Mean buffer contents and mean packet delay in statistical multiplexers with correlated train arrivals

Sabine Wittevrongel, Stijn De Vuyst and Herwig Bruneel

SMACS\textsuperscript{1} Research Group, Vakgroep Telecommunicatie en Informatieverwerking, University of Ghent, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium
Tel. : +32-9-2643415 Fax : +32-9-2644295 e-mail : sw@telin.rug.ac.be

Abstract

We consider a statistical multiplexer which is modeled as a discrete-time single-server queueing system. Messages consisting of a variable number of fixed-length packets arrive to the multiplexer at the rate of one packet per slot ("train arrivals"), which results in what we call a primary correlation in the packet arrival process. The distribution of the message lengths (in terms of packets) is general. Previous work put forward an analytic technique, based on the use of generating functions and an infinite-dimensional state description, for the analysis of the system in case the numbers of new messages generated by the user population in different slots are independent and identically distributed. Here we adapt this technique to handle the case where the arrival process contains an additional secondary correlation, which results from the fact that the distribution of the number of leading packet arrivals in a slot depends on some environment variable. We assume this environment to have two possible states, each with geometric sojourn times. Closed-form expressions are derived for the mean system contents and the mean packet delay. By means of some numerical examples we illustrate the effect of both primary and secondary correlation on the multiplexer performance.

Keywords : Discrete-time queueing model; Train arrivals; Variable-length messages; Markovian environment; Analytic study

1 Introduction

In the past several years there has been an increasing interest in discrete-time queueing models which take into account the correlation in traffic streams caused by the segmentation ([1]) at the edge of an ATM network of large external data frames, e.g. IP frames, into fixed-length ATM cells. In particular, a lot of effort has been devoted to the analysis of buffer systems where the customers are messages composed of multiple fixed-length packets, see e.g. [2]–[7]. In these models, time is divided into fixed-length intervals, slots, where one slot is the time period required to transmit exactly one packet from the buffer, and a message enters the buffer like a train at the rate of one packet per slot. Various distributions have been considered for the lengths of the messages (in terms of packets) : a geometric distribution in [2]–[4], constant-length messages in [5] and an arbitrary distribution in [6] and [7].

However, the analyses [2] to [7] all have one feature in common : the numbers of new messages generated by the user population during consecutive slots are independent and identically distributed (i.i.d.) random variables. In the present paper, we try to take a first step towards taking into account the possible correlation in the message generation process. Specifically, the distribution of the number of leading packet arrivals in a slot is assumed here to depend on the value of some environment variable, which represents the behavior of the user population. There are two possible environment states, each with geometrically distributed sojourn times.

The "correlated packet-train arrival process" considered in this paper clearly contains two types of correlation : (i) a primary correlation, which results from the fact that if the leading packet of a message of length $l$ arrives during the current slot, it is sure that the remaining $(l-1)$ packets of that message

\textsuperscript{1}SMACS: Stochastic Modeling and Analysis of Communication Systems
will consecutively enter the buffer in the next \((l-1)\) slots; (ii) a secondary correlation, which is due to the nonindependent generation of new messages. The purpose of the paper is to investigate the effect of both primary and secondary correlation on the behavior of the buffer system. To this end, we have adapted an analytic technique, developed in [6], which is based on the use of generating functions and an infinite-dimensional state description. As a result, closed-form expressions are obtained for the mean system contents and the mean packet delay.

The outline of the paper is as follows. The details of the mathematical model are explained in Section 2. The queueing analysis is carried through in Section 3. In Section 4, the influence of the primary and secondary correlation in the packet arrival stream on the multiplexer performance is discussed. Some concluding remarks are given in Section 5.

2 Mathematical model

The statistical multiplexer under study is modeled as a discrete-time single-server queueing system with infinite storage capacity and with a user population generating messages which consist of a variable number of fixed-length packets. These messages are then sent to the buffer at the rate of one packet per slot \((\text{train arrivals})\), a slot being the fixed amount of time it takes for the server to transmit one packet from the buffer. A new message, or a user becoming active, is thus seen by the buffer as the arrival of a leading packet in the current slot and one packet arrival in each of the following consecutive slots up to a total given by the length of the message. The message lengths (number of composing packets) are i.i.d. random variables with probability mass function \(l(n)\) and probability generating function (pgf) \(L(z) = \sum_{n=1}^{\infty} l(n) z^n\).

In each slot the user population is in one of two possible environment states, say \('1'\) and \('0'\), both of which have geometric sojourn times:

\[
\begin{align*}
\text{Prob}[\text{length of a '1'-period is } n] &= (1-\alpha) \alpha^{n-1}, n \geq 1; \\
\text{Prob}[\text{length of a '0'-period is } n] &= (1-\beta) \beta^{n-1}, n \geq 1.
\end{align*}
\] (1)

Because of the Markovian nature of the above distributions, we can determine \(t_k\), the environment state in slot \(k\), as

\[
t_k = \begin{cases} 
  a_k & \text{if } t_{k-1} = 1; \\
  b_k & \text{if } t_{k-1} = 0,
\end{cases}
\] (2)

whereby the \(a_k\) and \(b_k\) are rows of i.i.d. random variables with pgf \(a(z) = \alpha z + 1 - \alpha\) and \(b(z) = (1 - \beta) z + \beta\) respectively. Instead of by \(\alpha\) and \(\beta\), the user environment can also be characterized by the more comprehensible set of parameters \(\sigma\) and \(K\), to be understood as follows. Suppose the environment state is \('1'\) with probability \(\sigma\) and \('0'\) with probability \(1 - \sigma\), independently from slot to slot. The mean sojourn times are then given by \(1/(1-\sigma)\) and \(1/\sigma\) respectively. It is clear that the overall fraction of \('1'\)-slots remains equal to \(\sigma\) if the mean lengths of the \('1'\)-periods and the \('0'\)-periods are both multiplied by the same factor \(K\), i.e., if the parameters \(\alpha\) and \(\beta\) are chosen such that

\[
\begin{align*}
\text{E}[\text{length of '1'-period}] &= \frac{1}{1-\alpha} = \frac{K}{1-\sigma}, \\
\text{E}[\text{length of '0'-period}] &= \frac{1}{1-\beta} = \frac{K}{\sigma}.
\end{align*}
\] (3)

Here \(E[\cdot]\) denotes the expected value operator. The correlation factor \(K\) can be seen as a measure for the absolute lengths of the sojourn times, whereas the parameter \(\sigma\) characterizes their relative lengths.
Figure 1: State transition diagram of an active user

Let us now denote by \( m_{n,k} \) the number of users that sends the \( n \)th packet of a message during slot \( k \). The number of leading packet arrivals \( m_{1,k} \) in slot \( k \) is then either given by a random variable \( w_k^{(1)} \) or by \( w_k^{(0)} \), depending on whether the environment is in state ‘1’ or ‘0’. That is,

\[
m_{1,k} = \begin{cases} 
   w_k^{(1)} & \text{if } t_k = 1; \\
   w_k^{(0)} & \text{if } t_k = 0.
\end{cases}
\]  

(4)

Again, both rows \( w_k^{(1)} \) and \( w_k^{(0)} \) are i.i.d. with pgf’s \( M_1(z) \) and \( M_0(z) \) respectively. For simplicity, we make here the additional assumption that \( w_k^{(1)} \geq 1 \), thus interpreting ‘1’ as the ‘active’ environment state in which the users generate at least one new message per slot. Let us define \( q(n) \) as the probability that a message that is already \( n \) packets long in the current slot, will still continue in the next slot, i.e.,

\[
q(n) \triangleq \frac{1 - \sum_{r=1}^{n-1} l(r)}{1 - \sum_{r=1}^{n-1} l(r)}, \quad n \geq 1.
\]  

(5)

Each user – upon becoming active – runs through a state diagram as the one depicted in Fig. 1, which illustrates the meaning of the probabilities \( q(n) \), \( n \geq 1 \). Then the random variable \( m_{n+1,k} \) can be expressed as

\[
m_{n+1,k} = \sum_{i=1}^{m_{n,k}-1} d_{n,i}, \quad n \geq 1.
\]  

(6)

Indeed, the number of messages that generates its \( (n+1) \)th packet in slot \( k \) corresponds to the number of messages that generates its \( n \)th packet during slot \( k - 1 \) and continues in slot \( k \) by sending an \( (n+1) \)th packet. The \( d_{n,i} \)'s \( (n \geq 1) \) in (6) are, for given \( n \), i.i.d. random variables with common pgf \( D_n(z) = q(n) z + 1 - q(n) \). From (5) one can derive the following interesting properties of the pgf’s \( D_n(z) \):

\[
D_n(zD_{n+1}(...zD_N(zx) ...)) = \frac{\sum_{i=1}^{N} l(i) z^{i-n} + (1 - \sum_{r=1}^{n-1} l(r))}{1 - \sum_{r=1}^{n-1} l(r)} x z^{N-n+1},
\]  

(7)

and, hence,

\[
\lim_{N \to \infty} D_n(zD_{n+1}(...zD_N(zx) ...)) = z^{-n} \frac{L(z) - \sum_{r=1}^{n-1} l(i) z^{i}}{1 - \sum_{r=1}^{n-1} l(r)};
\]  

(8)

\[
\lim_{N \to \infty} D_n(D_{n+1}(...D_N(x) ...)) = 1.
\]  

(9)

The above properties make it possible to revert our analysis back in terms of \( L(z) \) later on.
Finally, let $s_{k+1}$ represent the system contents just after slot $k$ (i.e., at the beginning of slot $k+1$), where the term system contents indicates the number of packets that are either waiting or being transmitted. Since the server transmits a packet whenever one is available in the buffer, the evolution of the system contents from slot to slot is described by

$$s_{k+1} = e_k + (s_k - 1)^+,$$

(10)

where $(\cdot)^+ = \max(\cdot, 0)$ and $e_k$ is the total number of packet arrivals in slot $k$, which can be further expressed as

$$e_k = \sum_{m=1}^{\infty} m_n k .$$

(11)

Upon inspecting the slot to slot dependencies of the stochastic quantities in the system equations (2)–(11), we observe that the vectors $\{t_k, m_{n,k}(n \geq 1), s_{k+1}\}$ form an infinite-dimensional Markov chain. Stated otherwise, the values taken on by the variables in this vector are sufficient to fully describe the state of the multiplexer just after slot $k$.

### 3 Steady-state queueing analysis

As we will show in this section, our analysis of the queueing system at hand basically involves finding out as much as we can about the distribution of the joint process $\langle t_k, m_{n,k}(n \geq 1), s_{k+1}\rangle$ in the steady state. For this purpose, we establish a functional equation for the pgf of the system state vector.

Although we are not able to solve this equation, it is possible to derive from it various results regarding the marginal processes, such as closed-form expressions for the mean and variance of the number of packet arrivals in a slot, the mean system contents and the mean packet delay.

We start the analysis by defining the joint pgf of the random variables $t_k, m_{n,k}(n \geq 1)$ and $s_{k+1}$ as

$$P_k(x_1, y_1, y_2, \ldots, z) \triangleq \text{E}
\left[ x_1^{y_1} y_1^{m_{1,k}} y_2^{m_{2,k}} \cdots z^{s_{k+1}} \right].$$

(12)

Then, using the system equations introduced in the previous section, we find

$$P_{k+1}(x_1, y_1, \ldots, z) = M_0(y_1 z) b \left( x M_1(y_1 z) \right) \text{E}
\left[ a \left( x M_1(y_1 z) \right) \right] \left( D_1(y_2 z) \right) m_{1,k}^{y_1} \cdots z^{(s_{k+1}-1)^+}.$$

(13)

Next, we note that having an empty system at the beginning of slot $k+1$ implies that no packets have entered the system in slot $k$. Since we assumed that at least one leading packet arrives in the system during a ‘1’-slot, the system can only be empty at the start of slot $k+1$ when $t_k = 0$ and $e_k = 0$. Hence, in view of (11), if $s_{k+1} = 0$, then also $t_k = 0$ and $m_{n,k} = 0 (n \geq 1)$. This property allows us to rewrite equation (13) as

$$P_{k+1}(x_1, y_1, y_2, \ldots, z) = \frac{1}{z} M_0(y_1 z) b \left( x M_1(y_1 z) \right) \text{E}
\left[ a \left( x M_1(y_1 z) \right) \right] D_1(y_2 z) D_2(y_3 z) \cdots z^{(s_{k+1}-1)^+} + (z-1) \text{Prob}(s_{k+1} = 0).$$

(14)

Let us assume now that the condition for the queueing system to reach a stochastic equilibrium is met. Then, for large values of $k$, the functions $P_k$ and $P_{k+1}$ converge to the same limiting function $P$. 

4 / 12
From (14), it is easily seen that this steady-state joint pgf $P(x_1, y_1, y_2, \ldots, z)$ must satisfy the following functional equation:

$$
P(x_1, y_1, y_2, \ldots, z) = \frac{1}{z} M_0(y_1 z) \left( \frac{M_1(y_1 z)}{M_0(y_1 z)} \right) \cdot \left[ P \left( \frac{a(x)}{b(x)} \frac{M_1(y_1 z)}{M_0(y_1 z)} \right), D_1(y_2 z), D_2(y_3 z), \ldots, z \right] + p_0 (z - 1),
$$

(15)

where the quantity $p_0$ indicates the steady-state probability of having an empty buffer at the start of an arbitrary slot. Generally, finding an explicit solution for the joint equilibrium pgf $P(x_1, y_1, y_2, \ldots, z)$ from this equation is not a straightforward task. Nevertheless, equation (15) contains all relevant information concerning the equilibrium behavior of the multiplexer and it is possible to derive several interesting results from it.

**The environment state process**

First of all, equation (15) lets us obtain the steady-state pgf $T(x)$ corresponding to the environment state process if we let all arguments of the $P$-function other than $x$ be equal to one:

$$
T(x) \triangleq P(x, 1, 1, \ldots, 1) = b(x) T \left( \frac{a(x)}{b(x)} \right).
$$

(16)

Keeping in mind that $T(x)$ is a polynomial of degree one, we get

$$
T(x) = \frac{(1-\beta) x + 1 - \alpha}{2-\alpha-\beta}.
$$

(17)

**The packet arrival process**

In the same way, we can obtain the marginal pgf $\mathcal{E}_n(y_n)$ $(n \geq 1)$ of the number of users $m_n$ sending the $n$th packet of a message during an arbitrary slot in the steady state as

$$
\mathcal{E}_n(y_n) \triangleq P(1, 1, \ldots, 1, y_n, 1, \ldots, 1), \quad n \geq 1.
$$

(18)

Note in particular that for the pgf $\mathcal{E}_1(y_1)$ of the number of leading packet arrivals, the functional equation yields:

$$
\mathcal{E}_1(y_1) = M_0(y_1) T \left( \frac{M_1(y_1)}{M_0(y_1)} \right) = \frac{(1-\beta) M_1(y_1) + (1-\alpha) M_0(y_1)}{2-\alpha-\beta}.
$$

(19)

Furthermore, by repeatedly applying the relationship (15) and using the property (7) with $z = 1$, we find

$$
\mathcal{E}_n(y_n) = \mathcal{E}_1 \left( D_1(D_2(D_3(\ldots D_{n-1}(y_n) \ldots))) \right) = \mathcal{E}_1 \left( y_n + (1 - y_n) \sum_{r=1}^{\infty} l(r) \right),
$$

(20)

with the function $\mathcal{E}_1(z)$ as given in (19). Hence, the average number of users that delivers the $n$th packet of a message during an arbitrary slot is given by

$$
E[m_n] = \mathcal{E}_n'(1) = \mathcal{E}_1'(1) \left( 1 - \sum_{r=1}^{\infty} l(r) \right).
$$

(21)
The expected value of the total number of packet arrivals $e$ during an arbitrary slot in the steady state can be found by summation of the mean numbers of users $E_n'(1)$ ($n \geq 1$) that are present in each of the stages in Fig. 1. From (21) and (19) we find that

$$
E[e] = \sum_{n=1}^{+\infty} E_n'(1) = E_n'(1) L'(1) = \frac{(1-\alpha)M'_0(1) + (1-\beta)M'_1(1)}{2-\alpha-\beta} L'(1). \tag{22}
$$

Since the multiplexer has only one server, the quantity $E[e]$ also equals the load $\rho$ of the system, and the equilibrium condition of the system can be expressed as

$$
\rho = E_1'(1)L'(1) < 1. \tag{23}
$$

In view of (11), the second-order moment of $e$ can be calculated as

$$
E[e^2] = E\left[ \left( \sum_{n=1}^{+\infty} m_n \right)^2 \right] = \sum_{n=1}^{+\infty} E[m_n^2] + 2 \sum_{i=1}^{+\infty} \sum_{j=i+1}^{+\infty} E[m_i \cdot m_j], \tag{24}
$$

where

$$
E[m_i \cdot m_j] = \frac{\partial^2}{\partial y \partial y_j} P(1,1,\ldots,1) \quad \text{and} \quad E[m_n^2] = E_n'(1) + E_n''(1). \tag{25}
$$

To calculate the necessary partial derivatives of $P(x,y_1,y_2,\ldots,z)$, we resort to the functional equation (15) once more. Comparing the arguments of $P$ on the right-hand side to the ones on the left, we observe that the occurrences of $y_n$ ($n \geq 1$) have all shifted one place towards the left. It is this property that enables us to overcome the infinite dimensionality of $P$ and to calculate its consecutive partial derivatives in a recursive way. Eventually, for the variance $\text{Var}[e] = E[e^2] - E[e]^2$ of the number of packet arrivals $e$ during an arbitrary slot we obtain:

$$
\text{Var}[e] = (E_1''(1) - E_1'(1)^2) \sum_{i=1}^{+\infty} \left( 1 - \sum_{r=1}^{i-1} l(r) \right)^2 + E_1'(1)L'(1)
+ 2 \frac{(1-\alpha)(1-\beta)}{(2-\alpha-\beta)^2} \left( M'_1(1) - M'_0(1) \right)^2 \sum_{i=1}^{+\infty} \sum_{j=i+1}^{+\infty} \frac{(-1+\alpha+\beta)^{j-i} \left( 1 - \sum_{r=1}^{i-1} l(r) \right) \left( 1 - \sum_{r=1}^{j-1} l(r) \right)}{\left( 1 - \sum_{r=1}^{i} l(r) \right) \left( 1 - \sum_{r=1}^{j} l(r) \right)}.
\tag{26}
$$

**Mean system contents**

We now concentrate on the system contents $s$ at the beginning of an arbitrary slot in the steady state. Unfortunately, we are not able to derive from (15) an explicit expression for the distribution of $s$. However, to obtain the mean system contents $E[s]$, we can proceed as follows.

First, let us consider (15) for only those values of $x$, $y_n$ ($n \geq 1$) and $z$ for which the respective arguments of the functions $P$ on both sides of the equation are equal to each other, i.e., the values given by the equations

$$
y_n = D_n(y_{n+1}z), \quad n \geq 1; \tag{27}
$$

$$
x = \frac{a \left( x M'_1(y_1z) \right)}{b \left( x M'_1(y_1z) \right)}. \tag{28}
$$

The above equations appear to have more than one set of solutions. For our purposes, it is sufficient to consider only the solutions $x \triangleq \chi(z)$ and $y_n \triangleq \eta_n(z)$, which satisfy $\chi(1) = 1$ and $\eta_n(1) = 1$ ($n \geq 1$). By repeatedly applying (27) and using property (8), we then find

$$
\eta_n(z) = \frac{L(z) - \sum_{i=1}^{n-1} l(i) z^i}{z^n \left( 1 - \sum_{r=1}^{n-1} l(r) \right)}, \quad n \geq 1. \tag{29}
$$
Therefore, $\eta_n(z) \ (n \geq 1)$ can be interpreted as the pgf of the number of remaining packets in a message already containing $n$ packets. Note in particular that $\eta_1(z) = L(z)/z$. From this result, together with equation (28), it follows that the function $\chi(z)$ is determined implicitly by

$$(1-\beta)M_1(L(z))\chi(z)^2 - \left[\alpha M_1(L(z)) - \beta M_0(L(z))\right]\chi(z) - (1-\alpha)M_0(L(z)) = 0; \quad \chi(1) = 1. \quad (30)$$

When the arguments $(x_1, x_2, \ldots, z)$ are confined to the one-dimensional subspace $(\chi(z), \eta_1(z), \eta_2(z), \ldots, z)$, equation (15) becomes a linear equation for the function $P(\chi(z), \eta_1(z), \eta_2(z), \ldots, z)$, from which we get

$$P(\chi(z), \eta_1(z), \eta_2(z), \ldots, z) = \frac{p_0(z-1)G(z)}{z-G(z)}, \quad (31)$$

with

$$G(z) \triangleq M_0(L(z))b\left(\frac{\chi(z)M_1(L(z))}{M_0(L(z))}\right). \quad (32)$$

The quantity $p_0$ can now be calculated from the normalization condition $P(\chi(z), \eta_1(z), \eta_2(z), \ldots, z)|_{z=1} = 1$ as $p_0 = 1 - \rho$, where $\rho$ is the load of the multiplexer, given by (23). Next, derivation of both sides of (31) with respect to $z$ and evaluation of the result at $z = 1$ yields

$$\frac{\partial}{\partial x}P(1, \ldots, 1) \cdot \chi'(1) + \sum_{i=1}^{\infty} \left( \frac{\partial}{\partial x}P(1, \ldots, 1) \cdot M_0(1) \right) + \frac{\partial}{\partial z}P(1, \ldots, 1) = 1 - p_0 + \frac{G''(1)}{2p_0}. \quad (33)$$

Except for $E[s]$, all the quantities in this equation can easily be calculated from previous results; together with (17), (19), (21), (29), (30) and (32), we finally obtain:

$$E[s] = \frac{(1-\alpha)M''(1) + (1-\beta)M''(1)}{2-\alpha-\beta} \left( L'(1) - \frac{L''(1)}{2} \right) + \frac{(1-\beta)(1-\alpha-\beta)}{(2-\alpha-\beta)^2} \left( M''(1) - M'(1)L'(1) \right)
+ \frac{(2-\alpha-\beta)/2}{2-\alpha-\beta} - \left( (1-\alpha)M'(1) + (1-\beta)M'(1) \right) L''(1)
+ \left( \frac{(1-\alpha)M'(1) + (1-\beta)M'(1)}{2-\alpha-\beta} - 2 \frac{(1-\alpha)(1-\beta)}{(2-\alpha-\beta)^3} \left( M'(1) - M'(1) \right)^2 \right) L''(1). \quad (34)$$

It has been verified that this result for $E[s]$ reduces to the one obtained in [6], if abstraction is made of the environment state process by putting $\alpha = 0$ and $\beta = 1$, i.e., if the number of new messages is determined by the same pgf $M_0(z)$ in every slot.

The packet arrival process is determined by the following parameters: $\sigma$ and $K$ for the environment (instead of $\alpha$ and $\beta$), the distribution of the message length $l$ and the distributions of the numbers of new messages $w^{(1)}$ and $w^{(0)}$ in a ‘1’-slot and ‘0’-slot respectively. In order to show more clearly how the mean system contents is influenced by these parameters, we rewrite (34) as

$$E[s] = \frac{\sigma(M''(1) - M'(1))(M'(1)L'(1) - 1)L'(1)}{1 - E'(1)L'(1)} \cdot K + \frac{\sigma L'(1)^2E'(1)L'(1)}{2 - 2E'(1)L'(1)} \cdot \text{Var}[l]
+ \frac{\sigma L'(1)^2}{2 - 2E'(1)L'(1)} \cdot \text{Var}[w^{(1)}] + \frac{(1-\sigma)L'(1)^2}{2 - 2E'(1)L'(1)} \cdot \text{Var}[w^{(0)}]
+ \frac{L'(1)}{2} \left[ E'(1)(2 - L'(1)) + \sigma(M'(1) - M'(1)) \left( 2 - \frac{(1-\sigma)(M'(1) - M'(1)L'(1))}{1 - E'(1)L'(1)} \right) \right]. \quad (35)$$
where $E_l'(1) = \sigma M_l'(1) + (1-\sigma)M_0'(1)$.

First of all, we see that for given distributions of $I$, $w^{(1)}$ and $w^{(0)}$, and for a given fraction $\sigma$ of ’1’-slots, the mean system contents increases linearly with the correlation factor $K$ of the environment state process, this in spite of the fact that the mean number of packet arrivals in a slot is independent of $K$. An intuitive explanation for this effect can readily be found. Since we assumed that at least one packet enters the system per slot while the environment is in state ’1’, the server cannot keep up with the arrival stream during the ’1’-periods and the buffer contents will gradually increase. Of course, in stochastic equilibrium, these periods of buffer accumulation must alternate with periods during which the transmission of packets prevails over the arrival of packets and the buffer contents diminishes again. Such is the case, on the average, for the ’0’-periods. As explained in the previous section, the mean lengths of both the ’1’-periods (accumulation) and the ’0’-periods (transmission) increase linearly with the factor $K$. Now, for a given value of $\sigma$, one sees that the buffer contents, and also its expected value, will reach higher values as the accumulation periods last longer and are less interrupted by transmission periods, i.e., as $K$ becomes larger. This explains the observed impact of $K$ on $E[s]$.

Next, we notice that $E[s]$ is a linearly increasing function of the variances of both the message length and the numbers of new messages per slot in each of the environment states. Therefore, the mean system contents increases as the packet arrival process exhibits a greater ‘variability’, which is a common fact in queueing theory. Also, for a given load $\rho = E_l'(1)L'(1)$, the influence of $\text{Var}[l]$ on $E[s]$ increases as $L'(1)$ diminishes, whereas the influence of $\text{Var}[w^{(1)}]$ and $\text{Var}[w^{(0)}]$ decreases. In other words, for a given $\rho$, a larger number of, hence, shorter messages results in a larger impact of $\text{Var}[l]$ and a smaller impact of $\text{Var}[w^{(1)}]$ and $\text{Var}[w^{(0)}]$ on the mean system contents.

### Mean packet delay

We define the packet delay as the number of slots between the end of the packet’s arrival slot and the end of the slot during which the packet is transmitted from the buffer. Let $v$ be the random variable denoting the delay of an arbitrary packet in the steady state. In [5] and [8] it was shown that for any discrete-time single-server queueing system, with a first-come-first-served (FCFS) queueing discipline and constant service times of one slot, the following relationship exists between $V(z)$ and the pgf $S(z)$ of the system contents $s$ at the start of an arbitrary slot in the steady state:

$$V(z) = \frac{S(z) - S(0)}{1 - S(0)},$$  

(36)

irrespectively of the (possibly correlated) nature of the packet arrival process. Since in our model the above conditions are met, the mean packet delay relates to the mean system contents obtained in (34) as ($S(0) = p_0$)

$$E[v] = V'(1) = \frac{E[s]}{1-p_0} = \frac{2-\alpha-\beta}{(1-\alpha)M_0'(1) + (1-\beta)M_l'(1)L'(1)}E[s],$$  

(37)

in accordance with Little’s theorem ([19]).

### 4 Effect of primary and secondary correlation: numerical examples

As mentioned before, the correlated train arrivals model for the packet arrival process considered in this paper exhibits a twofold correlation: a primary correlation due to the arrival of messages in trains at the rate of one packet per slot and a secondary correlation due to the nonindependent generation of new messages. In this section, we will demonstrate the importance of taking into account both types of correlation. For this purpose, we compare the results obtained for the correlated train arrivals model
Figure 2: Mean system contents versus the total load $\rho$ for various values of $K$. The message-length distribution is a mixture of two geometrics ($p = 0.5$, mean 5 and variance 50), $M_0(z) = 1$ and $M_1(z) = 0.5z/(1 - 0.5z)$.

Figure 3: Mean system contents versus the total load $\rho$ for various values of $K$. The message-length distribution is negative binomial (mean 5), $M_0(z) = \exp[0.15(z - 1)]$ and $M_1(z) = z$. 
with the results that would be found if a model without secondary correlation or an uncorrelated model for the packet arrival process were used.

For the model without secondary correlation but with primary correlation (uncorrelated train arrivals), we assume that new messages are generated by the user population independently from slot to slot according to the pgf $E(z)$ given by equation (19); the message lengths are i.i.d. with pgf $L(z)$. In this case, the mean steady-state system contents $E_{prim}$ is given by (see [6])

$$E_{prim} = E_{i}(1)L'(1) + \left[ \frac{E_{i}(L(1)) + E_{i}(1)\sigma L(1)}{1 - E_{i}(1)L'(1)} \right] \frac{L'(1)}{2}. \quad (38)$$

By using (19), we can express $E_{prim}$ in terms of the basic parameters of our correlated train arrivals model as

$$E_{prim} = \frac{E_{i}(1)^2L'(1)}{2 - 2E_{i}(1)L'(1)} \cdot \text{Var}[\beta] + \frac{\sigma L(1)^2}{2 - 2E_{i}(1)L'(1)} \cdot \text{Var}[w^{(1)}] + \frac{(1-\sigma)L(1)^2}{2 - 2E_{i}(1)L'(1)} \cdot \text{Var}[w^{(0)}]$$

$$+ \frac{L'(1)}{2} \left[ E_{i}(2 - L'(1)) + \sigma(1-\sigma)\frac{M'(1) - M_{0}(1)^2L'(1)}{1 - E_{i}(1)L'(1)} \right]. \quad (39)$$

Comparing the expressions (39) and (35) for $E_{prim}$ and $E[z]$, we find

$$E[z] = E_{prim} + \frac{\sigma(M'(1) - M_{0}(1))(M'(1)L'(1) - 1)L'(1)}{1 - E_{i}(1)L'(1)} \cdot (K - 1). \quad (40)$$

This clearly shows that for given $L(z), M_{0}(z), M_{1}(z)$ and $\sigma \neq 0$, $E[z]$ is always greater than $E_{prim}$ when $K > 1$, i.e., in case of a positive correlation coefficient $\gamma = 1 - 1/K$ between the environment states in
two consecutive slots; the expressions only agree when $K = 1$ ($\gamma = 0$), i.e., in case of an uncorrelated message generation process. Therefore, neglecting the (positive) correlation in the message generation always leads to an underestimation of the mean system contents.

In the uncorrelated model, packets arrive to the multiplexer independently from slot to slot. In order to make a fair comparison we assume the mean and the variance of the number of packet arrivals per slot to be the same as in the correlated train arrivals model, i.e., equal to $\rho$ and $\text{Var}[e]$, given in equations (23) and (26). The mean steady-state system contents $E[s_{un}]$ in the uncorrelated model is then given by (see e.g. [9], [10])

$$E[s_{un}] = \frac{\rho}{2} + \frac{\text{Var}[e]}{2(1-\rho)}. \quad (41)$$

We now consider some practical examples. First, let us introduce four possible choices for the pgf $L(z)$ of the message length $l$:

$$L_1(z) = \frac{(1-\lambda)z^\gamma}{1-\lambda z}, \quad L_2(z) = z^m,$$

$$L_3(z) = \frac{(1-\theta)z^2}{(1-\theta z)^2}, \quad L_4(z) = p \frac{(1-\lambda_1)z^\gamma}{1-\lambda_1 z} + (1-p) \frac{(1-\lambda_2)z^\gamma}{1-\lambda_2 z}, \quad (42)$$

i.e., a geometric distribution, fixed-length messages, a negative binomial distribution and a mixture of two geometric distributions, respectively. The parameters of the distributions are chosen such that the mean message length $L'(1)$ is equal to a given value $m$. Additionally, in case of $L_4(z)$, a value for $p$ and $\text{Var}[l_4]$ must be specified. The variances of the other message-length distributions are given by

$$\text{Var}[l_1] = m(m-1), \quad \text{Var}[l_2] = 0, \quad \text{Var}[l_3] = \frac{1}{2}(m-1)(m+1). \quad (43)$$

In the Figs. 2, 3 and 4, the mean system contents for the three considered arrival models, i.e., $E[s]$ (correlated train arrivals), $E[s_{prim}]$ (uncorrelated train arrivals) and $E[s_{un}]$ (uncorrelated packet arrivals), are plotted versus the total load $\rho = (\sigma M'(1) + (1-\sigma)M_0'(1))L'(1)$. For each of the curves in these plots, a particular fixed choice is made for the distributions of $l$, $w^{(1)}$ and $w^{(0)}$. The variation of the load along the horizontal axis is brought about only by varying $\sigma$ between 0 and a maximum value implied by the stability condition. Hence, $\rho$ can range from $M_0'(1)L'(1)$ to 1.

In the Figs. 2 and 3, $E[s]$, $E[s_{prim}]$ and $E[s_{un}]$ are plotted for different values of the environment correlation factor $K$, namely $K = 1, 2, 5, 10$. In Fig. 2, the message-length distribution is a mixture of two geometrics according to $L_4(z)$ with $p = 0.5, m = 5$ and a variance of 50. The number of new messages in the ‘1’-slots is assumed to have a geometric distribution with an expected value of 2, whereas no new messages are generated during the ‘0’-slots. In Fig. 3, we give an example in which new messages are generated during ‘0’-slots too: $M_0(z) = \exp[q(z-1)]$, i.e., a Poisson distribution with intensity $q = 0.15$. Furthermore, exactly one new message starts per ‘1’-slot ($M_1(z) = z$) and the message lengths are negative binomially distributed with $m = 5$. Note that in this case, the load $\rho$ cannot become less than $mq = 0.75$ by merely varying $\sigma$. The figures clearly demonstrate the severe underestimation of the system contents when the different levels of correlation in the arrival process are neglected. First, we observe the rapid growth of $E[s]$ as $K$ increases, i.e., as the absolute (mean) lengths of the ‘1’-periods and the ‘0’-periods increase, even though the ratio $\sigma$ of these lengths remains unchanged. As is expected from (39) and (40), all the curves for $E[s_{prim}]$ coincide with the one representing $E[s]$ for $K = 1$ (bold curve). The dashed curves represent $E[s_{un}]$, showing what happens if the correlation in the packet arrival stream is neglected altogether. We see that $E[s_{un}]$ also slightly increases with higher values of $K$, although not in the same drastic way as $E[s]$. 
In Fig. 4, the environment correlation factor is chosen to be $K = 3$ for all curves. In the '1'-slots, the number of new messages has a geometric distribution with an expected value of 2, while no new messages are generated during the '0'-slots. To illustrate the influence of the distribution of the message lengths, we plotted the mean system contents for the four pgf’s given in (42). Their parameters are chosen such that $L'(1) = 5$, and for the mixed geometric distribution: $p = 0.5$, $\text{Var}[L_4] = 50$. For $E[s]$ and $E[s_{\text{prim}}]$, we know from (35) and (39) that the difference between the four curves is due only to the linear impact of the message-length variance, which here has the values: $\text{Var}[L_1] = 20$, $\text{Var}[L_2] = 0$, $\text{Var}[L_3] = 12$ and $\text{Var}[L_4] = 50$. For $E[s_{\text{un}}]$, however, we see that the curves are not ordered in the same predictable way as for the correlated arrivals models. In view of the rather complex sums occurring in (26), a simple relation between $E[s_{\text{un}}]$ and the moments of the message length distribution is not straightforward.

5 Conclusions

In this paper, we have analyzed a discrete-time buffer system with correlated variable-length packet-train arrivals. The use of an infinite-dimensional state description and generating functions yielded explicit formulas for the mean system contents and the mean packet delay. This allowed us to investigate the impact of the correlation in the packet arrival stream on the buffer behavior. The results show that the required buffer storage for a statistical multiplexer may be severely underestimated if the correlation in the packet arrival process is not taken into account properly.

Acknowledgement

The first author is postdoctoral fellow of the Fund for Scientific Research-Flanders (Belgium) (F.W.O.).

References