell \ .$$

In summary, the above procedure defines the function

$$B_{s\ell} = U_{s\ell}(\boldsymbol{\nu}_\ell) \ , \quad (s = 1, 2, \ldots, S;\ \ell = 1, 2, \ldots, L) \tag{5.1}$$

which contrasts with (3.4a).

The iterative, successive approximations procedure for solving the fixed point equations, which is described in (3.5), is also used here. Since a network design and optimization problem requires many network solutions, the reduction of complexity in analyzing a single link from $O(C)$ to $O(1)$ translates to a very substantial reduction of the numerical burden.

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 UAA, which is given in (4.14),

$$\frac{1}{(1 - B_{t\ell})} \frac{\partial B_{s\ell}}{\partial \nu_{t\ell}} = a_{t\ell} B_{s\ell} \tag{5.2}$$

where

$$a_{t\ell} = \frac{(z_\ell^*)^{d_{t\ell}}}{(1 - B_{t\ell})} - 1 \ . \tag{5.3}$$

This gives

$$c_{t\ell} = a_{t\ell} \sum_{\sigma,s} \sum_{r \in \mathcal{R}(s,\sigma):\ell \in r} \nu_{s\ell;r} B_{s\ell} \left(e_{sr} - \sum_{k \in r - \{\ell\}} c_{sk} \right) . \tag{5.4}$$

Now define

$$\xi_\ell \triangleq \sum_{\sigma,s} \sum_{r \in \mathcal{R}(s,\sigma):\ell \in r} \nu_{s\ell;r} B_{s\ell} \left(e_{sr} - \sum_{k \in r - \{\ell\}} c_{sk} \right) . \tag{5.5}$$

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

$$c_{t\ell} = a_{t\ell} \xi_\ell . \tag{5.6}$$

Now introduce (5.6) in (5.5) to obtain

$$\xi_\ell = \sum_{\sigma,s} \sum_{r \in \mathcal{R}(s,\sigma):\ell \in r} \nu_{s\ell;r} B_{s\ell} \left(e_{sr} - \sum_{k \in r - \{\ell\}} a_{sk} \xi_k \right) . \tag{5.7}$$

Equation (5.7) is a complete system of equations in $\{\xi_\ell\}$. The parameters in this system of equations, namely, $\{\nu_{s\ell;r}\}$, $\{B_{s\ell}\}$ and $\{a_{sk}\}$ are all obtained directly from the solution of the fixed point equations in Section 5.1. Equation (5.7) has the remarkable property of being a system of only L linear equations. Once the solution of (5.7) has been obtained then all the implied costs $\{c_{t\ell}\}$ are obtained from (5.6). This procedure for obtaining the implied costs has complexity $O(L^3)$, and its independence from the number of services constitutes one of the main contributions of this paper. The reader may verify that central to the above reduction is the product-form structure (4.14) implied by UAA for links.

6. 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 UAA. 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 blocking probabilities as well as the implied costs. Let

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

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

Each link is an element of one of these sets.

Let us first consider the system of linear equations for the implied costs. From (5.5) and (5.6), for $\ell \in \mathcal{L}_A$,

$$\xi_\ell = \sum_{\sigma,s} \sum_{r \in \mathcal{R}(s,\sigma):\ell \in r} \nu_{s\ell;r} B_{s\ell} \left[e_{sr} - \sum_{k \in \mathcal{L}_E \cap (r-\{\ell\})} c_{sk} - \sum_{k \in \mathcal{L}_A \cap (r-\{\ell\})} a_{sk}\xi_k \right] , \tag{6.1}$$

and $c_{t\ell} = a_{t\ell}\xi_\ell$. For $\ell \in \mathcal{L}_E$, from (3.7) and (3.10),

$$\begin{aligned} c_{t\ell} = \sum_{\sigma,s} \sum_{r \in \mathcal{R}(s,\sigma):\ell \in r} & \{\mathbb{L}_s(\mathbf{d}_\ell, \boldsymbol{\nu}_\ell, C_\ell - d_{t\ell}) - \mathbb{L}_s(\mathbf{d}_\ell, \boldsymbol{\nu}_\ell, C_\ell)\}\nu_{s\ell;r} \\ & \times \left[e_{sr} - \sum_{k \in \mathcal{L}_E \cap (r-\{\ell\})} c_{sk} - \sum_{k \in \mathcal{L}_A \cap (r-\{\ell\})} a_{sk}\xi_k \right] . \end{aligned} \tag{6.2}$$

Note that (6.1) and (6.2) form a complete set of equations in ξ_ℓ ($\ell \in \mathcal{L}_A$) 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 blocking probabilities. In (3.4),

$$B_{s\ell} = U_{s\ell}(\boldsymbol{\nu}_\ell), \quad \ell \in \mathcal{L}_A, \tag{6.3a}$$

$$= \phi_{s\ell}(\boldsymbol{\nu}_\ell), \quad \ell \in \mathcal{L}_E, \tag{6.3b}$$

where $U_{s\ell}$ is defined in (5.1). As before, $\boldsymbol{\nu} = \boldsymbol{\Psi}(\mathbf{B})$. The iterative method of successive approximations applies in a straightforward manner.

7. BOUNDS

The goal of the techniques in this paper is to maximize network performance, more precisely, its proxy "network revenue". The network revenue function (2.4) is generally not concave, as our experience (see Section 9) confirms. To calibrate the quality of the solutions obtained by the procedures for maximizing revenue, it is highly desirable to have bounds.

7.1. Flow Bounds

The simplest upper bound is based on stripping the traffic processes of their stochasticity, i.e., by considering such processes to be deterministic fluid or flow processes, which have rates uniform over time. In this framework network revenue is maximized by solving the following multirate, multicommodity maximum-flow problem, which is a general linear programming problem.

$$\max\ W_F = \sum_{\sigma,s} \sum_{r\in\mathcal{R}(s,\sigma)} e_{sr} x_{sr} \tag{7.1}$$

subject to

$$\left.\begin{array}{ll}\displaystyle\sum_{r\in\mathcal{R}(s,\sigma)} x_{sr} \le \bar{\rho}_{s\sigma} & \forall s, \forall\sigma \\ \displaystyle\sum_{\sigma,s} \sum_{r\in\mathcal{R}(s,\sigma):\ell\in r} d_{s\ell} x_{sr} \le C_\ell & \forall \ell \\ x_{sr} \ge 0 & \forall s, \forall r\end{array}\right\} \tag{7.2}$$

Let W_F^* denote the solution to (7.1) and (7.2). Then the intuitively plausible result is

$$W^* \le W_F^*\ , \tag{7.3}$$

where W^* is the solution to the problem in (2.4) and (2.5). The formal proof is also straightforward. Gibbens [29] (see also [14]) has given a proof for the single rate version of the problem.

7.2. Erlang Bounds

These are lower bounds on the probability of blocking bandwidth on the network, which take into account the stochasticity of traffic. The bounds are based on the notions of cuts and the aggregation of the links in the cut into a single link, which is amenable to standard stochastic analysis. The approach described below differs from prior work [29, 14] in the multirate feature and, more importantly, in the use of multiple cuts, each making an additive contribution to the estimate of total lost bandwidth in the network. The latter is consistent with the unidirectionality of traffic on links, a feature of our model, see Section 2.1. In the interest of brevity, we make the following simplifying assumptions: admission control is absent, i.e., all exogenous traffic of stream (s,σ) is offered to the network, so that $p_{s\sigma} = 1$ (see (2.3)); also, $d_{s\ell} = d_s$ for all links ℓ. Extensions to the general case exist.

A cut is a set of links which separate a set of nodes $\mathcal{D}$ from its complement $\mathcal{D}^c = \mathcal{N} \setminus \mathcal{D}$. By convention we associate with the cut traffic flowing from originating nodes in $\mathcal{D}^c$ to destination nodes in $\mathcal{D}$, and denote by ν_s the traffic intensity of service type s across this cut,

$$\nu_s = \sum_{\sigma_1 \in \mathcal{D}^c} \sum_{\sigma_2 \in \mathcal{D}} \bar{\rho}_{s,(\sigma_1,\sigma_2)} \ . \tag{7.4}$$

Also let $\bar{C}$ denote the sum of the capacities of all the links in the cut.

We claim the following lower bound on lost bandwidth in traffic which originates in nodes in $\mathcal{D}^c$ and has destinations in $\mathcal{D}$. For any $\{\rho_{sr}\}$ such that all constraints in (2.2) are satisfied with equality,

$$\sum_s \sum_{\sigma_1 \in \mathcal{D}^c} \sum_{\sigma_2 \in \mathcal{D}} \sum_{r \in \mathcal{R}(s,(\sigma_1,\sigma_2))} d_s \rho_{sr} L_{sr} \geq \sum_s d_s \nu_s \mathbb{L}_s(\mathbf{d}, \boldsymbol{\nu}, \bar{C}) \tag{7.5}$$

where the function $\mathbb{L}_s$ has been defined in (3.1) for single links. The bound is obtained by assuming that there is infinite capacity in links other than the ones in the cut and that the total lost bandwidth due to the latter links is no less than the total loss in a link of aggregate capacity.

Recalling that links are unidirectional, now consider improving the lower bound in (7.5) by making several cuts. For this we need the notion of "mutually exclusive cuts". Cuts with destination node sets $\mathcal{D}_1$, $\mathcal{D}_2, \ldots, \mathcal{D}_K$, where K is the number of cuts, are mutually exclusive if and only if each origin-destination node pair $\sigma \in \mathcal{D}_k$ for at most one k. For example, a set of N mutually exclusive cuts is obtained by considering, for each node in the network, the cut composed of links which terminate at the node. Another example has two mutually exclusive cuts: for a particular node, one cut is composed of all links terminating at the node and the other cut is composed of all links originating at the node.

Denoting by $\bar{\Omega}^{(k)}_{lost}$ the quantity on the right side of (7.5) for cut k in a set of mutually exclusive cuts, and $\bar{\Omega}_{lost} = \sum_k \bar{\Omega}^{(k)}_{lost}$, a lower bound on the probability of blocking bandwidth is

$$\bar{\Omega}_{lost} / \sum_{\sigma,s} d_s \bar{\rho}_{s\sigma} \ . \tag{7.6}$$

8. ADAPTATION

The equations for the implied costs in the general hybrid form, (6.1) and (6.2), may be put in the form

$$\mathbf{A}\mathbf{x} = \mathbf{y} \ , \tag{8.1}$$

where the elements of $\mathbf{x}$ are ξ_ℓ ($\ell \in \mathcal{L}_A$) and $c_{s\ell}$ ($\ell \in \mathcal{L}_E$; $s = 1, 2, \ldots, S$). It is shown in Section 6 that all the implied costs may be obtained from $\mathbf{x}$. Equation (8.1) has the important property that $\mathbf{y} > \mathbf{0}$ and $\mathbf{A} \geq \mathbf{0}$ (element-wise). We give an iterative algorithm here which seeks a nonnegative solution to (8.1), i.e., $\mathbf{x} \geq \mathbf{0}$. An attraction of the algorithm is that it is well-suited for network optimization by steepest ascent procedures in which (8.1) needs to be solved repeatedly for $(\mathbf{A}, \mathbf{y})$ not greatly different from the preceding instance. By starting the iterative procedure at the preceding solution, the number of iterations is reduced, sometimes quite substantially. This has obvious benefits in real-time applications, such as the adaptation of the logical network of virtual paths to changing traffic conditions.

The algorithm is an application of the EM algorithm. We briefly recapitulate a few salient observations made by Vardi and Lee [26, 27]. First, by appropriate scaling (8.1) may be put into a suitable form for statistical inference, wherein $\mathbf{x}$ and $\mathbf{y}$ are probability vectors, i.e., with nonnegative elements which sum to unity, and the column sums of $\mathbf{A}$ are unity. To establish an analogy between solving (8.1) and estimation from incomplete data, consider the statistical model in which X, $Y|X$ and Y are distributed as multinomials with parameters $\mathbf{x}$, $\mathbf{A}_x$ (column x of $\mathbf{A}$) and $\mathbf{y}$. The analogy is completed by considering (X, Y) as the "complete" data and its projection Y as the "incomplete" (observed) data. The EM algorithm may now be applied to this canonical estimation algorithm. The sequence generated by the iterative algorithm always converges to a point which minimizes the Kullback-Liebler information divergence between the iterated and target probability vectors. Moreover, the convergence is monotone in the sense that the distance is strictly decreasing, except on convergence. Finally, the nonnegativity constraints are satisfied throughout the procedure.

The EM algorithm for solving (8.1) is

$$x_i^{(n+1)} = x_i^{(n)} \sum_j \frac{\alpha_{ji}}{\sum_k A_{jk} x_k^{(n)}}, \quad (n = 0, 1, \ldots) \tag{8.2}$$

where $\alpha_{ji} = y_j A_{ji} / \sum_k A_{ki}$. The initial values $x_i^{(0)}$ are positive and otherwise arbitrary.

9. COMPUTATIONAL RESULTS

We describe the results of our computational experiments. The results were obtained by exercising TALISMAN, which, as described in Section 1, is a software package written in C++. The software structure of TALISMAN has several levels, with each level having a plug-in plug-out module and standard socket interface so that each module may run in a location and hardware transparent manner. The levels in the current version are optimization (level 3), implied cost calculations (level 2), fixed point equation solutions and link analyses (level 1).

This section is organized as follows. In Section 9.1 we give the basic facts of the network on which the experiments were conducted. In Section 9.2 we give the results from the revenue maximization problem for four different route sets, which are parameterized by the number of hops in the admissible routes. Section 9.2 also gives computed flow and Erlang bounds. Finally, Section 9.3 gives results on the number of iterations of the EM algorithm, which are required to compute all the implied costs as the traffic between pairs of nodes (selected for the heaviness of their mutual traffic) is changed by steps of 1%.

9.1. Core NSF Network

This fictitious network has 8 nodes ($N = 8$) and, as shown in Figure 1, 10 pairs of nodes are directly connected. Each such node pair is connected by two unidirectional links carrying traffic in opposite directions. Thus the network has 20 unidirectional links ($L = 20$). The typical bandwidth of a unidirectional link is 45 Mbps, with the exceptions being the links in both directions connecting Argonne (3) and Princeton (4), and also Houston (8) and Atlanta (7), which have bandwidths of 90 Mbps.

There are 6 services ($S = 6$). The transmission rates associated with these services are 16, 48, 64, 96, 384 and 640 Kbps respectively, uniformly over all links. Service 1 may be thought of as voice, services 2–5 as data and service 6 as video. It is convenient to rescale and let 16 Kbps be the bandwidth unit. Table 1 gives the service bandwidths in this unit, which then define $d_s \equiv d_{s\ell}$ for all s and ℓ. Also, link bandwidths of 45 Mbps and 90 Mbps translate to 2812 and 5625 units, respectively. Hence $C_\ell \in \{2812, 5625\}$ for $\ell = 1, 2, \ldots, L$.

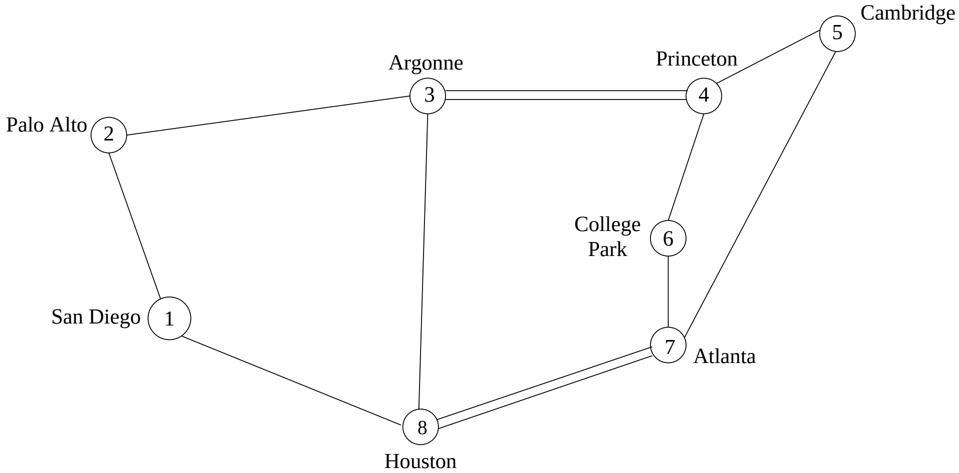

Figure 1: Core NSF Network. The double lines indicate double bandwidth on links.

services, s	1 (voice)	2 (data 1)	3 (data 2)	4 (data 3)	5 (data 4)	6 (video)
rates (Kbps)	16	48	64	96	384	640
d_s	1	3	4	6	24	40

Table 1: Service bandwidth d_s (unit = 16 Kbps), where $d_s \equiv d_{s\ell}$.

To keep the description of exogenous offered traffic intensities compact, we employ a base *traffic intensity matrix* $\mathbf{T}$, which is $N \times N$, and a multiplier m_s for service s ($s = 1, 2, \ldots, S$). Hence, $m_s T_{ij}$ gives the offered traffic intensity of service s from node i to node j, i.e., $\bar{\rho}_{s\sigma}$ where $\sigma = (i, j)$. The unit of measurement for traffic intensity is calls. The matrix $\mathbf{T}$ is given in Table 2 and the multipliers $\{m_s\}$ are in Table 3. Note the asymmetry in the offered traffic. Also,

$$T \triangleq \sum_{i,j} T_{ij} = 527 \; . \tag{9.1}$$

9.2. Revenue Maximization, Routing

We give the results obtained by exercising TALISMAN, which includes various "expert system" features to automatically select either one of the asymptotic formulas or the exact method of link analysis. Throughout, we let the revenue earned per carried call per unit time of service s on route r,

$$e_{sr} \equiv d_s \; , \qquad \forall r \; . \tag{9.2}$$

nodes	1	2	3	4	5	6	7	8
1	–	6	7	1	9	5	2	3
2	7	–	24	3	31	15	6	9
3	8	25	–	4	37	18	7	11
4	1	3	3	–	4	2	1	1
5	11	33	39	5	–	24	9	15
6	5	14	16	2	21	–	4	6
7	2	5	6	1	8	4	–	2
8	3	8	10	1	12	6	2	–

Table 2: Base traffic intensity matrix $\mathbf{T}$.

service, s	1	2	3	4	5	6
multiplier, m_s	0.5	0.5	1.0	0.5	0.5	0.5

Table 3: Offered traffic intensity matrix for service $s = m_s\mathbf{T}$.

The results are for four distinct sets of route sets. In each case the route sets are generated by a simple algorithm, which is described below.

$$\begin{aligned}
&\text{Route sets 1:} && \mathcal{R}(s,\sigma) = \{r \mid \text{min-hop}(\sigma) = \#(\text{hops}) \text{ in route } r\} \\
&\text{Route sets 2:} && \mathcal{R}(s,\sigma) = \{r \mid 3 \geq \#(\text{hops}) \text{ in route } r\} \\
&\text{Route sets 3:} && \mathcal{R}(s,\sigma) = \{r \mid \text{min-hop}(\sigma) + 1 \geq \#(\text{hops}) \text{ in route } r\} \\
&\text{Route sets 4:} && \mathcal{R}(s,\sigma) = \{r \mid 4 \geq \#(\text{hops}) \text{ in route } r\}\ ,
\end{aligned} \tag{9.3}$$

where min-hop(σ) is the minimum number of hops for node pair σ. Since this quantity does not exceed 3 (see Figure 1) for all node pairs, $\mathcal{R}_1(s,\sigma) \subseteq \mathcal{R}_2(s,\sigma) \subseteq \mathcal{R}_4(s,\sigma)$, for all s and σ, where the subscript indexes the method of generating the route sets; also, $\mathcal{R}_3(s,\sigma) \subseteq \mathcal{R}_4(s,\sigma)$ for all s and σ. In fact, the number of routes thus generated are, respectively, 68, 102, 108 and 160. Also note from (9.3) that for a given origin-destination pair σ, the route set $\mathcal{R}(s,\sigma)$ is common for all services s.

Table 4 summarizes the results on revenue maximization from TALISMAN, together with the flow bounds and Erlang bounds. The table also gives data on blocking. Blocking probabilities are computed on the basis of the offered and carried bandwidth traffic in units of bandwidth (not calls). Thus, the offered and carried bandwidth traffic are given by

Ω_{offered} and Ω_{carried}, respectively, where

$$\Omega_{\text{offered}} \triangleq \sum_{\sigma,s} d_s \bar{\rho}_{s\sigma}, \quad \text{and} \quad \Omega_{\text{carried}} = \sum_{\sigma,s} \sum_{r \in \mathcal{R}(s,\sigma)} d_s \rho_{sr}(1 - L_{sr}) \tag{9.4}$$

It follows from (9.1) and Tables 1 and 3 that $\Omega_{\text{offered}} = 21,607$. On account of (9.2) this quantity is also the revenue that may be earned if there is no blocking.

Route	ROUTING: Uniform loading of routes in set		OPTIMAL ROUTING					
			Revenue			Blocking (%)		
Sets	Revenue	Blocking (%)	TALISMAN	Erlang Bound	Flow Bound	TALISMAN	Erlang Bound	Flow Bound
1	19,167	11.3	20,401	21,287	21,081	5.6	1.5	2.4
2	19,050	11.8	21,012	21,287	21,607	2.8	1.5	0
3	19,443	10.0	21,020	21,287	21,607	2.7	1.5	0
4	18,065	16.4	21,026	21,287	21,607	2.7	1.5	0

Table 4: Summary of results from revenue maximizations and bounds.

Table 4 also gives revealing performance benchmarks, also obtained from TALISMAN, for routing based on balancing the offered traffic for each stream (s, σ) uniformly over all routes in the route set $\mathcal{R}(s, \sigma)$. An important special case has the offered traffic divided uniformly over the routes with the minimum number of hops ("min-hop routing"). In this case the revenue is 19,167 and the blocking is 11.3%. Persisting with uniform loading, a noticeable improvement is achieved by enlarging the sets to include routes with one hop in excess of the minimum. The data in the table shows that performance is substantially improved by the combination of optimization and enlarged route sets. In contrast, with uniform loading, enlarging the route sets can have the opposite effect, since excessive traffic may be offered to long, and therefore inefficient, routes.

The Erlang bound in Table 4 is obtained by considering two mutually exclusive cuts which involve, respectively, the links terminating and originating at Cambridge (5). Note that the best revenue result from TALISMAN differs from the Erlang bounds by only 1.2%. There is only one case where the flow bound does better than the Erlang bound, and this occurs when certain links are relatively heavily loaded.

We should mention that the network revenue function is not concave, and therefore the steepest ascent Algorithm A given in Section 2 terminates at local maxima, which depend on initial conditions. Thus, the results reported in Table 4 under "Optimal Routing" are the best that were obtained from various initial conditions. Here are some observations on the search procedures. Somewhat surprisingly, the range of final network revenue obtained from various initial conditions is quite narrow, of the order of only a few percent. Selecting the solution of the flow problem (see Section 7), which gives the flow bound, as initial conditions gives disappointing results upon termination. In fact, surprisingly, the best final results were obtained from obviously poor initial conditions, such as $\rho_{sr}^{(0)} = 0$ for all (s, r).

The performance results differ slightly depending upon whether the UAA or the exact method of link analyses is used. For instance, with Route sets 1, the revenues for "Optimal Routing" obtained by TALISMAN were 20,387 and 20,401, respectively, for the two methods. (Here we forced all links to employ a common method of calculation.) Also, for Route sets 1, there were about 450 network evaluations in the optimization procedure from initial conditions to termination.

9.3. Adaptation

We give results on the performance of the EM algorithm (see Section 8) for computing the exact implied costs (see (3.10)) under changing traffic conditions. Two scenarios are considered. In the first scenario the traffic from Palo Alto (2) to Cambridge (5) is incremented in steps of 1% (of the base traffic) up to 10% for all 6 services. We compute the incremental number of EM iterations required for the L_∞ norm of the vector of relative differences in successive iterates to be less than 10^{-4}. Table 5 gives the results, which shows that at most 3 iterations are required in each instance. This may be calibrated by comparing with the number of iterations of the EM algorithm required to compute all the implied costs starting from arbitrary initial conditions, which is about 23.

Incremental offered traffic	+1%	+1%	+1%	+1%	+1%	+1%	+1%	+1%	+1%	+1%
Incremental EM iterations	2	3	3	3	3	3	3	3	3	2

Table 5: Implied cost calculations for changing traffic from Palo Alto (2) to Cambridge (5).

Table 6 presents results for the second scenario in which traffic from Argonne (3) to

Cambridge (5) in all services is decremented in steps of 1% (of the base traffic) up to 10%.

Incremental offered traffic	−1%	−1%	−1%	−1%	−1%	−1%	−1%	−1%	−1%	−1%
Incremental EM iterations	9	10	9	9	10	10	10	10	10	10

Table 6: Implied cost calculations for changing traffic from Argonne (3) to Cambridge (5).

Note that the number of equations in (3.10) for the implied costs is $SL = 120$. The relatively small numbers of incremental iterations required for solving the equations is noteworthy. We do not have an explanation for the disparity in the numbers in Tables 5 and 6.

10. CONCLUSIONS

The paper has developed an approach to multirate ATM network design and optimization, which unifies nonasymptotic and asymptotic methods of analyses. The asymptotic techniques are aimed at alleviating the numerical complexity of handling links of large bandwidth in the presence of many services with distinct rates, which are intrinsic features of emerging ATM networks.

The directions of current and future research are briefly noted.

(i) Various refinements to the basic asymptotic techniques described here have been obtained. We expect to report on the results shortly.

(ii) There is work in progress in translating the output of the canonical design problem considered in this paper to real-time routing algorithms.

(iii) Similarly, we propose to demonstrate the real-time adaptation of the logical network of virtual paths to changing traffic conditions based on the techniques given here.

(iv) As mentioned in Section 1, higher level canonical design problems have been identified, which incorporate the problem considered here. These complex problems call for further investigation.

Acknowledgment. We are grateful to the anonymous reviewers for several helpful observations.

References

[1] *IEEE J. Selected Areas in Communications*. Special issue on *Advances in the Fundamentals of Networking*, Parts I and II, 13(6) and 13(7). Aug. and Sept. 1995.

[2] G. R. Ash, R. H. Cardwell, and R. P. Murray. Design and optimization of networks with dynamic routing. *Bell Syst. Tech. J.*, 60(8):1787–1820, Oct. 1981.

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

[4] A. Elwalid, D. Mitra, and R. H. Wentworth. A new approach for allocating buffers and bandwidth to heterogeneous, regulated traffic in an ATM node. *IEEE J. Selected Areas in Communications*, 13(6):1115–1127, Aug. 1995.

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

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

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

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

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

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

[11] J. W. Roberts. Teletraffic models for the Telecom 1 integrated services network. ITC-10, Session 1.1, paper #2.

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

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

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

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

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

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

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

[19] 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. Selected Areas in Communications*, 13(7):1199–1206, Sept. 1995.

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

[21] A. Girard and M. T. Ho. Optimization of ISDN networks with partial sharing. In J. Labetoulle and J. W. Roberts, editors, *ITC 14*, pages 1251–1260. Elsevier, 1994.

[22] J.-F. P. Labourdette and G. W. Hart. Blocking probabilities in multirate loss systems: insensitivity, asymptotic behavior and approximations. *IEEE Trans. Commun.*, COM-40:1355–1366, 1992.

[23] R. Siebenhaar. Optimized ATM virtual path bandwidth management under fairness constraints. *IEEE Globecom 94*, pages 924–928, 1994.

[24] L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist.*, 41:141–171, 1970.

[25] A. P. Dempster, N. M. Laird, and D. P. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B*, 39:1–38, 1977.

[26] Y. Vardi and D. Lee. From image deblurring to optimal investments: maximum likelihood solutions for positive linear inverse problems. *J. R. Statist. Soc. B*, 55(3):569–612, 1993.

[27] Y. Vardi. Applications of the EM algorithm to linear inverse problems with positivity constraints. Preprint, 1994.

[28] 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 Modelling and Measurement in Broadband and Mobile Communication*. Leidschendam, The Netherlands, Nov. 1995.

[29] R. J. Gibbens. *Dynamic Routing in Circuit-Switched Networks: The Dynamic Alternative Routing Strategy*. PhD thesis, University of Cambridge, UK, July 1988.

APPENDIX A: Proof of Proposition 3.1

We suppress the link index ℓ, let $n_s (1 \le s \le S)$ be nonnegative integers, and define

$$\mathcal{G}(\mathbf{d}, \boldsymbol{\nu}, C) = \sum_{\sum_{s=1}^{S} d_s n_s \le C} \prod_{s=1}^{S} \frac{\nu_s^{n_s}}{n_s!} \ . \tag{A1}$$

Now [10], $B_s = \mathbb{L}_s(\mathbf{d}, \boldsymbol{\nu}, C) = 1 - \mathcal{G}(\mathbf{d}, \boldsymbol{\nu}, C - d_s)/\mathcal{G}(\mathbf{d}, \boldsymbol{\nu}, C)$. From (A1), $\partial \mathcal{G}(\mathbf{d}, \boldsymbol{\nu}, C)/\partial \nu_t = \mathcal{G}(\mathbf{d}, \boldsymbol{\nu}, C - d_t)$. It follows that

$$\begin{aligned} \frac{\partial B_s}{\partial \nu_t} &= \frac{\mathcal{G}(\mathbf{d}, \boldsymbol{\nu}, C - d_s)\mathcal{G}(\mathbf{d}, \boldsymbol{\nu}, C - d_t)}{[\mathcal{G}(\mathbf{d}, \boldsymbol{\nu}, C)]^2} - \frac{\mathcal{G}(\mathbf{d}, \boldsymbol{\nu}, C - d_s - d_t)}{\mathcal{G}(\mathbf{d}, \boldsymbol{\nu}, C)} \\ &= (1 - B_t)[\mathbb{L}_s(\mathbf{d}, \boldsymbol{\nu}, C - d_t) - \mathbb{L}_s(\mathbf{d}, \boldsymbol{\nu}, C)] \ . \end{aligned} \tag{A2}$$

APPENDIX B: Proof of Proposition 3.2 and (3.13)

To simplify the notation, we define $\rho_{sr} = 0$ if $r \notin \mathcal{R}(s, \sigma) \forall \sigma$. Then,

$$W = \sum_s \sum_r e_{sr} \rho_{sr} (1 - L_{sr}) \ . \tag{B1}$$

We first prove (3.10). For this appendix only, we define

$$\nu_{sk} = \sum_{r:k \in r} \rho_{sr} \prod_{m \in r - \{k\}} (1 - B_{sm}) + \kappa_{sk}, \quad (1 \le s \le S,\ 1 \le k \le L) \ , \tag{B2}$$

where eventually κ_{sk} will be set equal to zero, so that (B2) will be consistent with (3.2). However, we will consider infinitesimal increments $d\boldsymbol{\kappa}_k$ of $\boldsymbol{\kappa}_k = (\kappa_{1k}, \ldots, \kappa_{Sk})$, and the corresponding derivatives of W, as in [18].

We consider a link ℓ, and for $k \neq \ell$ we choose $d\boldsymbol{\kappa}_k$ in terms of $d\boldsymbol{\kappa}_\ell$ so that $d\boldsymbol{\nu}_k = 0$, and hence $dB_{sk} = 0$, $s = 1, \ldots, S$, from (3.1). Then, since $\nu_{s\ell}$ does not depend on $B_{s\ell}$, $d\boldsymbol{\nu}_\ell = d\boldsymbol{\kappa}_\ell$. Hence, for $k \neq \ell$, since $dB_{sk} = 0$, $d\nu_{sk} = 0$ implies, from (B2) and (3.9), that

$$d\kappa_{sk} = \sum_{r:\{k,\ell\} \in r} \frac{\nu_{s\ell;r}}{(1 - B_{sk})} dB_{s\ell}, \quad k \neq \ell \ . \tag{B3}$$

Since $d\boldsymbol{\nu}_\ell = d\boldsymbol{\kappa}_\ell$, $dB_{s\ell} = \sum_t (\partial B_{s\ell}/\partial \nu_{t\ell}) d\kappa_{t\ell}$, and we obtain

$$d\kappa_{sk} = \sum_{r:\{k,\ell\} \in r} \frac{\nu_{s\ell;r}}{(1 - B_{sk})} \sum_t \frac{\partial B_{s\ell}}{\partial \nu_{t\ell}} d\kappa_{t\ell}, \quad k \neq \ell \ . \tag{B4}$$

Next, from (3.6), (3.9) and (B1), we have

$$dW = -\sum_s \sum_{r:\ell \in r} e_{sr} \nu_{s\ell;r} \sum_t \frac{\partial B_{s\ell}}{\partial \nu_{t\ell}} d\kappa_{t\ell} \ . \tag{B5}$$

If we substitute for $d\kappa_{sk}$, $k \neq \ell$, from (B4), we obtain

$$dW = \sum_t \left[\frac{\partial W}{\partial \kappa_{t\ell}} + \sum_s \sum_{k \neq \ell} \sum_{r:\{k,\ell\} \in r} \frac{\nu_{s\ell;r}}{(1 - B_{sk})} \frac{\partial B_{s\ell}}{\partial \nu_{t\ell}} \frac{\partial W}{\partial \kappa_{sk}} \right] d\kappa_{t\ell} \ . \tag{B6}$$

Finally, from (B5) and (B6), it follows that

$$\frac{\partial W}{\partial \kappa_{t\ell}} + \sum_s \sum_{r:\ell \in r} \nu_{s\ell;r} \frac{\partial B_{s\ell}}{\partial \nu_{t\ell}} \left[e_{sr} + \sum_{k \in r - \{\ell\}} \frac{1}{(1 - B_{sk})} \frac{\partial W}{\partial \kappa_{sk}} \right] = 0 \ . \tag{B7}$$

If we define the implied costs as

$$c_{t\ell} = -\frac{1}{(1 - B_{t\ell})} \frac{\partial W}{\partial \kappa_{t\ell}} \ , \tag{B8}$$

we obtain (3.10), since $\nu_{s\ell;r} = 0$ if $\rho_{sr} = 0$, and hence if $r \notin \mathcal{R}(s, \sigma)\ \forall \sigma$.

We next establish (3.8). We now consider an infinitesimal increment $d\rho_{tp}$ of ρ_{tp}. For $k = 1, \ldots, L$, we choose $d\boldsymbol{\kappa}_k$ so that $d\boldsymbol{\nu}_k = 0$, and hence $dB_{sk} = 0$, $s = 1, \ldots, S$. Hence, from (B2), with the help of (3.6), we obtain

$$d\kappa_{sk} = -\delta_{st} I(k \in p) \frac{(1 - L_{tp})}{(1 - B_{tk})} d\rho_{tp} \ . \tag{B9}$$

where $I(\ )$ is the indicator function and $\delta_{st} = 0$ if $s \neq t$, and $\delta_{tt} = 1$. It follows that

$$dW = \frac{\partial W}{\partial \rho_{tp}} d\rho_{tp} - \sum_{k \in p} \frac{\partial W}{\partial \kappa_{tk}} \frac{(1 - L_{tp})}{(1 - B_{tk})} d\rho_{tp} \ . \tag{B10}$$

But, from (3.6) and (B1), since $dB_{sk} = 0 (1 \leq s \leq S, 1 \leq k \leq L)$,

$$dW = e_{tp}(1 - L_{tp}) d\rho_{tp} \ . \tag{B11}$$

Hence,

$$\frac{\partial W}{\partial \rho_{tp}} = (1 - L_{tp}) \left[e_{tp} + \sum_{k \in p} \frac{1}{(1 - B_{tk})} \frac{\partial W}{\partial \kappa_{tk}} \right] \ . \tag{B12}$$

We obtain (3.8) from (B8) and (B12).

Lastly, for (3.13), consider an infinitesimal increment dC_ℓ of C_ℓ, for fixed ℓ, and let

$$d\kappa_{sk} = \frac{\delta_{st} \delta_{k\ell} dC_\ell}{d_{t\ell}(1 - B_{t\ell})}, \quad (1 \leq s \leq S,\ 1 \leq k \leq L) \ . \tag{B13}$$

We show that this is consistent with $dB_{sk} = 0$ $(1 \leq s \leq S, 1 \leq k \leq L)$. Thus, from (B2) and (B13), $d\boldsymbol{\nu}_k = 0$, $k \neq \ell$, which is consistent with $dB_{sk} = 0$, $k \neq \ell$, $1 \leq s \leq S$. Also,

$$d\nu_{s\ell} = \frac{\delta_{st} dC_\ell}{d_{t\ell}(1 - B_{t\ell})}, \quad (1 \leq s \leq S) \ . \tag{B14}$$

Hence, from (3.1), (3.7) and (3.12),

$$\mathbb{L}_s(\mathbf{d}_\ell, \boldsymbol{\nu}_\ell + d\boldsymbol{\nu}_\ell, C_\ell + dC_\ell) - \mathbb{L}_s(\mathbf{d}_\ell, \boldsymbol{\nu}_\ell, C_\ell) = \frac{\partial B_{s\ell}}{\partial \nu_{t\ell}} \frac{dC_\ell}{d_{t\ell}(1 - B_{t\ell})} + \frac{\partial \mathbb{L}_s}{\partial_t C_\ell}(\mathbf{d}_\ell, \boldsymbol{\nu}_\ell, C_\ell) dC_\ell = 0 \ . \tag{B15}$$

This is consistent with $dB_{s\ell} = 0$, $1 \leq s \leq S$. Finally, from (3.6) and (B1),

$$0 = dW = \frac{\partial W}{\partial \kappa_{t\ell}} \frac{dC_\ell}{d_{t\ell}(1 - B_{t\ell})} + \frac{\partial W}{\partial_t C_\ell} dC_\ell \ , \tag{B16}$$

and (3.13) follows from (B8) and (B16).

APPENDIX C: Proof of Proposition 4.3

We establish (4.14) by differentiating the asymptotic approximation (4.12) for B_s with respect to ν_t. First, from (4.4)–(4.6), we obtain $\partial z^*/\partial \nu_t = -d_t(z^*)^{d_t+1}/V(z^*) = O(1/C)$, since $V(z^*) = O(C)$. It follows that $\partial V(z^*)/\partial \nu_t = O(1)$. Also, since $F'(z^*) = 0$,

$$\frac{\partial F(z^*)}{\partial \nu_t} = (z^*)^{d_t} - 1 \ . \tag{C1}$$

It follows, from (4.7)–(4.9), that

$$\frac{\partial b_s}{\partial \nu_t} = O\left(\frac{1}{C}\right) \ , \tag{C2}$$

and the main contributions to $\partial B_s/\partial \nu_t$ arise from the terms involving $\exp[F(z^*)]$ and $\mathrm{Erfc}[\mathrm{sgn}(1 - z^*)\sqrt{-F(z^*)}]$.

The term in (4.9) which multiplies $\exp[F(z^*)]$ is analytic at $z^* = 1$, since $F(1) = 0$, $F'(z^*) = 0$ and $V(z^*) = (z^*)^2 F''(z^*)$. Also, $y = \mathrm{sgn}(1 - z^*)\sqrt{-F(z^*)}$ is analytic at $z^* = 1$, and $y^2 = -F(z^*)$, so that $2y \ \partial y/\partial \nu_t = 1 - (z^*)^{d_t}$, from (C1). Hence, from (4.9) and (4.11),

$$\frac{\partial M}{\partial \nu_t} \sim -\frac{[1 - (z^*)^{d_t}]}{(1 - z^*)} \frac{e^{F(z^*)}}{\sqrt{2\pi V(z^*)}} \ , \tag{C3}$$

and, from (4.8),

$$\frac{\partial B}{\partial \nu_t} \sim \frac{e^{F(z^*)}}{M\sqrt{2\pi V(z^*)}} \left[(z^*)^{d_t} - 1 - \frac{1}{M} \frac{\partial M}{\partial \nu_t}\right] \ . \tag{C4}$$

Finally, from (4.7), (4.8), (4.12) and (C2)–(C4), $\partial B_s/\partial \nu_t \sim B_s[(z^*)^{d_t} - 1 + B_t]$, which establishes (4.14).