Diffusion Process Approximations of a Statistical Multiplexer with Markov Modulated Bursty Traffic Sources*

Qiang Ren†
NEC USA, Inc.
C&C Research Laboratories
4 Independence Way
Princeton, NJ 08540

Hisashi Kobayashi
Department of Electrical Engineering
Princeton University
Princeton, NJ 08544

Abstract

We consider a statistical multiplexer model, in which each of N sources is a Markov modulated rate process (MMRP). This formulation allows a more realistic source model than the well studied but simple "on-off" source model in characterizing variable bit rate (VBR) sources such as compressed video, which is of increasing importance to ATM networks. In our model we allow an arbitrary distribution for the duration of each of M states (or levels) that the source can take on. We formulate Markov modulated sources as a closed queueing network with M infinite-server stations. By extending our earlier results [13, 19] we introduce an M-dimensional diffusion process to approximate the aggregated traffic of such Markov modulated sources. Under a set of reasonable assumptions we then show that this diffusion process can be expressed as an M-dimensional Ornstein-Uhlenbeck (O-U) process.

The behavior of buffer content is also approximated by a diffusion process, which is characterized by the aggregated traffic process and the output process.

We show some numerical examples which illustrate typical sample paths, and auto-correlations of the aggregated traffic from the Markov modulated sources and its diffusion process representation. Simulation results are provided to compare with our diffusion model for queueing analysis.

1 Introduction

In the future B-ISDN (Broadband Integrated Services Digital Network), multiple types of information (e.g. video, image, voice and high-speed data) services will be provided by means of fast packet switching with statistical multiplexers. The traffic into a statistical multiplexer is a superposition of packet streams from a large number of sources of differing types.

There have been a number of noteworthy efforts to characterize multiple "on-off" sources that are statistically multiplexed in a packet-switching node. A *fluid-flow* model with single type "on-off" sources was formulated by Hashida and Fujiki [6] and Kosten [15]. Anick, Mitra and Sondhi [1] present a comprehensive analysis of this fluid model. Kosten [16] extends [1] to multiple types of traffic. The "on-off" source model will be an appropriate model, when the source is a voice or data source. As an alternative approach to model voice and data traffic in packet-switching environment, Heffes and Lucantoni [7] and others discuss applications of a Markov Modulated Poisson Process (MMPP) representation for the superposed traffic. The Markov modulated source model discussed by Elwalid, Mitra and Stern [5] is a generalization of [7] with multiple Markovian states for each source.

In the B-ISDN environment, however, video traffic will make a significant part of the network traffic. Unlike voice and data sources, the well studied two-valued "on-off" representations will not be appropriate to characterize variable bit rate (VBR) traffic such as compressed video. It has been known [17, 18] that video traffic can be modeled more appropriately as a multi-valued rate process.

The traffic model we propose in this paper is distinct from the previous work in two ways. First, we are introducing and considering Markov Modulated Rate Process (MMRP) as source model. Second and more importantly, we introduce a diffusion process approximation for the superposition of such Markov modulated rate processes. This approach has an advantage that the model is not restricted to cases with the exponential distribution as for the duration of each state. The diffusion approximation model also allows us to obtain the transient solution as well as the steady state solution more efficiently than the previous solution methods, where computational complexity grows exponentially as the number of states and the number of multiplexed sources increase.

The diffusion process model we present in this paper is a multivariate Ornstein-Uhlenbeck (O-U) process, which was originally introduced as a refinement of the Brownian motion (see e.g. Feller [4]). Kobayashi et al. [10] discusses the O-U representation of the multiple access scheme of the ALOHA channel. In our recent [13, 19] work, we have shown that the O-U process provides a good approximation to characterize the superposition of traffic from a heterogeneous set of "on-off" sources.

This paper is to generalize the O-U process approximation to the Markov-modulated rate source models. We formulate the model by

*This research has been supported, in part, by a grant from the Ogasawara Foundation for the Promotion of Science and Technology; by the Innovative Partnership Program of the New Jersey Commission on Science and Technology; and by the NCIPT (National Center for Integrated Photonic Technology) through its ARPA grant.

†This research was conducted when the author was with Department of Electrical Engineering, Princeton University.

0-7803-1820-X/94 $4.00 © 1994 IEEE

observing that the behavior of the superposition of the N Markov-modulated rate processes with M states can be represented as a closed queueing network with N customers and M stations with each station being represented by infinite servers (or equivalently N parallel servers). We then further to approximate the buffer content process by a diffusion process as in [14] and analyze its queueing behavior.

2 Formulation of the Model

The system is composed of a statistical multiplexer and N independent homogeneous sources (Figure 1). Each source is governed by an M-state Markov chain $\mathbf{P} = \{p_{lm}\},\ l, m = 1, 2, \cdots, M$. When a source is in state m it generates packets at rate R_m [packets/second]. The duration or burst period of state m has a general distribution with mean α_m^{-1} and variance σ_m^2. When the source exits state l, it moves to state m with probability p_{lm}. Figure 2 depicts state transition diagram of a single source, and data stream from such source.

Let us define an M-dimensional process $\mathbf{n}(t)$

$$\mathbf{n}(t) = [n_1(t), n_2(t), \cdots, n_M(t)]^T, \qquad (1)$$

where $n_m(t)$ denote the number of sources in state m at time t, $m = 1, 2, \cdots, M$. Clearly we have

$$\sum_{m=1}^{M} n_m(t) = N. \qquad (2)$$

The superposed traffic to the multiplexer input at time t can be defined as

$$R(t) = \sum_{m=1}^{M} R_m n_m(t). \qquad (3)$$

The transmission link has a constant capacity C [packets/second]. Hence the change of the buffer content $Q(t)$ can be represented as

$$\frac{dQ(t)}{dt} = \begin{cases} R(t) - C, & \text{when } R(t) > C \text{ or } Q(t) > 0, \\ 0, & \text{otherwise.} \end{cases} \qquad (4)$$

As discussed in [12], a source in state m can be viewed as a customer attended by one of the N parallel servers at "station" m with mean service time α_m^{-1}. In this closed queueing network representation, there will be no queue at any station, since, N, the number of parallel servers is the same as the numbers of the customers in the network. This is depicted in Figure 3, in which $A_m(t)$ and $D_m(t)$ are the arrival and departure counting processes to station m (i.e. the total numbers of arrivals at and departures from station m up to time t) of the queueing network. The queueing network of Figure 3 is a variant of the machine servicing model well-discussed in the literature (see e.g. [11] pp.144). When there are n_m "customers" at "station" m, the mean rate of departure is given by $\alpha_m n_m$.

Upon the completion of "service" at "station" l, customers route to state m with probability p_{lm}, $m = 1, 2, \cdots, M$. Thus the arrival process at station m, $A_m(t)$, is the aggregation of those departures from other stations which route to station m.

Let $D_{l,m}(t)$ represent the counting process of customers which move from station l to station m. Clearly

$$A_m(t) = \sum_{l=1}^{M} D_{l,m}(t), \text{ and } D_m(t) = \sum_{i=1}^{M} D_{m,i}(t). \qquad (5)$$

Let an M-dimensional process

$$\mathbf{X}(t) = [X_1(t), X_2(t), \cdots, X_M(t)]^T$$

be a continuous-state Markov process approximation of the discrete-level function $\mathbf{n}(t) = [n_1(t), n_2(t), \cdots, n_M(t)]^T$. The process $\mathbf{X}(t)$ must satisfy the same constraint equation as Eq.(2), i.e., $\sum_{m=1}^{M} X_m(t) = N$. When $X_m(t) = x_m$, its mean departure rate is given by

$$u_m^{-1}(x_m) = \alpha_m x_m. \qquad (6)$$

We denote the variance of departure rate as $\sigma_m^2(x_m)$, and its coefficient of variation as $c_m(x_m) = \sigma_m^2(x_m)/u_m^2(x_m)$. Then the infinitesimal mean vector $\mathbf{b}(\mathbf{x})$ of the vector process $\mathbf{X}(t)$ is given by

$$\begin{aligned} \mathbf{b}(\mathbf{x}) &= \begin{bmatrix} b_1(\mathbf{x}) \\ \vdots \\ b_M(\mathbf{x}) \end{bmatrix} \\ &= \begin{bmatrix} -\alpha_1 & \alpha_2 p_{21} & \cdots & \alpha_M p_{M1} \\ \alpha_1 p_{12} & -\alpha_2 & \cdots & \alpha_M p_{M2} \\ \vdots & \vdots & & \vdots \\ \alpha_1 p_{1M} & \alpha_2 p_{2M} & \cdots & -\alpha_M \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix} \\ &= \mathcal{B}\mathbf{x}, \end{aligned} \qquad (7)$$

where $\mathcal{B} = \{\beta_{mm'}\}_{M \times M}$.
Similarly the infinitesimal covariance matrix $\mathcal{A}(\mathbf{x}) = \{a_{mm'}(\mathbf{x})\}_{M \times M}$ is given by

$$\begin{aligned} a_{m,m'}(\mathbf{x}) &= \sum_{l=1}^{M} \{(c_l(x_l) - 1)/u_l(x_l)\} p_{lm} p_{lm'} \\ &+ \left\{ c_m(x_m)/u_m(x_m) + \sum_{l=1}^{M} (p_{lm}/u_l(x_l)) \right\} \delta_{mm'} \\ &- \left(\frac{c_m(x_m)}{u_m(x_m)} \right) p_{mm'} - \left(\frac{c_{m'}(x_{m'})}{u_{m'}(x_{m'})} \right) p_{m'm}. \end{aligned} \qquad (8)$$

Note that in the Markov modulated source, we have $p_{mm} = 0$ for $m = 1, \cdots, M$. In other words, when the holding time in one state expires, the source always shifts to a different state. Then the expression for $\mathcal{A}(\mathbf{x})$ can be simplified:

$$\mathcal{A}(\mathbf{x}) = \sum_{l=1}^{M} \left(\frac{c_l(x_l)}{u_l(x_l)} \right) \mathbf{v}_l \cdot \mathbf{v}_l^T + \mathcal{W}(\mathbf{x}), \qquad (9)$$

where $\mathbf{v}_l$ is an M-dimensional column vector whose l-th element is unity and the m-th element ($m \neq l$) is $-p_{lm}$, i.e.

$$\mathbf{v}_l = [-p_{l1}, \cdots, 1, \cdots, -p_{lM}]^T.$$

Matrix $\mathcal{W}(\mathbf{x})$ is an $M \times M$ matrix whose element is

$$w_{mm'}(\mathbf{x}) = \sum_{l=1}^{M} [p_{lm}(\delta_{mm'} - p_{lm'})/u_l(x_l)]$$

for $1 \le m, m' \le M$. It is easy to show that $\mathcal{W}$ is nonnegative-definite.

Note that $\mathcal{W}(\mathbf{x})$ depends only on the Markov chain $\mathbf{P} = \{p_{lm}\}$, the mean burst periods α_m^{-1}, $1 \le m \le M$, and the system state $\mathbf{x}$.

For more detailed derivations of the above arguments, readers are referred to [9] and references therein.

3 Derivation of the Diffusion Approximation Model

With the formulation given in Section 2, we can show that the M-dimensional diffusion process $\mathbf{X}(t)$ is governed by the stochastic differential equation

$$d\mathbf{X}(t) = \mathcal{B} \cdot \mathbf{X}(t)dt + \sqrt{\mathcal{A}(\mathbf{x})} \cdot d\mathbf{W}(t), \qquad (10)$$

where $\mathbf{W}(t)$ is an M-dimensional Brownian motion with zero mean and the covariance (matrix) function $I\delta(t)$, where I is the $M \times M$ identity matrix and $\delta(t)$ is the impulse function.

Both $\mathcal{B}$ and $\mathcal{A}(\mathbf{x})$ are singular matrices due to the fact $\sum_{m=1}^{M} x_m = N$. By definition, $\mathcal{A}(\mathbf{x})$ is a positive semi-definite matrix. Hence $\sqrt{\mathcal{A}(\mathbf{x})}$ always exists and is uniquely defined.

Let $\mathbf{x}^* = (x_1^*, x_2^*, \cdots, x_M^*)$ be the equilibrium state of process $\mathbf{X}(t)$ such that

$$\mathbf{b}(\mathbf{x}^*) = 0. \qquad (11)$$

Then we can write

$$\mathbf{b}(\mathbf{x}) = \mathcal{B} \cdot (\mathbf{x} - \mathbf{x}^*). \qquad (12)$$

If we consider a narrow region around $\mathbf{x} = \mathbf{x}^*$, we can approximate $\mathcal{A}(\mathbf{x})$, the infinitesimal variance matrix, by its value at $\mathbf{x} = \mathbf{x}^*$:

$$\mathcal{A}(\mathbf{x}) \approx \mathcal{A}(\mathbf{x}^*) \stackrel{def}{=} \mathcal{A}. \qquad (13)$$

With the linear infinitesimal mean $\mathbf{b}(\mathbf{x})$ of Eq.(12) and the approximated constant infinitesimal variance $\mathcal{A}$ of Eq.(13), the stochastic differential equation (10) becomes

$$d\mathbf{X}(t) = \mathcal{B} \cdot (\mathbf{X}(t) - \mathbf{x}^*)dt + \sqrt{\mathcal{A}} \cdot d\mathbf{W}(t), \qquad (14)$$

which is a multi-variate *Ornstein-Uhlenbeck* process, leading to the following differential equation for f, the conditional probability density function of $\mathbf{X}(t)$:

$$\frac{\partial f}{\partial t} = -\sum_{m=1}^{M}\sum_{n=1}^{M} \beta_{mn} \frac{\partial}{\partial x_m}[(x_n - x_n^*)f] + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M} a_{mn} \frac{\partial^2 f}{\partial x_m \partial x_n}. \qquad (15)$$

The solution for Eq.(15) has a multi-dimensional Gaussian distribution with mean $\bar{\mathbf{x}}(t) = \mathbf{x}^* + e^{t\mathcal{B}}\mathbf{x}_0$ and variance $\Xi(t) = \int_0^t e^{(t-\tau)\mathcal{B}} \mathcal{A} e^{(t-\tau)\mathcal{B}^T} d\tau$. In steady state, we have

$$\bar{\mathbf{x}} = \mathbf{x}^*, \quad \Xi = \int_0^{+\infty} e^{\mathcal{B}t} \mathcal{A} e^{\mathcal{B}^T t} dt, \qquad (16)$$

with Ξ satisfying

$$\mathcal{B}\Xi + \Xi\mathcal{B}^T + \mathcal{A} = 0, \qquad (17)$$

which is a special case of *Lyapunov equation* [2].

As we noted earlier, $\mathcal{B}$ is a singular matrix, thus Ξ is not the unique solution of Eq.(17). To overcome this problem, we consider $(M-1)$ free variates $x_1, x_2, \cdots, x_{M-1}$ and find the following relationship between the density functions using the indicator function $\mathcal{X}_{\{\cdot\}}$:

$$f(x_1, \cdots, x_M) = f(x_1, \cdots, x_{M-1}) \cdot \mathcal{X}_{\{x_M = (N - \sum_{m=1}^{M-1} x_m)\}}. \qquad (18)$$

Thus $\mathcal{B}$ and $\mathcal{A}$ in Eqs.(7) and (13) are modified to $(M-1) \times (M-1)$ matrices $\mathcal{B}_1 = \{\beta'_{mn}\}$ and $\mathcal{A}_1 = \{a'_{mn}\}$, where

$$\beta'_{mn} = \beta_{mn} - \beta_{mM}, \quad m, n = 1, \cdots, M-1; \qquad (19)$$

$$a'_{mn} = a_{mn}, \quad m, n = 1, \cdots, M-1. \qquad (20)$$

Then we have a unique solution of a symmetric positive-definite Ξ' for $f(x_1, \cdots, x_{M-1})$ from Eq.(17).

4 Aggregated Traffic Process

As we have assumed earlier, a source at state m generates packets at constant rate R_m. So we can define the superposed traffic to the multiplexer input at time t as

$$R(t) = \sum_{m=1}^{M} R_m X_m(t). \qquad (21)$$

Without loss of generality we can further assume $R_M = 0$ (i.e., a source is always off at state M) so that we only consider processes $X_1(t), \cdots, X_{M-1}(t)$.

Note that Ξ' is a symmetric positive-definite matrix, thus it can be diagonalized as $\Xi' = Q^T \Lambda Q$, where Λ is a diagonal matrix with element λ_i, $i = 1, \cdots, M-1$, being the eigenvalues of Ξ', and the row vectors of matrix Q consist of the corresponding orthonormal eigenvectors .

Define $\mathbf{Z}(t) = Q\mathbf{X}(t)$, then $\mathbf{Z}(t)$ is a Gaussian process with mean $Q\mathbf{x}^*$ and covariance matrix Λ, which implies $Z_1(t), Z_2(t), \cdots, Z_{M-1}(t)$ are orthogonal (hence independent) Gaussian processes. Therefore the superposed traffic

$$R(t) = \sum_{m=1}^{M-1} R_m X_m(t) = [R_1, R_2, \cdots, R_{M-1}] \cdot \mathbf{X}(t) = [R_1, \cdots, R_{M-1}]Q^T \cdot \mathbf{Z}(t) = \sum_{m=1}^{M-1} R'_m Z_m(t), \qquad (22)$$

where $R'_m = \sum_{i=1}^{M-1} R_i Q_{mi}$. $R(t)$ is a weighted summation of independent Gaussian processes, therefore it is also a stationary Gaussian process, whose mean and variance are

$$\begin{aligned} \lim_{t\to\infty} E[R(t)] &= \sum_{m=1}^{M-1} R_m x_m^*, \\ \lim_{t\to\infty} Var[R(t)] &= \sum_{m=1}^{M-1} \lambda_m R_m'^2. \end{aligned} \tag{23}$$

But generally $R(t)$ is no longer a Markov process.

We define $A(t)$ as the accumulated arrival process for the aggregated traffic $R(t)$:

$$A(t) \stackrel{def}{=} \int_0^t R(t)dt = \sum_{m=1}^{M-1} R_m \int_0^t X_m(t)dt, \tag{24}$$

and $\mathbf{Y}(t) \stackrel{def}{=} [X_1(t), \cdots, X_{M-1}(t)]^T$, $\mathbf{y}^* \stackrel{def}{=} [x_1^*, \cdots, x_{M-1}^*]^T$. Then, from Eq.(14), $\mathbf{Y}(t)$ satisfies the following stochastic differential equation

$$d\mathbf{Y}(t) = \mathcal{B}_1(\mathbf{Y}(t) - \mathbf{y}^*)dt + \sqrt{\mathcal{A}_1}d\mathbf{W}(t). \tag{25}$$

We can show that for a sufficiently large t, the mean and variance of $\int_0^t \mathbf{Y}(t)dt$

$$E[\int_0^t \mathbf{Y}(t)dt] \sim \mathbf{y}^* t + o(t), \tag{26}$$

$$Var[\int_0^t \mathbf{Y}(t)dt] \sim \mathcal{B}_1^{-1}\mathcal{A}_1(\mathcal{B}_1^{-1})^T t + o(t), \tag{27}$$

which implies

$$\lim_{t\to\infty} \frac{E[A(t)]}{t} \sim \sum_{m=1}^{M-1} R_m x_m^*, \tag{28}$$

$$\lim_{t\to\infty} \frac{Var[A(t)]}{t} \sim \mathcal{R}\mathcal{B}_1^{-1}\mathcal{A}_1(\mathcal{B}_1^{-1})^T\mathcal{R}^T. \tag{29}$$

where $\mathcal{R} \stackrel{def}{=} [R_1, \cdots, R_{M-1}]$.

5 A Diffusion Model for Queue Process $Q(t)$

In this section we form a diffusion process $q(t)$ to approximate the buffer content process $Q(t)$ by using the approach in [14].

From Eqs.(28) and (29) we approximate the accumulated arrival process $A(t)$ by a diffusion process which is still denoted as $A(t)$ for convenience. By doing so, the underlying diffusion process captures the original process through its first and second order statistics in equilibrium, which are given in Eqs. (28) and (29),

$$dA(t) = b \cdot dt + \sqrt{a} \cdot dW(t), \tag{30}$$

where a and b are constants defined as

$$b = \sum_{m=1}^{M-1} R_m x_m^*, \tag{31}$$

$$a = \mathcal{R}\mathcal{B}_1^{-1}\mathcal{A}_1(\mathcal{B}_1^{-1})^T\mathcal{R}^T. \tag{32}$$

From Eq.(4) the output process $c(t)$ for transmission can be approximated by

$$dc(t) = \begin{cases} Cdt, & \text{when } Q(t) > 0 \text{ or } R(t) > C, \\ \eta dt, & \text{otherwise,} \end{cases} \tag{33}$$

where η is an unknown constant and may be approximated by $E[R|R < C]$ as suggested in [14].

The diffusion process $q(t)$ therefore satisfies the following stochastic differential equation[1]

$$\begin{aligned} dq(t) &= dA(t) - dc(t) \\ &= \begin{cases} (b - C)dt + \sqrt{a}dW(t), & q(t) > 0, \\ (b - \eta)dt + \sqrt{a}dW(t), & q(t) \le 0. \end{cases} \end{aligned} \tag{34}$$

It can be shown [14] that the stationary probability $Prob[q > x]$ ($= \lim_{t\to\infty} Prob[q(t) > x]$) is given by

$$Prob[q > x] = \frac{b - \eta}{C - \eta} exp[-2\frac{C - b}{a}x]. \tag{35}$$

The exponent approximates the ***tail-end*** distribution (i.e., for a large value of x) of $Prob[Q > x]$ and the coefficient $\frac{b-\eta}{C-\eta}$ is used to approximate $Prob[Q > 0]$. As discussed in [13], $Prob[R > C]$ can serve as a lower-bound approximation of $Prob[Q > 0]$.

6 Numerical Examples and Simulation Results

We give some numerical examples to illustrate and verify our modeling and analysis methods. The simulations are conducted for both the superposed (or aggregated) traffic of Markov modulated sources and its corresponding diffusion process approximation.

We consider the superposed traffic of 100 independent four-state Markov modulated sources, i.e. $N = 100$ and $M = 4$. The Markov generator for each source is given by

$$\mathbf{P} = \{p_{lm}\}_{4\times4} = \begin{bmatrix} 0, & \frac{1}{3}, & \frac{1}{2}, & \frac{1}{6} \\ \frac{1}{5}, & 0, & \frac{2}{5}, & \frac{2}{5} \\ \frac{3}{10}, & \frac{1}{2}, & 0, & \frac{1}{5} \\ 0, & \frac{1}{2}, & \frac{1}{2}, & 0 \end{bmatrix}.$$

Let

$$R_1 = 1.4,\ R_2 = 5.0,\ R_3 = 2.7 \text{ and } R_4 = 0.0;$$
$$\alpha_1 = 1.0,\ \alpha_2 = 3.0,\ \alpha_3 = 2.0 \text{ and } \alpha_4 = 7.0,$$

and we assume that the burst periods are exponentially distributed (with means $\alpha_1^{-1}, \alpha_2^{-1}, \alpha_3^{-1}$ and α_4^{-1}, respectively).

[1]Note that $q(t)$ is a diffusion approximation of $Q(t)$ and can take negative values.

Figure 4 shows a simulated sample path of a typical packet generation process from a single source as defined above. Figure 5 shows a simulated sample path of *superposed* traffic stream from 100 independent sources, each of which generates the four-level bursty traffic similar to Figure 4. Figure 6 shows a simulated sample path of the diffusion process (as defined in Eq.(14) with corresponding $\mathcal{B}$ and $\mathcal{A}$), which is an approximation of the superposed traffic in Figure 5. The distributions of the superposed traffic from the Markov modulated sources and its diffusion process representation are plotted in Figure 7. They are compared with the Gaussian distribution which we have obtained analytically from Eq.(23) and has a mean of 260.06 and a variance of 228.65. Figure 8 shows the auto-correlations of the superposed traffic from the Markov modulated sources and of its diffusion process representation. Figure 9 provides some simulation results of buffer overflow probabilities from the 100 MMRP sources with their parameters given as above. They are compared with the tail-end distributions obtained from the diffusion approximation by using Eq.(35).

In Figure 10 we plot buffer overflow probabilities of MMRP sources (from simulations) and compare them with the ones by using diffusion approximations (from Eq.(35)). In addition to the case of exponentially distributed burst periods, we show the results for two other holding-time distributions: (1) 2-stage Erlangian distribution; (2) 2-stage hyper-exponential distribution, with same means $\alpha_1^{-1}, \alpha_2^{-1}, \alpha_3^{-1}$ and α_4^{-1} as the exponential case.

7 Conclusion and Discussion

We have formulated diffusion process models to analyze the superposed traffic streams from many MMRP sources in a statistical multiplexer and then analyze its queueing behavior. Our source model is more general than those assumed by most of the previous studies: each source has a finite number of states which are governed by a discrete Markov chain. At each Markovian state, its duration (i.e., burst period) has a *general* distribution and packets are generated at the constant rate determined by that state. We showed that the total number of multiplexed traffic sources, N, introduces no extra computational complexity (compared with a single source model). Furthermore, the accuracy of the diffusion process model will improve as N becomes larger.

Although we treated the case of single type homogeneous sources, our approach can be generalized directly to a system with multiple types of traffic — each type of traffic is modeled as treated in this paper and the overall process is simply a sum of these components, i.e., the quantities in (28) and (29) for the heterogeneous case can be represented as the sums of the corresponding quantities for each single type traffic case.

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): pp.1871-1894.

[2] Bellman, R. [1970], *Introduction to Matrix Analysis*, McGraw-Hill Book Company, 1970.

[3] Cox, D.R. and H.D. Miller [1965], *Theory of Stochastic Process*, Methuen & Co Ltd.

[4] Feller, W. [1966], *An Introduction to Probability Theory and Its Applications*, Vol. II. Wiley, New York, 1966.

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

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

[7] Heffes, H. and D.M. Lucantoni [1986], "A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance", *IEEE Journal on Selected Areas in Communications*, **SAC-4**(6): pp.856-868.

[8] Van Kampen, N.G. [1992], *Stochastic Processes in Physics and Chemistry*, North-Holland.

[9] Kobayashi, H. [1974], "Application of the Diffusion Approximation to Queueing Networks, Part I: Equilibrium Queue Distribution", *Journal of the ACM*, **21**(2): pp.316-328.

[10] Kobayashi, H., Y. Onozato and D. Huynh [1977], "An Approximation Method for Design and Analysis of An ALOHA System", *IEEE Trans. on Communications*, **COM-25**(1), pp.148-157. Also in *Multiple Access Communications: Foundations for Emerging Technologies*, Norman Abramson (ed.), IEEE Press 1992, pp.262-271.

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

[12] Kobayashi, H. [1990], "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.

[13] Kobayashi, H. and Q. Ren [1993], "A Diffusion Approximation Analysis of an ATM Statistical Multiplexer with Multiple Types of Traffic, Part I: Equilibrium State Solutions", *Proc. the 1993 IEEE International Conference on Communications*, Vol.2, pp.1047-1053, Geneva, Switzerland.

[14] Kobayashi, K. [1994], "Steady State Approximation Analysis For ATM Multiplexer By Diffusion Process Without Reflection Barrier", *Proc. the Symposium on Performance Models for Information Communications Networks*, pp.309-320, Hakone, Japan, Jan. 19-21, 1994.

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

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

[17] Maglaris, B., D. Anastassiou, P. Sen, G. Karlsson and J.D. Robbins [1988], "Performance Models of Statistical Multiplexing in Packet Video Communications", *IEEE Trans. on Communications*, **36**(7): pp.834-844.

[18] Melamed, B. and B. Sengupta [1992], "TES Modeling of Video Traffic", *IEICE Trans. on Communications*, Vol. E75-B, No.12, December 1992, pp.1292-1300.

[19] Ren, Q. and H. Kobayashi [1993], "A Diffusion Approximation Analysis of an ATM Statistical Multiplexer with Multiple Types of Traffic, Part II: Time-Dependent Solutions", *Proc. the 1993 IEEE Global Telecommunications Conference*, Vol.2, pp.772-777, Houston, Texas, Nov.29-Dec.2, 1993.

[20] Sriram, K. and W. Whitt [1986], "Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data", *IEEE Journal on Selected Areas in Communications*, **4**(6): pp.1124-1132.

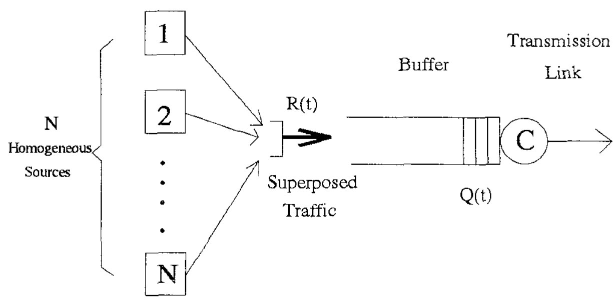
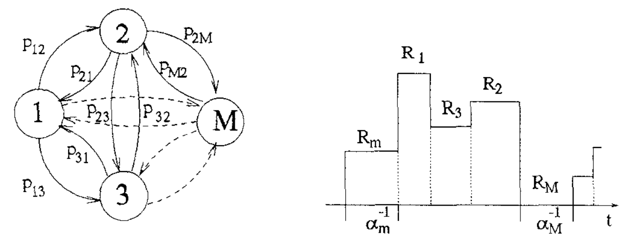


Figure 1: A buffered statistical multiplexer and N homogeneous sources.



(a) The state transition diagram for a single source

(b) A typical packet generation process from a single source

Figure 2: The state transition diagram and packet generation rate process from a single source.

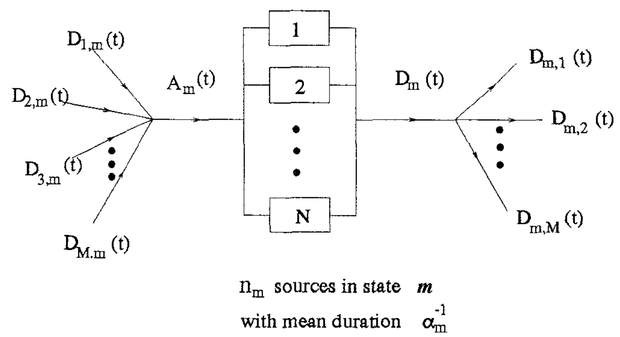


Figure 3: Queueing network representation of $n_m(t)$ –the number of sources in state m.

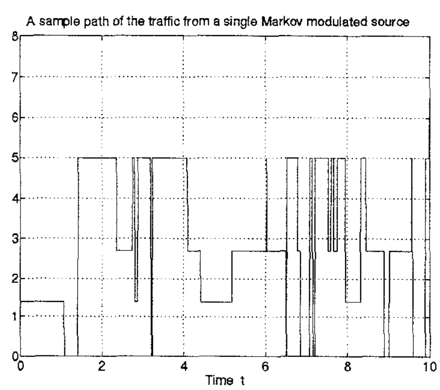


Figure 4: A sample path of the traffic process from a single 4-state Markov modulated source with the parameters provided in Section 6.

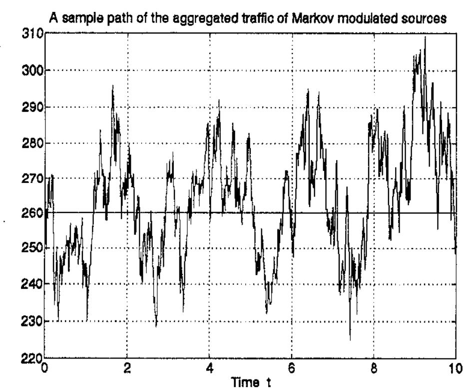


Figure 5: A sample path of the aggregated traffic process of 100 Markov modulated sources, of which has a pattern statistically similar to the one depicted in Figure 4.

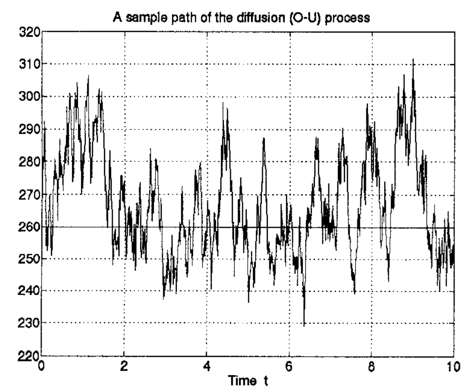


Figure 6: A sample path of the diffusion process, which approximates the aggregated traffic stream depicted in Figure 5.

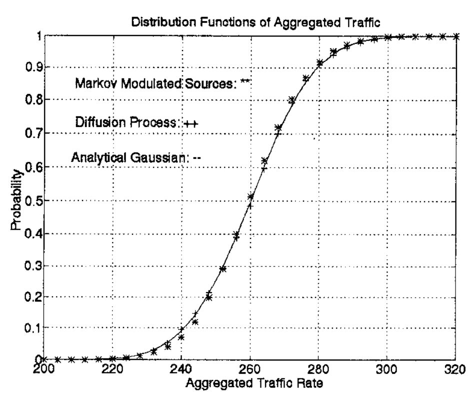


Figure 7: The cumulative distribution functions of the aggregated traffic process of Markov modulated sources ('$*$'), its diffusion process representation ('$+$') and the analytically derived Gaussian process (solid line).

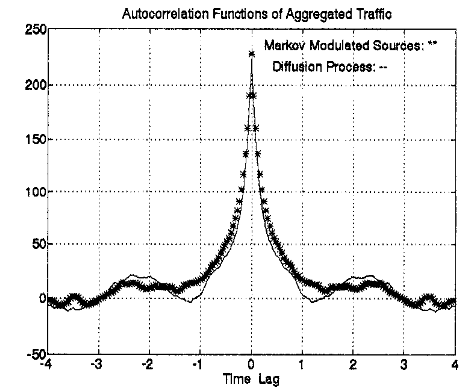


Figure 8: The auto-correlation functions of the aggregated traffic process ('$*$') and its diffusion approximation process (solid line).

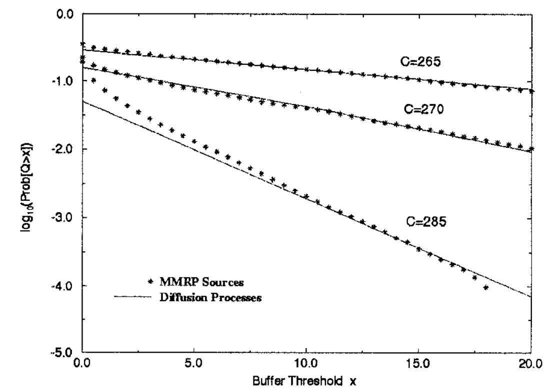


Figure 9: The buffer overflow probabilities $Prob[Q > x]$: diffusion approximation of tail-end distribution from Eq.(35) vs. simulation results with 100 MMRP sources in Section 6 under different link capacities, i.e., $C = 265, 270$ and 285.

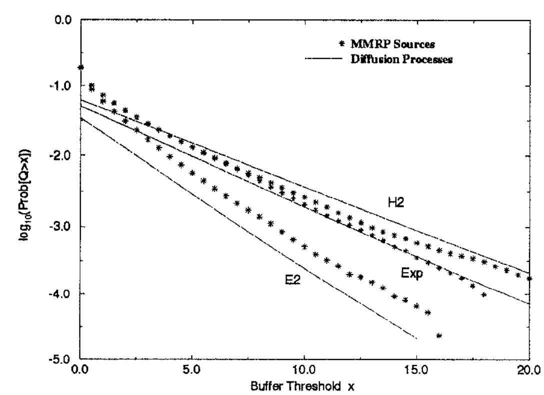


Figure 10: The buffer overflow probabilities $Prob[Q > x]$: diffusion approximation of tail-end distribution from Eq.(35) vs. simulation results with 100 MMRP sources in Section 6 under different holding-time distributions: 2-stage hyper-exponential, exponential, 2-stage Erlangian. Link capacity $C = 285$.