A mathematical model for discrete-time buffer systems with correlated output process

Herwig BRUNEL

Department of Computer Science, Ghent State University, Sint-Pietersnieuwstraat41, B-9000 Gent, Belgium

Abstract. This article considers an infinite buffer system with one or more input channels and multiple output channels. Transmission of messages from the buffer is synchronous and the arrival process of messages to the buffer is general. Each of the output channels is subjected to a random interruption process, i.e., the number of available output channels varies in time and is stochastic. The analysis of this system is carried out under the assumption that the output process can be described as a first order Markov process, i.e., the probability distribution of the number of available output channels during some clock time interval depends only on the number of available output channels during the previous clock time interval.

A set of equations describing the behavior of this buffer system is derived. For a number of interesting special cases this set is solved and explicit expressions are obtained for the probability generating function of the number of messages in the buffer. Several prior studies are found as special cases of the present one. An illustrative example is treated and the results are compared to the ones obtained for an uncorrelated output process with the same equilibrium distribution. Some considerable deviations from these results are found.

Keywords: Queues, Markov processes, communications, performance

1. Introduction

In recent years the study of discrete-time buffer systems has received considerable attention. Both systems with finite waiting room [1–4] and systems with infinite waiting room [5–15] have been considered. Systems where the output channels are always available have been treated [5–7] as well as systems where the output lines are subjected to random interruptions in time [1–4,8–15]. Not only has the case been studied where the buffer has only one single output line [1–6,8–15], but a considerable research effort has also been spent on multiple output buffers [7,14–15]. The present paper belongs to this latter category.

In [14–15] a discrete-time buffer system with \( m \) output channels is studied. The output channels are subjected to random interruptions in time so that during each clock time interval a stochastic number of them (between 0 and \( m \)) is available for the transmission of data. In [14] it is assumed that the number of available output lines is 0 or \( m \), i.e., the output channels are all blocked or all available. In [15] the model is more general in that any integer number between 0 and \( m \) of output channels may be available, i.e., the

The author is with the Belgian National Fund for Scientific Research (NFWO) at the department of Computer Science (ICI), Ghent State University, Ghent, Belgium.

This work was supported by the Belgian National Fund for Scientific Research (N.F.W.O.).

Received May 1983; revised August 1983

North-Holland

0377-2217/84/$3.00 © 1984, Elsevier Science Publishers B.V. (North-Holland)
output lines need not be interrupted by the same switch. There remains, however, one important restriction: the numbers of available output lines during consecutive clock time intervals are treated as i.i.d. stochastic variables, i.e., no correlation from clock time interval to clock time interval is allowed.

In the present paper this restriction is eliminated by modelling the output process of the buffer system as a first order Markov chain. Thus it is assumed that the number of available output channels during each clock time interval is a nonnegative integer random variable which takes values between 0 and \( m \), the density function of which is dependent on the number of available output lines in the immediately preceding clock time interval. This assumption is equivalent to assuming that the sojourn times of the output states of the buffer system (the system is in output state \( i \) \((0 \leq i \leq m)\) when there are \( i \) output channels available) are geometrically distributed random variables with independently chosen parameters, whereas in the model [15] the sum of these parameters must equal unity, which is a very severe restriction.

The specific assumptions of the analysis are given in Section 2, along with some terminology. In section 3, the basic equations describing the system are derived. Section 4 treats three interesting special cases for which explicit expressions for the probability generating function of the equilibrium buffer occupancy can be obtained. Finally, an illustrative example of the method is treated in Section 5.

2. The mathematical model

The buffer system is depicted in Fig. 1. It consists of one or more input channels, an infinite waiting room and \( m \) output channels subjected to random interruptions in time. The messages arrive in the buffer in a random fashion via one of the input channels, wait in the buffer for some time and are finally taken out via one of the available output channels. It is assumed that all messages have the same constant length and that the output channels transmit data with constant speed. This implies that the time required to transmit each message is constant; this constant (transmission) time will be indicated by the term clock time period, or, alternatively, (time) slot. The message transmission is synchronous, i.e., the messages are taken out from the buffer for transmission at each discrete clock time (provided at least one output channel is available). A consequence of the above assumption is that a message cannot leave the buffer at the end of the time slot during which it has entered the buffer, because one full time slot is needed for its transmission. Therefore, when the buffer contains \( n \) messages at the beginning of a slot, no other messages than these \( n \) can leave the buffer at the end of this slot, even if there have been some arrivals during this slot.

The arrival process of the messages is described in terms of a sequence of i.i.d. random variables, which denote the numbers of arriving messages during the consecutive time slots. Let \( a(n) \) and \( A(z) \) denote the (common) probability density function and probability generating function respectively of these stochastic variables. That is

\[
a(n) = \text{Prob}[n \text{ messages arrive to the buffer during a time slot}], \quad n = 0, 1, 2, \ldots,
\]

and

\[
A(z) = \sum_{n=0}^{\infty} a(n) z^n.
\]

No restrictions concerning the shape of \( a(n) \) are made, i.e., the arrival process is general.

The output process of the buffer system is characterized through a first order Markov chain with transition probabilities \( r_j(n) \), defined as

\[
r_j(n) = \text{Prob}[n \text{ output lines available during a time slot given there were } j \text{ output lines available during the immediately preceding time slot}], \quad j, n \in \{0, 1, \ldots, m\}.
\]

The (conditional) probability generating functions corresponding to the densities \( r_j(n) \) are denoted by \( R_j(z) \), i.e.,

\[
R_j(z) = \sum_{n=0}^{m} r_j(n) z^n, \quad j \in \{0, 1, \ldots, m\}.
\]
Fig. 1. An infinite buffer system with multiple output channels subjected to random interruptions.

Fig. 2. Illustration to the definition of the stochastic sequences \( \{ u_k \} \), \( \{ a_k \} \) and \( \{ r_k \} \).

Fig. 3. An infinite buffer system with multiple output channels, a part of which is subjected to random interruptions.

Fig. 4. Mean buffer occupancy \( \bar{N} \) versus the mean arrival rate \( \lambda \) (per time slot) for a buffer system with two output channels, one of which is subjected to random interruptions. All curves correspond to a value of 0.5 for \( \sigma \). Curves (a), (b), (c), (d) and (e) for \( \alpha = \beta = 0.5, 0.8, 0.9, 0.95 \) and 0.99 respectively.
The equilibrium probabilities of the Markov chain will be denoted by \( r(n) \) and the symbol \( R(z) \) will be used to indicate the corresponding probability generating function. Thus

\[
r(n) = \text{Prob}[n \text{ output channels available during a time slot}], \quad n \in \{0, 1, \ldots, m\},
\]

and

\[
R(z) = \sum_{n=0}^{m} r(n)z^n.
\]

Here the \( r(n) \)'s follow from the set of equilibrium equations

\[
r(n) = \sum_{j=0}^{m} r_j(n)r(j), \quad n \in \{0, 1, \ldots, m\}.
\]

Note that the above assumptions imply that the sojourn time of the \( i \)th output state (the buffer system is in output state \( i \) whenever \( i \) output channels are available) is a geometrically distributed random variable with parameter \( r_i(i) \), i.e.,

\[
\text{Prob}[\text{output state remains } i \text{ during exactly } n \text{ consecutive slots}]
= \left[1 - r_i(i)\right]\left[r_i(i)\right]^{n-1}, \quad n = 1, 2, \ldots.
\]

3. The basic system equations

Let us define the stochastic variables \( a_k, r_k \) and \( u_k \) as (Fig. 2):

\[
a_k = \text{number of messages that arrive to the buffer during the } k \text{th time slot},
\]
\[
r_k = \text{number of output channels which are available during the } k \text{th time slot},
\]
\[
u_k = \text{buffer occupancy (i.e. number of messages in the buffer) just after the } k \text{th time slot}.
\]

Then the following equations hold:

\[
u_{k+1} = \begin{cases} u_k - r_{k+1} + a_{k+1} & \text{if } r_{k+1} < u_k, \\ a_{k+1} & \text{if } r_{k+1} \geq u_k, \end{cases}
\]

or, equivalently,

\[
u_{k+1} = (u_k - r_{k+1})^+ + a_{k+1}
\]

(1)

where the notation \((\ldots)^+\) denotes \(\max(0, \ldots)\).

In [15] the same system was studied under the assumption that the random variables \( \{r_k\} \) were uncorrelated. A consequence of this assumption was that the stochastic variables \( u_k \) and \( r_{k+1} \) were statistically independent, which implied that the sequence \( \{u_k\} \) formed a Markov chain, which, in turn, made the analysis rather easy. This is no longer true if statistical dependence is allowed between the \( r_k \)'s. Therefore we define a new, two-dimensional, state description of our system which does have the Markov property: the state of the buffer system just after the \( k \)th time slot is described by the pair \( (u_k, r_k) \). The transition probabilities of the Markov chain \( \{(u_k, r_k)\} \) are denoted by \( p(l, n | i, j) \), i.e.,

\[
p(l, n | i, j) = \text{Prob}[u_{k+1} = l \text{ and } r_{k+1} = n \mid u_k = i \text{ and } r_k = j],
\]

\[
l, i = 0, 1, 2, \ldots; \quad n, j \in \{0, 1, \ldots, m\}.
\]

They can be expressed as

\[
p(l, n | i, j) = \text{Prob}[r_{k+1} = n | u_k = i \text{ and } r_k = j] \text{Prob}[u_{k+1} = l | r_{k+1} = n \text{ and } u_k = i \text{ and } r_k = j]
= r_f(n) \times \text{Prob}[\left(i - n\right)^+ + a_{k+1} = l]
\]
where we have used our definition of the output process, and Equation (1); hence

\[ p(l, n | i, j) = \begin{cases} r_j(n) \times a(l - (i - n)^+) & \text{if } l \geq (i - n)^+, \\ 0 & \text{if } l < (i - n)^+. \end{cases} \]  
\[ p(l, n) = \lim_{k \to \infty} \Pr[u_k = l \text{ and } r_k = n], \quad l = 0, 1, 2, \ldots; n \in \{0, 1, \ldots, m\}. \]

They can be computed from the set of equations

\[ p(l, n) = \sum_{i=0}^{\infty} \sum_{j=0}^{m} p(i, j) p(l, n | i, j) \]

which, using (2), can be written as

\[ p(l, n) = \sum_{i=0}^{l+n} \sum_{j=0}^{m} p(i, j) r_j(n) a(l - (i - n)^+), \quad l = 0, 1, 2, \ldots; n \in \{0, 1, \ldots, m\}. \]  
\[ p(l, n) = \sum_{l=0}^{\infty} \left[ \sum_{n=0}^{m} p(l, n) \right] z^l = \sum_{n=0}^{m} p_n(z). \]  

Let us introduce the partial generating functions

\[ P_n(z) = \sum_{l=0}^{\infty} p(l, n) z^l \]

then we can find the probability generating function \( P(z) \) of the equilibrium number of messages in the buffer, \( u \), from

\[ P(z) = \sum_{l=0}^{\infty} \Pr[u = l] z^l = \sum_{l=0}^{\infty} \left[ \sum_{n=0}^{m} p(l, n) \right] z^l = \sum_{n=0}^{m} P_n(z). \]  

Combining (3)–(4) we obtain the following set of equations for the \( P_n(z) \)'s:

\[ P_n(z) = \sum_{l=0}^{l+n} \sum_{i=0}^{m} p(i, j) r_j(n) a(l - (i - n)^+) \]

\[ = \sum_{j=0}^{m} r_j(n) \sum_{i=0}^{\infty} p(i, j) \sum_{l=(i-n)^+}^{\infty} a(l - (i - n)^+) z^l \]

\[ = A(z) \sum_{j=0}^{m} r_j(n) \sum_{i=0}^{\infty} p(i, j) z^{(i-n)^+} \]

\[ = A(z) \sum_{j=0}^{m} r_j(n) \left[ z^{-n} P_j(z) + \sum_{i=0}^{n} p(i, j)(1 - z^{i-n}) \right] \]

\[ = z^{-n} A(z) \left[ \sum_{j=0}^{m} r_j(n) P_j(z) + \sum_{i=0}^{n} (z^n - z^i) g(i, n) \right], \]

\[ n \in \{0, 1, \ldots, m\}, \]

where the (unknown) quantities \( g(i, n) \) are defined by

\[ g(i, n) = \sum_{j=0}^{m} r_j(n) p(i, j), \quad n \in \{0, 1, \ldots, m\}; i \in \{0, 1, \ldots, n - 1\} \text{ for each } n. \]
The Equations (6) can finally be written as
\[ z^n P_n(z) - A(z) \sum_{j=0}^{m} r_j(n) P_j(z) = A(z) g_n(z), \quad n \in \{0, 1, \ldots, m\}, \] (7)
where the function \( g_n(z) \) is a polynomial of degree \( n \) in \( z \), defined by
\[ g_n(z) = \sum_{i=0}^{n} (z^n - z') g(i, n), \quad n \in \{0, 1, \ldots, m\}. \]

It is easily seen that the functions \( g_n(z) \) can be expressed as
\[ g_0(z) = 0, \quad g_n(z) = (z - 1) f_{n-1}(z) \quad \text{if} \quad n \geq 1. \] (8)
Here \( f_{n-1}(z) \) denotes an unknown polynomial of degree \( n - 1 \) in \( z \).

The Equations (7) compose a set of \((m + 1)\) linear equations in the \((m + 1)\) unknown functions \( P_n(z) \) \((n \in \{0, 1, \ldots, m\})\), which can thus be derived by solving this set for the \( P_n(z)\)'s. The resulting expressions for the \( P_n(z)\)'s contain known quantities on one hand and the unknown polynomials \( f_k(z) \) \((k \in \{0, 1, \ldots, m - 1\})\) on the other hand. A similar expression for \( P(z) \) is then obtained by using the expressions for the \( P_n(z)\)'s in Equation (5). The complete determination of the unknown polynomials \( f_k(z) \) is done by use of Rouche's theorem (see e.g. [16]) and the fact that \( P(z) \) is an analytic function of \( z \) inside the unit disk of the complex plane. The method is explained in detail in the next section, where we present a number of interesting special cases of our model, which allow explicit solution.

4. Special cases for which explicit expressions for \( P(z) \) can be obtained

4.1. Uncorrelated output process

In a system with an uncorrelated output process we have
\[ r_j(n) = r(n), \quad R_j(z) = R(z) \quad \forall j \in \{0, 1, \ldots, m\}. \]
The Equations (7) can be written as
\[ P_n(z) = z^{-n} A(z) \left[ r(n) P(z) + g_n(z) \right] \]
which, upon summation over \( n \) between 0 and \( m \), yields
\[ P(z) = A(z) \left[ P(z) R(1/z) + \sum_{n=0}^{m} z^{-n} g_n(z) \right]. \]

It follows that \( P(z) \) is given by
\[ P(z) = \frac{A(z) \sum_{n=0}^{m} z^{-n} g_n(z)}{z^{-m} - z^{-m} R(1/z) A(z)} \]
or, using (8),
\[ P(z) = \frac{(z - 1) A(z) f_{m-1}(z)}{z^{-m} - z^{-m} R(1/z) A(z)}. \] (9)

Here \( f_{m-1}(z) \) denotes an unknown polynomial of degree \( m - 1 \) in \( z \). This expression corresponds exactly to the one derived in [15], i.e., the whole analysis in [15] is found as a special case of the present one. Notice
that the determination of the $m$ unknown parameters of $f_{m-1}^+(z)$ can be carried out as described in [15] as well.

### 4.2. Partially correlated output process

Suppose that the number of available output channels during a time slot is described by the density functions $s_1(n)$ or $s_2(n)$ depending on whether the number of available output lines in the previous slot was greater than a number $K$ ($K < m$) or not, i.e., not the actual number ($j$) of available output channels during the previous time slot is important, but the range in which $j$ lies. Formally we then have

$$ r_j(n) = \begin{cases} s_0(n) & \text{if } j \leq K, \\ s_1(n) & \text{if } j > K. \end{cases} $$

Let $S_0(z)$ and $S_1(z)$ denote the probability generating functions corresponding to the densities $s_0(n)$ and $s_1(n)$ respectively.

The Equations (7) for this case reduce to

$$ P_n(z) = z^{-n}A(z) \left[ s_0(n) \sum_{j=0}^{K} P_j(z) + s_1(n) \sum_{j=K+1}^{m} P_j(z) + g_n(z) \right]. \quad (10) $$

Introducing the functions

$$ Q_0(z) = \sum_{j=0}^{K} P_j(z) \quad \text{and} \quad Q_1(z) = \sum_{j=K+1}^{m} P_j(z) $$

and summing (10) over $n$ between 0 and $K$ on one hand and between $K+1$ and $m$ on the other hand, the following equations can be derived:

$$ Q_0(z) = A(z) \sum_{n=0}^{K-1} z^{-n} \left[ s_0(n) Q_0(z) + s_1(n) Q_1(z) + g_n(z) \right], $$

$$ Q_1(z) = A(z) \sum_{n=K+1}^{m} z^{-n} \left[ s_0(n) Q_0(z) + s_1(n) Q_1(z) + g_n(z) \right]. $$

Introducing

$$ S_{00}(z) = \sum_{j=0}^{K} s_0(j) z^j, \quad S_{01}(z) = \sum_{j=K+1}^{m} s_0(j) z^j, $$

$$ S_{10}(z) = \sum_{j=0}^{K} s_1(j) z^j, \quad S_{11}(z) = \sum_{j=K+1}^{m} s_1(j) z^j $$

and

$$ g_0^*(z) = \sum_{n=0}^{K} z^{-n} g_n(z), \quad g_1^*(z) = \sum_{n=K+1}^{m} z^{-n} g_n(z) $$

(12)

this set of equations can be rewritten as

$$ [1 - A(z) S_{00}(1/z)] Q_0(z) - A(z) S_{10}(1/z) Q_1(z) = A(z) g_0^*(z), $$

$$ -A(z) S_{01}(1/z) Q_0(z) + [1 - A(z) S_{11}(1/z)] Q_1(z) = A(z) g_1^*(z). $$
The functions $Q_0(z)$ and $Q_1(z)$ are easily obtained from this set. Summation of the resulting expressions yields an expression for $P(z)$:

$$
P(z) = \frac{A(z) \left[ 1 + A(z) \left[ S_{01}(1/z) - S_{11}(1/z) \right] \right] g_0^\dagger(z) + \left[ 1 + A(z) \left[ S_{10}(1/z) - S_{00}(1/z) \right] \right] g_1^\dagger(z)}{1 - A(z) \left[ S_{00}(1/z) + S_{11}(1/z) \right] + A^2(z) \left[ S_{00}(1/z) S_{11}(1/z) - S_{01}(1/z) S_{10}(1/z) \right]}.
$$

(13)

It now remains for us to determine the unknown functions $g_0^\dagger(z)$ and $g_1^\dagger(z)$, defined in (12). Therefore, further mathematical manipulations on the expression (13) are necessary.

First, notice that $S_{00}(1/z), S_{01}(1/z), S_{10}(1/z), S_{11}(1/z)$ can be expressed as follows:

$$
S_{00}(1/z) = z^{-K} \hat{S}_{00}(z), \quad S_{01}(1/z) = z^{-m} \hat{S}_{01}(z),
$$

$$
S_{10}(1/z) = z^{-K} \hat{S}_{10}(z), \quad S_{11}(1/z) = z^{-m} \hat{S}_{11}(z).
$$

(14)

Here the functions $\hat{S}_{00}(z), \hat{S}_{01}(z), \hat{S}_{10}(z)$ and $\hat{S}_{11}(z)$ are known polynomials in $z$.

Second, $g_0^\dagger(z)$ and $g_1^\dagger(z)$ can be written as

$$
g_0^\dagger(z) = \sum_{n=1}^{K} z^{-n} (z - 1)f_{n-1}(z) = z^{-K} (z - 1)h_{K-1}(z),
$$

$$
g_1^\dagger(z) = \sum_{n=K+1}^{m} z^{-n} (z - 1)f_{n-1}(z) = z^{-m} (z - 1)h_{m-1}(z)
$$

(15)

where $h_{K-1}(z)$ and $h_{m-1}(z)$ are unknown polynomials in $z$ of degrees $K - 1$ and $m - 1$ respectively.

Using Equations (14)--(15) in expression (13), we are led to an expression for $P(z)$ which contains only analytic functions of $z$ in numerator and denominator:

$$
P(z) = \frac{(z - 1)A(z) \left[ \left[ z^m + A(z) \left[ \hat{S}_{00}(z) - \hat{S}_{11}(z) \right] \right] h_{K-1}(z) + \left[ z^K + A(z) \left[ \hat{S}_{10}(z) - \hat{S}_{00}(z) \right] \right] h_{m-1}(z) \right]}{z^{m+K} - A(z) \left[ z^m \hat{S}_{00}(z) + z^K \hat{S}_{11}(z) \right] + A^2(z) \left[ \hat{S}_{00}(z) \hat{S}_{11}(z) - \hat{S}_{01}(z) \hat{S}_{10}(z) \right]}.
$$

(16)

Using Rouche’s theorem as explained in the appendix of [15] one can prove that the denominator of (16) has exactly $(m + K)$ zeros inside the unit disk ($z : |z| < 1$) of the complex plane, one of which equals unity, whenever the condition for the existence of a stochastic equilibrium of the buffer system under consideration is fulfilled, i.e., whenever the average number of arrivals per time slot is strictly less than the average number of available output channels per time slot. Since the whole analysis was carried through under this condition, the result applies here. However $P(z)$ must be an analytic function of $z$ inside the unit disk of the complex plane, since it denotes a probability generating function. Thus the $(m + K)$ zeros of the denominator of (16) must be zeros of the numerator of (16) as well. This observation yields $(m + K - 1)$ linear equations in the unknown coefficients of the powers of $z$ in $h_{K-1}(z)$ and $h_{m-1}(z)$—no equation is obtained for the zero $z = 1$—which, together with the normalisation equation $P(1) = 1$, can be used for the computation of these unknown parameters. Once these parameters are found, $h_{K-1}(z)$ and $h_{m-1}(z)$ are known polynomials in $z$, and Equation (16) provides an expression for the probability generating function $P(z)$ of the buffer occupancy, which contains known quantities only.

The equilibrium distribution of the number of available output channels per time slot, for this case, is easily shown to be characterized by the generating function

$$
R(z) = \frac{S_{10}(1)S_0(z) + S_{01}(1)S_1(z)}{S_{10}(1) + S_{01}(1)}.
$$
If we use this expression for $R(z)$ in (9), we obtain an expression for the probability generating function $P_n(z)$ of the buffer occupancy for a buffer system with uncorrelated output process but with the same equilibrium distribution for the number of available output channels per clock time interval. This expression reads

$$P_n(z) = \frac{[S_{01}(1) + S_{10}(1)](z - 1)A(z) f_{n-1}^*(z)}{[S_{01}(1) + S_{10}(1)] z^m - A(z)[S_{10}(1)[z^{m-k} S_{00}(z) + \hat{S}_{01}(z)] + S_{01}(1)[z^{m-k} S_{10}(z) + \hat{S}_{11}(z)]]}.$$ (17)

Comparison of (16) and (17) allows an evaluation of the influence of the correlation of the output process on the statistics of the buffer. A special case will be treated in detail in the next section.

4.3. Only two possible values for the number of available output channels

Suppose we have a buffer system with $M$ output channels, $N$ of which are always available, and $M - N$ of which are randomly interrupted by one common switch (Fig. 3). In this case $r_j(n)$ is defined only if $j = N$ or $M$; moreover $r_j(n) = 0$ unless $n = N$ or $M$. For simplicity let us introduce new notations for the four remaining non-zero transition probabilities:

$$r_M(N) = 1 - \alpha; \quad r_M(M) = \alpha;$$
$$r_N(N) = \beta; \quad r_N(M) = 1 - \beta.$$

Equations (7) for this case reduce to

$$[z^N - \beta A(z)] P_N(z) - (1 - \alpha) A(z) P_M(z) = A(z) g_N(z),$$
$$- (1 - \beta) A(z) P_N(z) + [z^M - \alpha A(z)] P_M(z) = A(z) g_M(z)$$

from which $P_N(z)$ and $P_M(z)$, and hence $P(z)$, can easily be derived. The resulting expression for $P(z)$ reads:

$$P(z) = \frac{(z - 1) A(z) \left( [z^N + (1 - \alpha - \beta)A(z)] f_{m-1}(z) + [z^M + (1 - \alpha - \beta)A(z)] f_{n-1}(z) \right)}{z^{M+N} - (\alpha z^N + \beta z^M) A(z) - (1 - \alpha - \beta) A^2(z)},$$ (18)

where we have used (8), with the additional agreement that $f_{-1}(z) = 0$.

Notice that the result (18) could have been derived as a special case of (16) as well. The direct derivation from (7), however, is less complicated.

The equilibrium distribution of the number of available output channels per time slot is now characterized by

$$R(z) = \frac{(1 - \alpha) z^N + (1 - \beta) z^M}{2 - \alpha - \beta}$$

and the probability generating function $P_n(z)$ of the buffer occupancy in the 'corresponding' buffer system with uncorrelated output process reads

$$P_n(z) = \frac{(2 - \alpha - \beta)(z - 1)A(z) f_{n-1}(z)}{(2 - \alpha - \beta) z^M - A(z)[1 - \beta + (1 - \alpha) z^{M-N}]}.$$ (19)

5. An application

Consider a buffer system with two output channels and an infinite waiting room. One of the output channels is permanently available for the transmission of messages from the buffer. The other output
channel, however, is shared by the buffer and some other device (e.g., another buffer or a speech signal) and is only available part of the time. Let us define \( A \)-times and \( B \)-times as the periods of time during which the second output channel is available or blocked respectively for the transmission of data from the buffer under consideration. The \( A \)-times and \( B \)-times are considered as two mutually independent sequences of i.i.d. random variables which take only positive integer values (both types of time intervals are expressed in time slots). If \( \sigma \) denotes the fraction of time during which the second output line is available, then

\[
\sigma = \frac{E[\text{A-time}]}{E[\text{A-time}]+E[\text{B-time}]}.
\]

(20)

Clearly if \( \sigma \) is given, the ratio of \( E[\text{A-time}] \) and \( E[\text{B-time}] \) is known, but not the actual values of \( E[\text{A-time}] \) and \( E[\text{B-time}] \), i.e., the same value is found for \( \sigma \) for all the systems which give rise to the same ratio of \( E[\text{A-time}] \) and \( E[\text{B-time}] \).

If only \( \sigma \) is known, then, since \( \sigma \) denotes the (long-run) probability of having two output channels available, the best one can do is to assume that \( \sigma \) is also the instantaneous probability of having two output channels available in each individual time slot, no matter what the number of available output channels was during previous time slots. This assumption is equivalent to assuming that both \( A \)-times and \( B \)-times are geometrically distributed random variables with parameters \( \sigma \) and \( 1-\sigma \) respectively. However, this also implies that \( E[\text{A-time}] = 1/(1-\sigma) \) and \( E[\text{B-time}] = 1/\sigma \) which may be far from realistic.

An improvement of the one parameter (\( \sigma \)) model would be the following: let \( A \)-times and \( B \)-times be geometrically distributed random variables with two independent parameters \( \alpha \) and \( \beta \) which are chosen such that

\[
E[\text{A-time}] = 1/(1-\alpha), \quad E[\text{B-time}] = 1/(1-\beta).
\]

In this case the probabilities of having one or two output channels available during a time slot are dependent on the number of available output channels in the immediately preceding time slot, and hence the model presented in this paper is applicable, provided we allow only two values (1 and 2) for the number of available output channels in a time slot and we choose the transition probabilities as follows:

\[
\begin{align*}
    r_1(1) &= \beta, & r_1(2) &= 1-\beta, \\
    r_2(1) &= 1-\alpha, & r_2(2) &= \alpha.
\end{align*}
\]

Clearly, the results derived in the third paragraph of the previous section apply here, for \( N = 1 \) and \( M = 2 \). Equation (18) gives the probability generating function of the buffer occupancy which corresponds to the improved two-parameter model, whereas (19) gives the corresponding result for the simplified one parameter model (notice that (19) depends on \( \alpha \) and \( \beta \) only through the combination \( (1-\beta)/(2-\alpha-\beta) = \sigma \)).

In the case \( N = 1, M = 2 \), (18) and (19) reduce to

\[
P(z) = \frac{(z-1)A(z)[(z+(1-\alpha-\beta)A(z))(c_1+c_2z)]}{z^3-(\alpha z + \beta z^2)A(z)-(1-\alpha - \beta)A^2(z)},
\]

(21)

\[
P_a(z) = \frac{(z-1)A(z)(c_4+c_5z)}{z^2-A(z)[\sigma+(1-\sigma)z]}
\]

(22)

where the \( c_i \)'s are unknown parameters.

Notice that (22) can be found from (21) by putting \( \alpha = \sigma \) and \( \beta = (1-\sigma) \); we will thus concentrate on (21).

As an example let us assume that the arrivals are governed by a geometric arrival process, i.e.,

\[
    a(n) = \frac{1}{1+\lambda} \left( \frac{\lambda}{1+\lambda} \right)^n, \quad n = 0, 1, 2, \ldots, \quad \text{and} \quad A(z) = \frac{1}{1+\lambda(1-z)}.
\]
Here \( \lambda \) denotes the average number of arrivals per time slot.

In this case (21) can be rewritten as

\[
P(z) = \frac{(z - 1) \left( (c_1 + c_2 z) \left( z(1 + \lambda (1 - z)) + 1 - \alpha - \beta \right) + c_3 \left( z^2 (1 + \lambda (1 - z)) + 1 - \alpha - \beta \right) \right)}{z^3 (1 + \lambda (1 - z))^2 - (\alpha z + \beta z^2)(1 + \lambda (1 - z)) - (1 - \alpha - \beta)}.
\]

The numerator \( n(z) \) and the denominator \( d(z) \) of this expression are both polynomials in the variable \( z \) of degrees 4 and 5 respectively. Furthermore, we know that \( d(z) \) has exactly 3 zeros inside the unit disk \( |z| < 1 \) of the complex plane, one of which equals unity; let \( u \) and \( v \) denote the two other zeros of \( d(z) \) inside the unit disk. Since \( P(z) \) is an analytic function of \( z \) inside the unit disk, the zeros 1, \( u \) and \( v \) of \( d(z) \) must be zeros of \( n(z) \) as well.

It follows that, in this special case, \( n(z) \) and \( d(z) \) can be expressed as

\[
n(z) = (z - 1)(z - u)(z - v)(N_0 + N_1 z)
\]

and

\[
d(z) = (z - 1)(z - u)(z - v)(D_0 + D_1 z + D_2 z^2)
\]

so that the following expression for \( P(z) \) results:

\[
P(z) = \frac{N_0 + N_1 z}{D_0 + D_1 z + D_2 z^2}.
\]  

(23)

Applying the normalisation equation \( P(1) = 1 \) to (23) a relationship between \( N_0 \) and \( N_1 \) can be found, so that \( P(z) \) finally can be written as

\[
P(z) = \frac{D_0 + D_1 + D_2 + N_1 (z - 1)}{D_0 + D_1 z + D_2 z^2}.
\]  

(24)

it is easily shown that the constants \( D_0, \ D_1, \ D_2 \) and \( N_1 \) are given by

\[
D_0 = 1 + \beta \lambda - \lambda (\lambda + 2)(u + v) + \lambda^2 (u^2 + uv + v^2), \quad D_2 = \lambda^2,
\]

\[
D_1 = \lambda^2 (u + v) - \lambda (\lambda + 2), \quad N_1 = -\lambda (c_2 + c_3).
\]

Thus, \( P(z) \) is completely specified as a function of the zeros \( u \) and \( v \) of \( d(z) \) inside the unit disk, if we can determine the unknown constants \( c_2 \) and \( c_3 \). These can be found in the way we described in the previous section, i.e., by expressing that \( u \) and \( v \) are zeros of \( n(z) \) and by using the normalising equation \( P(1) = 1 \), or equivalently, \( d(1) = n(1) \). As a result, \( N_1 \) turns out to be

\[
N_1 = \frac{\lambda (\lambda - 1 - \sigma) \left[ 2 + \alpha + \beta + \lambda uv - \lambda (u + v) \right]}{1 + \lambda + 2(\alpha + \beta) + [1 + 2 \lambda - \lambda (u + v)] (u^2 + uv + v^2) - (1 + 2 \lambda) (u + v)},
\]

where \( \sigma \) denotes the quantity defined in (20).

The above makes clear that it suffices to determine the two zeros \( u \) and \( v \) in order to completely specify the probability generating function \( P(z) \) of the buffer occupancy in stochastic equilibrium. From this function important performance measures of the buffer, such as mean value and variance of the buffer occupancy can be derived, by using the moment generating property of \( P(z) \). For instance, the mean buffer occupancy \( \bar{N} \) is given by

\[
\bar{N} = \left. \frac{dP(z)}{dz} \right|_{z=1} = \frac{N_1 - (D_1 + 2D_2)}{D_0 + D_1 + D_2}.
\]  

(25)

Using a digital computer the values of \( u \) and \( v \), and hence of \( \bar{N} \), can be determined in a numerical way for
each set of values for $\alpha$, $\beta$ and $\lambda$. In Fig. 4 plots of $\bar{N}$ versus $\lambda$ are given for different choices of $\alpha$ and $\beta$. All the curves, however, correspond to the same constant value of 0.5 for $\sigma = (1 - \beta)/(2 - \alpha - \beta)$. Curve (a) corresponds to the set of values $\alpha = 0.5$ and $\beta = 0.5$, i.e., $\alpha = \sigma$ and $\beta = 1 - \sigma$ and thus represents the case where the output process is uncorrelated. Curves (b), (c), (d) and (e) are for $\alpha = \beta = 0.8, 0.9, 0.95, 0.99$ respectively, i.e., for $E[A\text{-time}] = E[B\text{-time}] = 5, 10, 20, 100$ respectively. The plots make clear that the actual values of $E[A\text{-time}]$ and $E[B\text{-time}]$ have a considerable influence on the mean queue length, and that the use of an uncorrelated output process (one parameter) instead of a correlated output process (two parameters) may lead to serious errors in the determination of the actual (finite) length of the buffer.

Remark. Other types of arrival distributions than the one chosen here, can be treated in a similar manner. In particular, the numerical investigations are conceptually identical to the ones presented here (for a geometric arrival process), for all arrival processes having a generating function $A(z)$ which is a rational function of $z$, i.e., for all types of Coxian arrival processes. The reason for this is the fact that in all these cases the generating function $P(z)$ can be written as the ratio of two polynomials in $z$.

6. Conclusion

A mathematical model for discrete-time buffer systems with correlated output process was developed. A general arrival process was assumed and the output process was described as a first order Markov chain. A method of analysis of this model was presented. Several prior models for this kind of buffer system were found as special cases of the model described here. A practical example was worked out in detail, which showed that neglecting the correlation of the output process, i.e., treating the output process as uncorrelated, as in [15], may lead to results which are considerably different from the ‘actual’ results, obtained with a model which does account for the correlation in the output process. It is this observation which makes the analysis worthwhile.

References

