Non-stationary Behavior of Statistical Multiplexing for Multiple Types of Traffic*

Hisashi Kobayashi† **and Qiang Ren**
Department of Electrical Engineering
School of Engineering and Applied Science
Princeton University
Princeton, NJ 08544

Abstract

We develop a general theory to derive the non-stationary (or transient) behavior of a statistical multiplexer , which combines data that are generated from a heterogeneous set of information sources. The problem is motivated by performance analysis and control of a multi-media communication network in the future B-ISDN (Broadband Integrated Services Digital Network) environment, where multiple types of information traffic will be integrated and transported over high speed links.

Our work extends earlier results by Kosten [1984] and others by solving the general non-stationary behavior of the multiplexer, whereas the earlier work dealt with only the equilibrium solutions, and in most cases, for a single type of traffic. The transient solution will be useful to better understand various performance problems that may arise in future high-speed networks which will carry bursty traffic of various types.

Our analysis is based on a linear operator theory and its spectral expansion method applied to the transform domain (the joint z-transform and double Laplace transforms) of the partial differential equation that governs the stochastic behavior of the statistical multiplexer. Computational aspects of this method is left for a further investigation.

1 Introduction

With recent progress in optical fiber and related technologies, the architectural design and prototype implementation of broadband ISDN (B-ISDN) are actively pursued on the international level. In such future broadband networks, multiple types of communication media — voice, data, graphics, image, video and TV (high definition TV in the future) – will be integrated and transmitted over high speed links, by using a new type of digital switching and multiplexing technique, known as fast packet switching in asynchronous transfer mode (ATM). The basic principle of this fast packet switching scheme is *statistical multiplexing*.

There have been a number of studies that report analytic model of statistical multiplexing, but most studies have been limited to models with one type of information source (e.g., voice sources) or two types of sources (e.g, voice and data).

Anick, Mitra and Sondhi [1982], Cohen [1974], Hashida and Fujiki [1973], Kosten [1974], Mitra [1988], Stern [1984] and others discuss *fluid approximation* models that are relevant to statistical multiplexing systems and obtain the *equilibrium state* solutions for *single type* of traffic sources. For an expository treatment of these earlier results, the reader is referred to Kobayashi [1990a]. Ren and Kobayashi [1992] recently obtained transient solutions for such statistical multiplexors by using the double Laplace transform method.

Kosten [1984] presents some analytic and simulation methods to derive the equilibrium solution for *multiple types* of traffic. Kobayashi [1990b, 1991] discusses the case of *infinite sources* with multiple types, and characterizes an asymptotic behavior of the buffer contents in terms of simple parameters of what he terms the "dominant" type traffic. Elwalid, Mitra and Stern [1991] and Stern and Elwalid [1991] discuss equilibrium state solutions when the sources are modeled as Markov modulated sources, and derive both theoretical results and computational methods.

Our paper presents a general theory to deal with the transient analysis of multiple types of traffic. Our analytic results will be useful to understand dynamic behavior of statistical multiplexors in high-speed ATM networks, which will carry bursty traffic of various types. The non-stationary solution will also provide a mathematical basis for formulating preventive congestion control algorithms.

2 A Mathematical Formulation of Multiple Types of Sources

2.1 A Finite Source Model with Multiple Types

Let there be N_m sources of type m, where $m = 1, 2, \ldots, M$, and let $J_m(t)$ denote the number of types m sources in "on" state (or in "burst", or "talk spurt" mode in the case of voice sources). Therefore, there are $N_m - J_m(t)$ sources which are in "off" state (i.e., "silence" mode).

We assume that successive "on" and "off" periods of each source form an alternating renewal process. For mathematical simplicity, we further assume that the "on" and "off" periods of type m sources are both exponentially distributed with parameters α_m and β_m, respectively:

$$\alpha_m^{-1} = \text{The mean } \textit{off} \text{ period of a type } m \text{ source;} \quad (1)$$

$$\beta_m^{-1} = \text{The mean } \textit{on} \text{ period of a type } m \text{ source.} \quad (2)$$

Let $\mathbf{j}$ be a vector defined by

$$\mathbf{j} = [j_1, j_2, \ldots, j_M], \quad (3)$$

where j_m is an integer that $J_m(t)$ can take on. Let us define the time-dependent probability distribution

$$P(\mathbf{j}; t) = P[J_m(t) = j_m, 1 \leq m \leq M]. \quad (4)$$

*This research has been supported in part by the Ogasawara Foundation for the Promotion of Science and Technology, 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.

†Currently on sabbatical leave to the University of Tokyo, NEC Computer and Communications Chair, Research Center for Advanced Science and Technology (RCAST).

Then by applying the well known birth-and-death process model (see e.g. Syski [1960, 1986], Kobayashi [1978]), we obtain

$$\begin{aligned}\frac{dP(\mathbf{j};t)}{dt} = & - \sum_{m=1}^{M}[(N_m - j_m)\alpha_m + j_m\beta_m]P(\mathbf{j};t) \\ & + \sum_{m=1}^{M}(N_m - j_m + 1)\alpha_m P(\mathbf{j} - \mathbf{1}_m;t) \\ & + \sum_{m=1}^{M}(j_m+1)\beta_m P(\mathbf{j} + \mathbf{1}_m;t), \end{aligned} \tag{5}$$

where $\mathbf{1}_m$ is a vector that has unity in its mth entry and is zero elsewhere.

We define the following M-dimensional Z *transform* or probability generating function:

$$\begin{aligned} G(\mathbf{z};t) &= \mathcal{Z}\{P(\mathbf{j};t)\} \\ &= \sum_{j_1=0}^{N_1}\sum_{j_2=0}^{N_2}\cdots\sum_{j_M=0}^{N_M} P(\mathbf{j};t)z_1^{j_1}z_2^{j_2}\cdots z_M^{j_M}. \end{aligned} \tag{6}$$

Then Eq.(5) can be transformed into the following partial differential equation with respect to time t and the variable $\mathbf{z}$:

$$\frac{\partial}{\partial t}G = \mathcal{M}G, \tag{7}$$

where we drop, for the sake of notational brevity, the arguments $\mathbf{z}$ and t in the function $G(\mathbf{z};t)$. The symbol $\mathcal{M}$ represents a *linear operator* defined by

$$\mathcal{M}G = \sum_{m=1}^{M}(z_m - 1)[N_m\alpha_m G - (\alpha_m z_m + \beta_m)\frac{\partial}{\partial z_m}G]. \tag{8}$$

We choose the simple representation of equation (7) because the operator $\mathcal{M}$ can be interpreted as a matrix operated on vector G. We will use the symbol $\mathcal{M}$ to represent the linear operator of (8) and the corresponding matrix interchangeably. If we interpret the operand G as a vector of dimension L, where $L = \prod_{m=1}^{M}(N_m+1)$, then the operator $\mathcal{M}$ of (8) should be interpreted as the following matrix that acts on the vector G:

$$\mathcal{M} = \mathcal{M}_1 \oplus \mathcal{M}_2 \oplus \cdots \oplus \mathcal{M}_M, \tag{9}$$

where $\oplus$ represents Kronecker sum (see e.g., Bellman [1960], Neuts [1981]) and will be explicitly shown in the example of Section 4. Lexicographical ordering of the elements of vector G and those of $\mathcal{M}$ should be chosen consistently and the example will clarify this point. The matrix $\mathcal{M}_m$ is an $(N_m+1)\mathrm{x}(N_m+1)$ tridiagonal matrix whose elements $\mathcal{M}_{m_{ij}}$ are given by

$$\begin{aligned} \mathcal{M}_{m_{jj-1}} &= (N_m - j + 1)\beta_m, \\ \mathcal{M}_{m_{jj}} &= -(N_m - j + 1)\alpha_m - j\beta_m, \\ \mathcal{M}_{m_{jj+1}} &= (j+1)\beta_m \quad \text{for } j = 0,1,\ldots,N_m, \end{aligned} \tag{10}$$

$$\text{and } \mathcal{M}_{m_{ij}} = 0 \text{ for all other } i \text{ and } j,$$

By generalizing the birth-and-death process model of one dimension (see e.g. Syski [1960]) we can derive the following *product form* solution:

$$G(\mathbf{z};t) = \prod_{m=1}^{M}[q_m(t)(z_m-1)+1]^{N_m}, \tag{11}$$

where we assume that the system is in state $\mathbf{0}$ at time $t=0$. The function $q_m(t)$ is the binomial distribution parameter and is given by

$$q_m(t) = \frac{\alpha_m}{\alpha_m+\beta_m}[1 - e^{-(\alpha_m+\beta_m)t}]. \tag{12}$$

Hence we find that the joint probability is the product of M binomial distributions.

$$P(\mathbf{j};t) = \prod_{m=1}^{M}\binom{N_m}{j_m} q_m^{j_m}(t)[1-q_m(t)]^{N_m - j_m}. \tag{13}$$

In the limit $t \to \infty$, the above distribution converges to the equilibrium distribution with the constant binomial parameters $q_m = \alpha_m/(\alpha_m+\beta_m)$.

2.2 A Statistical Multiplexer Model for Multiple Types of Traffic

Now we analyze the behavior of a statistical multiplexer in fast packet switching. Each source of type m in its *burst* state generates packets (or *cells* in the ATM terminology) at the rate of R_m [packets/unit time]. The aggregate rate of packet arrivals at time t is therefore

$$R(t) = \sum_{m=1}^{M} R_m J_m(t). \tag{14}$$

Suppose that the buffer content is initially empty. Then while $R(t) < C$, the link capacity of the multiplexer output, the arriving packets are processed immediately, thus no queue of packets will develop in the buffer. Once $R(t)$ exceeds C, however, the output link can no longer handle all the packets instantaneously, and buffer contents will grow or deplete at the rate $R(t)-C$, depending on whether this quantity is positive or negative at a given instant.

Let us define $Q(t)$ as the total amount of packets found in the buffer of multiplexer output link. Strictly speaking, $Q(t)$ is an integer-valued function, but we approximate it by time-continuous function, assuming that a series of packets arrive like fluid flows. This assumption is well justified in modeling a multiplexer for a high speed link.

By extending Eq.(4) we define the following probability distribution function:

$$P(\mathbf{j};t,x) = P[J_m(t) = j_m, 1 \le m \le M; \text{ and } Q(t) \le x]. \tag{15}$$

Then by generalizing the differential-difference equation (5), we obtain the following partial differential equation:

$$\begin{aligned} & \frac{\partial P(\mathbf{j};t,x)}{\partial t} + (\sum_{m=1}^{M} R_m j_m - C)\frac{\partial}{\partial x}P(\mathbf{j};t,x) \\ = & - \sum_{m=1}^{M}[(N_m - j_m)\alpha_m + j_m\beta_m]P(\mathbf{j};t,x) \\ & + \sum_{m=1}^{M}(N_m - j_m + 1)\alpha_m P(\mathbf{j} - \mathbf{1}_m;t,x) \\ & + \sum_{m=1}^{M}(j_m+1)\beta_m P(\mathbf{j} + \mathbf{1}_m;t,x), \end{aligned} \tag{16}$$

Similarly we generalize Eq.(6) and define

$$\begin{aligned} G(\mathbf{z};t,x) &= \mathcal{Z}\{P(\mathbf{j};t,x)\} \\ &= \sum_{j_1=0}^{N_1}\cdots\sum_{j_M=0}^{N_M} P(\mathbf{j};t,x)z_1^{j_1}\cdots z_M^{j_M}. \end{aligned} \tag{17}$$

Then Eq.(16) can be transformed into

$$\frac{\partial}{\partial t}G + \mathcal{D}\frac{\partial}{\partial x}G = \mathcal{M}G, \tag{18}$$

where we now drop the arguments z, t and x in the function $G(z;t,x)$. $\mathcal{M}$ is the linear operator defined by Eq.(8) and operator $\mathcal{D}$ is defined as

$$\mathcal{D}G = \sum_{m=1}^{M}(z_m - 1)[R_m z_m \frac{\partial}{\partial z_m}G - CG]. \quad (19)$$

The matrix interpretation of the operator $\mathcal{D}$ is given by the following diagonal matrix of dimension L:

$$\mathcal{D} = diag[< \mathbf{j}, \mathbf{R} > -C] \quad (20)$$

where $< \mathbf{j}, \mathbf{R} >$ is an inner product of the vector of Eq.(3) and the packet generation rate vector defined by

$$\mathbf{R} = [R_1, R_2, \ldots, R_M] \quad (21)$$

The elements of $\mathbf{R}$ are arranged in such a way that lexicographical ordering of the diagonal matrix $\mathcal{D}$ and the operand vector G are consistent. Note that $\mathcal{D}$ is a natural generalization of the matrix D discussed by Kosten [1974], Anick et al. [1982] and others.

We define the Laplace transform of $G(z;t,x)$ with respect to time t by

$$G^*(z;s,x) = \mathcal{L}_t\{G(z;t,x)\}. \quad (22)$$

Then Eq.(18) can be transformed into

$$sG^*(z;s,x) - G(z;0,x) + \mathcal{D}\frac{\partial}{\partial x}G^*(z;s,x) = \mathcal{M}G^*(z;s,x). \quad (23)$$

We then take the Laplace transform of $G^*(z;s,x)$ with respect to the buffer content variable x:

$$sG^{**}(z;s,u) - G^*(z;0,u) + \mathcal{D}\{uG^{**}(z;s,u) - G^*(z;s,0)\} = \mathcal{M}G^{**}(z;s,u), \quad (24)$$

from which we obtain

$$[sI + u\mathcal{D} - \mathcal{M}]G^{**}(z;s,u) = G^*(z;0,u) + \mathcal{D}G^*(z;s,0). \quad (25)$$

We formally define an operator $\mathcal{R}(s)$ by

$$\mathcal{R}(s) = [sI + u\mathcal{D} - \mathcal{M}]^{-1}, \quad (26)$$

which can be viewed as a *resolvent* defined in the theory of semigroups (see e.g., Friedman [1956], Feller [1966] and Syski [1960]). The inverse operator (26) exists for those s which are not equal to eigenvalues of the operator $\mathcal{M} - u\mathcal{D}$. The set of all s for which $\mathcal{R}(s)$ exists is called the *resolvent set*. The set of eigenvalues $\{s_j; 0 \le j \le L-1\}$ is called the *spectrum*. Then by using the spectral expansion method (see e.g. Syski [1960]), we can represent the operator $\mathcal{R}(s)$ as follows:

$$\mathcal{R}(s) = \sum_{j=0}^{L-1}\frac{\mathcal{E}_j}{s - s_j}, \quad (27)$$

where the operator $\mathcal{E}_j$ is called a *projection operator*, and is obtained as

$$\mathcal{E}_j = \lim_{s \to s_j}(s - s_j)\mathcal{R}(s). \quad (28)$$

As we discuss later, however, the operator $\mathcal{E}_j$ can be often more easily obtained, once we find the jth left and right eigenvectors associated with the eigenvalue s_j.

Let s be one of the eigenvalues and let $V(z;u)$ be the associated right eigenvector. Then $V(z;u)$ should satisfy the following characteristic equation:

$$sV(z;u) = [\mathcal{M} - u\mathcal{D}]V(z;u) \quad (29)$$

or

$$\sum_{m=1}^{M}[\alpha_m z_m^2 + (uR_m + \beta_m - \alpha_m)z_m - \beta_m]\frac{\partial}{\partial z_m}\{\ln V(z;u)\} = -s + uC + \sum_{m=1}^{M}[N_m\alpha_m(z_m - 1)]. \quad (30)$$

Let $z_{m1}(u)$ and $z_{m2}(u)$ be two roots of the quadratic equation

$$\alpha_m z_m^2 + (uR_m + \beta_m - \alpha_m)z_m - \beta_m = 0, \quad (31)$$

which leads to the following *product form* for the eigenvector $V(z;u)$:

$$V(z;u) = \prod_{m=1}^{M}(z_m - z_{m1}(u))^{k_m}(z_m - z_{m2}(u))^{N_m - k_m}, \quad (32)$$

where k_m is an integer parameter between 0 and N_m. Then taking the logarithm of Eq.(32), and substituting it into Eq.(30), we find the following explicit formula for the eigenvalue:

$$s = uC - \sum_{m=1}^{M}\alpha_m[N_m - k_m z_{m2}(u) - (N_m - k_m)z_{m1}(u)]. \quad (33)$$

By substituting

$$z_{m1}(u), z_{m2}(u) = \frac{-(uR_m + \beta_m - \alpha_m) \pm \sqrt{D_m(u)}}{2\alpha_m}, \quad (34)$$

$$D_m(u) = (uR_m + \beta_m - \alpha_m)^2 + 4\alpha_m\beta_m, \quad (35)$$

into the last equation, we obtain

$$s = u(C - \frac{1}{2}\sum_{m=1}^{M}R_m N_m) - \frac{1}{2}\sum_{m=1}^{M}(\alpha_m + \beta_m)N_m - \sum_{m=1}^{M}\sqrt{D_m}(k_m - \frac{N_m}{2}). \quad (36)$$

Therefore, for a given integer vector

$$\mathbf{k} = [k_1, k_2, \ldots, k_M] \quad (37)$$

we uniquely determine the corresponding eigenvalue $s_{\mathbf{k}}$ as a function of u, and the associated eigenvector $V_{\mathbf{k}}(z;u)$, which takes the form (32).

Now we write the kth eigenvector as

$$V_{\mathbf{k}}(z;u) = \prod_{m=1}^{M}V_{k_m}(z_m;u), \quad (38)$$

where

$$V_{k_m}(z;u) = (z - z_{m1}(u))^{k_m}(z - z_{m2}(u))^{N_m - k_m}. \quad (39)$$

The coefficient of z^j term, $0 \le j \le N_m$, can be obtained by enumerating the z^h term in $(z - z_{m1}(u))^{k_m}$(for the range $j - \min\{j, N_m - k_m\} \le h \le \min\{j, k_m\}$) and z^{j-h} term in $(z - z_{m2}(u))^{N_m - k_m}$(for the range $j - \min\{j, k_m\} \le j - h \le \min\{j, N_m - k_m\}$). Because we can represent

$$(z - z_{m1}(u))^{k_m} = \sum_{h=0}^{k_m}\binom{k_m}{h}z^h(-z_{m1}(u))^{k_m - h}, \quad (40)$$

and

$$(z - z_{m2}(u))^{N_m - k_m} \quad (41)$$

$$= \sum_{j-h=0}^{N_m - k_m}\binom{N_m - k_m}{j - h}z^{j-h}(-z_{m2}(u))^{N_m - k_m - j + h},$$

the coefficient of $z^j = z^h z^{j-h}$ term is computed as

$$\begin{aligned} V_{k_m j}(u) &= \sum_{h=\max\{0,j+k_m-N_m\}}^{\min\{j,k_m\}} \binom{k_m}{h}\binom{N_m-k_m}{j-h} \\ &\quad \cdot(-z_{m1}(u))^{k_m-h}(-z_{m2}(u))^{N_m-k_m-j+h} \\ &= (-1)^{N_m-j}\sum_{h=\max\{0,j+k_m-N_m\}}^{\min\{j,k_m\}} \binom{k_m}{h} \\ &\quad \cdot \binom{N_m-k_m}{j-h} z_{m1}(u)^{k_m-h} z_{m2}(u)^{N_m-k_m-j+h}. \end{aligned} \tag{42}$$

Once the right eigenvector, denoted $\mathbf{V_k}$, is obtained, the corresponding left eigenvector $\mathbf{U_k}$ is given by

$$\mathbf{U_k} = (\mathcal{Q}^{-1})^2 \mathbf{V_k}, \tag{43}$$

where $\mathcal{Q}$ is a diagonal matrix that transforms the matrix $\mathcal{M}$ into a symmetric matrix. It is not difficult to show that

$$\mathcal{Q} = \mathcal{Q}_1 \otimes \mathcal{Q}_2 \otimes \cdots \otimes \mathcal{Q}_M, \tag{44}$$

where $\otimes$ represents Kronecker product, and the j element of the diagonal matrix $\mathcal{Q}_m$ is given by

$$\mathcal{Q}_{m_j} = \sqrt{(\frac{\alpha_m}{\beta_m})^j \binom{N_m}{j}}, \tag{45}$$

which corresponds to τ_j defined by Anick et al [1982].

We normalize these eigenvectors, which are orthogonal to each other, so that

$$\mathbf{U'_k V_{k'}} = \delta_{\mathbf{kk'}}, \tag{46}$$

where $\delta_{\mathbf{kk'}}$ is Kronecker delta. We find that the projection operators $\mathcal{E}_\mathbf{k}$ is representable, in its matrix interpretation, as

$$\mathcal{E}_\mathbf{k} = \mathbf{V_k U'_k}. \tag{47}$$

Therefore, from Eqs.(25), (27) and (47) we obtain the following:

$$\begin{aligned} \mathbf{P}^{**}(s,u) &= \mathcal{Z}^{-1}\{G^{**}(\mathbf{z};s,u)\} \qquad (48) \\ &= \sum_\mathbf{k} \frac{\mathbf{V_k U'_k}}{s-s_\mathbf{k}}[\mathbf{P}^*(0,u) + \mathcal{D}\mathbf{P}^*(s,0)], \end{aligned}$$

where $\mathcal{Z}^{-1}$ is the inverse Z transform. Thus the $\mathbf{P}^{**}(s,u)$ is the probability vector of dimension $L = \prod_{m=1}^M (N_m+1)$, and its kth element is the coefficient of $z_1^{k_1} z_2^{k_2} \cdots z_M^{k_M}$ term in $G^{**}(\mathbf{z};s,u)$, and is equivalent to $\mathcal{L}_t\mathcal{L}_x\{P(\mathbf{k};t,x)\}$.

In Eq.(48) the initial condition $P^*(0,u)$ is known. For example, if the system is initially empty, then all the entries of $P^*(0,u)$ are zero except for the first term, which is $1/u$. The boundary condition $\mathbf{P}^*(s,0)$, on the other hand, is unknown, and must be solved by finding a proper set of constraints. Ren and Kobayashi [1992] discuss the single type traffic case, and show that the method developed by Kosten [1974, 1982], Anick et al [1982] and Mitra [1988] is generalizable to determine the unknown boundary condition. The solution technique discussed with numerical examples in Ren and Kobayashi [1992] can be extended to the multiple type case under discussion.

Taking the inverse Laplace transform of Eq.(48) with respect to the variable s, we obtain the following time-dependent solution:

$$\begin{aligned} \mathbf{P}^*(t,u) &= \mathcal{L}_s^{-1}\{\mathbf{P}^{**}(s,u)\} \qquad (49) \\ &= \sum_\mathbf{k} \mathcal{E}_\mathbf{k}[\exp\{s_\mathbf{k}t\}\mathbf{P}^*(0,u) + \mathcal{D}\mathcal{L}_s^{-1}\{\frac{\mathbf{P}^*(s,0)}{s-s_\mathbf{k}}\}] \end{aligned}$$

The above summation should be taken over only those values of k for which $s_\mathbf{k} \le 0$, since positive eigenvalues would yield unstable solutions. Then taking the inverse Laplace transform of Eq.(49) we finally obtain

$$\mathbf{P}(t,x) = \mathcal{L}_u^{-1}\{\mathbf{P}^*(t,u)\} \tag{50}$$

In many practical problems we need to resort to numerical methods to perform the inverse Laplace transform. See, for example, Kobayashi [1978] and references therein that discuss the numerical inversion methods.

3 Special Cases

We now discuss three special cases of the above model and the corresponding solution and illustrate how they reduce to previously known results.

3.1 Equilibrium State Solution

If we define

$$G^*(\mathbf{z};u) = \lim_{t\to\infty} \mathcal{L}_s^{-1}\{G^{**}(\mathbf{z};s,u)\} = \lim_{s\to 0} sG^{**}(\mathbf{z};s,u) \tag{51}$$

and substitute it into Eq.(25) and let s approach zero, we obtain

$$[uI - \mathcal{D}^{-1}\mathcal{M}]G^*(\mathbf{z};u) = G(\mathbf{z};0). \tag{52}$$

We now take the spectral expansion with respect to the variable u, and write

$$G^*(\mathbf{z};u) = \sum_{j=0}^{L-1} \frac{\mathcal{E}_j}{u-u_j} G(\mathbf{z};0), \tag{53}$$

where $\{u_j; 0 \le j \le L-1\}$ are eigenvalues of the operator $\mathcal{D}^{-1}\mathcal{M}$ and $\mathcal{E}_j$'s are the corresponding projection operators.

The characteristic equation that the eigenvalues u must satisfy is obtained by setting $s=0$ in Eq.(36).

$$\begin{aligned} & u(C - \frac{1}{2}\sum_{m=1}^M R_m N_m) \qquad (54) \\ &= \frac{1}{2}\sum_{m=1}^M (\alpha_m+\beta_m)N_m + \sum_{m=1}^M \sqrt{D_m}(k_m - \frac{N_m}{2}). \end{aligned}$$

Kosten [1984] showed that

1. There are exactly $L (= \prod_{m=1}^M (N_m+1))$ eigenvalues, all real.
2. One eigenvalue is zero, which corresponds to the vector $\mathbf{k} = \mathbf{0}$.
3. The number of positive eigenvalues is one less than the number of integer vectors $\mathbf{k}$ that satisfy

$$\sum_{m=1}^M R_m k_m < C. \tag{55}$$

If we further specialize in this model and set $M=1$, then the model reduces to the case discussed by Anick et al [1982]. A minor difference from their characteristic equation, which is a quadratic equation in u (or z in Equation (20) in Anick et al [1982]), is that the characteristic equation (54) relates the integer vector $\mathbf{k}$ and the eigenvalues $\{u_\mathbf{k}\}$ one to one.

3.2 Multidimensional Birth-and-Death Process

If we just focus on the on-off behavior of the traffic sources, and ignore the buffer content of the multiplexer, the problem reduces to the model discussed in Section 2.1. Noting that the probability distribution (4) is the limiting case of (15), we define

$$G^*(\mathbf{z};s) = \lim_{x\to\infty} \mathcal{L}_u^{-1}\{G^{**}(\mathbf{z};s,u)\} = \lim_{u\to 0} uG^{**}(\mathbf{z};s,u). \tag{56}$$

Multiply Eq.(25) by u and let u approach zero, then we obtain

$$[sI - \mathcal{M}]G^*(\mathbf{z};s) = G(\mathbf{z};0). \tag{57}$$

Then the quadratic equation (31) yields

$$z_{m1}, z_{m2} = 1, \quad -\frac{\beta_m}{\alpha_m}. \tag{58}$$

The eigenvalue $s_{\mathbf{k}}$ is then

$$s_{\mathbf{k}} = -\sum_{m=1}^{M} k_m(\alpha_m + \beta_m) = - < \mathbf{k}, \alpha + \beta >, \tag{59}$$

where $<,>$ represents an inner product of M dimensional vectors. The associated right eigenvector is obtained from Eqs. (32) and (58). Equation (48) then becomes

$$\mathbf{P}^{**}(s) = \sum_{\mathbf{k}} \frac{\mathbf{V_k U_k'}}{s + < \mathbf{k}, \alpha + \beta >} \mathbf{P}(0), \tag{60}$$

where $\mathbf{P}(0)$ is the value of the probability vector process $\mathbf{P}(t)$ at $t = 0$.

Thus

$$\mathbf{P}(t) = \sum_{\mathbf{k}} a_{\mathbf{k}} \exp\{- < \mathbf{k}, \alpha + \beta > t\} \mathbf{V_k}, \tag{61}$$

where

$$a_{\mathbf{k}} = \mathbf{U_k'} \mathbf{P}(0). \tag{62}$$

It should not be difficult to reduce the above to Eq.(13), when the system is initially empty, i.e. $\mathbf{P}(0) = [1, 0, \ldots, 0]$, although this closed form expression is more directly derivable by solving the differential equation (6) in the $z - t$ domain, instead of dealing with the matrix representations.

3.3 Infinite Source Models

We now consider the limiting case, where

$$N_m \to \infty, \quad \alpha_m \to 0, \quad \text{while} \quad N_m \alpha_m \to \lambda_m, \tag{63}$$

for $m = 1, 2, \cdots, M$.

Then each traffic type becomes an infinite source model, and type m bursts arrive according to a Poisson process with rate λ_m and lasts on the average for $1/\beta_m$ seconds. The marginal probability vector in the product solution (13) now takes the form

$$\begin{aligned} P(\mathbf{j}; t) &= \prod_{m=1}^{M} P_m(j_m; t) \\ P_m(j_m; t) &= \frac{\{\frac{\lambda_m}{\beta_m}(1 - e^{-\beta_m t})\}^{j_m}}{j_m!} \exp\{-\frac{\lambda_m}{\beta_m}(1 - e^{-\beta_m t})\}, \\ m &= 1, 2, \ldots, M \quad \text{and} \quad j_m = 0, 1, \ldots. \end{aligned}$$

The last expression represents the probability that j_m bursts of type m traffic are found at time t, and it is equivalent to the time-dependent solution for an $M/M/\infty$ queueing system. Using a result in Takaćs [1962], we can assume that the burst period of type m traffic has a general distribution $G_m(t)$, and find the following transient solution:

$$P_m(j_m; t) = \frac{\{\lambda_m \int_0^t G_m^c(y) dy\}^{j_m}}{j_m!} \exp\{-\lambda_m \int_0^t G_m^c(y) dy\},$$

where $G_m^c(t)$ is the complement of the distribution function $G_m(t)$.

Now returning to the original statistical multiplexer specified by (63), we can show that the analysis for each traffic type reduces to the case originally discussed by Kosten [1974] for a single type.

The operator $\mathcal{M}_m$ represents an infinite matrix given by

$$\mathcal{M}_m = \begin{bmatrix} -\lambda_m & \beta_m & 0 & 0 & \cdots \\ \lambda_m & -(\lambda_m + \beta_m) & 2\beta_m & 0 & \cdots \\ 0 & \lambda_m & -(\lambda_m + 2\beta_m) & 0 & \cdots \\ 0 & 0 & \lambda_m & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Then the mth component $V_m(z_m; u)$ of the eigenvector should be of the following form (Kosten [1974]):

$$V_m(z; u) = \exp\{\frac{\lambda_m z}{u R_m + \beta_m}\} [\beta_m(1 - z) - z R_m u]^{k_m}, \tag{64}$$

where k_m is an integer: $k_m = 1, 2, \ldots$

4 Further Discussion

The main contribution of this paper is that we developed a general theory to derive exact expressions for the non-stationary behavior of the joint distribution of $\mathbf{j}$, the vector variable representing the numbers of *on* sources of different types, and x, the buffer content variable, as a function of time t. Once the joint distribution is obtained, a number of performance measures are directly obtainable. Among them is the probability of cell (or packet) blocking, which is of significant importance in designing a multimedia high-speed networks. A cell loss occurs since multiplexer capacity is in reality finite. The cell loss probability due to buffer overflows can be approximated by computing the tail end of the marginal distribution of x that exceeds the buffer capacity.

Certainly computational complexity grows exponentially as the size of the problem becomes large, but in practice it will be sufficient to compute the first few dominant exponential terms in (49) that correspond to those negative eigenvalues $s_{\mathbf{k}}$ which are close to zero. Kobayashi [1991b] discusses a simple method to identify *dominant types* of traffic that provide a tight bound on the probability of cell blocking due to buffer overflows. Such procedures should be extended to deal with the time-dependent case as well.

The multiplexer transient analysis that we obtained in this paper should be useful in developing a mathematical basis for formulating flow control models, such as admission control. In a high-speed network, the conventional feed-back control scheme based on the *steady state* analysis will fail, since by the time some information on traffic congestion (such as cell blocking at the multiplexer level or call blocking at the network level) is detected and sent to the originating sources, it will be too late for the network to take corrective actions. Thus it is clear that some type of *predictive* control must be formulated, and our result of transient analysis will be valuable in developing such design and control procedures.

There are a few important areas for further investigations which are necessary to make our results useful to practical applications. One area is computational complexity aspects: e.g., to efficiently identify the dominant terms to compute a given performance measure. The second area is a generalized traffic source model instead of the simple exponential model that we have dealt with. Elwalid, Mitra and Stern [1991], and Stern and Elwalid [1991] discuss a Markov model with many states. An extension of our time-dependent solution technique to their source model should be of significant value. In order to cope with the computational complexity a general theory of asymptotic approximations and bounding arguments should be explored.

Another promising avenue will be to formulate the multiplexer model based on the diffusion approximation as has been discussed by Kobayashi [1983] in dealing with multiple access protocols such as Aloha random access protocol.

References

[1] Anick, D., D. Mitra, and M. M. Sondhi [1982]. "Stochastic Theory of a Data-Handling System with Multiple Sources." *The Bell System Technical Journal* 61(8): 1871-1894.

[2] Bellman, R. [1960]. *Introduction to Matrix Analysis.* Mc-Graw Hill, Inc. 1960.

[3] Cohen, J. W. [1974]. "Superimposed Renewal Process and Storage with Gradual Input." *Stochastic Process and Their Applications* 2(January): 31-58.

[4] Elwalid, A. I., D. Mitra, and T. E. Stern [1991]. "Statistical Multiplexing of Markov Modulated Sources: Theory and Computational Algorithms" *Proceedings of 13th International Teletraffic Congress (ITC-13)*: North-Holland Publishing Co., pp. 495-500.

[5] Feller, W. [1966]. *An Introduction to probability Theory and Its Applications: Volume II.* John Wiley & Sons, Inc. 1966.

[6] Friedman, B. [1956]. *Principles and Techniques of Applied Mathematics.* John Wiley & Sons, Inc. 1956.

[7] Hashida, O., and M. Fujiki [1973]. "Queueing Models for Buffer Memory in Store-and-Forward System." *Proceedings of 7th International Teletraffic Congress (ITC-7)*, Stockholm, June 1973: 323/1-323/7.

[8] Kobayashi, H. [1978]. *Modeling and Analysis: An Introduction to System Performance Evaluation Methodology.* Addison-Wesley Publishing Co.

[9] Kobayashi, H. [1983]. "Diffusion Approximations in Queueing Analysis." In Louchard, G., and G. Latouche (eds.), *Probability Theory and Computer Science*, pp.86-98. Academic Press.

[10] Kobayashi, H. [1990a]. "Performance Issues of Broadband ISDN, Part I: Traffic Characterization and Statistical Multiplexing." *Proc. the 10th International Conference on Computer Communication*, pp. 349-355. New Delhi: Narosa Publishing House.

[11] Kobayashi, H. [1990b]. "Performance Issues of Broadband ISDN, Part II: Statistical Multiplexing of Multiple Types of Traffic." *Proc. the 10th International Conference on Computer Communication*, pp. 355-361. New Delhi: Narosa Publishing House.

[12] Kobayashi, H. [1991]. "A Spectral Representation Approach to Statistical Multiplexing of Multiple Types of Traffic." *Proc. 1991 IEEE International Symposium on Information Theory*, p.156. Budapest, Hungary, IEEE Catalog Number 91CH3003-1

[13] Kosten, L. [1974]. "Stochastic Theory of a Multi-Entry Buffer (1)." *Delft Progress report* 1 (1974): 10-18.

[14] Kosten, L. [1984]. "Stochastic Theory of Data Handling Systems with Groups of Multiple Sources." In H. Rudin and W. Bux (eds.), *Performance of Computer-Communication Systems*, pp. 321-331. North-Holland Publishing Co.

[15] Mitra, D. [1988]. "Stochastic Theory of a Fluid Model of Producers and Consumers Coupled by a Buffer." *Advances in Applied Probability* 20: 646-676.

[16] Neuts, M. F. [1981]. *Matrix Geometric Solutions in Stochastic Models.* John Hopkins University Press, Baltimore 1981.

[17] Ren, Q. and H. Kobayashi [1992]. "Transient Solutions for the Buffer Behavior in Statistical Multiplexing." *Proc. Twenty-Sixth Annual Conference on Information Sciences and Systems*, Princeton University, Princeton, N.J., March 1992.

[18] Stern, T. E. [1984]. "Analysis and Computational Techniques for Packet Voice System Performance Evaluation." *Proc. Conference on Information Sciences and Systems*, (1984): 415-419.

[19] Stern, T. E. and A. I. Elwalid [1991]. "Analysis of Separable Markov-Modulated Rate Models for Information-Handling Systems." *Advances in Applied Probability* 23. March 1991.

[20] Syski, R. [1960, 1986]. *Introduction to Congestion Theory in Telephone Systems:* First edition published by Oliver and Boyd, Edinburgh (1960). Second edition published by North-Holland Publishing Co. as *Studies in Telecommunication, Vol. 4* (1986).

[21] Takács, L. [1962]. *Introduction to the Theory of Queues*, Oxford University Press, Inc., New York (1962).