ELSEVIER

Computer Communications 22 (1999) 1266–1273

computer communications

Research note

Methodology for traffic modeling using two-state Markov-modulated Bernoulli process

C.-H. Ng*, L. Yuan, W. Fu, L. Zhang

School of Electrical & Electronic Engineering, Nanyang Technological University, Blk S2, Nanyang Technological University, Nanyang Avenue, Singapore, Singapore 639798

Received 9 June 1998; received in revised form 20 April 1999; accepted 20 April 1999

Abstract

In this paper, we systematically present the methodology of modeling bursty traffic sources using the 2-state Markov Modulated Bernoulli Process (called MMBP-2). The technique used can be easily extended to an *m*-state MMBP though the numerical calculation is complicated. We first defined the parameters of the model and some processes associated with it, and subsequently examined the queue length distribution of an infinite buffer driven by an MMBP-2 with batch arrivals. We next looked at the case where the same buffer is fed by a group of two identical MMBP-2 sources. Instead of deriving the queue length expression, we cast the problem in the framework of the previous case and made use of the previous results with some modification. Lastly, we looked at the case of a finite buffer driven by two MMBP-2 sources with different parameters. We formulated the queue length solution in the framework of Markov theory and calculated the Cell Loss Probability (CLP) for this case. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Cell loss probability; Markov theory; Asynchronous transfer mode

1. Introduction

Traffic models and queueing analysis are the two main pillars of the analytical performance evaluation of telecommunications networks. Traditionally, renewal type of traffic models, such as Poisson or Bernoulli processes, were often used to characterize the arrival process of data packets in continuous-time or discrete-time queueing analysis due to their elegant analytical properties and mathematical simplicity. With the advent of Asynchronous Transfer Mode (ATM) networks that are capable of delivering data, voice and video in an integrated manner, these models fail to reflect the correlation structure and bursty nature that are inherently present in multi-media traffic. Realistic source models reflecting the correlation chracteristics of a multimedia traffic need to be found for the study of ATMs performance.

Many kinds of traffic models and associated approaches have been proposed to model and analyze the various components of a multi-media traffic to ATM networks. Auto-Regressive (AR) model and its variants was used to model the output bit rate of variable-bit-rate encoders [1,2]. Transform-Expansion-Sample (TES) was introduced in Refs. [3,4]. Recently, self-similar model was used in Ref. [5].

By far, the Markov-modulated arrival models are the most commonly used models in recent literature pertaining to ATM networks. They constitute an extremely important class of traffic models. Unlike renewal traffic, Markov-modulated traffic models introduce dependence into the random arriving stream, and hence they can capture traffic burstiness because of non-zero autocorrelations in the arriving stream.

2. Markov-modulated arrival model

Markov-modulated arrival (MMA) [6,7] process is a term coined for a broad class of statistical traffic models in which the intensity of the arrival process to a queueing system depends on the states of an independent irreducible Markov chain. The idea is to introduce an explicit notion of dependency among the arrivals of a traffic stream—i.e. an auxiliary Markov process is evolving in time and its current state controls the probability law of the traffic mechanism. This class of models allows for generalizations that cover a wide range of applications where the arrival processes are bursty and exhibit correlations among arrivals. Time scale in these models can be continuous or discrete in nature, and arrivals

---

* Corresponding author. Tel.: + 65-790-5634; fax: + 65-792-0415.
E-mail address: echng@ntu.edu.sg (C.-H. Ng)

PII: S0140-3664(99)00114-0

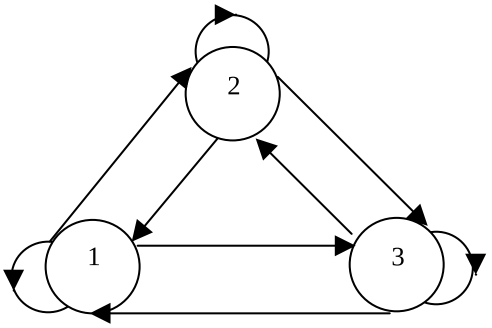

Fig. 1. A 3-state discrete-time Markov chain.

may be single units, a batch of units or a continuous quantity. Depending on the combination of the two, we have the following broad categories of models that appear frequently in literature:

- Markov-modulated Poisson Process (MMPP): Discrete arrival unit, continuous-time.
- Markov-modulated Bernoulli Process (MMBP): Discrete arrival unit, discrete-time.
- Markov-modulated Fluid-Flow (MMFF): Continuous arrival unit, continuous-time.

Markov-modulated Bernoulli Process (MMBP) [8–10] is the discrete-time version of the MMA. Time in MMBP is discretized into fixed-length slots (hereafter referred as time slots). Arrivals in discrete units (messages, packets or cells) are generated by the source whose stochastic behavior is governed by a finite (m-state) irreducible discrete-time Markov chain. The Markov chain makes a transition from state i to state j with a probability p_{ij} after spending a geometric duration of time slots in state i or it may stay in the same state in the next time slot as shown in Fig. 1. A transition probability matrix as given below governs the movement of the Markov Chain:

$$\tilde{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ p_{m1} & \cdots & \cdots & p_{mm} \end{bmatrix}. \tag{1}$$

As the probability of going from one state to all other states should sum to one, we have $\sum_{j=1}^{m} p_{ij} = 1$ for the above matrix. The steady state probability distribution $\tilde{\pi} = [\pi_1, \pi_2, \cdots, \pi_m]$ of the Markov chain is then given by $\tilde{\pi}\tilde{P} = \tilde{\pi}$ with the usual condition $\sum_{i=0}^{m} \pi_i = 1$.

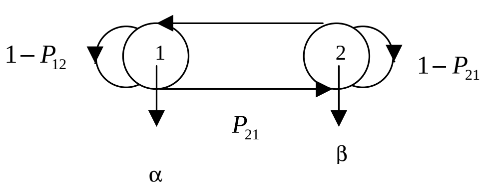

Fig. 2. An MMBP-2.

The arrivals generated in a time-slot is a Bernoulli process with a probability α_i, which depends solely on the current state (state i) of the process. Hence, an MMBP is jointly characterized by the above transition probability matrix $\tilde{P}$ and a diagonal matrix $\tilde{\Lambda}_{\mathrm{MMBP}}$ of arrival probabilities:

$$\tilde{\Lambda}_{\mathrm{MMBP}} = \mathrm{diag}(\alpha_1, \alpha_2, \cdots, \alpha_m). \tag{2}$$

It is assumed that the evolution of the underlying modulating Markov chain is independent of the arrival process.

3. Two-state MMBP

In this paper, we will systematically present the modeling of bursty traffic sources such as variable-bit-rate (VBR) traffic as MMBPs and its associated queueing analysis. We restrict ourselves to the case of two-state MMBP (hereafter, referred as MMBP-2) because it greatly facilitates the analysis and presentation. The extension to any m-state Markov chain is possible using the same techniques, but the expression is complicated and the numerical calculation is tedious.

We model each individual traffic source as an MMBP-2 [10] as shown in Fig. 2. When the source is in state-1, it generates an arrival with probability α and may go to state-2 in the next time slot with probability p_{12}. Similarly, when it is in state-2, it generates an arrival with probability β and may transit to state-1 in the next time slot with probability p_{21}. Without loss of generality, we further assume that arrivals occur more frequently in state-1, i.e. $\alpha \gg \beta$. The width of the time slots is taken to be the time required to transmit a data unit (hereafter, we use cells interchangeably).

The described model is a generalization of the well-known On/Off model [11]. On/Off model is quite successful in modeling voice traffic soruces due to their active/silent nature. But silent state is almost non-existing in VBR video traffic sources. The above model with a high-level activity state (state-1) and a low-level activity state (state-2) can represent this situation more gracefully. Furthermore, if the α and β in the above model are set to 1 and 0 respectively, the model will become the On/Off model.

From the discussion in Section 2, we know that an MMBP-2 is fully characterized by the following two matrices:

$$\tilde{P} = \begin{bmatrix} 1-p_{12} & p_{12} \\ p_{21} & 1-p_{21} \end{bmatrix} \quad \text{and} \quad \tilde{\Lambda} = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}. \tag{3}$$

The duration that the Markov chain spends in both states are geometrically distributed as

$$\left.\begin{array}{ll} \Pr[\text{duration} = x] = (1-p_{12})^{x-1} p_{12} & \text{state-1} \\ \Pr[\text{duration} = y] = (1-p_{12})^{y-1} p_{12} & \text{state-2} \end{array}\right\} \tag{4}$$

where x and y are in terms of time slots.The steady-state probability distribution $\tilde{\pi} = [\pi_1, \pi_2]$ is given by

$$\pi_1 = \frac{p_{21}}{p_{12} + p_{21}} \qquad \text{and} \qquad \pi_2 = \frac{p_{12}}{p_{12} + p_{21}}. \tag{5}$$

And the source utilization ρ, which is the probability that a slot contains an arrival, is then given by

$$\rho = \pi_1 \alpha + \pi_2 \beta = \frac{\alpha p_{21} + \beta p_{12}}{p_{21} + p_{12}}. \tag{6}$$

The generating function $T(z)$ of the inter-arrival time of an MMBP is given by [12] as

$$T(z) = bz\{I - z\tilde{P}(I - \tilde{\Lambda})\}^{-1}\tilde{P}\tilde{\lambda} \tag{7}$$

where I is the identity matrix and $b = \pi\tilde{\Lambda}/\pi\tilde{\lambda}$ with $\tilde{\lambda} = (\alpha, \beta)^{\mathrm{T}}$. Expanding $T(z)$ with the corresponding matrices, we have

$$\begin{aligned} T(z) &= bz\{I - z\tilde{P}(I - \tilde{\Lambda})\}^{-1}\tilde{P}\tilde{\lambda} \\ &= \frac{z[\pi_1\alpha \quad \pi_2\beta]}{\pi_1\alpha + \pi_2\beta}\left\{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - z\begin{bmatrix} 1 - p_{12} & p_{12} \\ p_{21} & -p_{21} \end{bmatrix}\right. \\ &\quad \left. \times \begin{bmatrix} 1-\alpha & 0 \\ 0 & 1-\beta \end{bmatrix}\right\}^{-1} \times \begin{bmatrix} 1-p_{12} & p_{12} \\ p_{21} & 1-p_{21} \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \\ &= \frac{c_2 z^2 + c_1 z}{d_2 z^2 + d_1 z + d_0} \end{aligned} \tag{8}$$

where

$$c_2 = (1 - p_{21} - p_{12})[p_{21}(1-\beta)\alpha^2 + p_{12}(1-\alpha)\beta^2]$$

$$c_1 = -p_{21}\alpha[(1-p_{12})\alpha + p_{12}\beta] + p_{12}\beta[(1-p_{21})\beta + \alpha p_{21})]$$

$$d_2 = (1-\alpha)(1-\beta)(p_{12} + p_{21} - 1)(\alpha p_{12} + \beta p_{21})$$

$$d_1 = [(1-\beta)(1-p_{21}) + (1-\alpha)(1-p_{12})](\alpha p_{21} + \beta p_{12})$$

$$d_0 = -(\alpha p_{21} + \beta p_{12}).$$

4. Matching traffic parameters to MMBP-2

Voice and video traffic sources are two typical VBR traffic sources with different traffic generation characteristics. A voice source alternates between active and silent periods, while a video source generates a continuous bit stream at varying rates and almost has no silent period. Video is presented to users as a series of frames in which the motion of the scene is reflected in small changes in sequential displayed frames. Therefore, scenes with rapid motion will cause much higher bit rates than those with slow motion. This characteristic lands video traffic source well to this model. Various parameters frequently used to characterise a bursty source such as video include:

- Peak cell rate ($\hat{p}$), which is the maximum number of cells the source generates during one time period.
- Mean cell rate($\bar{m}$), which is the average number of cells the source generates during one time period.
- Traffic burstiness ($\hat{B}$), which is usually defined as the ratio of the peak cell rate to the mean cell rate, i.e. $\hat{B} = \hat{p}/\bar{m}$.

To match the source parameters to the model, we first observe that the peak cell rate ($\hat{p}$) occurs when the source is in state 1, and can be shown to be equal to α if a time slot is taken as a measuring unit as follows:

$$\hat{p} = \frac{1}{\sum_{k=1}^{\infty} k(1-\alpha)^{k-1}\alpha} = \alpha. \tag{9}$$

Secondly, the mean cell rate ($\bar{m}$) is equal to the normalized load generated by the source or in other words the source utilisation (ρ) as given by expression (6). The mean cell rate can also be interpreted as the expected probability of cell arrivals in a time slot. The burstiness ($\hat{B}$) is by above definition defined as

$$\hat{B} = \frac{\hat{p}}{\bar{m}} = \frac{\hat{p}}{\rho} = \frac{\alpha(p_{21} + p_{12})}{\alpha p_{21} + \beta p_{12}}. \tag{10}$$

There are four independent model parameters $(p_{12}, p_{21}, \alpha, \beta)$ but there are only two independent source parameters because $\hat{B}$ is dependent on $\hat{p}$ and $\bar{m}$. Hence, corresponding to one set of source parameters, there is more than one set of possible model parameters. As long as the model parameters do not introduce any conflict with the source parameters provided, they are acceptable.

If we define a new variable R as the ratio of the time a traffic source spent in high-activity state (state-1) to the time spent in low-activity state (state-2),

$$R = \frac{\pi_1}{\pi_2} = \frac{p_{21}}{p_{12}} \tag{11}$$

then R determines the upper boundary of burstiness in the parameter matching process as

$$\hat{B} = \frac{1+R}{R + (\beta/\alpha)} < B_{\max} = \frac{1+R}{R} \quad \text{as} \quad 1 \gg \beta/\alpha > 0. \tag{12}$$

Using the definitions of R, $\hat{B}$, $\hat{p}$ and $\bar{m}$, the source parameters can be matched to model parameters in the following steps:

1. Select a group of α and β according to the values of R and $B_{\max}$.
2. $\beta = \bar{m} \times \hat{B}$.
3. $\alpha = (1 + R) \times \bar{m} - R \times \beta$.

It should be noted that the model parameters are only used for the purpose of calculation. They do not have direct physical meaning in real life. Instead, it is the relationship between analysis results and the source parameters that can give the intuition on traffic management.

5. Batch MMBP-2/D/1 queue

We now consider the case where an infinite buffer is driven by a traffic source that is modeled as an MMBP-2. Cells arrive at the buffer according to the dynamics of an MMBP-2, queue in the buffer for some time, and finally get transmitted according to their order of arrivals. It is assumed that the MMBP-2 source is synchronized with the transmission and cells arrive at the beginning of a time slot. One and only one cell is transmitted at the end of each time slot if there is any in the buffer.

Without loss of generality, let us assume that the single output channel takes one unit of time to transmit a cell, so the service time is deterministic and equal to one. From the model of MMBP-2, we see that if the arrivals are from one MMBP-2 source only and the source utilization is less than one, then there is no queue because every cell can be served within the arriving time-slot.

To have a meaningful analysis of the buffer queue, let us introduce the notion of a batch arrival. We assume that the MMBP-2 generates an arrival with a batch size of n ($n = 0, 1, \ldots, \infty$) with probability G_n while in state-1 and correspondingly in state-2 with a probability F_n. The generating functions of the batch size in both states are assumed to be

$$G(z) = \sum_{n=0}^{\infty} G_n z^n \qquad \text{while in state-1} \tag{13}$$

$$F(z) = \sum_{n=0}^{\infty} F_n z^n \qquad \text{while in state-2.} \tag{14}$$

Let the queue length of the system $q(k+1) = j, j = 0, 1, \ldots, \infty$ and the phase of the Markov chain $ph(k+1) = i, i = 1, 2$ at the end of slot $(k+1)$. We define the z-transform of the joint probability $\Pr[q(k+1) = j, ph(k+1) = i]$ as

$$Q_i^{k+1}(z) = \sum_{j=0}^{\infty} P[q(k+1) = j, ph(k+1) = i] z^j \tag{15}$$

$$= \sum_{j=0}^{\infty} \pi_{ji}^{k+1} z^j.$$

In matrix form, we have

$$Q^{k+1}(z) = \begin{bmatrix} Q_1^{k+1}(z) \\ Q_2^{k+1}(z) \end{bmatrix} = \begin{bmatrix} \pi_{01}^{k+1} z^o & \pi_{11}^{k+1} z^o & \pi_{21}^{k+1} z^o & \cdots & \pi_{\infty 1}^{k+1} z^o \\ \pi_{02}^{k+1} z^o & \pi_{12}^{k+1} z^o & \pi_{22}^{k+1} z^o & \cdots & \pi_{\infty 2}^{k+1} z^o \end{bmatrix}. \tag{16}$$

According to the queueing model described earlier, the buffer queue will evolve according to the following:

$$q(k+1) = [q(k) - 1]^+ + b(k+1) \tag{17}$$

where $b(k+1)$ is the number of cells arrived during slot $(k+1)$. If we expand $Q_1^{k+1}(z)$ and $Q_2^{k+1}(z)$ in terms of those quantities in slot k using (16), we will arrive at the following expression after some lengthy algebraic manipulation:

$$Q^{k+1}(z) = \begin{bmatrix} p_{11} G(z) \left(\frac{1}{z} Q_1^k(z) + \pi_{01} - \frac{\pi_{01}}{z} \right) + p_{21} F(z) \left(\frac{1}{z} Q_2^k(z) + \pi_{02} - \frac{\pi_{02}}{z} \right) \\ p_{12} G(z) \left(\frac{1}{z} Q_1^k(z) + \pi_{01} - \frac{\pi_{01}}{z} \right) + p_{22} F(z) \left(\frac{1}{z} Q_2^k(z) + \pi_{02} - \frac{\pi_{02}}{z} \right) \end{bmatrix} \tag{18}$$

$$= \frac{1}{z} \begin{bmatrix} G(z) & 0 \\ 0 & F(z) \end{bmatrix} \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix} Q^k(z) \begin{bmatrix} \pi_{01} z - \pi_{01} \\ \pi_{02} z - \pi_{02} \end{bmatrix}.$$

If we assume that the total arrivals during a slot is less than one,

$$\pi_1 \alpha \frac{\mathrm{d}}{\mathrm{dz}} G(z)\big|_{z=1} + \pi_2 \beta \frac{\mathrm{d}}{\mathrm{dz}} F(z)\big|_{z=1} < 1 \tag{19}$$

then the queue will reach an equilibrium and we have

$$\lim_{k\to\infty} Q^{k+1}(z) = \lim_{k\to\infty} Q^k(z) = \lim_{k\to\infty} Q(z). \tag{20}$$

Hence, expression (18) can be re-written as follows after going through the limit:

$$Q(z) = \left[I - \frac{1}{z} \Lambda_B \tilde{P}^{\mathrm{T}} \right]^{-1} \frac{1}{z} \Lambda_B \tilde{P}^{\mathrm{T}} \Phi \tag{21}$$

where

$$\Lambda_B = \mathrm{diag}(G(z), F(z)) = \begin{bmatrix} G(z) & 0 \\ 0 & F(z) \end{bmatrix} \tag{22}$$

$$\Phi = \begin{bmatrix} \pi_{01} z - \pi_{01} \\ \pi_{02} z - \pi_{01} \end{bmatrix}. \tag{23}$$

The z-transform of the queue length can be found as

$$q(z) = \sum_{i=0}^{\infty} \Pr[q = j] z^j \tag{24}$$

$$= \sum_{i=1}^{2} Q_i(z) = eQ(z)$$

where $e = (1, 1)$. The average queue length can then be obtained by taking the derivative of $eQ(z)$ at $z = 1$.

$$E[q] = eQ'(z)\big|_{z=1} \tag{25}$$

$$= e[I - \tilde{P}^{\mathrm{T}}]^{-1}\tilde{P}^{\mathrm{T}}\Phi'$$

where $\Phi' = [\pi_{01}, \pi_{02}]^{\mathrm{T}}$ is a column vector of a group of initial conditions.

6. Multiple identical MMBP-2 sources

We next look at the case where an infinite buffer is fed by a group of ω identical MMBP-2 sources. For simplicity and illustration purposes, we again look at the simplest case where there are only two identical sources ($\omega = 2$). Instead of doing the analysis from scratch [10], we will cast the problem in the framework of the previous analysis and make use of the results with some modification.

If we examine expressions (3) and (21) carefully, we see that the matrix $\tilde{P}$ basically describes the motion of the underlying Markov chain from slot k to $k + 1$. If we define $a(k + 1)$ as the number of MMBP-2 sources in state-1 in slot $(k + 1)$, and $a(k)$ the corresponding quantity in slot k, then we can describe the dynamics of this group of sources by relating $a(k + 1)$ to $a(k)$ in a matrix A similar to $\tilde{P}$. Basically, $a(k + 1)$ is made up of two components:

1. $\sum_{j=1}^{a(k)} c_j$: A group of $a(k)(= 0, 1, 2)$ sources which were in state-1 during slot k, remain in state-1 when transit to slot $(k + 1)$. Here, c_j is an indicator function defined as

$$c_j = \begin{cases} 1 & c_j \in \text{state-1} \quad \text{with prob.} \quad 1 - p_{12} \\ 0 & c_j \notin \text{state-1} \quad \text{with prob.} \quad p_{12} \end{cases}. \tag{26}$$

2. $\sum_{j=1}^{N-a(k)} d_j$: A group of $(2 - a(k))$ sources which were in state-2 during slot k, but makes a transition to state-1 in slot $(K + 1)$. Again, d_j is an indicator function defined as

$$d_j = \begin{cases} 1 & d_j \in \text{state-1} \quad \text{with prob.} \quad p_{21} \\ 0 & d_j \notin \text{state-1} \quad \text{with prob.} \quad (1 - p_{21}) \end{cases}. \tag{27}$$

Define matrix A whose (i,j) terms as

$$A(i,j) = \Pr[a(k + 1) = j \big| a(k) = i] \tag{28}$$

and we have

$$A = \begin{bmatrix} (1 - p_{21})^2 & 2(1 - p_{21})p_{21} & p_{21}^2 \\ p_{12}(1 - p_{12}) & (1 - p_{12})(1 - p_{21}) + p_{12}p_{21} & (1 - p_{12})p_{21} \\ p_{12}^2 & 2(1 - p_{12})p_{12} & (1 - p_{12})^2 \end{bmatrix}. \tag{29}$$

Next, we see that $G(z)$ and $F(z)$ basically describe the number of cells arrive at the buffer per time-slot when the MMBP-2 is in state-1 and state-2, respectively. Let us define $b(k)$ as the total number of cells arrive during slot k. Again, $b(k)$ is made up of two terms:

1. $\sum_{j=1}^{a(k} s_j$: Cells generated by those sources in state-1 during slot k. Here, s_j is an indicator function defined as

$$s_j = \left\{ \begin{array}{lll} 1 & s_j \in \text{a cell generated} & \text{with prob. } \alpha \\ 0 & \textit{otherwise} & \text{with prob. } (1 - \alpha) \end{array} \right]. \tag{30}$$

2. $\sum_{j=1}^{N-a(k} t_j$: Cells generated by those sources in state-2 during slot k. Again, t_j is an indicator function defined as

$$t_j = \left\{ \begin{array}{lll} 1 & t_j \in \text{a cell generated} & \text{with prob. } \beta \\ 0 & \text{otherwise} & \text{with prob. } (1 - \beta) \end{array} \right]. \tag{31}$$

Define matrix B whose i row has an z-transform as

$$B_i(z) = \sum_{j=0}^{2} \Pr[b(k) = j \big| a(k) = i]z^j \tag{32}$$

then we have

$$\left.\begin{array}{c} B_0(z) = [(1 - \beta) + \beta z]^2 \\ B_1(z) = [(1 - \alpha) + \alpha z][(1 - \beta) + \beta z] \\ B_2(z) = [(1 - \alpha) + \alpha z]^2 \end{array}\right\} \tag{33}$$

and

$$B = \begin{bmatrix} (1 - \beta)^2 & 2\beta(1 - \beta) & \beta^2 \\ (1 - \beta)(1 - \alpha) & \beta(1 - \alpha) + \alpha(1 - \beta) & \alpha\beta \\ (1 - \alpha)^2 & 2\alpha(1 - \alpha) & \alpha^2 \end{bmatrix}. \tag{34}$$

If we define a diagonal matrix Λ_ω as

$$\Lambda_\omega = \mathrm{diag}(B_0(z), B_1(z), B_2(z)) = \begin{bmatrix} [(1 - \beta) + \beta z]^2 & 0 & 0 \\ 0 & [(1 - \alpha) + \alpha z][(1 - \beta) + \beta z] & 0 \\ 0 & 0 & [(1 - \alpha) + \alpha z]^2 \end{bmatrix} \tag{35}$$

then we readily see that $A\&\Lambda_\omega$ play the roles of $\tilde{P}\&\Lambda_B$ and hence the queue length solution in matrix form for this case is the same as before and given by expression (21) with matrices $\Lambda_B \& \tilde{P}$ replaced by $\Lambda_\omega \& A$. The queue length expression and the mean can then be found using (24) and (25).

7. Finite buffer with multiple heterogeneous MMBP-2

The superposition of a group of MMBP-2s can be viewed as traffic on a single queuing buffer fed by multiple bursty traffic sources. These traffic sources are independent of each other, they may have different parameters such as traffic burstiness and load (source utilization). Therefore, we now turn our attention to the case where a buffer is fed by a group of MMBP-2 sources with different parameters. We further limit the capacity of the buffer to L ($L > \omega$). Because of the finite queue length, we will not be able to derive a similar matrix equation due to the discontinuity in the queue length (truncation at L); instead we resort to Markov theory.

Unlike the case of identical sources in which we care only about the number of sources in State-1, we need to differentiate which source is in state-1 in this situation. Hence, we use transition state $a'(k)$ to replace $a(k)$ as shown below.

State of source 1	State of source 2	No. of sources in state-1 $a(k)$	Tansition state $a'(k)$
State 2	State 2	0	0
State 2	State 1	1	1
State 1	State 2	1	2
State 1	State 1	2	3

Similar to matrix A, we define matrix A' whose (i,j) terms as

$$A'(i,j) = \Pr[a'(k+1) = j | a'(k) = 1] \tag{36}$$

then we have

$$A = \begin{bmatrix} p_{22}p''_{22} & p_{22}p''_{21} & p_{21}p''_{22} & p_{21}p''_{21} \\ p_{22}p''_{12} & p_{22}p''_{11} & p_{21}p''_{12} & p_{21}p''_{11} \\ p_{12}p''_{22} & p_{12}p''_{21} & p_{11}p''_{22} & p_{11}p''_{21} \\ p_{12}p''_{12} & p_{12}p''_{11} & p_{11}p''_{12} & p_{11}p''_{11} \end{bmatrix} \tag{37}$$

where $p_{11} = 1 - p_{12}$ and $p_{22} = 1 - p_{21}$ as shown in Fig. 2. Parameters of source 1 are marked with $'$ while parameters of source 2 are marked with $''$.

The cells arrive at the buffer can be differentiated using the generation state $b'(k)$ as shown below:

Cell from source 1	Cell from source 2	No. of cells arrived $b(k)$	Generation state $b'(k)$
0	0	0	0
0	1	1	1
1	0	1	2
1	1	2	3

Define matrix B' whose (i,j) terms as

$$B' = \Pr[b'(k) = j | a'(k) = i]. \tag{38}$$

Then, we have

$$B' = \begin{bmatrix} (1-\beta')(1-\beta'') & (1-\beta')\beta'' & \beta'(1-\beta'') & \beta'\beta'' \\ (1-\beta')(1-\alpha'') & (1-\beta')\alpha'' & \beta'(1-\alpha'') & \beta'\alpha'' \\ (1-\alpha')(1-\beta'') & (1-\alpha')\beta'' & \alpha'(1-\beta'') & \alpha'\beta'' \\ (1-\alpha')(1-\alpha'') & (1-\alpha')\alpha'' & \alpha'(1-\alpha'') & \alpha'\alpha'' \end{bmatrix}. \tag{39}$$

Note that A and B are 3×3 matrices while A' and B' are 4×4 matrices.

As the buffer size is limited at L, the queue length $q(k+1)$ in slot $(k+1)$ of the system will evolve as follows:

$$q(k+1) = \min\{([q(k) - 1]^+ + b(k+1)), L\}. \tag{40}$$

As the evolution of $q(k)$ and $a'(k)$ completely describe the system, we define $\pi_{ji} = P[q(k) = j, a'(k) = i]$ as the probability of observing the queue with j cells and the two sources at transition state i. The transition probability of this bivariate Markov Chain $\{q(k), a'(k)\}$ is then given by:

$$P\{q(k+1) = m, a'(k+1) = j | q(k) = n, a'(k) = i\}. \tag{41}$$

If we linearize the state probability to form a vector

$$\tilde{\pi} = (\pi_0, \pi_1, \ldots \pi_L) \tag{42}$$

$$= (\pi_{00}, \pi_{01}, \pi_{02}\pi_{03}, \pi_{10}, \pi_{11}, \pi_{12}, \pi_{13}, \ldots \pi_{L0}, \pi_{L1}, \pi_{L2}, \pi_{L3})$$

then we have $\tilde{\pi}(k+1) = \tilde{\pi}(k)\tilde{T}$, where $\tilde{T}$ is a $4L \times 4L$ matrix defined below:

$$\tilde{T} = \begin{bmatrix} C_0 & C_{1+2} & C_3 & 0 & \cdots & \cdots & 0 & 0 & 0 & 0 \\ C_0 & C_{1+2} & C_3 & 0 & \cdots & \cdots & 0 & 0 & 0 & 0 \\ 0 & C_0 & C_{1+2} & C_3 & \cdots & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & C_0 & C_{1+2} & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & C_0 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & C_3 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & C_{1+2} & C_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & \cdots & C_0 & C_{1+2} & C_3 & 0 \\ 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & C_0 & C_{1+2} & C_3 \\ 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & C_0 & C_{1+2+3} \end{bmatrix} \tag{43}$$

where

$C_{1+2} = C_1 + C_2$ and $C_{1+2+3} = C_1 + C_2 + C_3$

$C_i (i = 0, 1, 2, 3)$ are 4×4 sub-matrices defined as follows:

$$C_i = A' \times E \qquad \text{and} \qquad E = \mathrm{diag}(b'_{1i}, b'_{2i}, b'_{3i}, b'_{4i}) \tag{44}$$

where $(b'_{1i}, b'_{2i}, b'_{3i}, b'_{4i})$ are the corresponding entries in (39). The sub-matrices C_i describe the changes of the number of cell arrivals when the transition state is i.

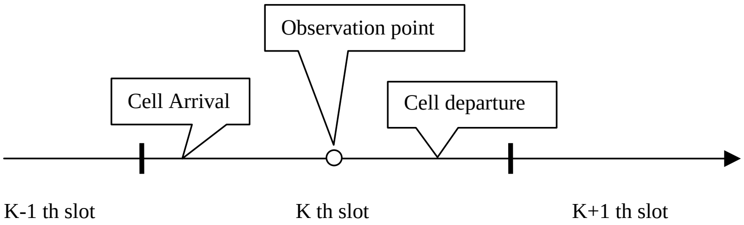

Fig. 3. The time slot diagram.

Hence, C_i is calculated with the columns of A' which represent the probability of having i sources in state-1, with the jth column weighted by the $B'(i,j)$ which is the probability of observing j arrivals with i sources in state-1. Using the parameters for an MMBP-2 described earlier; the matrices C_i are given as follows:

$$C_0 = \begin{bmatrix} P'_{22}P''_{22}(1-\beta')(1-\beta'') & P'_{22}P''_{21}(1-\beta')(1-\alpha'') & P'_{21}P''_{22}(1-\alpha')(1-\beta'') & P'_{21}P''_{21}(1-\alpha')(1-\alpha'') \\ P'_{22}P''_{12}(1-\beta')(1-\beta'') & P'_{22}P''_{11}(1-\beta')(1-\alpha'') & P'_{21}P''_{12}(1-\alpha')(1-\beta'') & P'_{21}P''_{11}(1-\alpha')(1-\alpha'') \\ P'_{12}P''_{22}(1-\beta')(1-\beta'') & P'_{12}P''_{21}(1-\beta')(1-\alpha'') & P'_{11}P''_{22}(1-\alpha')(1-\beta'') & P'_{11}P''_{21}(1-\alpha')(1-\alpha'') \\ P'_{12}P''_{12}(1-\beta')(1-\beta'') & P'_{12}P''_{11}(1-\beta')(1-\alpha'') & P'_{11}P''_{12}(1-\alpha')(1-\beta'') & P'_{11}P''_{11}(1-\alpha')(1-\alpha'') \end{bmatrix}$$

$$C_1 = \begin{bmatrix} P'_{22}P''_{22}(1-\beta')\beta'' & P'_{22}P''_{21}(1-\beta')\alpha'' & P'_{21}P''_{22}(1-\alpha')\beta'' & P'_{21}P''_{21}(1-\alpha')\alpha'' \\ P'_{22}P''_{12}(1-\beta')\beta'' & P'_{22}P''_{11}(1-\beta')\alpha'' & P'_{21}P''_{12}(1-\alpha')\beta'' & P'_{21}P''_{11}(1-\alpha')\alpha'' \\ P'_{12}P''_{22}(1-\beta')\beta'' & P'_{12}P''_{21}(1-\beta')\alpha'' & P'_{11}P''_{22}(1-\alpha')\beta'' & P'_{11}P''_{21}(1-\alpha')\alpha'' \\ P'_{12}P''_{12}(1-\beta')\beta'' & P'_{12}P''_{11}(1-\beta')\alpha'' & P'_{11}P''_{12}(1-\alpha')\beta'' & P'_{11}P''_{11}(1-\alpha')\alpha'' \end{bmatrix}$$

$$C_2 = \begin{bmatrix} P'_{22}P''_{22}\beta'(1-\beta'') & P'_{22}P''_{21}\beta'(1-\alpha'') & P'_{21}P''_{22}\alpha'(1-\beta'') & P'_{21}P''_{21}\alpha'(1-\alpha'') \\ P'_{22}P''_{12}\beta'(1-\beta'') & P'_{22}P''_{11}\beta'(1-\alpha'') & P'_{21}P''_{12}\alpha'(1-\beta'') & P'_{21}P''_{11}\alpha'(1-\alpha'') \\ P'_{12}P''_{22}\beta'(1-\beta'') & P'_{12}P''_{21}\beta'(1-\alpha'') & P'_{11}P''_{22}\alpha'(1-\beta'') & P'_{11}P''_{21}\alpha'(1-\alpha'') \\ P'_{12}P''_{12}\beta'(1-\beta'') & P'_{12}P''_{11}\beta'(1-\alpha'') & P'_{11}P''_{12}\alpha'(1-\beta'') & P'_{11}P''_{11}\alpha'(1-\alpha'') \end{bmatrix}$$

$$C_3 = \begin{bmatrix} P'_{22}P''_{22}\beta'\beta'' & P'_{22}P''_{21}\beta'\alpha'' & P'_{21}P''_{22}\alpha'\beta'' & P'_{21}P''_{21}\alpha'\alpha'' \\ P'_{22}P''_{12}\beta'\beta'' & P'_{22}P''_{11}\beta'\alpha'' & P'_{21}P''_{12}\alpha'\beta'' & P'_{21}P''_{11}\alpha'\alpha'' \\ P'_{12}P''_{22}\beta'\beta'' & P'_{12}P''_{21}\beta'\alpha'' & P'_{11}P''_{22}\alpha'\beta'' & P'_{11}P''_{21}\alpha'\alpha'' \\ P'_{12}P''_{12}\beta'\beta'' & P'_{12}P''_{11}\beta'\alpha'' & P'_{11}P''_{12}\alpha'\beta'' & P'_{11}P''_{11}\alpha'\alpha'' \end{bmatrix}.$$

Again, if we assume that the total arrivals per slot is less than 1, then the system will reach equilibrium and we have

$$\tilde{\pi}\tilde{T} = \tilde{\pi} \qquad \text{and} \qquad \sum_{ji} \pi_{ji} = 1. \tag{45}$$

By solving this set of linear equations of (45), the stationary queue length distribution can be found as

$$P[q = j] = \pi_{j0} + \pi_{j1} + \pi_{j2} + \pi_{j3}. \tag{46}$$

8. Cell loss probability

For the case of a finite buffer fed with two MMBP-2 sources, a single arriving cell has no chance to encounter a full buffer and be lost since one cell in the queue is served at the beginning of the time slot. As shown in Fig. 3, loss of cells can only happen when simultaneous arrivals occur from both sources when the buffer is full in the previous time slot. In this case, one cell will be dropped and we must decide which cell to drop. The following scheme will be used: drop the cell from source 1 with probability ε' and the cell from source 2 with probability $\varepsilon'' = 1 - \varepsilon'$.

We define the Cell Loss Probability (CLP) as the ratio of the number of lost cells in a time slot to the number of arriving cells in a time slot [12]. The number of lost cells in a time slot (P_{Loss}) can be obtained as

$$P_{\text{Loss}} = \sum_{i=1}^{4} M_{1,i} \tag{47}$$

where M is a 1×4 matrix defined as

$$M = [\pi_{L0}, \pi_{L1}, \pi_{L2}, \pi_{L3}]C_3. \tag{48}$$

Hence, the number of lost cells in a time slot for source 1 and source 2 are

$$P'_{\text{Loss}} = \varepsilon' \times P_{\text{Loss}} \tag{49}$$

$$P''_{\text{Loss}} = \varepsilon'' \times P_{\text{Loss}}. \tag{50}$$

The number of arriving cells from a source in a time slot is equal to the source utilization ρ defined in Eq. (6). Therefore, the number of arriving cells from the two sources can be expressed as

$$P'_{\text{Arrival}} = \pi'_1\alpha' + \pi'_2\beta' \tag{51}$$

$$P''_{\text{Arrival}} = \pi''_1\alpha'' + \pi''_2\beta''. \tag{52}$$

And the total number arriving cells is

$$P_{\text{Arrival}} = P'_{\text{Arrival}} + P''_{\text{Arrival}}. \tag{53}$$

According to the definition of CLP, the CLP for source 1 and source 2 are then given by

$$\text{CLP}' = \frac{p'_{\text{Loss}}}{P'_{\text{Arrival}}} \qquad \text{source 1} \tag{54}$$

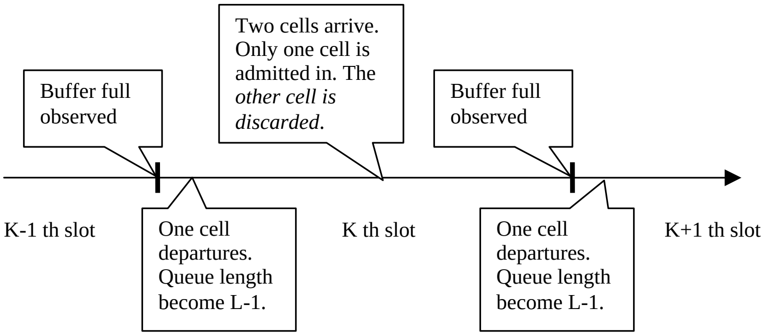

Fig. 4. The condition where cell loss happens.

$$\mathrm{CLP}'' = \frac{p''_{\mathrm{Loss}}}{P''_{\mathrm{Arrival}}} \qquad \text{source 2.} \tag{55}$$

The overall CLP for the arrivals as a whole is then given by

$$\mathrm{CLP}_{\mathrm{Overall}} = \frac{p_{\mathrm{Loss}}}{P_{\mathrm{Arrival}}} = \frac{P'_{\mathrm{Loss}} + P''_{\mathrm{Loss}}}{P'_{\mathrm{Arrival}} + P''_{\mathrm{Arrival}}}. \tag{56}$$

9. Conclusion

In this paper, we have systematically presented the methodology of modeling bursty traffic with 2-state MMBPs (called MMBP-2). We first defined the parameters of the model and some processes associated with it, and subsequently looked at the case of an infinite buffer driven by an MMBP-2 with batch arrivals. We derived a matrix equation for the queue length in z-transform and also the mean queue length expression.

We next looked at the case where the same buffer was fed by two identical MMBP-2 sources. Instead of deriving the queue length expression, we cast the problem in the framework of the previous case and made use of the results with some modification. Lastly, we looked at the case of a finite buffer which was fed with two MMBP-2 with different parameters and formulated the queue length solution in the framework of Markov chain. We also showed the cell loss probability in this case (Fig. 4).

References

[1] M. Nomura, T. Fuji, N. Ohta, Basic characteristics of variable rate video coding in ATM environment, IEEE JSAC June (1989) 752–760.

[2] C. Shim, et al., Modeling and call admission control algorithm of variable bit rate video in ATM networks, IEEE JSAC 12 (1994) 1.

[3] B. Melamed, TES: a class of methods for generating autocorrelated uniform varieties, ORSA Journal 3 (4) (1991) 317–329.

[4] B. Melamed, et al., TES-based traffic modelling for performance evaluation of integrated networks, INFOCOM'92 (1992) 75–84.

[5] N. Likhanov, B. Tsybakov, N. Georganas, Analysis of an ATM buffer with self-similar (fractal) input traffic, Proceedings of the IEEE INFOCOM'95, April 1995.

[6] H. Heffes, D. Lucantoni, A Markov-modulated characterization of packetized voice and data traffic and related statistical multiplexer performance, IEEE JSAC September (1986) 856–868.

[7] W. Fischer, H.K. Meier, The Markov-modulated Poisson process (MMPP) Cookbook, Performance Evaluation 18 (1992) 149–171.

[8] S.Q. Li, A general solution technique for discrete queueing analysis of multimedia traffic on ATM, IEEE Trans. Communications 39 (7) (1991) 1115–1132.

[9] N.Y. Le Boudec, An efficient solution method for Markov models of ATM links with loss priorities, IEEE Journal on Selected Areas in Comm. 9 (3) (1991) 408–417.

[10] C.H. Ng, L.Z. Bai, L.R. Zhang, Queue length solutions for an ATM buffer with MMBP arrivals, Comp. Comm. Journal 20 (10) (1997) 878–883.

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

[12] S. Monaghan, Exact analysis in discrete time of 2 × 2 ATM switch serving correlated traffic at both inputs, Electronics Letters August (1997).