ELSEVIER

Performance Evaluation 23 (1995) 65-87

PERFORMANCE EVALUATION
An International Journal

Transient solutions for the buffer behavior in statistical multiplexing ☆

Qiang Ren [a], Hisashi Kobayashi [b,*]

[a] C & C Research Laboratories, NEC USA Inc., Princeton, NJ 08540, USA

[b] Department of Electrical Engineering, School of Engineering and Applied Science, Princeton University, Princeton, NJ 08544, USA

Abstract

In this paper we present time-dependent (or transient) solutions for a mathematical model of statistical multiplexing. The problem is motivated by the need to better understand the performance of fast packet switching in asynchronous transfer mode (ATM), which will be adopted in the broadband ISDN. The transient solutions will be of critical value in understanding dynamic behavior of the multiplexer, and loss probabilities at the cell (or packet) level.

We use the double Laplace transform method, and reduce the partial differential equation that governs the multiplexer behavior to the eigenvalue problem of a matrix equation in the Laplace transform domain. We derive important properties of these eigenvalues, by extending earlier results discussed by Anick, Mitra and Sondhi (1982) for the equilibrium solutions.

A most critical step in our analysis is to identify sets of linear equations that uniquely determine the time-dependent probability distributions at the buffer boundaries. These boundary conditions are in turn used to solve the general transient solutions. For the infinite buffer case, we show that a closed form solution is given in terms of explicitly identified eigenvalues and eigenvectors. When the buffer capacity is finite, the determination of boundary conditions requires us to solve a matrix equation.

We also observe that the statistical multiplexing not only achieves the effective bandwidth gain (i.e., a multiplexing gain), but also reduces the system's packet loss probability and shorten transient periods.

We present some numerical results to illustrate our solution technique. A potential application of the time-dependent solution is in the area of preventive congestion control in a high speed network.

Keywords: Transient analysis; Buffer behavior; Statistical multiplexing; ATM; Fluid flow model

☆ An earlier version of this paper was presented at the 26th Annual Conference on Information Sciences and Systems, Princeton University, Princeton, N.J., March 18-20, 1992 (see [21]).

* Corresponding author.

SSDI 0166-5316(93)E0064-C

1. Introduction

There have been a number of studies reported on queueing theoretic models of statistical multiplexing. As for earlier results, the reader is referred, for example, to [8,10]. The fluid flow model for buffer behavior analysis, which can be viewed as a generalization of the birth-and-death process model, has been discussed by Hashida and Fujiki [7], Cohen [4], Kosten [15,16], Anick et al. [2], Mitra [18], Morrison [19], Kobayashi [12], Elwalid et al. [5], Coffman et al. [3], and others. This class of probabilistic models, often referred to as *data handling systems* in deference of Kosten's series of papers on the subject, is recognized increasingly important, because it provides a practical mathematical framework to analyze the buffer behavior of statistical multiplexing, which is the basic principle of fast packet switching or ATM (asynchronous transfer mode) switching in the B-ISDN (broadband integrated services digital network) environment.

All of the above studies, however, have exclusively dealt with the steady-state solutions of this fluid model, whereas the present work is, to the best of the authors' knowledge, a first result on time-dependent solutions. Simonian and Virtamo [22] obtained the transient distribution of queue size with continuous input process by using Benes' method, but their transient solution is given in an implicit form. Zhang [25], independent of our work, formulated a time-dependent solution for the statistical multiplexing, where he assumes a continuous flow Markov modulated source model, and discusses general properties of the solution for an infinite buffer case.

The dynamics of the statistical multiplexer are characterized by a set of linear partial differential equations. Although the initial condition is clearly known, the transient behavior at the buffer boundaries are unknown functions of time. Anick et al. [2] and Mitra [18] presented elegant ways to treat the boundary conditions for the steady-state case solutions. By generalizing their approach, we develop a procedure to determine the time-dependent boundary functions.

2. The model of mathematical analysis: fluid flow model

We assume that there are N statistically independent and identical sources and each source alternates between the *on* (or *burst*) state and the *off* (or *silence*) state. We also assume that successive *burst* and *silence* periods of a source form an alternating renewal process, and their durations are exponentially distributed with mean β^{-1} and α^{-1} respectively. If a source is in the on-state, it will generate packets (or *cells* in the ATM terminology) and its rate is assumed, without loss of generality, to be one packet per unit time. While a source is in off-state, it generates no packets.

Let C [packets/unit time] denote the multiplexer's output *link capacity*, and $Q(t)$ the total amount of packets outstanding at the multiplexer output link at time t. Let $J(t)$ denote the number of sources in the on-state at time t (see Fig. 1). Then, while $J(t) < C$, all arriving packets are transmitted immediately over the output link, thus there will be no packets left in the output link buffer. When $J(t)$ exceeds C, a queue will be built up at the rate of $J(t) - C$. Although $Q(t)$ is, strictly speaking, an integer-valued function, we approximate it by a time-continuous function, assuming that a series of packets arrive in a continuous stream of bits. In other words, we represent the stream of packets as a fluid flow. This should indeed be an excellent approximation, when $Q(t)$ represents the unprocessed amount of data residing in the output buffer of fast packet switching in an ATM network, in which packets or cells are on the order of several hundred bits: the cell transmission time will be of the order

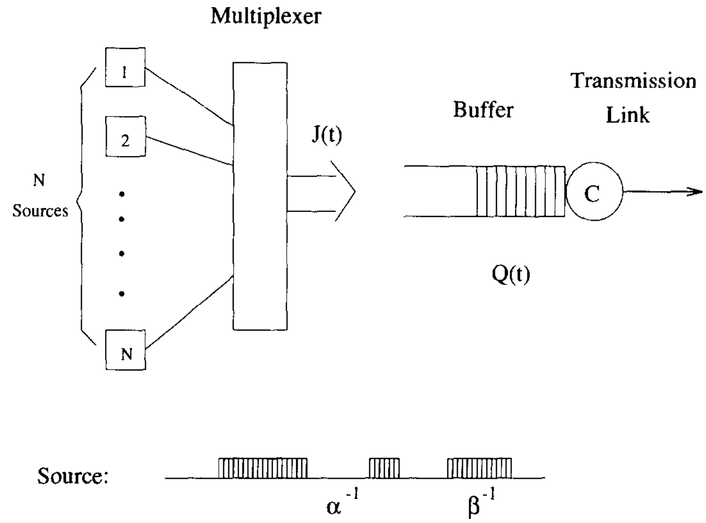

Fig. 1. The statistical multiplexer and the source model.

of a few microseconds or less over the link capacity of a B-ISDN network, and will be negligibly small compared with a typical burst period of the source.

Then we can write

$$\frac{\mathrm{d}Q(t)}{\mathrm{d}t} = \begin{cases} J(t) - C, & \text{if } Q(t) > 0 \text{ or } J(t) > C, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

A typical behavior of the random process $J(t)$ and the corresponding $Q(t)$ are depicted in Fig. 2. Thus while $Q(t) > 0$, it is the integral of the process $J(t) - C$:

$$Q(t) = \int_{t_0}^{t} J(u)\,\mathrm{d}u - C(t - t_0), \tag{2}$$

where t_0 is the most recent moment when $Q(t) = 0$.

We assume here that the buffer capacity is either infinite or finite with the upper limit X, and that in the infinite buffer case the following stability condition is satisfied:

$$\rho \stackrel{\text{def}}{=} \frac{N\alpha}{C(\alpha + \beta)} < 1, \tag{3}$$

where ρ is the traffic intensity.

The probabilistic behavior of $J(t)$ can be characterized by a well-studied birth-and-death process model, in which the birth rate and the death rate are set as $\lambda(j) = (N - j)\alpha$ and $\mu(j) = j\beta$, respectively, where j is the system state that $J(t)$ can take on, $0 \le j \le N$.

In order to analyze the behavior of the buffer content $Q(t)$, we need to consider the *pair process* $\{J(t), Q(t)\}$ and thus we define

$$P_j(t, x) \stackrel{\text{def}}{=} \Pr[J(t) = j, Q(t) \le x], \quad 0 \le j \le N,\ t \ge 0,\ x \ge 0. \tag{4}$$

The transient behavior of the pair process $\{J(t), Q(t)\}$ depends on its initial value $\{J(0), Q(0)\}$, but we omit this in the notation.

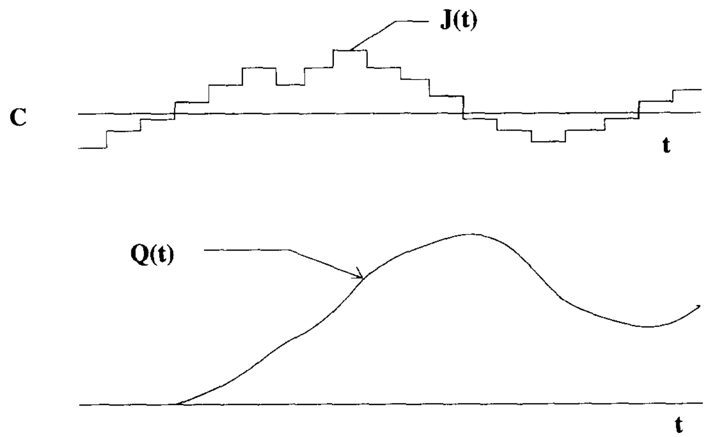

Fig. 2. The number of "on" sources $J(t)$ and the number of queued packets $Q(t)$.

By extending the birth-and-death process model equation, we can derive the following set of equations that the joint probability function $P_j(t,x)$ must satisfy:

$$\frac{\partial P_j(t,x)}{\partial t} + (j-C)\frac{\partial P_j(t,x)}{\partial x} = -[\lambda(j)+\mu(j)]P_j(t,x) + \lambda(j-1)P_{j-1}(t,x) + \mu(j+1)P_{j+1}(t,x) \quad \text{for } 0 \le j \le N,\ x>0, \tag{5}$$

with

$$P_{-1}(t,x) = P_{N+1}(t,x) = 0 \quad \text{for all } t \text{ and } x \tag{6}$$

where in the multiplexer problem, the birth and death rates are given by

$$\lambda(j) = (N-j)\alpha, \qquad \mu(j) = j\beta\,. \tag{7}$$

If we write Eq. (5) in matrix form, it becomes

$$\frac{\partial \boldsymbol{P}(t,x)}{\partial t} + \mathcal{D}\frac{\partial \boldsymbol{P}(t,x)}{\partial x} = \mathcal{M}\boldsymbol{P}(t,x), \tag{8}$$

where

$$\boldsymbol{P}(t,x) = \begin{bmatrix} P_0(t,x) \\ P_1(t,x) \\ \vdots \\ P_N(t,x) \end{bmatrix}, \tag{9}$$

and $\mathcal{D}$ is a $(N+1)\times(N+1)$ diagonal matrix defined by

$$\mathcal{D} = \mathrm{diag}[-C, 1-C, \ldots, j-C, \ldots, N-C]\,.$$

We assume that the *link capacity* C is a non-integer number between 0 and N, and satisfies the stability condition Eq. (3) for the infinite buffer case. The matrix $\mathcal{M}$, an $(N+1)\times(N+1)$ tridiagonal matrix, is the infinitesimal generator of the Markov process $J(t)$:

$$\mathcal{M} = \begin{bmatrix} -N\alpha & \beta & & & & \\ N\alpha & -(N-1)\alpha-\beta & 2\beta & & & \\ & (N-1)\alpha & -(N-2)\alpha-2\beta & \ddots & & \\ & & \vdots & & & \\ & & & & -\alpha-(N-1)\beta & N\beta \\ & & & & \alpha & -N\beta \end{bmatrix}.$$

It will be instructive to remark that the limiting case of Eq. (8), as $x \to +\infty$, corresponds to the differential equation that governs the birth-and-death process [24]. Similarly, the equilibrium state of Eq. (8) where $t \to +\infty$ corresponds to the steady-state of statistical multiplexing, which was thoroughly analyzed by Anick et al. [2]:

(1) The time-dependent birth-and-death process model, i.e., $x = +\infty$, [24],

$$\frac{\mathrm{d}\boldsymbol{P}(t,+\infty)}{\mathrm{d}t} = \mathcal{M}\boldsymbol{P}(t,+\infty). \tag{10}$$

(2) The steady-state statistical multiplexer model, i.e., $t = +\infty$, [2],

$$\mathcal{D}\frac{\mathrm{d}\boldsymbol{P}(+\infty,x)}{\mathrm{d}x} = \mathcal{M}\boldsymbol{P}(+\infty,x). \tag{11}$$

3. Transient analysis

3.1. Double Laplace transforms

Let us first apply the Laplace transform to $\boldsymbol{P}(t,x)$ with respect to the time variable t:

$$\boldsymbol{P}^*(s,x) = \begin{bmatrix} P_0^*(s,x) \\ P_1^*(s,x) \\ \vdots \\ P_N^*(s,x) \end{bmatrix} = \mathcal{L}_t\{\boldsymbol{P}(t,x)\} = \int_0^{+\infty} \mathrm{e}^{-st}\boldsymbol{P}(t,x)\,\mathrm{d}t. \tag{12}$$

Then Eq. (8) can be transformed into

$$\mathcal{D}\frac{\mathrm{d}\boldsymbol{P}^*(s,x)}{\mathrm{d}x} = (\mathcal{M}-sI)\boldsymbol{P}^*(s,x) + \boldsymbol{P}(0,x), \tag{13}$$

where I is the $(N+1) \times (N+1)$ identity matrix.

We then take the Laplace transforms of $\boldsymbol{P}(t,x)$ and $\boldsymbol{P}^*(s,x)$ with respect to the buffer content variable x:

$$\boldsymbol{P}^*(t,u) = \mathcal{L}_x\{\boldsymbol{P}(t,x)\} = \int_0^{+\infty} \mathrm{e}^{-ux}\boldsymbol{P}(t,x)\,\mathrm{d}x, \tag{14}$$

$$\boldsymbol{P}^{**}(s,u) = \mathcal{L}_x\{\boldsymbol{P}^*(s,x)\} = \int_0^{+\infty} \mathrm{e}^{-ux}\boldsymbol{P}^*(s,x)\,\mathrm{d}x. \tag{15}$$

Then Eq. (13) will be transformed into the following matrix equation:

$$(u\mathcal{D} + sI - \mathcal{M})\boldsymbol{P}^{**}(s,u) = \boldsymbol{P}^{*}(0,u) + \mathcal{D}\boldsymbol{P}^{*}(s,0). \tag{16}$$

The boundary value $\boldsymbol{P}^*(s,0)$, which characterizes the time-dependent behavior of the empty buffer case $(x = 0)$, is unknown here, but will be determined later. If we assume that j_0 $(0 \le j_0 \le N)$ sources are on and the buffer content $Q(t) = x_0$ [1] at $t = 0$, the initial condition $\boldsymbol{P}^*(0,u)$ is given as

$$\boldsymbol{P}^*(0,u) = \mathcal{L}_x\{\boldsymbol{P}(0,x)\} = \frac{\mathrm{e}^{-ux_0}}{u}\boldsymbol{e}_{j_0}, \tag{17}$$

where $\boldsymbol{e}_{j_0}$ is a unit vector with its (j_0+1)th entry being 1 and all other components being zero.

Then Eq. (16) leads to

$$\begin{aligned}\boldsymbol{P}^{**}(s,u) &= [u\mathcal{D} + sI - \mathcal{M}]^{-1}[\boldsymbol{P}^*(0,u) + \mathcal{D}\boldsymbol{P}^*(s,0)] \\ &= \frac{\mathcal{A}(s,u)}{C(s,u)}\boldsymbol{B}(s,u),\end{aligned} \tag{18}$$

where

$$\mathcal{A}(s,u) = \text{Adjugate matrix of } [u\mathcal{D} + sI - \mathcal{M}], \tag{19}$$

$$\boldsymbol{B}(s,u) = \boldsymbol{P}^*(0,u) + \mathcal{D}\boldsymbol{P}^*(s,0), \tag{20}$$

$$C(s,u) = \det[u\mathcal{D} + sI - \mathcal{M}]. \tag{21}$$

Let $u_0(s), u_1(s), \ldots, u_N(s)$ be the $N+1$ roots that can be obtained by solving, with respect to u, the characteristic equation

$$C(s,u) = \det[u\mathcal{D} + sI - \mathcal{M}] = 0. \tag{22}$$

Note that $\boldsymbol{B}(s,u)$ contains a factor u in its denominator, as given by Eq. (17). Thus the denominator of Eq. (18) is an $(N+2)$-degree polynomial of u: the first $(N+1)$ roots are the characteristic roots defined above, and the $(N+2)$th root is

$$u_{N+1}(s) = 0. \tag{23}$$

We can show (see Appendix A) that these $(N+2)$ roots are all distinct.

3.2. Properties of eigenvalues and eigenvectors

The problem of finding the roots of the characteristic equation (22) can be reduced to the one of finding eigenvalues (or *spectrums*) of the following matrix equation:

$$u\mathcal{D}V(s) = [\mathcal{M} - sI]V(s). \tag{24}$$

The Laplacian variable u introduced in Eq. (14) is now interpreted as an eigenvalue of the matrix

[1] Note that the following arguments hold for both the infinite buffer case and a finite buffer case. If the buffer capacity is finite with its upper limit X, then obviously $Q(t) \le X$ and $x_0 \le X$. (See Section 3.3.2 for detailed discussions on the finite buffer case.)

$\mathcal{D}^{-1}[\mathcal{M} - sI]$, and $V(s)$ is the associated right eigenvector, whose $(j+1)$th element is denoted by $V_j(s)$, $0 \le j \le N$.

Then the generating function, defined by

$$V(z,s) = \sum_{j=0}^{N} V_j(s) z^j, \tag{25}$$

satisfies

$$\begin{aligned}\frac{\partial}{\partial z} \ln V(z,s) &= \frac{uC + N\alpha(z-1) - s}{\alpha z^2 + (u+\beta-\alpha)z - \beta} \\ &= \frac{K}{z - z_1} + \frac{N-K}{z - z_2}.\end{aligned} \tag{26}$$

The last equation is obtained by substituting Eq. (25) into Eq. (24). Equation (26) readily leads to the following solution (except for a scaling constant):

$$V(z,s) = (z - z_1)^K (z - z_2)^{N-K}, \tag{27}$$

where

$$z_1, z_2 = \frac{-(u+\beta-\alpha) \mp \sqrt{(u+\beta-\alpha)^2 + 4\alpha\beta}}{2\alpha} \tag{28}$$

and

$$K = \frac{uC + N\alpha(z_1 - 1) - s}{\alpha(z_1 - z_2)}. \tag{29}$$

Observe that by its definition of Eq. (25), $V(z,s)$ is a polynomial in z of degree N. Since z_1 and z_2 are distinct, this is possible if and only if K is an integer $\in [0,N]$.

From Eqs. (28) and (29), we obtain the following quadratic equation:

$$a(K,s)u^2 + b(K,s)u + c(K,s) = 0, \tag{30}$$

where

$$\begin{aligned}a(K,s) &= (\tfrac{1}{2}N - K)^2 - (\tfrac{1}{2}N - C)^2, \\ b(K,s) &= 2(\beta-\alpha)(\tfrac{1}{2}N - K)^2 - 2(\tfrac{1}{2}N - C)[s + \tfrac{1}{2}N(\alpha+\beta)], \\ c(K,s) &= (\alpha+\beta)^2(\tfrac{1}{2}N - K)^2 - [\tfrac{1}{2}N(\alpha+\beta) + s]^2.\end{aligned}$$

Then for a given integer $K \in [0,N]$, the above quadratic equation yields, after some algebraic manipulation, the following root:

$$u(K,s) = \alpha - \beta + \frac{(\frac{1}{2}N - C)[s + (N-C)\alpha + C\beta] + (\frac{1}{2}N - K)\sqrt{\Delta(K,s)}}{(\frac{1}{2}N - K)^2 - (\frac{1}{2}N - C)^2}, \tag{31}$$

where

$$\triangle(K,s)=[s+(N-C)\alpha+C\beta]^2+4\alpha\beta(K-C)(N-K-C). \tag{32}$$

We now obtain the following theorem, which is a generalization of the main theorem in [2].

Theorem 1.
(1) As K varies from 0 to N, Eq. (31) produces a total of $N+1$ non-zero, distinct roots $\{u(K,s); K=0,1,\dots,N\}$.
(2) Among the above $N+1$ distinct roots above:
- *There are $N-\lfloor C\rfloor$* negative *roots*[2]*, which we relabel and denote $\{u_k(s);\ 1\le k\le N-\lfloor C\rfloor\}$, i.e.,* $\mathrm{Re}\{u_k(s)\}<0$ *for all* $\mathrm{Re}\{s\}\ge 0$*. Furthermore, each $u_k(s)$ is a* strictly decreasing *function of $s\ge 0$.*
- *There are $\lfloor C\rfloor$* positive *roots, denoted by $\{u_k(s);\ N-\lfloor C\rfloor+1\le k\le N\}$, i.e.,* $\mathrm{Re}\{u_k(s)\}>0$ *for all* $\mathrm{Re}\{s\}\ge 0$*. Furthermore, each $u_k(s)$ is a* strictly increasing *function of all $s\ge 0$.*
- *$u_0(s)$, the remaining root which corresponds to $u(0,s)$, is a* strictly increasing *function of $s\ge 0$ and is strictly positive, i.e.,* $\mathrm{Re}\{u_0(s)\}>0$ *for all* $\mathrm{Re}\{s\}>0$*, but $u_0(0)=0$.*

The proof of the theorem is given in Appendix A.

For each $u_k(s)$, the corresponding right eigenvector, denoted by $\boldsymbol{V}_k(s)$, can be obtained from its generating function Eq. (27). The corresponding left eigenvector, denoted by $\boldsymbol{U}_k(s)$ and defined by $u_k(s)\boldsymbol{U}_k'(s)\mathcal{D}=\boldsymbol{U}_k'(s)(\mathcal{M}-sI)$, is given by

$$\boldsymbol{U}_k(s) = (\mathcal{Q}^{-1})^2\boldsymbol{V}_k(s), \tag{33}$$

where $\mathcal{Q}$, denoted as τ in [2], is a diagonal matrix that transforms the tridiagonal matrix $\mathcal{M}$ into a symmetric matrix, and the $(j+1)$th element of $\mathcal{Q}$ is given by $\sqrt{(\alpha/\beta)^j\binom{N}{j}}$.

We normalize these eigenvectors with respect to $\boldsymbol{U}_k(s)$, which are orthogonal to each other via $\mathcal{D}$, so that [3]

$$\boldsymbol{U}_k'(s)\mathcal{D}\boldsymbol{V}_l(s) = \delta_{kl}, \tag{34}$$

where δ_{kl} is Kronecker delta.

We find (see [6]) that the inverse matrix $[u\mathcal{D}+sI-\mathcal{M}]^{-1}$ in Eq. (18) has the following *spectral expansion*:

$$[u\mathcal{D}+sI-\mathcal{M}]^{-1} = \sum_{k=0}^{N}\frac{\boldsymbol{V}_k(s)\boldsymbol{U}_k'(s)}{u-u_k(s)}. \tag{35}$$

Equation (18) can then be equivalently written as

$$\boldsymbol{P}^{**}(s,u) = \sum_{k=0}^{N}\frac{\boldsymbol{V}_k(s)\boldsymbol{U}_k'(s)}{u-u_k(s)}[\boldsymbol{P}^*(0,u)+\mathcal{D}\boldsymbol{P}^*(s,0)]. \tag{36}$$

By taking the inverse Laplace transform of $\boldsymbol{P}^{**}(s,u)$ with respect to Laplacian variable u, we obtain

[2] We let $\lfloor C\rfloor$ denote the integer part of C.

[3] Equations (28)–(30) in our earlier paper [21] are incorrect and should be replaced by Eqs.(39)–(41) given here.

$$P^*(s,x) = -\sum_{k=0}^{N} \frac{V_k(s)U_k'(s)}{u_k(s)} e_{j_0}$$
$$+\sum_{k=0}^{N} U_k'(s)[P^*(0,u_k(s))U(x-x_0) + \mathcal{D}P^*(s,0)]V_k(s)e^{u_k(s)x}, \quad (37)$$

where $U(\cdot)$ is the unit step function.

Note that all the terms in Eq. (37) are known except for $P^*(s,0)$ which is the boundary value of $P^*(s,x)$ at $x=0$. We will focus on the derivation of $P^*(s,0)$ in the following section.

3.3. Transient boundary conditions

3.3.1. Infinite buffer case: $x = 0$

If we assume the buffer capacity to be infinite, the unknown boundary condition $P^*(s,0)$ or $\mathcal{L}_t\{P(t,0)\}$ characterizes the transient behavior of the empty buffer. We now solve $P^*(s,0)$ by using conditions required for the stable solution.

From Eq. (37) and Theorem 1, we find that the solution is stable, if

$$U_k'(s)[P^*(0,u_k(s)) + \mathcal{D}P^*(s,0)] = 0 \quad \text{for } k=0 \text{ and } N-\lfloor C\rfloor+1 \le k \le N. \quad (38)$$

Now we make the following observation regarding the empty buffer, by generalizing the properties discussed by Kosten [15] and Anick et al. [2].

When the incoming traffic is greater than the network link capacity at any given time t, i.e., $J(t) > C$, the buffer content necessarily increases and the buffer cannot stay empty. It follows that

$$P_j(t,0) = 0 \quad \text{for } \lfloor C\rfloor+1 \le j \le N, \text{ and } t \ge 0. \quad (39)$$

The last equation implies

$$P_j^*(s,0) = 0, \quad \text{for } \lfloor C\rfloor+1 \le j \le N. \quad (40)$$

Now we can write

$$P^*(s,0) = [P_0^*(s,0), P_1^*(s,0), \ldots, P_{\lfloor C\rfloor}^*(s,0), 0, \ldots, 0]'. \quad (41)$$

A total of $\lfloor C\rfloor+1$ unknown elements $P_0^*(s,0), P_1^*(s,0), \ldots, P_{\lfloor C\rfloor}^*(s,0)$ can now be determined by the $\lfloor C\rfloor+1$ linear constraint equations given in Eq. (38).

After we determine $P^*(s,0)$ by solving the matrix equation of dimension $\lfloor C\rfloor+1$, we can then obtain the final solution Eq. (37). But as we shall discuss in Section 4, we find computationally convenient formulae to obtain the time-dependent solution.

3.3.2. Finite buffer case: $x = 0$ and $x = X$

We now consider the case where the buffer is finite with its upper limit X. By generalizing observations made for the steady-state case by Hashida and Fujiki [7] and Mitra [18], we have the following two sets of linear constraint equations on the buffer boundary conditions, i.e., $P^*(s,0)$ and $P^*(s,X)$.

(1) For $x = 0$: When the incoming traffic rate is greater than the network link capacity at any time t, i.e., $J(t) > C$, the buffer cannot stay empty. Thus it follows that

$$P_j^*(s,0) = 0 \quad \text{for } \lfloor C \rfloor + 1 \le j \le N. \tag{42}$$

(2) For $x = X$: When the incoming traffic rate is less than the network link capacity at any time t, i.e., $J(t) < C$, the buffer cannot stay at its upper limit. Thus it follows that

$$P_j^*(s,X) = P_j^*(s,+\infty) \quad \text{for } 0 \le j \le \lfloor C \rfloor. \tag{43}$$

$P_j^*(s,+\infty)$ here is the Laplace transform of $P_j(t,+\infty)$, the probability that j sources are on at time t. It can be easily obtained in closed form from the time-dependent solution for the birth-and-death process model, which we referred to by Eq. (10). In fact, $P_j^*(s,+\infty)$ is the $(j+1)$th entry of $-\sum_{k=0}^{N}(\boldsymbol{V}_k(s)\boldsymbol{U}_k'(s)/u_k(s))\boldsymbol{e}_{j_0}$ in Eq. (37).

A total of $\lfloor C \rfloor + 1$ unknown elements $P_0^*(s,0), P_1^*(s,0), \ldots, P_{\lfloor C \rfloor}^*(s,0)$ can be now uniquely determined by $\lfloor C \rfloor + 1$ linear constraint equations given at the upper boundary $x = X$, i.e., Eq. (43). This in turn gives the complete solution which is represented by Eq. (37).

4. Transient solutions

4.1. *Infinite buffer case: closed form solution*

Although the solution in Eq. (37) requires us to solve an $(\lfloor C \rfloor + 1)$-dimensional matrix equation as stated in Section 3.3.1, we now derive a closed form solution by extending [2].

From Eq. (37), we can see that $\boldsymbol{P}^*(s,x)$ can be represented as

$$\begin{aligned}\boldsymbol{P}^*(s,x) = \{ &\boldsymbol{H}_{N+1}(s) + \sum_{k=1}^{N-\lfloor C \rfloor} [a_k(s) + b_k(s)]\boldsymbol{V}_k(s)\mathrm{e}^{u_k(s)x} \} U(x - x_0) \\ &+ \sum_{k=0}^{N} a_k(s)\boldsymbol{V}_k(s)\mathrm{e}^{u_k(s)x}[U(x) - U(x - x_0)],\end{aligned} \tag{44}$$

where

$$\boldsymbol{H}_{N+1}(s) = [h_{N+1,0}(s), \ldots, h_{N+1,N}(s)]' = -\sum_{k=0}^{N} \frac{\boldsymbol{V}_k(s)\boldsymbol{U}_k'(s)}{u_k(s)}\boldsymbol{e}_{j_0}, \tag{45}$$

$$a_k(s) = \boldsymbol{U}_k'(s)\mathcal{D}\boldsymbol{P}^*(s,0), \quad 0 \le k \le N, \tag{46}$$

$$b_k(s) = \frac{\boldsymbol{U}_k'(s)\boldsymbol{e}_{j_0}}{u_k(s)}\mathrm{e}^{-u_k(s)x_0}, \quad 0 \le k \le N. \tag{47}$$

Note that from Eq. (38), the stability condition, we find

$$a_k(s) = -b_k(s) \quad \text{for } k = 0 \text{ and } N - \lfloor C \rfloor + 1 \le k \le N. \tag{48}$$

Thus only $\{a_k(s); 1 \le k \le N - \lfloor C \rfloor\}$ are left as the unknowns to be determined.

Define

$$\underline{a}(s) = [a_1(s), \ldots, a_{N-\lfloor c \rfloor}(s)]', \tag{49}$$

$$\underline{b}(s) = [b_0(s), b_{N-\lfloor C \rfloor+1}(s), \ldots, b_N(s)]' \tag{50}$$

and observe the last entry of the lth derivative of $\boldsymbol{P}^*(s,x)$ at $x=0$

$$\frac{\mathrm{d}^{(l)} \boldsymbol{P}^*(s,x)}{\mathrm{d}x^{(l)}}\Big|_{x=0}, \tag{51}$$

for $l = 0, \ldots, N - \lfloor C \rfloor - 1$. From Eqs.(13) and (44) (see Appendix B), we have

$$\mathcal{T}\underline{a}(s) = \mathcal{T}_1\underline{b}(s) \overset{\text{def}}{=} \underline{c}(s), \tag{52}$$

where

$$\mathcal{T} = \begin{bmatrix} 1 & 1 & \cdots 1 \\ u_1(s) & u_2(s) & \cdots u_{N-\lfloor C \rfloor}(s) \\ \vdots & \vdots & \vdots \\ (u_1(s))^{N-\lfloor C \rfloor-1} & (u_2(s))^{N-\lfloor C \rfloor-1} & \cdots (u_{N-\lfloor C \rfloor}(s))^{N-\lfloor C \rfloor-1} \end{bmatrix} \tag{53}$$

and

$$\mathcal{T}_1 = \begin{bmatrix} 1 & 1 & \cdots 1 \\ u_0(s) & u_{N-\lfloor C \rfloor+1}(s) & \cdots u_N(s) \\ \vdots & \vdots & \vdots \\ (u_0(s))^{N-\lfloor C \rfloor-1} & (u_{N-\lfloor C \rfloor+1}(s))^{N-\lfloor C \rfloor-1} & \cdots (u_N(s))^{N-\lfloor C \rfloor-1} \end{bmatrix}. \tag{54}$$

Note that $\mathcal{T}$ is a *Vandermonde* matrix, the inverse matrix $\mathcal{T}^{-1}$ can be represented as (see [9])

$$\mathcal{T}^{-1} = \begin{bmatrix} L_1(y) & L_1^{(1)}(y) & \cdots L_1^{(N-\lfloor C \rfloor-1)}(y) \\ L_2(y) & L_2^{(1)}(y) & \cdots L_2^{(N-\lfloor C \rfloor-1)}(y) \\ \vdots & \vdots & \vdots \\ L_{N-\lfloor C \rfloor}(y) & L_{N-\lfloor C \rfloor}^{(1)}(y) & \cdots L_{N-\lfloor C \rfloor}^{(N-\lfloor C \rfloor-1)}(y) \end{bmatrix}_{y=0} \tag{55}$$

where $L_k^{(l)}(y) = \mathrm{d}^{(l)} L_k(y)/\mathrm{d}y^{(l)}$, $l = 1, \ldots, N - \lfloor C \rfloor - 1$, $k = 1, \ldots, N - \lfloor C \rfloor$. And

$$L_k(y) = \frac{\prod_{j=1, j\neq k}^{N-\lfloor C \rfloor}(y - u_j(s))}{\prod_{j=1, j\neq k}^{N-\lfloor C \rfloor}(u_k(s) - u_j(s))}, \quad k = 1, \ldots, N - \lfloor C \rfloor \tag{56}$$

are *Lagrange interpolating polynomials.*

Thus the $N - \lfloor C \rfloor$ unknowns of Eq. (49), can be uniquely determined by the following closed form expression

$$\underline{a}(s) = \mathcal{T}^{-1}\underline{c}(s) = \mathcal{T}^{-1}\mathcal{T}_1\underline{b}(s). \tag{57}$$

For a special case where $x_0 = 0$ and $j_0 < C$, we have from Eqs. (44), (55) and (57)

$$a_k(s) = -b_k(s) - h_{N+1,N}(s) \prod_{j=1, j\neq k}^{N-\lfloor C\rfloor} \frac{u_j(s)}{u_j(s) - u_k(s)}, \tag{58}$$

for $1 \leq k \leq N - \lfloor C \rfloor$.

4.2. *Finite buffer case*

For the finite buffer case, the *positive* eigenvalues, i.e., $\{u_k(s); k = 0, N - \lfloor C\rfloor + 1, \ldots, N\}$ can be allowed. Thus our transient solution is given from Eq. (36) as

$$\boldsymbol{P}^*(s,x) = \boldsymbol{H}_{N+1}(s)U(x-x_0) + \sum_{k=0}^{N} [\, a_k(s) + b_k(s)U(x-x_0)\,]\, \boldsymbol{V}_k(s)\mathrm{e}^{u_k(s)x}U(x), \tag{59}$$

where $\boldsymbol{H}_{N+1}(s), a_k(s)$ and $b_k(s)$ are same as those defined in Eqs. (45), (46) and (47).

Note that the two sets of equations (42) and (43) give exactly $(N+1)$ linear constraint equations on $\{a_k(s); 0 \leq k \leq N\}$. Thus, the $N+1$ unknown $a_k(s)$'s of Eq. (46) can be uniquely determined by solving a matrix equation of dimension $(N+1)$.

4.3. *Computational complexity*

We have obtained the transient solutions of Eq. (8) for both the infinite and finite buffer cases. Our final solution $\boldsymbol{P}^*(s,x)$ is given in the form of Laplace transform with respect to the time domain. A numerical-inversion method of the Laplace transform must be applied to obtain $\boldsymbol{P}(t,x) = \mathcal{L}_s^{-1}\{\boldsymbol{P}^*(s,x)\}$.

For the infinite buffer case, a closed form solution is given by Eqs. (44)–(57). The complexity of main computations (i.e., solving Eq. (57)) is at most of the order of N^2 when C is comparable with N. This is a significant saving compared with the computation method suggested in Section 3.3.1, which is of the order of N^3 by standard Gaussian elimination. The latter case has an advantage only when $C << N$, because its actual computational complexity is of the order of $(C+1)^3$.

For the finite buffer case, the solution is also represented in a closed form, i.e., Eq. (59), except that a matrix equation needs to be solved as given by Eq. (48) in Section 3.3.2. The computational complexity is of the order of $(C+1)^3$, if we apply the standard Gaussian elimination method to solve that matrix equation.

5. Asymptotics and time scales of transient periods

In this section we examine the behavior of the buffer overflow probability $G(t,x) \stackrel{\text{def}}{=} \Pr[Q(t) > x]$ for a *large* value of x. This quantity is often called the *tail-end distribution*.

From Eq. (44) and the *Theorem*, we can see that $G^*(s,x)$ $(= \mathcal{L}_t\{G(t,x)\})$ will be dominated by the term with the largest negative exponent $u_k(s)$. This dominant root is given by setting $K = N$ in $u(K,s)$ of Eq. (31) and is denoted here $u_{dom}(s)$. Hence,

$$G^*(s,x) \approx g_N(s) e^{u_{dom}(s)x} \tag{60}$$

where

$$g_N(s) = -(a_N(s) + b_N(s))[1 - z_1(u_{dom}(s))]^N. \tag{61}$$

In the steady state, we have

$$\begin{aligned} G(+\infty, x) &\approx [s \cdot g_N(s)]_{s=0} e^{u_{dom}(0)x} \\ &= (\frac{\alpha}{\alpha+\beta})^N \prod_{j=N-\lfloor C \rfloor+1}^{N-1} \frac{u_j(0)}{u_j(0) - u_{dom}(0)} [1 - z_1(u_{dom}(0))]^N e^{u_{dom}(0)x}, \end{aligned} \tag{62}$$

with

$$u_{dom}(0) = -\frac{(\alpha+\beta)(1-\rho)}{1 - C/N} \tag{63}$$

It is instructive to note that the dominant exponent $u_{dom}(0)$ remains unchanged, as we change N, the degree of multiplexing insofar as α, β and ρ are kept constant, hence the ratio C/N stays constant. This means that the asymptotic decay of the overflow probability, i.e., $G(+\infty, x)$ as $x \to +\infty$, is independent of N, as long as the traffic intensity remains unchanged.

Now we examine how the convergence time scale of $G(t,x)$ varies as the total number of sources changes, that is, how fast the system will reach its steady state under the same traffic intensity but for different values of N.

Denote by $u_{dom}^{(1)}(s)$ the dominant eigenvalue when $N = 1$ (single source case) and by $u_{dom}^{(M)}(s)$ the dominant eigenvalue when $N = M$. Then we have from Eq. (31)

$$u_{dom}^{(M)}(s) = u_{dom}^{(1)}(\frac{s}{M}). \tag{64}$$

For large values of x, the exponential term in Eq. (60) dominates the transient behavior of $G(t,x)$. Then we obtain the following approximate relationship between $G^{(M)}(t,x)$ and $G^{(1)}(t,x)$

$$\begin{aligned} G^{(M)}(t,x) &\sim \mathcal{L}_s^{-1}\{g^{(M)}(s) e^{u_{dom}^{(M)}(s)x}\} \\ &\sim \mathcal{L}_s^{-1}\{g^{(1)}(\frac{s}{M}) e^{u_{dom}^{(1)}(s/M)x}\} \\ &\sim G^{(1)}(Mt, x). \end{aligned} \tag{65}$$

From (65) we can make the following important observation: the total traffic load fixed, the more sources (hence the less traffic per source) are multiplexed, the faster the system will reach its steady state, and its convergence rate is proportional to M, the number of multiplexed sources. Thus, we can predict asymptotically (i.e. for a sufficiently large buffer threshold x) the transient period of a system with a large number of sources by examining a system with a relatively small number of sources. The numerical examples we discuss in the next section support this key observation.

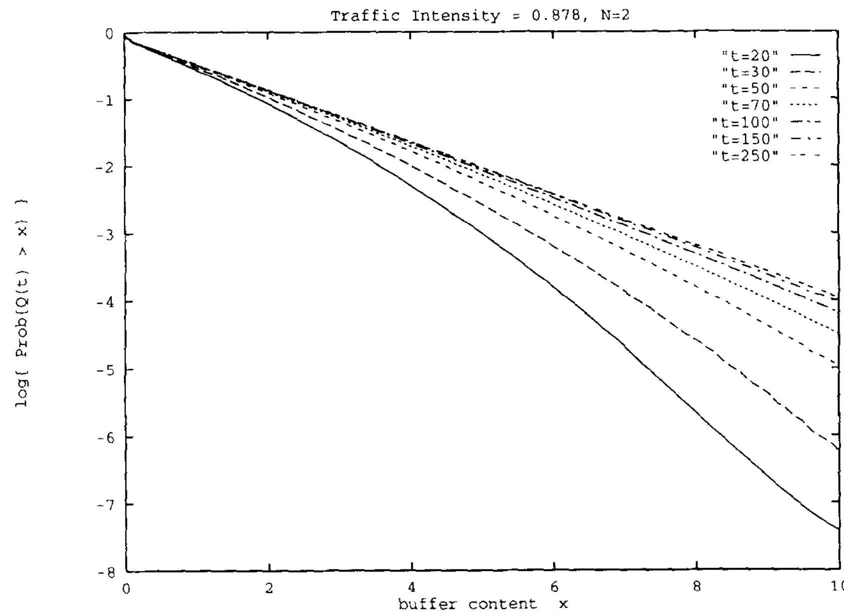

Fig. 3. $\log_{10} \Pr[Q(t) > x]$ versus x. $C = 0.8$ (or $\rho = 0.878$). $t = 20, 30, 50, 70, 100, 150$ and 250.

6. Numerical examples

We present here some numerical results to illustrate our solution technique. The numerical-inversion method of the Laplace transform used here is given by (see [11, pp. 73–74])

$$P(t,x) \simeq \frac{e^{ct}}{T}\Big[\frac{1}{2}\mathrm{Re}\{P^*(c,x)\} + \sum_{k=1}^{+\infty} \mathrm{Re}\{P^*(c + \frac{\pi i k}{2T}, x)\} \cos\frac{\pi k t}{2T}\Big], \tag{66}$$

where $[0, T]$ is the finite range over which we wish to evaluate $P(t,x)$, and c is an arbitrary number in the convergence region of $P^*(s,x)$, which is the entire open right-half plane, i.e., $\mathrm{Re}\{s\} > 0$, to be shown in Appendix A.

Abate and Whitt [1] give an extensive review and discussion on the above Fourier-series method for inverting transforms of probability distribution functions.

The parameters used here are given by $\alpha^{-1} \approx 650\,\mathrm{ms}$, $\beta^{-1} \approx 352\,\mathrm{ms}$ which correspond to the empirical values of a voice source model discussed by Sriram and Whitt [23] except that in Figs. 5 and 6, we choose $\alpha = 0.4$ and $\beta = 1.0$ to compare our transient results with the steady-state results in [2].

(1) *Infinite buffer case.* We show some numerical results for the cases of $N = 2$, $N = 50$ and $N = 100$. For simplicity, we assume that all sources are initially off, i.e., $j_0 = 0$.
Figure 3 shows the probability that $Q(t)$ exceeds x at fixed time t. The vertical axis is the logarithm with base 10 of the survivor function. The link capacity $C = 0.8$ is fixed and $N = 2$. The probability is observed at different times t.
Figure 4 shows how fast $Q(t)$ exceeds $x = 2.0$ when the initial buffer content x_0 takes on various values. A heavy traffic intensity $\rho = 0.95$ (or $C = 0.74$) and $N = 2$ are assumed.
Figure 5 shows how fast $Q(t)$ exceeds various thresholds x (x = 0, 3.0 and 5.0) with the

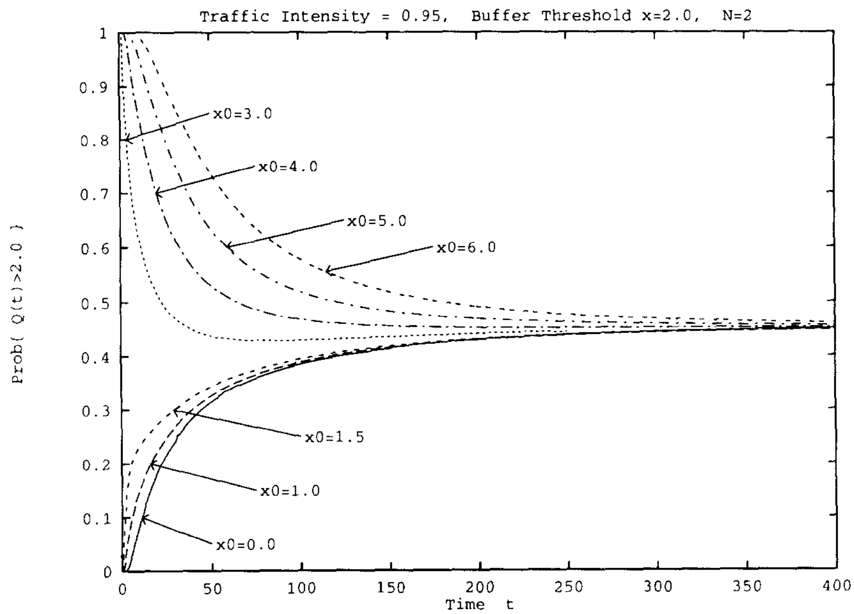

Fig. 4. $\Pr[Q(t) > 2.0]$ versus t for different values of the initial buffer contents $x_0 = 0.0, 1.0, 1.5, 3.0, 4.0, 5.0$ and 6.0. $\rho = 0.95$ (i.e. $C = 0.74$) is assumed.

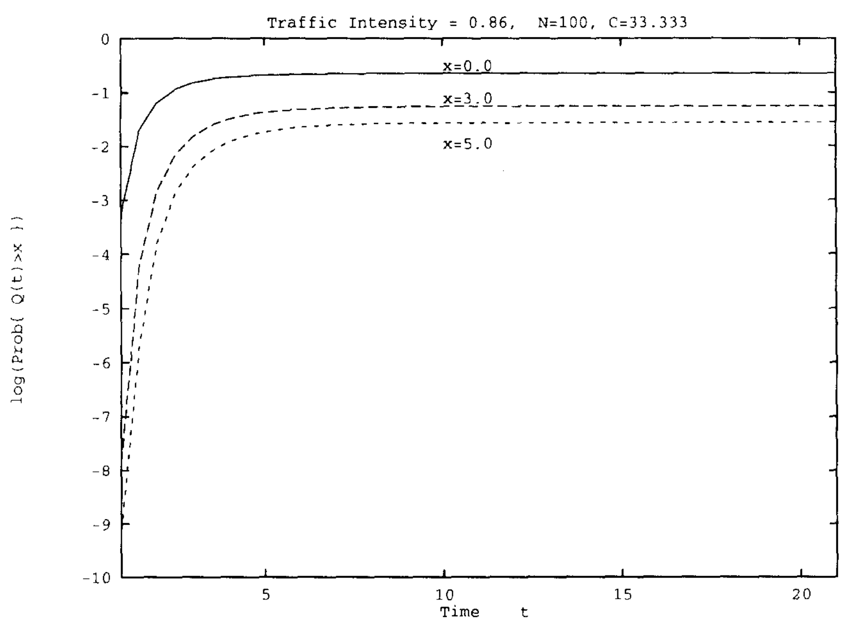

Fig. 5. $\log_{10} \Pr[Q(t) > x]$ versus time t for $x = 0.0, 3.0$ and 5.0.

source number $N = 100$. The buffer is assumed initially empty.

In Fig. 6, we keep the traffic intensity constant and show how the time scales of transient periods change with the number of sources. $N = 2$, 10, 50, 100 are considered. These curves verify the general observation we made earlier based on Eq. (65).

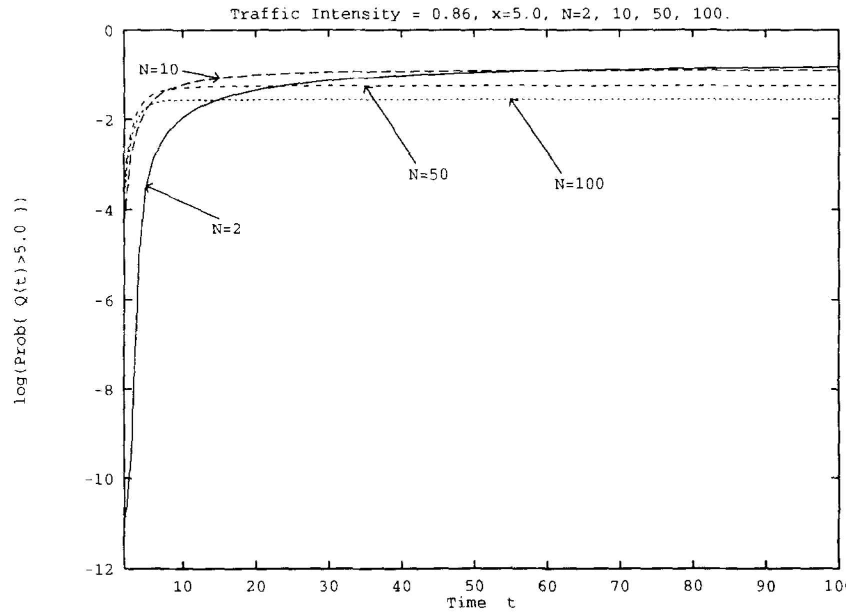

Fig. 6. $\log_{10}\Pr[Q(t) > 5.0]$ versus time t for $N = 2$, 10, 50 and 100, with same traffic intensity 0.86.

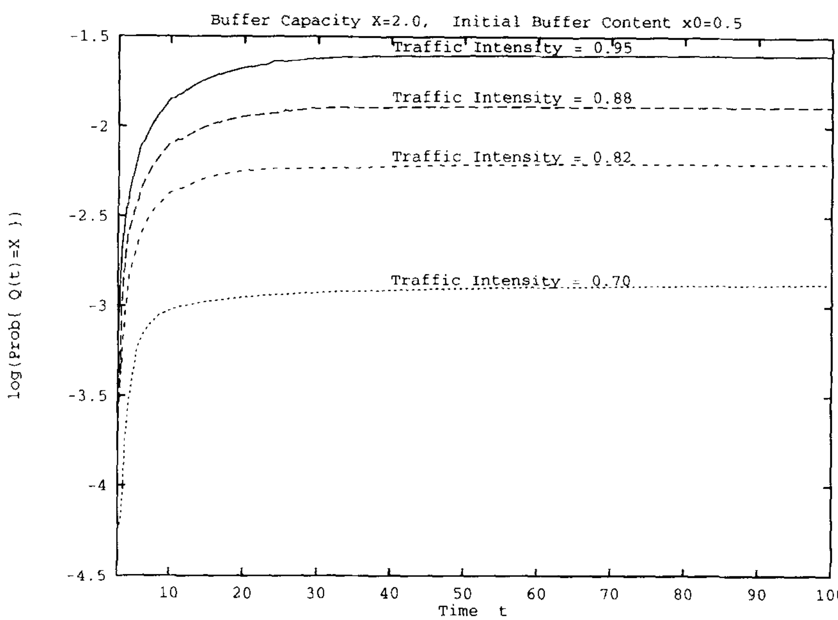

Fig. 7. $\log_{10}\Pr[Q(t) = X]$ versus t for $X = 2.0$ and the traffic intensity $\rho = 0.95, 0.88, 0.82$ and 0.70. $x_0 = 0.5$ is fixed.

(2) *Finite buffer case.* We show some numerical results for the single source example, assuming that the source is initially off, i.e., $N = 1$ and $j_0 = 0$.
In Fig. 7 we fix the buffer capacity $X = 2.0$ and the initial value $x_0 = 0.5$ and assume different values of the traffic intensity ρ.

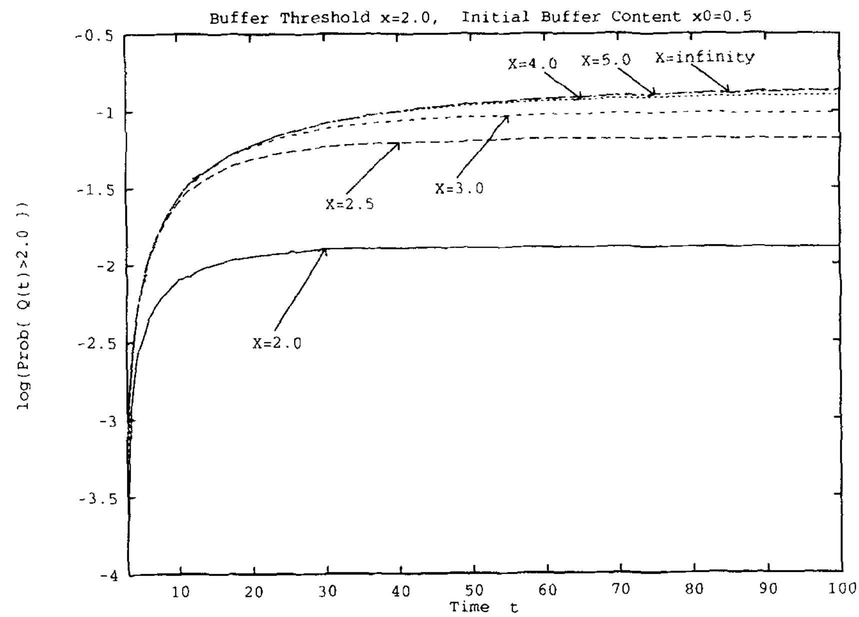

Fig. 8. $\log_{10} \Pr[Q(t) > 2.0]$ versus time for different values of the buffer capacity $X = 2.0, 2.5, 3.0, 4.0, 5.0$ and infinity. $x_0 = 0.5$ and $\rho = 0.88$ are assumed for all cases.

Figure 8 shows how soon $Q(t)$ exceeds buffer level $x = 2.0$. The different curves correspond to different values of the buffer capacity X. The case with infinite buffer capacity is also shown. The initial buffer content $x_0 = 0.5$ and the traffic intensity $\rho = 0.88$ are assumed throughout. While the cell loss probability decreases as the buffer capacity increases, Fig. 8 shows that a increase in the buffer capacity leads to a longer queue, i.e., a larger *delay time* $D(t)$ $(\stackrel{\text{def}}{=} (Q(t)/C))$. In other words, for fixed x or τ, $\Pr[Q(t) > x|X]$ and $\Pr[D(t) > \tau|X]$ increase as the buffer capacity X increases. These curves show quantitatively the trade-off between the packet loss probability and delay time, that is, if a larger buffer is provided to reduce the packet loss probability, then the packet delay time will become larger and vice versa.

7. Conclusion

We have presented mathematical results on the time-dependent behavior of the fluid flow model of statistical multiplexing. In Sections 3.2 and 3.3, we showed how to determine the time-dependent boundary conditions by solving a set of linear equations, and then explicitly obtain the solution in the form of a Laplace transform. In Section 4, we presented our final solutions in closed forms which involve less computational complexity. In Section 5, the asymptotic behavior and transient periods of the system were discussed.

By extending our analysis, we should be able to develop a new method to predict the network load in real time. A prediction model will help us gain some insight into the design and analysis of network congestion control. We expect that in a high-speed network, most existing control strategies will fail due to its large propagation delay as compared with the small transmission time. An accurate prediction of the transient network load will enable us to develop a preventive control of network congestion at the cell level by regulating traffic, or dynamically assigning the link capacity. Such new control strategies are under our current investigation and will be reported in a separate paper.

Appendix A. Proof of Theorem 1

The joint probability distribution function $P_j(t,x)$ ($j \in [0,n]$) is bounded by 1 for arbitrary $t \geq 0, x \geq 0$. Thus its Laplace transform $P^{**}(s,u)$ defined by Eq. (15) must be complex analytic for $\mathrm{Re}(s) > 0$ and $\mathrm{Re}(u)0$, which is the region of convergence. On the other hand, the region of convergence of the Laplace transform consists of strips parallel to the imaginary axis in the s-plane. In order to show $\mathrm{Re}\{u_k(s)\} > 0$ or $\mathrm{Re}\{u_k(s)\} < 0$ in the entire open right-half plane, i.e., $\mathrm{Re}\{s\} > 0$, it is sufficient to consider them along the nonnegative real axis, i.e., $\mathrm{Re}\{s\} \geq 0$ and $\mathrm{Im}\{s\} = 0$.

In the case $s = 0$, $\{u(K,0); 0 \leq K \leq N\}$ in Eq. (31) reduce to those of the steady-state case discussed by Anick et al. [2].

From Eq. (36)

$$u(K,s) = \alpha - \beta + \frac{(\frac{1}{2}N - C)\{s + (N-C)\alpha + C\beta\} + (\frac{1}{2}N - K)\sqrt{\Delta(K,s)}}{(\frac{1}{2}N - K)^2 - (\frac{1}{2}N - C)^2} \tag{A.1}$$

where

$$\Delta(K,s) = \{s + (N-C)\alpha + C\beta\}^2 + 4\alpha\beta(K-C)(N-K-C). \tag{A.2}$$

We now define

$$\nabla(K,s) = \frac{\mathrm{d}u(K,s)}{\mathrm{d}s} = \frac{(\frac{1}{2}N - C) + (\frac{1}{2}N - K)\frac{s + (N-C)\alpha + C\beta}{\sqrt{\Delta(K,s)}}}{(C-K)(N-K-C)}. \tag{A.3}$$

- *Case 1:* $C \leq \frac{1}{2}N$. When $0 \leq K \leq \lfloor C \rfloor$, the roots $\{u(K,0);\ 0 \leq K \leq \lfloor C \rfloor\}$ are *non-negative*. We find that $N - K - C > 0$, $C - K > 0$, $\frac{1}{2}N - K > 0$, and

  $$\frac{s + (N-C)\alpha + C\beta}{\sqrt{\Delta(K,s)}} \geq 1 \quad \text{for all } s \geq 0.$$

  Then

  $$\nabla(K,s) \geq \frac{(\frac{1}{2}N - C)}{(C-K)(N-K-C)} > 0. \tag{A.4}$$

  When $\lfloor C \rfloor + 1 \leq K \leq N$, the roots $\{u(K,0);\ \lfloor C \rfloor + 1 \leq K \leq N\}$ are *negative*.

  If $\lfloor C \rfloor + 1 \leq K \leq N - \lfloor C \rfloor$, it follows that $N - K - C > 0$, $C - K < 0$, $\frac{1}{2}N - C > 0$, and

  $$0 \leq \frac{s + (N-C)\alpha + C\beta}{\sqrt{\Delta(K,s)}} \leq 1, \quad \text{for all } s \geq 0.$$

  Then

  $$\nabla(K,s) \leq \frac{(\frac{1}{2}N - C) + (\frac{1}{2}N - K)}{(C-K)(N-K-C)} < 0. \tag{A.5}$$

If $N-\lfloor C\rfloor+1 \le K \le N$, it follows that $N-K-C<0$, $C-K<0$, $\frac{1}{2}N-K<0$, and

$$\frac{s+(N-C)\alpha+C\beta}{\sqrt{\Delta(K,s)}} \ge 1 \quad \text{for all } s \ge 0.$$

Then

$$\nabla(K,s) \le \frac{(\frac{1}{2}N-C)+(\frac{1}{2}N-K)}{(C-K)(N-K-C)} < 0. \tag{A.6}$$

- *Case 2:* $\frac{1}{2}N < C < N$. When $0 \le K \le \lfloor C\rfloor$, the roots $\{u(K,0); 0 \le K \le \lfloor C\rfloor\}$ are *non-negative*. If $0 \le K \le N-\lfloor C\rfloor-1$, it follows $N-C-K>0$, $C-K>0$, $\frac{1}{2}N-K>0$, and

$$\frac{s+(N-C)\alpha+C\beta}{\sqrt{\Delta(K,s)}} \ge 1 \quad \text{for all } s \ge 0.$$

Then

$$\nabla(K,s) \ge \frac{(\frac{1}{2}N-C)+(\frac{1}{2}N-K)}{(C-K)(N-K-C)} > 0. \tag{A.7}$$

If $N-\lfloor C\rfloor \le K \le \lfloor C\rfloor$, it follows $N-C-K<0$, $C-K>0$, and

$$0 < \frac{s+(N-C)\alpha+C\beta}{\sqrt{\Delta(K,s)}} \le 1 \quad \text{for all } s \ge 0.$$

Then

$$\nabla(K,s) \ge \frac{(\frac{1}{2}N-C)+(\frac{1}{2}N-K)}{(C-K)(N-K-C)} > 0. \tag{A.8}$$

When $\lfloor C\rfloor+1 \le K \le N$, the roots $\{u(K,0); \lfloor C\rfloor+1 \le K \le N\}$ are *negative*. We find that $N-K-C<0$, $C-K<0$, $\frac{1}{2}N-K<0$, and

$$\frac{s+(N-C)\alpha+C\beta}{\sqrt{\Delta(K,s)}} \ge 1 \quad \text{for all } s \ge 0.$$

Then

$$\nabla(K,s) \le \frac{(\frac{1}{2}N-C)}{(C-K)(N-K-C)} < 0. \tag{A.9}$$

The above observations can be summarized in the following:

- Consider those roots which are negative at $s=0$. There exist $N-\lfloor C\rfloor$ such roots, which we denote by $u_1(s),\ldots,u_{N-\lfloor C\rfloor}(s)$. Their derivatives $\mathrm{d}u_k(s)/\mathrm{d}s$ are always negative.
- Consider those roots which are positive at $s=0$. There exist $\lfloor C\rfloor$ such roots, which we denote by $u_{N-\lfloor C\rfloor+1}(s),\ldots,u_N(s)$. Their derivatives $\mathrm{d}u_k(s)/\mathrm{d}s$ are always positive.
- The root that corresponds to $K=0$ intersects the origin at $s=0$: $u(0,0)=0$. Let $u_0(s)$ denote this root, and we find $\mathrm{d}u_0(s)/\mathrm{d}s>0$.

None of them is identical to any other, because they are all distinct at $s=0$. None of them is zero because $|sI-\mathcal{M}| \ne 0$ for some s.

This concludes the proof of the theorem.

Appendix B

From Eq. (41), the last $N - \lfloor C \rfloor$ entries of $P^*(s,0)$ are zero. $\mathcal{D}^{-1}(M - sI)$ is a tri-diagonal matrix. Note that a multiplication of $\mathcal{D}^{-1}(M - sI)$ on $P^*(s,0)$ will reduce by one the number of entries that are zero, and that each additional multiplication will have the same effect until $[\mathcal{D}^{-1}(M - sI)]^{N-\lfloor C \rfloor -1} P^*(s,0)$ has only its last entry equal to zero.

The generating function of $V_k(s)$ takes the form of Eq. (27) for each k, so the $(N+1)$th (or last) entry of each $V_k(s)$ is one.

Since the term $P(0,x)$ makes Eq. (13) a non-homogeneous differential equation, we derive Eq. (52) in the following two cases, depending on the value of x_0, the initial buffer content.

(1) $x_0 > 0$. Taking the lth derivative of Eq. (13) and evaluating it at $x = 0$, we find

$$\frac{\mathrm{d}^{(l)} P^*(s,x)}{\mathrm{d}x^{(l)}}\Big|_{x=0} = [\mathcal{D}^{-1}(M - sI)]^l P^*(s,0), \quad l = 0, \ldots, N - \lfloor C \rfloor - 1. \tag{B.1}$$

Similarly from Eq. (44), we obtain

$$\frac{\mathrm{d}^{(l)} P^*(s,x)}{\mathrm{d}x^{(l)}}\Big|_{x=0} = \sum_{k=0}^{N} a_k(s)[u_k(s)]^l V_k(s), \quad l = 0, \ldots, N - \lfloor C \rfloor - 1. \tag{B.2}$$

Comparing the last entries in the right-hand sides of Eqs.(B.1) and (B.2), we have

$$\sum_{k=0}^{N} a_k(s)[u_k(s)]^l = 0, \quad l = 0, \ldots, N - \lfloor C \rfloor - 1, \tag{B.3}$$

from which and Eq. (48), we obtain the following expression:

$$\sum_{k=1}^{N-\lfloor C \rfloor} a_k(s)[u_k(s)]^l = b_0(s) u_0^l(s) + \sum_{k=N-\lfloor C \rfloor+1}^{N} b_k(s)[u_k(s)]^l, \tag{B.4}$$

for $l = 0, \ldots, N - \lfloor C \rfloor - 1$.

This in turn gives Eq. (52).

(2) $x_0 = 0$. In this case, Eq. (13) is equal to

$$\frac{\mathrm{d}P^*(s,x)}{\mathrm{d}x} = \mathcal{D}^{-1}(M - sI) P^*(s,x) + \mathcal{D}^{-1} e_{j_0}. \tag{B.5}$$

Then taking the lth derivative of the above equation and evaluating it at $x = 0$, we obtain

$$\frac{\mathrm{d}^{(l)} P^*(s,x)}{\mathrm{d}x^{(l)}}\Big|_{x=0} = [\mathcal{D}^{-1}(M - sI)]^l P^*(s,0) + [\mathcal{D}^{-1}(M - sI)]^{l-1} \mathcal{D}^{-1} e_{j_0} \tag{B.6}$$

for $l = 1, \ldots, N - \lfloor C \rfloor - 1$, where we denote the last entry of $[\mathcal{D}^{-1}(M - sI)]^{l-1}\mathcal{D}^{-1} e_{j_0}$ as $c_l(s)$. From Eq. (44) we have

$$P^*(s,0) = H_{N+1}(s) + \sum_{k=1}^{N-\lfloor C \rfloor} [a_k(s) + b_k(s)] V_k(s), \tag{B.7}$$

$$\frac{d^{(l)}\boldsymbol{P}^*(s,x)}{dx^{(l)}}|_{x=0} = \sum_{k=1}^{N-\lfloor C\rfloor} [a_k(s)+b_k(s)][u_k(s)]^l \boldsymbol{V}_k(s) \tag{B.8}$$

for $l = 1,\ldots,N-\lfloor C\rfloor-1$. Comparing the last entries in the right-hand sides of Eqs. (B.6), (B.7) and (B.8), we have

$$\sum_{k=1}^{N-\lfloor C\rfloor} [a_k(s)+b_k(s)][u_k(s)]^l = c_l(s) \quad l=0,\ldots,N-\lfloor C\rfloor-1, \tag{B.9}$$

or

$$\mathcal{T}\underline{a}(s) = \underline{c}(s), \tag{B.10}$$

where the $\underline{a}(s)$ is modified as

$$\underline{a}(s) = [a_1(s)+b_1(s),\ldots,a_{N-\lfloor C\rfloor}(s)+b_{N-\lfloor C\rfloor}(s)]', \tag{B.11}$$

and

$$\underline{c}(s) = [-h_{N+1,N}(s), c_1(s),\ldots,c_{N-\lfloor C\rfloor-1}]'. \tag{B.12}$$

When the number of sources that are initially on is less than the link capacity, i.e., $j_0 < C$, then the last expression is simplified to

$$\underline{c}(s) = [-h_{N+1,N}(s), 0,\ldots,0]'. \tag{B.13}$$

Acknowledgements

The authors would like to thank Mr. John Paz for carrying out the numerical computations of our examples that are presented in Figs. 5 and 6. This research has been supported in part by the Ogasawara Foundation for the Promotion of Science and Technology, Sony Corporation's graduate fellowship and by the New Jersey Commission on Science and Technology "Broadband Communication Network for Multiple Services", Grant No. 89-240510-4 and No. 90-240510-6.

References

[1] Abate, J. and W. Whitt, The Fourier-series method for inverting transforms of probability distributions, *Queueing Systems* **10** (1992).

[2] Anick, D., D. Mitra and M.M. Sondhi, Stochastic theory of a data-handling system with multiple sources, *Bell System Tech. J.* **61**(8) (1982) 1871–1894.

[3] Coffman, Jr., E.G., B.M. Igelnik and Y.A. Kogan, Controlled stochastic model of a communication system, *Proc. 13th Int. Teletraffic Congress (ITC-13) Workshop on Queueing, Performance and Control in ATM*, Studies in Telecommunication Vol. 15, North-Holland, Amsterdam, 1991, pp. 188–191.

[4] Cohen, J.W., Superimposed renewal process and storage with gradual input, *Stoch. Process Appl.* **2** (1974) 31–58.

[5] Elwalid, A.I., D. Mitra and T.E. Stern, Statistical multiplexing of Markov modulated sources: theory and computational algorithms, *Proc. 13th Int. Teletraffic Congress (ITC-13)*, June 1991.

[6] Friedman, B., *Principles and Techniques of Applied Mathematics*, Wiley, New York (1956).

[7] Hashida, O. and M. Fujiki, Queueing models for buffer memory in store-and-forward system, *Proc. 7th Int. Teletraffic Congress (ITC-7)*, Stockholm, June 1973, pp. 323/1-323/7.
[8] Hayes, J.F., *Modeling and Analysis of Computer Communications Networks.*, Plenum Press, New York (1984).
[9] Horn, R.A. and C.A. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge (1985).
[10] Kobayashi, H. and A.G. Konheim, Queueing models for computer communications system analysis, (Invited paper) *IEEE Trans. Comm.* **25** (1977) 2-29.
[11] Kobayashi, H., *Modeling and Analysis: An Introduction to System Performance Evaluation Methodology*, Addison-Wesley, Reading, Mass. (1978).
[12] Kobayashi, H., Performance issues of broadband ISDN, Part I: Traffic characterization and statistical multiplexing, and performance issues of broadband ISDN; Part II: Statistical multiplexing of multiple types of traffic, *Proc. ICCC'90, 10th Int. Conf. on Computer Communication*, November 1990, pp. 349-361.
[13] Kobayashi, H. and Q. Ren, Non-stationary behavior of statistical multiplexing for multiple types of traffic, *Proc. 26th Annual Conf. on Information Sciences and Systems*, Princeton University, Princeton, N.J., March 18-20, 1992.
[14] Kobayashi, H. and Q. Ren, A mathematical theory for transient analysis of communication networks, (Invited Paper), *IEICE Trans. Comm.* **E75-B** (12) (1992) 1266-1276.
[15] Kosten, L., Stochastic theory of a multi-entry buffer (1), *Delft Progress Report* **1** (1974) 10-18.
[16] Kosten, L., Stochastic theory of data handling systems with groups of multiple sources, in: H. Rudin and W. Bux (Eds.), *Performance of Computer-Communication Systems*, North-Holland, Amsterdam (1984) 321-331.
[17] Kulkarni, V.G., V.F. Nicola and K.S. Trivedi, On modeling the performance and reliability of multimode computer systems, *J. Systems and Software* **6** (1/2) (1986) 175-183.
[18] Mitra, D., Stochastic theory of a fluid model of producers and consumers coupled by a buffer, *Adv. Appl. Probab.* **20** (1988) 646-676.
[19] Morrison, J.A., Asymptotic analysis of a data-handling system with many sources, *SIAM J. Appl. Math.* **49** (2) (1988) 617-637.
[20] Reibman, A., R. Smith and K. Trivedi, Markov and Markov reward model transient analysis: an overview of numerical approaches, *Eur. J. Op. Res.* **40** (1989) 257-267.
[21] Ren, Q. and H. Kobayashi, Transient solutions for the buffer behavior in statistical multiplexing, *Proc. 26th Annual Conf. on Information Sciences and Systems*, Princeton University, Princeton, N.J., March 18-20, 1992.
[22] Simonian, A. and J. Virtamo, Transient and stationary distributions for fluid queues and input processes with a density, *SIAM J. Appl. Math.* **51** (6) (1991) 1732-1739.
[23] Sriram, K. and W. Whitt, Characterizing superposition arrival processes in packet multiplexers for voice and data, *IEEE J. Select. Areas Comm.* **4** (6) (1986) 833-846.
[24] Syski, R., *Introduction to Congestion Theory in Telephone Systems*, North-Holland Studies in Telecommunication Vol. 4, North-Holland, Amsterdam (1986).
[25] Zhang, J., Transient solution of continuous flow Markov modulated rate models, *Proc. 26th Annual Con. on Information Sciences and Systems*, Princeton University, Princeton, N.J., March 18-20, 1992.

Qiang Ren was born in Shanghai, China, in 1968. He received the B.S. degree in Mathematics from Beijing University in July 1989, the M.A. and Ph.D. degrees in electrical engineering from Princeton University in 1991 and 1994, respectively.
He was a Teaching Assistant and a Research Assistant at Princeton University. He was the recipient of Sony Graduate Fellowship in the academic year 1992-1993. His research interests include mathematical modeling and performance analysis of broadband communication networks, queueing theory and simulation methods. His thesis topic was *Modeling of Traffic Sources and Performance Analysis of Statistical Multiplexing in an ATM Network*.
He is currently a research staff member at the C & C Research Laboratories, NEC USA, Inc., in Princeton, N.J., USA.

Hisashi Kobayashi holds the title of Sherman Fairchild University Professor at Princeton University, where he is currently Professor of Electrical Engineering and Computer Science. From 1986 till 1991 he served as the Dean of the School of Electrical Engineering and Applied Science. He received his B.S. and M.S. degrees in Electrical Engineering from the University of Tokyo in 1961 and 1963, resp. He was a radar system engineer at Toshiba Corporation in 1963–1965, and came to the U.S. in 1965 as a recipient of the Orson Desaix Munn Fellowship of Princeton University where he received his M.A. and Ph.D. degrees in 1966 and 1967, resp. From 1967 until 1986 he was with the IBM Research Division, Yorktown Heights, N.Y., where he conducted research in coding theory for high speed data transmission and high digital magnetic recording, seismic signal processing, image data compression, performance modeling of computer systems and communication networks, and queueing theory. He also held a number of managerial positions including Senior Manager of Systems Analysis and Algorithms, and Department Manager of VLSI Design. From 1982–1986 he was the founding director of IBM Japan Science Institute (now called IBM Tokyo research Lab.). He held visiting professorships at UCLA (1969–1970), the University of Hawaii (19750, Stanford University (1976), Technical University of Darmstadt, Germany (1979–1980), the Free University of Brussels, Belgium (1980) and at the University of Tokyo (1991–1992). He holds seven U.S. patents, authored more than 90 research articles and published a text book, *Modeling Analysis* (Addison-Wesley, 1978). Among the various honors and awards that he received are: the 1979 Humboldt Award from West Germany, IFIP Silver Core Award (1980), and IBM Outstanding Contribution Awards (1975, 1984), Fellow of IEEE (1977) and Engineering Academy of Japan (1992). He has served as a scientific and technical advisor to several research institutions and universities including NASA, Washington, D.C.; SRI International, Menlo Park, Calf.; Premier's Council, Ontario, Canada; the National University of Singapore; Advanced Systems Foundation, British Columbia, Canada; and Information Technology Research Center, Ontario, Canada.