A Versatile Queueing Model Applicable in IP Traffic Studies

Bart Steyaert and Herwig Bruneel
SMACS Research Group
Vakgroep TELIN (TW07V)
University of Ghent
Sint-Pietersnieuwstraat 41
B-9000 Gent, Belgium.
Phone: +32-9-264 34 15
Fax: +32-9-264 42 95

Guido H. Petit and Danny De Vleeschauwer
Alcatel Bell Corporate Research Center
Francis Wellesplein 1
B-2018 Antwerp, Belgium

Abstract

In this report we consider a single-server discrete-time queueing system, where each source is modeled as a correlated 2-state Markovian packet arrival process, and a generally distributed packet length. For this model, we will study three important quantities, namely the number of packets in the buffer, the amount of work in the buffer, and the packet delay. In particular, for each of these quantities, we will derive an expression for their steady-state probability generating function. From these results, we can then obtain closed-form expressions for a number of interesting performance measures such as the mean value, variance, and tail distribution of buffer contents, unfinished work, and packet delay. In addition to the analysis of the system, a lot of emphasis is put on finding closed-form expressions for these quantities that reduce all numerical calculations to an absolute minimum, at the expense of introducing approximations for a number of boundary probabilities that must normally be calculated by solving a set of linear equations.

(*) SMACS: Stochastic Modeling and Analysis of Communication Systems
This work was sponsored by the Flemish Institute for the Promotion of Scientific and Technological Research in the Industry (IWT).
1. Introduction

In this report we analyze a queueing model with a generally distributed packet length and a correlated packet arrival process. This kind of model will be useful in assessing the performance of a packet-switched communication infrastructure, where messages carried by the network may have a variable transmission time, such as the current Internet, or a (future) dedicated Voice-over-IP (VoIP) network. Multiplexer buffers and/or switches and routers in such a network can in general be modeled by means of a discrete-time queueing system where new packets are generated by a superposition of individual traffic sources. The service time of a packet equals its transmission time, which is distributed according to a generally distributed (i.i.d.) random variable. Analyzing such a system is mandatory in the design and evaluation of these networks. However, this can be a difficult task, due to the fact that the sources that generate traffic in a buffer may have completely traffic characteristics (such as voice, data and file transfer), and because of the time-correlated behavior that each of the individual sources may exhibit.

To facilitate the queueing analysis, a traffic source is usually modeled as a Markov modulated arrival process, such as the discrete-time batch Markovian arrival process (D-BMAP, [1]) with the Markov modulated Bernoulli process (MMBP) as an important special case, and the Markov modulated Poisson process (MMPP, [2]). In this report, we will confine ourselves to a two-state Markov modulated packet arrival process, and the traffic sources are assumed to be homogeneous, i.e., they are all described by the same stochastic process. Continuous-time models with generally distributed service times and a Markovian arrival process have received some attention in the past, such as [3] and [4] (and the references therein), where an interrupted Poisson process (IPP) and a MMPP packet arrival process are considered respectively. Discrete-time models where general service time distributions are considered are rare. In [5] a discrete-time single-server queue is studied with an uncorrelated i.i.d. packet arrival process. Contribution [6] considers a two-state MMPP arrival process for packets that have a geometrically distributed transmission time, whereas in [7] a generally distributed packet length is considered where the traffic sources are aggregated and described by a two-state Markov modulated batch Bernoulli arrival process. In [4] and [6] a matrix-analytic solution is presented, which requires handling matrices that can become quite large. In [3,5,7], a generating-functions solution method is applied, the drawback here being that (for a correlated arrival process) a number of boundary probabilities must be calculated numerically from solving a set of linear equations. As mentioned before, in addition to solving the model, we also focus on finding closed-form expressions for the relevant performance measures, which implies that approximations are introduced for the boundary probabilities.

2. System Description

We consider a discrete-time queueing model, i.e., a system where the time axis is divided into fixed-length contiguous intervals, referred to as slots. The system consists of one single server and an infinite waiting room for packets awaiting transmission. Packets arrive in the buffer according to a correlated Markovian arrival process, described further on. It is assumed that the transmission of a packet requires a positive integer number of slots, described by a general transmission time distribution, and can start (and end) at slot marks only, i.e., at the time points between consecutive slots, meaning that the packet transmission is synchronized with respect to the slot marks. This is not necessarily the case as far as packet arrivals are concerned: packets may enter the buffer at any continuous time instant. Nevertheless, due to the
synchronous nature of a packet transmission, it can commence at the earliest at the beginning of the slot following its arrival (if it arrives in an empty buffer). Therefore, the precise details of the position of the packet arrival instants within a slot are irrelevant, and it suffices to characterize the arrival process by a random variable describing the total number of packet arrivals during a slot. We will now discuss the packet arrival and transmission processes in more detail.

Consider a multiplexer model fed by \( N \) identical and independent traffic sources generating packets of variable length. Each source is modeled as a 2-state Markov modulated arrival process, where the state of a source during a slot is represented by \( S_i \), \( 1 \leq i, j \leq 2 \). The packet arrival process during a slot is completely characterized by the \( 2 \times 2 \) matrix

\[
Q(z) = \begin{bmatrix} q_{11}(z) & q_{12}(z) \\ q_{21}(z) & q_{22}(z) \end{bmatrix} .
\]  

(1)

State transitions occur at slot boundaries, and let us denote by \( p_{ij} \), \( 1 \leq i, j \leq 2 \), the one-step transition probability \( \Pr[S_i \rightarrow S_j] \) at the end of slot \( k-1 \), where of course \( p_{j1} + p_{j2} = 1 \), \( 1 \leq j \leq 2 \). The elements \( q_{ij}(z) \) of \( Q(z) \) are then given by

\[
q_{ij}(z) = G_{ij}(z)p_{ij} ,
\]  

(2)

where \( G_{ij}(z) \), \( 1 \leq i, j \leq 2 \), is the probability generating function (pgf) describing the number of cells generated during a slot by a source, given that the source is in state \( S_j \) during the tagged slot and was in state \( S_i \) during the preceding slot, i.e.,

\[
G_{ij}(z) = \lim_{k \to \infty} E[z^{e_{ik}(n)}|S_i \rightarrow S_j \text{ at the end of slot } k-1] ,
\]  

(3)

where \( e_{ik}(n) \), \( 1 \leq n \leq N \), represents the number of packet arrivals generated by source \( n \) during slot \( k \), and where \( E[.] \) denotes the expected value of the tagged quantity. For the Markov modulated Bernoulli Process (MMBP), the number of packets generated by a source during any slot is either zero or equal to one, meaning that each of the generating functions \( G_{ij}(z) \) is a linear function of \( z \), i.e.,

\[
G_{ij}(z) = 1 - g_{ij} + g_{ij}z ,
\]  

(4)

for some parameters \( g_{ij} \) satisfying \( 0 \leq g_{ij} \leq 1 \). Although attention is mostly focused on this specific arrival model, the analysis is general, and can also be applied when the \( G_{ij}(z)'s \) have a more complex form.

The aggregate cell arrival process is fully determined, once the probability generating matrix \( Q(z) \) has been specified. Let us define \( a_{kj} \), \( 1 \leq j \leq 2 \), as the number of sources in state \( S_j \) during slot \( k \). Note that, due to the fact that the total number of sources equals \( N \), these random variables satisfy

\[
\sum_{j=1}^{2} a_{k,j} = N .
\]  

(5)
In the rest of the paper, we will denote by $x$ the 2×1 column vector $(x_1, x_2)^T$ (where $(.)^T$ represents the matrix transposition operation), and similarly we write $a_k=(a_{k,1}, a_{k,2})$. Let us also denote by the 2×1 column vector $B(x,z)$ the matrix product

$$B(x, z) = (B_1(x, z), B_2(x, z))^T = Q(z)x ,$$

and, for a set of random variables $r=(r_1, r_2)$, define the joint generating function as

$$E[x^r] = E\left[\prod_{j=1}^{2} x_j^{r_j}\right] .$$

Then, with the previous definitions, it is not difficult to show that the joint probability generating function of the random variables denoting the number of packet arrivals and the state of the arrival process during a slot can be written in terms of the state of the arrival process during the previous slot, leading to

$$E\left[e^{e_{k+1}} x^{a_{k+1}}\right] = E\left[e^{e_{k+1}} \prod_{j=1}^{2} x_j^{a_{k+1,j}}\right] = E\left[\prod_{j=1}^{2} B_j(x, z)^{a_{k,j}}\right] = E\left[Q(z)x^{a_k}\right] ,$$

where $e_k$ represents the total number of packet arrivals during slot $k$, i.e.,

$$e_k \doteq \sum_{n=1}^{N} e_k(n) .$$

Equation (8) fully describes the number of cells generated during consecutive slots by the $N$ packet sources.

Starting from some initial state, the packet arrival process will evolve to a stochastic equilibrium after a sufficiently long period of time, and we define $A(x)$ as the joint probability generating function of the number of sources in state $S_1$ and $S_2$ during an arbitrary slot. If we define $\sigma_1$ and $\sigma_2$ as the steady-state probability that the Markov state of a source during an arbitrary slot is $S_1$ and $S_2$ respectively, and $\sigma$ as the 2×1 vector $(\sigma_1, \sigma_2)^T$, which is the solution of the matrix equation

$$\sigma^T = \sigma^T Q(1) , \quad \sigma^T I = I ,$$

leading to

$$\sigma_1 = \frac{p_{21}}{p_{12} + p_{21}} , \quad \sigma_2 = \frac{p_{12}}{p_{12} + p_{21}} ,$$

where $I$ is the 2×1 column vector with all elements equal to 1, then it can be shown that $A(x)$ equals

$$A(x) = (\sigma^T x)^N .$$
Packets that arrive in the buffer are of variable length. It is assumed that the transmission times of consecutive packets that arrive in the buffer form a set of independent and identically distributed random variables, denoted by $s$, with common probability mass function
\[ s(l) = \Pr[\text{transmission time of a packet }= l \text{ slots}], \quad l \geq 1, \]
and corresponding probability generating function
\[ S(z) = E[z^s] = \sum_{l=1}^{\infty} s(l)z^l. \]

Obviously, we assume that the transmission time of a packet is at least one slot.

Due to the infinite-buffer capacity, the system will reach a stochastic equilibrium only if the equilibrium condition is satisfied. Defining $\rho$ as the load of the system, this implies that
\[ \rho \equiv NpL_p < 1, \tag{12} \]
must hold, where $p$ denotes the mean number of packet arrivals per slot and per inlet, and $L_p = S'(1)$ represents the mean packet length. Taking into account the packet arrival process described in this section, it follows that $p$ can be calculated from
\[ p = \frac{d}{dz}\left[z^r Q(z)I\right]_{z=1} = \sum_{i=1}^{2} \sigma_i \sum_{j=1}^{2} q_{ij} (1), \tag{13} \]
where primes denote derivatives with respect to the argument.

3. The joint pgf of the state vector

3.1 System Equations

We first establish the system equations that control the evolution of the number of packets in the buffer during consecutive slots. Let us therefore define the random variable $v_k$ as the system contents at the beginning of slot $k + 1$, which is the number of packets in the buffer at the beginning of slot $k + 1$, including the one being transmitted (in case of a nonempty buffer). In addition to $v_k$, we also need information about the amount of transmission time the packet in service still requires at the beginning of slot $k + 1$ in order to be completely transmitted. We therefore define the random variable $h_k$ as the residual transmission time, which is the number of slots required at the beginning of slot $k + 1$ to complete the transmission of the packet being transmitted in case of a nonempty buffer; when $v_k = 0$, we automatically have that $h_k = 0$.

Together with the definitions of the previous section, we can now establish the system equations. We must distinguish between the three following cases:

1) $h_k = 0 \ (\Leftrightarrow v_k = 0)$. 

When the buffer is empty at the beginning of a slot, the number of packets in the buffer at the beginning of the next slot equals the number of new arrivals, and we find that

\[ v_{k+1} = v_k + e_{k+1} \]  

(14)

If no packets have entered the buffer during slot \( k+1 \), the residual transmission time remains zero, otherwise it is equal to the transmission time of a new packet, which leads to

\[ h_{k+1} = \begin{cases} 0 & \text{if } (e_{k+1} = 0) \text{ and } (v_k = 0) \\ s & \text{if } (e_{k+1} > 0) \text{ or } (v_k > 0) \end{cases} \]  

(15)

2) \( h_{k} \geq 1 \).

This implies that the packet in service is completely transmitted at the end of slot \( k+1 \)

\[ v_{k+1} = v_k + e_{k+1} - 1 \]  

(16)

Similar to the previous case, if no packets have entered the buffer during slot \( k+1 \) and the packet in service was the only one in the buffer, then the buffer becomes empty and the residual transmission time equals zero at the beginning of the next slot, otherwise the transmission of a new packet commences, i.e.,

\[ h_{k+1} = \begin{cases} 0 & \text{if } (e_{k+1} = 0) \text{ and } (v_k = 1) \\ s & \text{if } (e_{k+1} > 0) \text{ or } (v_k > 1) \end{cases} \]  

(17)

3) \( h_k > 1 \).

In this case, the packet being transmitted receives an extra slot of service without the transmission being completed, and we obtain

\[ \begin{cases} v_{k+1} = v_k + e_{k+1} \\ h_{k+1} = h_k - 1 \end{cases} \]  

(18)

3.2 Derivation of a Functional Equation

Clearly, in view of equations (14-18), we need to keep track of the random variables \( v_k \) and \( h_k \) in our state description. Consequently, due to the correlated first-order Markov nature of the process controlling the number of packet arrivals during a slot, it becomes clear that the set of random variables \((v_k, h_k, a_k)\) constitutes a four-dimensional Markov state description of the system at the beginning of consecutive slots; this triplet of random variables will be referred to as the state vector. Note that, in view of (5), one of the random variables \((a_{k,1}, a_{k,2})\) can be omitted from the state description. However, due to reasons of symmetry, we prefer to maintain both random variables in the state description.

Let us now define the joint probability generating function of \((v_k, h_k, a_k)\) as
\[ P_k(z, y, x) \doteq E\left[z^{v_k}y^{h_k}x^{a_k}\right] = E\left[z^{v_k}y^{h_k} \prod_{j=1}^{2} x^{a_{k,j}}\right]. \] (19)

Let us, for sake of notation, define a partial generating function of a random variable \( X \) if an event \( A \) occurs as

\[ E\left[z^{X}\mid A\right] \doteq E\left[z^{X}\mid A\right]\Pr[A]. \] (20)

From system equations (14) and (15), taking into account that \( h_k=0 \iff v_k=0 \), we then first of all derive that

\[ E\left[z^{v_{k+1}}y^{h_{k+1}}x^{a_{k+1}}\mid v_k=0, h_k=0\right] = E\left[z^{v_{k+1}}y^{h_{k+1}}x^{a_{k+1}}\mid v_k=0, h_k=0\right] \\
+ E\left[y^{0} - y^{s}\right]x^{a_{k+1}}\mid h_k=0, v_k=0, e_{k+1}=0]. \] (21)

In a similar way, we obtain from (16) and (17)

\[ E\left[z^{v_{k+1}}y^{h_{k+1}}x^{a_{k+1}}\mid h_k=1\right] = E\left[z^{v_{k+1}}y^{h_{k+1}}x^{a_{k+1}}\mid h_k=1\right] \\
+ E\left[y^{0} - y^{s}\right]x^{a_{k+1}}\mid h_k=1, v_k=1, e_{k+1}=0]. \] (22)

Finally, (18) can also be translated into a relation between z-transforms, leading to

\[ E\left[z^{v_{k+1}}y^{h_{k+1}}x^{a_{k+1}}\mid h_k>1\right] = E\left[z^{v_{k+1}}y^{h_{k+1}}x^{a_{k+1}}\right] \\
- y^{-1}E\left[z^{v_{k+1}}y^{h_{k+1}}x^{a_{k+1}}\mid h_k=1, v_k=0\right] - E\left[z^{v_{k+1}}y^{h_{k+1}}x^{a_{k+1}}\mid h_k=1\right]. \] (23)

Summation of (21–23) yields

\[ yP_{k+1}(z, y, x) = E\left[z^{v_{k+1}}y^{h_{k+1}}x^{a_{k+1}}\right] + (yS(y) - 1)E\left[z^{v_{k+1}}y^{h_{k+1}}x^{a_{k+1}}\mid h_k=0, v_k=0\right] \\
+ y(S(y) - z)E\left[z^{v_{k+1}}y^{h_{k+1}}x^{a_{k+1}}\mid h_k=1\right] \\
+ y \sum_{i=0}^{1} (1-S(y))E\left[z^{v_{k+1}}y^{h_{k+1}}x^{a_{k+1}}\mid h_k=i, v_k=i, e_{k+1}=0\right]. \]

Note that, given that \( a_i \) is known, the value of \( a_{i+1} \) (and the number of packet arrivals \( e_{k+1} \)) is independent of \( (v_k, h_k) \). Therefore, applying equation (8), the previous relation can be transformed into

\[ yP_{k+1}(z, y, x) = P_k(z, y, Q(z)x) + y(1-S(y))\phi_k(Q(0)x) + (yS(y) - 1)\phi_k(Q(z)x) \\
+ y(1-S(y))\Psi_k(0, Q(0)x) + y(S(y) - z)\Psi_k(z, Q(z)x). \] (24)

where
First of all, due to the foregoing definitions, we have that \( P_k(0,y,x) \equiv P_k(0,0,x) = \varphi_k(x) \). If we insert \( z=0 \) into equation (24), we derive that

\[
\varphi_{k+1}(x) = \varphi_k(Q(0)x) + \psi_k(0,Q(0)x) .
\]

Let us now assume that the equilibrium condition (12) is satisfied, implying that all the functions that occur in (24) evolve to a steady-state limit, which will be reflected in the remainder by suppressing the subscript \( k \) (which expressed the time-dependence in the previous derivations). Together with the previous relation, we finally obtain

\[
yP(z,y,x) = P(z,y,Q(z)x) + y(1-S(y))\varphi(x) + (yS(y)-1)\varphi(Q(z)x) + y(S(y)-z)\psi(z,Q(z)x)
\]

Equation (26), together with definition (25), defines a functional equation for \( P(z,y,x) \), the steady-state joint generating function of the random variables \((v_k, h_k, a_k)\), i.e., the system contents and residual transmission time at the beginning of an arbitrary slot, and the state of the Markovian packet arrival process during the preceding slot.

### 3.3 Solution of the functional equation

If we substitute consecutive arguments \( x = Q(z)^i x' \), \( 0 \leq i \leq n \), into (26), and sum the resulting equations, we obtain

\[
P(z,y,x) = y^{-n}P(z,y,Q(z)^n x) + (1-S(y))\sum_{i=1}^{n} y^{-i+1}\varphi(Q(z)^{i-1} x) + (yS(y)-1)\sum_{i=1}^{n} y^{-i}\psi(z,Q(z)^i x).
\]

We now let \( n \) approach infinity. Let us therefore express \( Q(z)^n \) in terms of its eigenvalues and –vectors as \( W(z)\Lambda(z)^N U(z) \), and define the functions \( F_{lm}(z) \), \( 1 \leq l,m \leq N \), as

\[
\left( \sum_{j=1}^{2} u_{1j}(z) x_j \right)^l \left( \sum_{j=1}^{2} u_{2j}(z) x_j \right)^{N-l} = \sum_{m=0}^{N} F_{lm}(z) x_1^m x_2^{N-m}.
\]

In the above expression, the functions \( u_{ij}(z) \) are the components of the column eigenvectors of \( Q(z) \), given by (A.5). For further details concerning the eigenvalues and –vectors of \( Q(z) \), we refer to the Appendix. One can easily derive that this relation implies that these functions can be written as

\[
\varphi_k(x) = E[x^{\alpha_i} \{ u_k = 0, h_k = 0 \}] = E\left[ \prod_{j=1}^{2} x_{j}^{\alpha_{i,j}} \{ u_k = 0, h_k = 0 \} \right].
\]
where \( \begin{pmatrix} x \\ y \end{pmatrix} = \frac{x!}{y!(x-y)!} \), and \((.)^+ = \max\{0, .\}\). In view of definition (28) and defining

\[
p_l \triangleq \lim_{k \to \infty} \Pr[u_k = 0, h_k = 0, a_{k,1} = l, a_{k,2} = N - l] \quad ,
\]

then for \( n \to \infty \), the first sum in the right-hand side of (27) can then be written as

\[
\sum_{i=0}^{\infty} y^{-i} \varphi(Q(z)^i x) = \sum_{i=0}^{\infty} y^{-i} \sum_{l=0}^{N} \sum_{m=0}^{N} \frac{2}{\sum_{j=1}^{N} u_{1j}(z) \lambda_{j}(z)^i w_{j}(z)x} \left( \sum_{j=1}^{N} u_{2j}(z) \lambda_{j}(z)^i w_{j}(z)x \right)^{N-l} p_l 
\]

where \( w_{j}(z) \) constitutes the \( j \)-th row eigenvector of \( Q(z) \). Due to definition (28), this can be transformed into

\[
\sum_{i=0}^{\infty} y^{-i} \varphi(Q(z)^i x) = \sum_{i=0}^{\infty} y^{-i} \sum_{l=0}^{N} \sum_{m=0}^{N} F_{lm}(z) \left( \lambda_1(z)^i w_1(z)x \right)^m \left( \lambda_2(z)^i w_2(z)x \right)^{N-m} p_l 
\]

Let us now consider values of \( y \) and \( z \) for which \( |\lambda_j(z)| < |y| \leq 1, 1 \leq j \leq 2 \). Such values of \( y \) and \( z \) exist; this is for instance the case if \( |y| = 1 \), and \( |z| = 1 \) and \( z \neq 1 \) (e.g. see [9]), since

\[
|\lambda_1(z)| \leq 1 \text{ and } |\lambda_2(z)| < 1, \text{ for } |z| = 1 \text{ and } z \neq 1 \quad ,
\]

unless some special cases for the generating functions \( G_{ij}(z) \) defined in (3) are considered, a situation that falls beyond the scope of this report. The sum for \( i \) in the above expression then converges, and we obtain

\[
\sum_{i=0}^{\infty} y^{-i} \varphi(Q(z)^i x) = \sum_{m=0}^{N} \frac{y(w_1(z)x)^m(w_2(z)x)^{N-m}}{y - \lambda_1(z)^m \lambda_2(z)^{N-m}} \sum_{l=0}^{N} F_{lm}(z)p_l 
\]

In a similar way we can derive that

\[
\sum_{i=1}^{\infty} y^{-i} \varphi(Q(z)^i x) = \sum_{m=0}^{N} \frac{(\lambda_1(z)w_1(z)x)^m(\lambda_2(z)w_2(z)x)^{N-m}}{y - \lambda_1(z)^m \lambda_2(z)^{N-m}} \sum_{l=0}^{N} F_{lm}(z)p_l 
\]

where

\[
\psi_l(z) \triangleq \lim_{k \to \infty} E\left[e^{\psi_k - 1}\{h_k = 1, a_{k,1} = l, a_{k,2} = N - l\}\right] 
\]
In addition, note that

\[
\lim_{n \to \infty} y^{-n} P(z, y, Q(z)^n x) = \sum_{j=0}^{\infty} z^j \sum_{k=0}^{\infty} y^k \sum_{l=0}^{N} \sum_{m=0}^{N} F_{lm}(z) \Pr[v = j, h = k, a_1 = l, a_2 = N - l] \]

\[
\lim_{n \to \infty} y^{-n} \left( \lambda_1(z)^n w_1(z)x \right)^m \left( \lambda_2(z)^n w_2(z)x \right)^{N-m},
\]

which, for \(|\lambda_j(z)| < |y|\leq 1, 1 \leq j \leq 2\), becomes equal to 0. Summarizing, in view of the previous results, equation (27) becomes

\[
P(z, y, x) = \sum_{m=0}^{N} \frac{\left( w_1(z)x \right)^m \left( w_2(z)x \right)^{N-m}}{y - A_m(z)} \sum_{l=0}^{N} F_{lm}(z) \left\{ [y(S(y) - z)\psi_l(z) + (yS(y) - 1)p_l]A_m(z) + y(1 - S(y))p_l \right\},
\]

where \(A_m(z)\) has been defined as

\[
A_m(z) = \lambda_1(z)^m \lambda_2(z)^{N-m}.
\]

Expression (33) for the joint probability generating function \(P(z,y,x)\) still contains the unknown constants \(p_l\), as well as the unknown functions \(\psi_l(z)\). Let us therefore, for each value of \(m\), consider values of \(y\) for which \(y = A_m(z)\), which will be represented by \(y_m(z)\). Note if \(|z| \leq 1\), due to (31) this solution for \(y\) also satisfies \(|y_m(z)| \leq 1\), and for these values of \(z\) and \(y\), \(P(z,y,x)\) is finite. This means that, if we multiply both hand sides of (33) by \(y - A_m(z)\) and consider \(y = y_m(z)\), then the left-hand side becomes equal to zero, implying that the same must hold for the right-hand side. For each value of \(0 \leq m \leq N\), this leads to the following equation:

\[
\sum_{l=0}^{N} F_{lm}(z) \left\{ A_m(z) \left( S(A_m(z)) - z \right) \psi_l(z) + S(A_m(z)) \left( A_m(z) - 1 \right) p_l \right\} = 0.
\]

If we insert this equation into (33), we finally obtain

\[
P(z, y, x) = \sum_{m=0}^{N} \left( w_1(z)x \right)^m \left( w_2(z)x \right)^{N-m} \left\{ 1 + \frac{(A_m(z) - 1)(S(y) - S(A_m(z)))}{(y - A_m(z))(z - S(A_m(z)))} \right\} \sum_{l=0}^{N} F_{lm}(z) p_l.
\]

This final result for \(P(z,y,x)\) expresses the joint generating function of the state of the arrival process during an arbitrary slot, and the buffer contents and residual transmission time at the beginning of the next slot, in terms of the unknown constants \(p_l\). In the next section, it will become clear how these can be calculated. Also, starting from this result, we will establish expressions for the generating functions of quantities such as the buffer contents, unfinished work, and message delay, and their corresponding moments and tail distribution.
3.4 Calculation of the unknowns $p_l$

We can calculate these probabilities, defined in (30), by expressing that $P(z,y,x)$ is analytic when the complex variables $y$ and $z$ both lie inside the unit circle, i.e., $|z| \leq 1$ and $|y| \leq 1$. We have already expressed in the previous section that for $|y| \leq 1$ and $y = A_m(z)$, $P(z,y,x)$ is finite. In view of expression (35), the only remaining potential singularities inside the complex unit disk are those values of $z$ (and $y$) for which $z = S(A_m(z))$. Indeed, taking into account the definition for $A_m(z)$, then from (31) and Rouché’s theorem it follows that this equation has exactly one solution inside the complex unit circle. Let us for each value of $0 \leq m \leq N$, denote this solution by $z_{[m]}$, i.e.,

$$z_{[m]} = S(A_m(z_{[m]})), \quad |z_{[m]}| \leq 1, \quad 0 \leq m \leq N. \quad (36)$$

Then, by expressing that $z_{[m]}$ must also be a zero of the corresponding numerator in expression (38) for $P(z,y,x)$, we obtain

$$\xi_m(z_{[m]}) = 0, \quad 0 \leq m \leq N-1, \quad (37)$$

where

$$\xi_m(z) \triangleq \frac{1}{1-\rho} \sum_{l=0}^{N} F_{lm}(z)p_l. \quad (38)$$

Note that $z_{[N]} = 1$ (since $\lambda_1(1)=1$, see (A.6), implying $A_N(1) = 1$), which leads to no additional equation for the unknowns. Together with the normalization condition, which yields $\xi_N(1) = 1$ (as we will see further on), equation (37) forms a set of $N+1$ linear equation for the $N+1$ unknowns $p_l$, that can easily be solved.

However, in deriving numerical results for the performance measures considered further in this paper, we want to avoid all numerical calculations as much as possible. Therefore we also present an approximation for the boundary probabilities $p_l$. As before, denote by the random variables $e$ and $v$ the number of packet arrivals during a tagged arbitrary slot and the buffer contents at the beginning of the following slot; also, $a$ represents the state of the Markovian arrival process during the tagged slot. We define the steady-state joint probability

$$q_l \triangleq \Pr[a_1 = l, a_2 = N-l, e = 0].$$

Obviously, $v=0$ implies that there have been no cell arrivals during the tagged slot, i.e. $v = 0 \Rightarrow e = 0$. It is therefore clear that the inequality $q_l \leq p_l$ holds. We will show by some numerical examples that approximating the boundary probabilities by

$$p_l \approx \frac{1-\rho}{\Pr[e = 0]} q_l, \quad (39)$$

(where we have taken into account that $\Pr[v = 0] = 1-\rho$) yields excellent approximations for the performance measures of interest.
An additional advantage of the approximation proposed here is that it avoids the explicit calculation of the \( q_i \)'s, as well as the functions \( F_{lm}(z) \) from equation (29), that occur in expression (35) for \( P(z,y,x) \). From definition (28) and expression (8) it is not difficult to show that

\[
\sum_{l=0}^{N} F_{lm}(z)q_l = \left( \frac{N}{m} \right) \left( \sigma^T Q(0)u_1(z) \right)^{m} \left( \sigma^T Q(0)u_2(z) \right)^{N-m},
\]

where \( u_j(z) \) represents the \( j \)-th column eigenvector of \( Q(z) \). In view of the previous, this implies that the following approximation may be applied:

\[
\xi_m(z) \equiv \frac{1}{\Pr[e = 0]} \left( \frac{N}{m} \right) \left( \sigma^T Q(0)u_1(z) \right)^{m} \left( \sigma^T Q(0)u_2(z) \right)^{N-m}. \tag{40}
\]

To conclude, note that it immediately follows from (8) that \( \Pr[e=0] \) satisfies

\[
\Pr[e = 0] = \left( \sigma^T Q(0)I \right)^N.
\]

4. The Buffer Contents

4.1 Derivation of the probability generating function

First of all, due to property (A.2), we have that \( w_j(z)I=1, 1\leq j\leq 2 \). Therefore, inserting \( y=1 \) and \( x=I \) into (38), we obtain the following expression for the probability generating function \( V(z) \) describing the number of packets in the buffer at the beginning of an arbitrary slot:

\[
V(z) = (1-\rho) \sum_{m=0}^{N} \frac{(z-1)S(A_m(z))\xi_m(z)}{z - S(A_m(z))}.
\tag{41}
\]

From (A.7) (see Appendix) and (12), it follows that \( S(1)A_\lambda(1)=NS(1)A_\lambda(1)=\rho \), implying that the normalization condition for \( V(z) \) indeed requires that \( \xi_\lambda(1)=1 \), i.e., \( \Pr[v=0]=1-\rho \). From (41), one can easily derive the performance measures of interest related to the buffer contents, as shown next.

4.2 Moments of the buffer contents

The mean and variance of the buffer contents \( v \) can be calculated by taking the appropriate derivatives of expression (41) for \( V(z) \). First, note that (A.8) implies that \( F_{lm}(1)=\delta(N-m) \) (where \( \delta(n) \) represents the discrete Kronecker-delta function, which equals 1 if \( n \) is 0, and 0 otherwise). Since this, in turn, implies that \( \xi_m(1)=\delta(N-m) \), and defining \( \zeta_m(z) \equiv S(A_m(z)) \), we obtain the following expression for mean and variance of the buffer contents:
\[ E[v] = \xi_N^\prime(1) + \frac{\xi_N^{\prime\prime}(1)}{2(1 - \xi_N(1))} + \xi_N(1) \]

\[ \text{Var}[v] = \xi_N^{\prime\prime}(1) + \xi_N(1) - \xi_N(1)^2 + \frac{\xi_N^{\prime\prime\prime}(1)}{3(1 - \xi_N(1))} + \frac{\xi_N^{\prime\prime}(1)}{2(1 - \xi_N(1))} \left( 1 + \frac{\xi_N^{\prime\prime}(1)}{2(1 - \xi_N(1))} \right) \]

\[ + \xi_N(1) + \xi_N(1) - \xi_N(1)^2 + T'(1) \]

where, with (A.7) and (12)

\[ \xi_N^\prime(1) = NpL_p = \rho \]

\[ \xi_N^{\prime\prime}(1) = N^2p^2S^{\prime\prime}(1) + (N-1)\rho p + NL_p \lambda^{\prime\prime}(1) \]

\[ \xi_N^{\prime\prime\prime}(1) = N^3p^3S^{\prime\prime\prime}(1) + 3Np(N-1)\rho p \lambda^{\prime\prime}(1) + N(N-1)S^{\prime\prime}(1)p^2 \]

\[ + \rho(N-1)(N-2)p^2 + NL_p \lambda^{\prime\prime\prime}(1) \]

In these expressions, primes denote derivatives with respect to the argument, as before. From definition (38) for \( \xi_m(z) \) and the observation that \( F_{IN}(z) = u_{11}(z)u_{12}(z)^{N-1} \), we also find that

\[ \xi_N^\prime(1) = \frac{1}{1 - \rho} \sum_{l=0}^{N} \left( l u_{11}(1) + (N-l)u_{21}(1) \right) p_l \]

\[ \xi_N^{\prime\prime}(1) = \frac{1}{1 - \rho} \sum_{l=0}^{N} \left( l u_{11}(1) + 2l(N-l)u_{11}(1)u_{21}(1) + (N-l)u_{21}(1) \right) p_l \]

\[ + l(l-1)u_{11}(1)^2 + (N-l)(N-l-1)u_{21}(1)^2 \]

If we choose to use the approximation (43) for \( \xi_m(z) \) in order to avoid the numerical calculation of the \( p_i \)'s, this becomes

\[ \xi_N^\prime(1) \equiv N \left( \sigma^T Q(0) u_{11}(1) \right) \left( \sigma^T Q(0) I \right)^{-1} \]

\[ \xi_N^{\prime\prime}(1) \equiv N \left( \sigma^T Q(0) u_{11}(1) \right) \left( \sigma^T Q(0) I \right)^{-1} + N(N-1) \left( \sigma^T Q(0) u_{11}(1) \right)^2 \left( \sigma^T Q(0) I \right)^{-2} \]

In addition, the term \( T'(1) \) in the expression for \( \text{Var}[v] \) is given by

\[ T'(1) = (1 - \rho) \sum_{m=0}^{N-1} \frac{A_m(1)}{1 - A_m(1)} \xi_m(1) \]

which can be considered as the first derivative with respect to \( z \) of \( T(z) \) for \( z=1 \), where, using (31) and the definition for \( \xi_m(z) \) and \( A_m(z) \), \( T(z) \) satisfies

\[ T(z) = \frac{1}{1 - \rho} \sum_{m=0}^{N-1} A_m(1) \xi_m(z) = \sum_{l=0}^{N} \sum_{k=1}^{N} \left( -F_{ln}(z) + \sum_{m=0}^{N} A_m(1)^k F_{lm}(z) \right) p_l \]
where we have also taken into account that $\lambda_1(1) = A_N(1) = 1$. Taking the first order derivative, we thus obtain, keeping in mind that $u_{j2}(1) = 0$ and $u_{j1}(1), 1 \leq j \leq 2$, (see (A.8-9))

$$T'(1) = -(1 - \rho) \frac{\lambda_2(1)}{1 - \lambda_2(1)} \xi_N'(1). \quad (46)$$

By using expressions (41-46), we have now expressed the moments of the buffer contents exclusively in terms of the system parameters and the derivatives of the P-F eigenvalue and – vector given in (A.7-10).

### 4.3 Tail distribution of the buffer contents

It has been observed in many cases (see e.g. [10]) that approximating the tail distribution of the buffer contents by a geometric form is quite accurate, if the poles of $V(z)$ have a different modulus and multiplicity equal to 1, which is, in general, the case for the model under consideration. Approximating the distribution of the buffer contents by a geometric form corresponds to approximating $V(z)$ by

$$V(z) \approx C_v \frac{(z - z_{0,v})^n}{z - z_{0,v}} = -C_v \sum_{n=0}^{\infty} \left( \frac{z}{z_{0,v}} \right)^n,$$

where we are particularly interested in sufficiently large values of $n$. In this expression, $z_{0,v}$ is the pole with the smallest modulus of $V(z)$, which is a zero of the denominator for $m=N$ in expression (41) for $V(z)$, which yields

$$z_{0,v} = S(A_N(z_{0,v})) = S(\lambda_1(z_{0,v})^N),$$

and $z_{0,v}$ is a real and positive number larger than 1. Furthermore, the constant $C_v$ can be calculated from the residue theorem, leading to

$$C_v = \lim_{z \to z_{0,v}} (z - z_{0,v})V(z)$$

$$= (1 - \rho) \frac{z_{0,v}^N - 1}{1 - N\lambda_1'(z_{0,v})^N} \xi_N(z_{0,v})$$

Combining these relations, we find the following approximate expression for the probability that the buffer contents exceeds a certain threshold $n$:

$$\Pr[v > n] \approx \frac{(1 - \rho)\xi_N(z_{0,v})}{1 - N\lambda_1'(z_{0,v})^N} \frac{z_{0,v}^{-n}}{\xi_N(z_{0,v})}.$$  \quad (47)
Again, we can either use the exact value for $\zeta_N(z_{0,\nu})$ calculated from (38), or the approximation calculated from (40).

5. The Unfinished Work

5.1 Derivation of the probability generating function

The unfinished work in the buffer at the beginning of a slot is defined as the amount of work in the buffer at the beginning of the slot, which is the number of slots required to empty the buffer if no more packets were to arrive during the subsequent slots. The amount of work required to process a packet whose transmission has not started yet is given by its transmission time, described by $S(z)$. On the other hand, at the beginning of a slot, the packet currently being transmitted (if any) still requires $h$ slots before it is entirely sent, where $h$ equals the residual transmission time. The unfinished work $u_k$ at the beginning of slot $k+1$ is therefore given by

$$u_k = \begin{cases} 
0 & \text{if } v_k = 0, h_k = 0 \\
\frac{v_k-1}{h_k} + \sum_{i=1}^{v_k} s_i & \text{if } v_k > 0
\end{cases}$$

where each of the $s_i$'s represents a packet transmission time, and is therefore described by the pgf $S(z)$. Consequently, the corresponding steady-state pgf $U(z)$ is given by

$$U(z) = S(z)^{-1}\{P(S(z), z, I) + (S(z)-1)P(0,0, I)\}.$$ 

The observation that $P(0,0, I) = \text{Pr}[i=0]=1-\rho$ can be written as

$$P(0,0, I) = \sum_{l=0}^{N} p_l = \sum_{m=0}^{N} \sum_{l=0}^{N} F_{lm}(S(z)) p_l,$$

together with the previous relations, leads to the following expression for the steady-state probability generating function describing the unfinished work at the beginning of an arbitrary slot:

$$U(z) = (1-\rho) \sum_{m=0}^{N} \frac{(z-1)A_m(S(z))z}{z-A_m(S(z))} \zeta_m(S(z)).$$

(48)

Again, due to $A_N'(1)S'(1) = NSp = \rho$ and $\zeta_N(1) = \delta(N-m)$, note that $U(1)=1$, as expected. We would like point out that each of the denominators in the above expression also has a zero inside the unit disk $|z|\leq 1$, which leads to a an additional linear equation for the boundary probabilities $p_l$ (since $U(z)$ is analytic inside the complex unit circle $|z|\leq 1$), as explained in Section 3.4. However, note that if $\tilde{z}_{(m)}$ is a solution of $\tilde{z}_{(m)} = A_m(S(\tilde{z}_{(m)}))$ with $|\tilde{z}_{(m)}|\leq 1$, we then have that $S(\tilde{z}_{(m)}) = S(A_m(S(\tilde{z}_{(m)})))$, implying that
\[
\begin{align*}
\tilde{z}_m &= S(z_m) \\
\hat{z}_m &= A_m(z_m)
\end{align*}
\] (49)

where \(z_m\) satisfies (39), and we obtain exactly the same set of linear equations as in (40) for the boundary probabilities.

5.2 Moments of the unfinished work

In a similar way as in Section 4.2, if we define \(\chi_m(z) \equiv A_m(S(z))\) and take the appropriate derivatives with respect to \(z\) of expression (48) for \(U(z)\), we find an expression for the mean and variance of the unfinished work:

\[
E[u] = \chi_N(1) + \frac{\chi_N(1)}{2(1-\chi_N(1))} + L_p \xi N(1)
\]

\[
\text{Var}[u] = \chi''_N(1) + \chi''_N(1) - \chi''_N(1)^2 + \frac{\chi''_N(1)}{3(1-\chi_N(1))} + \frac{\chi''_N(1)}{2(1-\chi_N(1))} 
\]

\[
\left(1 + \frac{\chi_N(1)}{2(1-\chi_N(1))}\right), \quad (50)
\]

\[
+ L_p \xi N(1) + S''(1)\xi N(1) + L_p \xi N(1) - L_p^2 \xi N(1)^2 + 2L_p T'(1)
\]

where

\[
\chi''_N(1) = NpL_p = \rho
\]

\[
\chi''_N(1) = \rho^2(1-1/N) + \rho(\lambda_1''(1)L_p/p + S''(1)/L_p)
\]

\[
\chi''''_N(1) = \rho^3(1-1/N)(1-2/N) + 3\rho^2(1-1/N)(\lambda_1''(1)L_p/p + S''(1)/L_p)
\]

\[
+ \rho(\lambda_1''(1)L_p^2/p + 3\lambda_1''(1)S''(1)/p + S''''(1)/L_p)
\]

The derivatives of \(\xi_N(z)\) and \(T'(1)\) in the previous expressions are given by (44) (or (45) in case the approximation for the boundary probabilities is applied) and (46) respectively.

5.3 Tail distribution of the unfinished work

A similar approach as in Section 4.3 can be applied to obtain a geometric tail approximation for the distribution of the unfinished work. We then obtain an approximation for the probability that the unfinished work exceeds a threshold \(n\):

\[
\Pr[u > n] \approx \frac{(1-\rho)\xi_N(S(z_{0,u}))}{1 - NS''(z_{0,u})\lambda_1(S(z_{0,u}))\lambda_1(S(z_{0,u}))N^{-1}z_{0,u}^{-n}}
\] (52)
where the smallest pole \( z_{0,u} \) of \( U(z) \) is the solution of \( z_{0,u} = \lambda_{1}(S(z_{0,u}))^{N} \), which is a real and positive quantity larger than 1, and similar to (49), can also be calculated from \( z_{0,u} = \lambda_{1}(z_{0,v})^{N} \). Also note that \( S(z_{0,u}) = z_{0,u} \), implying that the above formula can be expressed in terms of the smallest pole \( z_{0,v} \) of \( V(z) \) as well.

### 6. The Packet Delay

#### 6.1 Derivation of the probability generating function

We define the packet delay as the number of slots an arbitrary packet remains in the buffer, including its service time. Consider an arbitrary tagged packet that enters the buffer during slot \( k \). Note that slot \( k \) is not an arbitrary slot, but a slot where an arbitrary packet arrives (which for instance implies that at least one packet arrives). Due to the FCFS service discipline, the delay of a packet is determined by the amount of work in the buffer upon the tagged packet’s arrival. In particular, if we denote by \( \hat{u}_{k-1} \) the unfinished work just after slot \( k-1 \), i.e., at the beginning of slot \( k \), and by \( f_{k} \) the number of packets that have arrived during the same slot as the tagged one and will be served before it, then we find that

\[
d_k = (\hat{u}_{k-1} - 1)^+ + \sum_{i=1}^{f_k+1} s_i,
\]

where the random variables \( s_i, 1 \leq i \leq f_k+1 \), represent the service times of the tagged packet and the \( f_k \) packets that have arrived during its arrival slot and will be transmitted before it, which are all described by the pgf \( S(z) \). Let us denote by \( G(z,x) \) the steady-state joint generating function of the random variables \((u_{k-1},e_k)\) (with steady-state joint probability mass function \( g(i,j) \)) describing the unfinished work at the beginning and the number of packet arrivals during an arbitrary slot \( k \), i.e.,

\[
G(z,x) = \lim_{k \to \infty} E\left[z^{u_{k-1}}x^{e_k}\right] = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} z^i x^j g(i,j),
\]

and by \( H(z,x) \) the steady state joint pgf of \((\hat{u}_{k-1},f_k)\) (with joint probability mass function \( h(i,j) \)). Then, similar as in [10], we can establish the following relations :

\[
h(i,j) = \frac{1}{Np} \sum_{l=j+1}^{\infty} g(i,l)
\]

\[
H(z,x) = \frac{G(z,1) - G(z,x)}{Np(1-x)}.
\]

System equation (53) can be transformed into a relation between \( z \)-transforms, which together with (54), leads to an expression for the steady-state pgf \( D(z) \) describing the delay of an arbitrary packet

\[
D(z) = \frac{S(z)}{Npz} \left\{ H(z,S(z)) + (z-1) \sum_{l=0}^{\infty} \sum_{j=l+1}^{\infty} S(z)^j g(0,l) \right\}
\]

\[
= \frac{S(z)}{Npz} \left\{ \frac{G(z,1)-G(z,S(z))}{1-S(z)} + (z-1) \sum_{l=0}^{\infty} \frac{1-S(z)^l}{1-S(z)} g(0,l) \right\}.
\]
On the other hand, since the unfinished work satisfies the system equation

\[ u_k = (u_{k-1} - 1)^+ + \sum_{i=1}^{e_k} s_i \, , \]

assuming that a stochastic equilibrium is reached, this implies that the pgf \( U(z) \) (given by (48)) can also be written as (again, all \( s_i \)'s are described by \( S(z) \))

\[ U(z) = z^{-1} \left\{ G(z, S(z)) + (z-1) \sum_{l=0}^{\infty} S(z)^l g(0, l) \right\} \, . \]

In combination with the previous expression for \( D(z) \) (and \( U(z) = G(z, 1) \)) this finally leads to

\[ D(z) = \frac{S(z)}{z} \left( \frac{1}{z} S(1)(z-1) \right) U_0(z) \, . \] (55)

where, in view of \( U(0) = \text{Pr}[u=0] = \text{Pr}[v=0] = 1 - \rho \)

\[ U_0(z) = E\left[ z^u | u > 0 \right] = (U(z) - U(0))/\rho \] (56)

This formula for \( D(z) \), which expresses the delay of an arbitrary packet in terms of the unfinished work at the beginning of an arbitrary slot, and which apparently is independent of the details of the packet arrival process (correlation, etc.) was also derived in [11] via an alternative analysis. From this result, it is not difficult to express the performance measures concerning the packet delay in terms of the corresponding performance measures for the unfinished work, as shown next.

### 6.2 Moments of the packet delay

By taking the first two order derivatives of (55-56) with respect to \( z \) for \( z=1 \), we can establish the following relations between the mean and variance of the packet delay, and the mean and variance of the unfinished work:

\[ E[d] = L_p - 1 - \frac{S''(1)}{2L_p} + \frac{E[u]}{\rho} \]

\[ Var[d] = Var[s] - \frac{S''(1)}{3L_p} - \frac{S''(1)}{2L_p} \left( 1 - \frac{S''(1)}{2L_p} \right) + \frac{Var[u]}{\rho} - (1-\rho) \left( \frac{E[u]}{\rho} \right)^2 \] (57)
Through some simple calculations, one can verify that Little’s theorem indeed holds, i.e.,
\[ E[d] = E[v]/(Np). \]

### 6.3 Tail distribution of the packet delay

Again, applying an analogous technique as the one used for determining the tail distribution of the buffer contents and unfinished work, the asymptotic behavior of the packet delay tail distribution can be calculated from

\[
D(z) \approx \frac{C_d}{z - z_{0,d}} = -\frac{C_d}{z_{0,d}} \sum_{n=0}^{\infty} \left( \frac{z}{z_{0,d}} \right)^n, \quad C_d \triangleq \lim_{z \to z_{0,d}} (z - z_{0,d})D(z),
\]

where \( z_{0,d} \) is the smallest pole of \( D(z) \) which, in view of (55-56) leads to the observation that \( z_{0,d} \neq z_{0,u} \), and the probability that the packet delay exceeds a threshold \( n \) is approximated by

\[
\Pr[d > n] \approx \frac{1 - \rho}{\rho} \frac{(z_{0,u} - 1)}{L_p} \frac{S(z_{0,u})}{S(z_{0,u}) - 1} \frac{S(z_{0,u})}{N} \frac{\xi_N(S(z_{0,u}))}{\lambda_1(S(z_{0,u}))S'(z_{0,u})S(z_{0,u})\lambda_1(S(z_{0,u}))^{N-1}} z_{0,u}^{-n-1}. \quad (58)
\]
7. Numerical examples

We will illustrate the results derived throughout the previous sections by a few numerical examples. Let us therefore consider a two-state ON/OFF MMBP arrival process for each source, which means that

\[
G_{j1}(z) = 1 - \gamma + \gamma z, \quad 1 \leq j \leq 2,
\]

i.e., packets are generated at a rate \( \gamma \) during an ON period (corresponding with state \( S_1 \) in the modulating Markov chain, and no packets are generated during an OFF period (corresponding with state \( S_2 \)). Let us also denote by \( L_1 \) and \( L_2 \) the mean length of an ON and OFF period respectively, i.e., \( L_1 = \frac{1}{p_{12}} \) and \( L_2 = \frac{1}{p_{21}} \). Furthermore, we will assume that packets have a constant transmission time equal to \( L_p \) slots, implying that \( S(z) = z^{L_p} \). The load of the system is then given by

\[
\rho = N L_p \gamma \frac{L_1}{L_1 + L_2},
\]

and any subset containing five parameters of the set \( (N, \rho, \gamma, L_1, L_2, L_p) \) fully describes the packet arrival process in the buffer.

In the remainder we consider a multiplexer buffer fed by \( N=16 \) identical sources. For the arrival model described above, we have plotted the mean and variance of the buffer contents versus \( \rho \) in case of \( L_1=L_2=8 \) and for increasing packet length \( L_p=1,2,4,8,16 \) in Figs. 1 and 3 respectively. Varying the value of the load in this case means that the arrival rate \( \gamma \) is adjusted. From this figures we can conclude that the moments of the buffer contents only slightly depend...
on the packet length (and decrease for increasing values of $L_p$). In these figures, we have also included the curves when approximation (39) is used for the boundary probabilities, which leads to approximation (45) for the derivatives of $\xi_d(z)$; the absolute error between the exact and approximate results is so small that the latter are not visible in Figs. 1 and 3. Therefore, in Figs. 2 and 4, we have also plotted the relative error ($RE$) that is made when using the approximate formula for the mean and variance of the buffer contents. For $\rho$ approaching 0 or 1, the $RE$ of mean and variance approaches to zero, and reaches a maximum in between. In addition, the smaller the packets, the smaller the relative error tends to be. We also observed that, while for the mean value the approximation always forms an upper bound compared to the exact value, this is not necessarily the case for the variance. This explains why $RE[Var[v]]$ becomes zero around $\rho=0.84$, and then increase again. For the same set of system parameters, we have plotted the moments of the packet waiting time $w$ (which is defined as the number of slots a packet waits in the buffer before its transmission commences, and is therefore equal to the packet delay minus the packet transmission time) and their corresponding $RE$ in Figs. 5-8. Since the amount of packets in the buffer was roughly the same for any value of $L_p$, we expect that the packet waiting time increases for increasing values of $L_p$, which is readily observed from Figs. 5,7. The $RE$ of the mean and variance can become quite large (up to 30%); however, this is mainly the case in the area where the moments are small, and the difference between the approximation (dashed line) and exact results (full line) is again hardly visible in Figs. 5-7. The $RE$ now not necessarily tends to 0 for $\rho \to 0$; note that for $RE[E[w]]$ this is mainly due to the fact that we have plotted the moments of the packet waiting time instead of the packet delay (due to Little’s theorem, we would obtain exactly the same curves for $RE[E[d]]$ as for $RE[E[v]]$).

In Figs.9-10, we have confined ourselves to plotting the mean value of buffer contents and mean waiting time; the corresponding curves for the variance are completely similar, albeit on a different scale. In Figs. 9-10 we have plotted the mean buffer contents and packet waiting time versus $\rho$ in case of a constant packet length $L_p=10$, and increasing values of the mean length of an ON and OFF period, while keeping the probability that an arbitrary slot is an ON slot constant.
Again, under these circumstances, \( i \) the difference between and exact and approximate values of the mean buffer contents and packet waiting time is negligible and \( ii \) the mean buffer contents (and consequently the waiting time as well, since \( L_p=10 \) is constant in these curves) is roughly independent of the mean length of the ON and OFF periods. However, this is partly due to the fact that the packet arrival rate \( \gamma \) during an ON period is relatively small (note that \( \gamma<0.0125 \) for \( \rho<1 \)). In Figs. 11-12 we therefore consider the (rather extreme case) \( \gamma=0.05 \), \( L_p=10 \) and increasing values of the mean length of an active period. In this case, we have that \( L_2 \to \infty \) for \( \rho \to 0 \). It is now clear that the mean values of buffer contents and waiting time now depend on the mean length of an active period. In addition, we observe from these curves that the error that is made by using the approximation for the boundary probabilities (dashed line) can be significant when the load is low, and the arrival rate \( \gamma \) and \( L_1 \) are high. Another feature is that the mean waiting time not necessarily tends to 0 for \( \rho \to 0 \). This is due to the fact that under these conditions, we have sources that sporadically \( (L_2 \to \infty) \) generate packets at a rate \( \gamma \) during a geometrically distributed ON period with mean length \( L_1 \), and the intersection with the ordinate \( \rho=0 \) is given by the mean waiting time for packets generated by such a single source.

In Figs. 13-20 we have plotted the tail distribution of the buffer contents and packet waiting time for a number of system parameters. Each curve for the exact value of the parameters of the geometric-tail behavior (full line) is accompanied by the corresponding approximation (dashed line) obtained by using approximation (39) for \( \xi_v(z) \). We may conclude that for the wide range of system parameters considered in these figures, the approximate results are extremely accurate. Furthermore it is observed that, for the buffer contents as well as the packet waiting time, the approximation forms an upper bound, and consequently, is on the safe side. The only numerical calculation required to obtain this upper bound is that of the dominant pole \( z_{0,v} \) (and \( z_{0,a}=A_2(z_{0,v}) \)) of the pgf describing the steady-state buffer contents (packet waiting time) distribution. Insertion of the system parameters in the corresponding formulas then leads to these curves.
In Figs. 13-14, we have plotted the probability that the buffer contents and packet waiting time exceed a certain threshold $n$ versus $n$, for a multiplexer buffer with $N=16$ sources, a system load $\rho=0.8$, an arrival rate $\gamma=0.1$ during an ON period, and a mean length of an ON period $L_1=50$, for varying values of the packet length $L_1$. Again, as was also observed when discussing the moments, for these parameter values the results for the quantile of the buffer contents (defined as the smallest value of $n(x)$ for which $\Pr[v>n(x)]<10^{-x}$) only slightly depend on the packet length (and in fact slightly decrease for increasing $L_p$). From Figs. 13-14 (and the figures thereafter) it also follows that the quantile of the waiting time is roughly equal to the quantile of the buffer contents multiplied by the packet length $L_p$. This is of course mainly due to the fact that we consider constant-length packets equal to $L_p$. In Figs 15-16, we show the results for the buffer contents and packet waiting time for $N=16$, $\gamma=0.1$, $L_1=50$ and $L_p=10$, and increasing values of the load (which means that the arrival rate $\gamma$ during the ON periods increases), where, as expected, the quantiles of both quantities increase for increasing values of the load. In Figs. 17-18 the tail distribution of buffer contents and delay are shown for $N=16$, $\rho=0.8$, $\gamma=0.1$, and $L_p=10$, for various values of $L_1$; the longer the ON periods become, the larger the quantiles of buffer contents and packet waiting time will be. This is not the case if we increase $L_1$ while keeping $\gamma L_1$ constant (which means that the expected value of the total number of packets generated during an ON period remains constant), as can be deduced from Figs. 19-20. We observe that under these circumstances the quantiles of buffer contents and packet waiting time decrease as $L_1$ increases, which is a result of the traffic smoothing that is introduced as $L_1$ augments.
Figure 11: $E[v]$ versus $\rho$, $N=16$, $\gamma=0.05$, $L_p=10$

Figure 12: $E[w]$ versus $\rho$, $N=16$, $\gamma=0.05$, $L_p=10$

Figure 13: $\Pr[v>n]$ versus $n$, $N=16$, $\rho=0.8$, $\gamma=0.1$, $L_p=50$

Figure 14: $\Pr[w>n]$ versus $n$, $N=16$, $\rho=0.8$, $\gamma=0.1$, $L_p=50$
Figure 15: $\text{Pr}[v>n]$ versus $n$, $N=16$, $\gamma=0.1$, $L_p=10$, $L_1=50$

Figure 16: $\text{Pr}[w>n]$ versus $n$, $N=16$, $\gamma=0.1$, $L_p=10$, $L_1=50$

Figure 17: $\text{Pr}[v>n]$ versus $n$, $N=16$, $\rho=0.8$, $\gamma=0.1$, $L_p=10$

Figure 18: $\text{Pr}[w>n]$ versus $n$, $N=16$, $\rho=0.8$, $\gamma=0.1$, $L_p=10$
Figure 19: Pr[v > n] versus n, N=16, ρ=0.8, L_p=10

Figure 20: Pr[w > n] versus n, N=16, ρ=0.8, L_p=10
REFERENCES


APPENDIX: The eigenvalues and eigenvectors of $Q(z)$

As became clear from the analysis throughout this report, an important role is played by the eigenvalues and eigenvectors of $Q(z)$. Let us denote by $\Lambda(z)$ the $2\times2$ diagonal matrix containing the eigenvalues $\lambda_i(z)$, and by $U(z)$ and $W(z)$ the $2\times2$ matrices containing the corresponding right column and left row eigenvectors respectively, i.e.,

$$W(z)Q(z)U(z) = \Lambda(z), \quad U(z)W(z) = W(z)U(z) = I,$$

with $I$ the $2\times2$ diagonal unity matrix. As in [1], for reasons that become will clear during the analysis in this report, the normalization constant of the left row and right column eigenvectors are chosen such that

$$U(z)I = W(z)I = I ,$$

We will denote by $u_j(z)$ ($w_j(z)$) the right column (left row) eigenvector corresponding to $\lambda_j(z), 1 \leq j \leq 2$.

Taking into account expression (1) for $Q(z)$, the characteristic equation of this matrix can be written as

$$\lambda(z)^2 - (q_{11}(z) + q_{22}(z))\lambda(z) + q_{11}(z)q_{22}(z) - q_{12}(z)q_{21}(z) = 0 .$$

Solution of the characteristic equation of $Q(z)$ leads to

$$\begin{bmatrix}
\lambda_1(z) \\
\lambda_2(z)
\end{bmatrix} = \frac{q_{11}(z) + q_{22}(z) \pm \left( (q_{11}(z) - q_{22}(z))^2 + 4q_{12}(z)q_{21}(z) \right)^{1/2}}{2} .$$

In addition, solving (13) and (14) yields the following expression for $U(z)$:

$$U(z) = \begin{bmatrix}
\left( \lambda_1(z) - q_{11}(z) + q_{21}(z) \right)q_{12}(z) & \left( \lambda_2(z) - q_{11}(z) + q_{21}(z) \right)q_{12}(z) \\
\left( q_{11}(z) - \lambda_1(z) \right)(\lambda_2(z) - \lambda_1(z)) & \left( q_{11}(z) - \lambda_2(z) \right)(\lambda_1(z) - \lambda_2(z)) \\
\left( \lambda_1(z) - q_{11}(z) + q_{21}(z) \right) & \left( \lambda_2(z) - q_{11}(z) + q_{21}(z) \right) \\
\left( \lambda_1(z) - \lambda_2(z) \right) & \left( \lambda_2(z) - \lambda_1(z) \right)
\end{bmatrix} .$$

In a similar way, we can also calculate the elements of $W(z)$; however, the explicit expression of this matrix is not required in our derivations. Note that

$$\begin{align*}
\lambda_1(1) &= 1 \\
\lambda_2(1) &= 1 - p_{12} - p_{21} .
\end{align*}$$

$\lambda_1(z)$ is the Perron-Fröbenius (P-F) eigenvalue of $Q(z)$, and is an important quantity for the analysis in this report. Similarly, $u_1(z)$ will be referred to as the P-F eigenvector of $Q(z)$. 

In our expressions for the mean and variance of the buffer contents, unfinished work and packet delay, we also require the first, second and third order derivatives with respect to \( z \) for \( z=1 \) of the P-F eigenvalue. Taking the appropriate derivatives of the characteristic equation (A.3), we obtain with (A.6)

\[
\lambda'_1(1) = \sum_{i=1}^{2} \sum_{j=1}^{2} q'_{ij}(1) \approx p
\]

\[
\lambda''_1(1) = \sum_{i=1}^{2} \sum_{j=1}^{2} q''_{ij}(1) + 2 \frac{\lambda'(1)\lambda_2(1) + q_{12}(1)q_{21}(1) - q_{11}(1)q_{22}(1)}{p_{12} + p_{21}}
\]

\[
\lambda'''_1(1) = \sum_{i=1}^{2} \sum_{j=1}^{2} q'''_{ij}(1) + 3 \frac{\lambda''(1)\lambda'_2(1) + \lambda'(1)\lambda_2(1) + q_{12}(1)q_{21}(1) + q_{12}(1)q_{21}(1)}{p_{12} + p_{21}}
\]

\[
-3 \frac{q_{11}(1)q_{22}(1) + q_{11}(1)q_{22}(1)}{p_{12} + p_{21}}
\]

where \( p \) was defined in equation (13). The \( n \)-th derivative of \( \lambda_2(z) \) with respect to \( z \) for \( z=1 \) is readily calculated in terms of the \( n \)-th derivative of \( \lambda_1(z) \) from the relation

\[
\lambda'_1(z) + \lambda'_2(z) = q_{11}(z) + q_{22}(z)
\]

Similarly, we also require the first and second order derivatives with respect to \( z \) for \( z=1 \) of the P-F eigenvector. First of all, we point out that due to (A.5), \( U(1) \) satisfies

\[
U(1) = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \quad \text{\textsuperscript{(1)}}
\]

Then, taking the appropriate derivatives of the equations generated by (A.2) and the relation \( Q(z)U(z)=U(z)\Lambda(z) \), we obtain

\[
\dot{U}(1) = \frac{\delta}{p_{12}} \begin{bmatrix} -\sigma_2 & \sigma_2 \\ \sigma_1 & -\sigma_1 \end{bmatrix}
\]

\[
\ddot{U}(1) = \frac{\gamma_1}{p_{12}} \begin{bmatrix} -\sigma_2 & \sigma_2 \\ \sigma_1 & -\sigma_1 \end{bmatrix} + \frac{\gamma_2}{p_{12}} \begin{bmatrix} -\sigma_2 & \sigma_2 \\ \sigma_1 & -\sigma_1 \end{bmatrix}
\]

where

\[
\delta = p - (q'_{11}(1) + q'_{12}(1))
\]

\[
\gamma_1 = \lambda''_1(1) - 2 \frac{\delta p}{p_{12} + p_{21}} - (q''_{11}(1) + q''_{12}(1)) - 2(q_{11}(1)u_{11}(1) + q_{12}(1)u_{21}(1))
\]

\[
\gamma_2 = -2 \frac{\delta \lambda'_2(1)}{p_{12} + p_{21}} + 2(q_{11}(1)u_{11}(1) + q_{12}(1)u_{21}(1))
\]

\(^{(1)}\) Strictly speaking, the second column of \( U(z) \) is no longer an eigenvector of \( Q(z) \) for \( z=1 \). However, it is not difficult to verify that all the results derived throughout this report still hold when taking the limit \( z=1 \).