On Discrete Buffers in a Two-State Environment

HERWIG BRUNEEL

Abstract—A discrete buffered system with infinite buffer size, one single output channel, and periodic opportunities for service (synchronous transmission) is considered in a two-state environment. The output channel is subjected to a random interruption process, which is characterized by a Bernoulli sequence of independent random variables, with probabilities dependent on the environment state. The environment states have random sojourn times with "mixture of geometrics"-type distributions. The arrival process is dependent on the environment state, but arbitrary.

For this system, expressions are derived for the probability generating functions of the number of messages in the buffer at various time instants. A number of special cases and possible applications of the model are discussed, and an extended example is given as an illustration of the study.

I. INTRODUCTION

DISCRETE-time buffer systems with synchronous transmission have been investigated quite extensively in the past several years. Much work has been done on the case where the output lines of the buffer are available for all times, e.g., [1], [2]; however, lately, many researchers have concerned themselves with buffers where the output lines are subjected to random interruptions in time or where messages have to be retransmitted a stochastic number of times, due to errors, e.g., [3]–[15]. Both finite [1], [2], [6]–[9] and infinite [3]–[5], [10]–[15] waiting rooms have been treated. A Poisson arrival process has been considered in [3]–[6], a hybrid input process (Poisson and compound Poisson) has been treated in [7]–[9], whereas the arrival process is arbitrary in [10]–[15].

However, most of these analyses have one feature in common: the numbers of arrivals during consecutive clock time intervals are i.i.d. random variables, i.e., the arrival process is uncorrelated. Only a few researchers have used a different kind of arrival process. Bruneel [14] has treated a buffer system in which the arrival stream is subjected to random interruptions in time, i.e., where "available periods" (during which i.i.d. groups arrive in the consecutive clock time intervals) and "blocked periods" (during which no arrivals occur) of stochastic length occur alternately. The analysis remains, however, restricted to geometrically distributed available periods and arbitrarily distributed but finite blocked periods. Towsley [13] has considered a statistical multiplexer in a two-state environment, with different arrival processes for the two environment states. Unfortunately, his analysis is only valid for Markovian environments, i.e., for geometrically distributed sojourn times for the two environment states, except for a few particular choices for the parameters of the system.

The present analysis is the result of an attempt to generalize Towsley's results to more general distributions for the sojourn times of the environment states. We consider environment state sojourn times with probability density functions that can be written as mixtures of finite numbers of geometric densities. Although such mixtures do not yield the full generality that we would want, they can be used as approximations for a large class of discrete probability distributions whose coefficient of variation is larger than for a simple geometric distribution. The extension from a geometric density to a mixture of geometrics can be considered as the discrete-time equivalent of the generalization from an exponential to a hyperexponential distribution in continuous time, a technique which has been applied successfully to the interarrival time and service time distributions in continuous-time queuing systems (see, e.g., [16]).

The details of the model are explained in Section II. Some preliminary terminology is introduced in Section III, whereas the actual analysis is carried through in Section IV. Sections V and VI contain a discussion of the model and its analysis, along with a number of special cases and practical applications of the study. Finally, an example is worked out in detail in Section VII.

II. THE SYSTEM UNDER CONSIDERATION

A diagram of the investigated system is depicted in Fig. 1. Messages arrive in a stochastic manner via one of the input channels, wait in the buffer for some time, and are then taken out via the output channel. We assume that the messages have a fixed length and that the output channel transmits data with constant speed. Hence, the transmission time of the messages is constant. The clock time period is defined as the time required to transmit one message. The data transmission is synchronous, i.e., the data are taken out synchronously from the buffer for transmission at each discrete clock time. However, this can only happen if the switch S is closed, i.e., if the output channel is available. Whenever the switch S is open, no transmission can take place.

The "environment" of the buffer is described as follows: there are two possible environment states, which will be indicated by A and B. As time goes by, the environment state is A for a number of clock time periods, then B for a number of clock time periods, then A again, and so on. The time intervals (expressed in clock time periods) during which the environment state is A or B will be called A-times and B-times, respectively. The A-times and B-times occur alternately and take positive integer values only. We assume that the A-times are i.i.d. random variables with probability density function \( p_A(n) \), that the B-times are i.i.d. random variables with probability density function \( p_B(n) \), and that the A-times and B-times are mutually independent. Throughout this paper it will be assumed that both \( p_A(n) \) and \( p_B(n) \) can be written as mixtures of a finite number of geometric density functions:

\[
p_A(n) = \sum_{i=1}^{r_A} a_i (1 - \alpha_i) (\alpha_i)^{n-1}
\]

\[
p_B(n) = \sum_{i=1}^{r_B} b_i (1 - \beta_i) (\beta_i)^{n-1}
\]
for $n = 1, 2, \cdots$, where

\[ 0 \leq \alpha_i < 1 \quad \text{for } i = 1, 2, \cdots, r_A \]
\[ 0 \leq \beta_i < 1 \quad \text{for } i = 1, 2, \cdots, r_B \]

and

\[ \sum_{i=1}^{r_A} a_i = \sum_{i=1}^{r_B} b_i = 1. \quad (3) \]

The interruptions of the output channel are described by a Bernoulli sequence of i.i.d. random variables, i.e., it is assumed that the probability of having an available or blocked output channel during an arbitrary clock time interval is constant and independent of the availability of the output channel during previous clock time intervals. More specifically, we assume

\[ \text{Prob} \{ \text{output channel available during a clock time interval where the environment state is } A \} = \sigma_A \]

and

\[ \text{Prob} \{ \text{output channel available during a clock time interval where the environment state is } B \} = \sigma_B. \]

Furthermore, the arrival process is assumed to be of the following type: the number of arrivals to the buffer (expressed in messages) during an arbitrary clock time interval is a random variable with an arbitrary probability distribution, which is only dependent on the environment state, and not on the number of arrivals during previous clock time intervals. The following notations will be used:

\[ e_A(n) = \text{Prob} \{ n \text{ arrivals during a clock time interval where the environment state is } A \} \]

and

\[ e_B(n) = \text{Prob} \{ n \text{ arrivals during a clock time interval where the environment state is } B \} \]

for $n = 0, 1, 2, \cdots$. The probability generating functions corresponding to the densities $e_A(n)$ and $e_B(n)$ will be denoted by $E_A(z)$ and $E_B(z)$, respectively.

Finally, we make the assumption that a message cannot leave the buffer at the end of the clock time interval during which it entered the buffer. Therefore, when the buffer is empty at the beginning of a clock time interval, no message can leave the buffer at the end of this particular clock time interval, even if there have been some arrivals during the interval.

### III. Definitions and Terminology

Let us consider the buffer system a long time after the first arrivals, when a stochastic equilibrium has been reached. Let $c_0$ and $d_0$ denote the random variables which indicate the equilibrium number of messages in the buffer at the beginning of an $A$-time or a $B$-time, respectively, and $C(z)$ and $D(z)$ the corresponding probability generating functions.

Let $c_1, c_2, \cdots, c_k, \cdots$ be the random variables which equal the equilibrium number of messages in the buffer after the first, second, $\cdots$, $k$th, $\cdots$ clock time interval of an $A$-time, and $d_1, d_2, \cdots, d_k, \cdots$ the random variables which equal the equilibrium number of messages after the first, second, $\cdots$, $k$th, $\cdots$ clock time interval of a $B$-time. The corresponding probability generating functions are denoted by $C_i(z)$, $C_i(z)$, $\cdots$, $C_i(z)$, $\cdots$, and $D_i(z)$, $D_i(z)$, $\cdots$, $D_i(z)$, $\cdots$, respectively.

Fig. 2 illustrates these definitions.

Furthermore, we define for each value of $x$ in the interval $[0, 1)$ the functions $f_A(x, z)$ and $f_B(x, z)$ by

\[ f_A(x, z) = \sum_{k=1}^{\infty} (1-x)x^{k-1}C_k(z) \quad (4) \]

and

\[ f_B(x, z) = \sum_{k=1}^{\infty} (1-x)x^{k-1}D_k(z). \quad (5) \]

Finally, we define the functions $N_A(z)$, $N_B(z)$, and $N(z)$ as the generating functions of the equilibrium number of messages in the buffer, just after a random clock time interval belonging to an $A$-time, just after a random clock time interval belonging to a $B$-time, and at a random clock time, respectively.

### IV. Analysis of Buffer Behavior

#### A. Derivation of $C_k(z)$ and $D_k(z)$

First, let us concentrate on $d_0$ and $D_0(z)$. From its definition, it follows that $d_0$ can be viewed as the buffer occupancy after the last clock time interval of an $A$-time. Thus, using the notations of Section III, $d_0 = c_k$, if the $A$-time has length $k$, an event which occurs with probability $p_A(k)$. It follows that

\[ D_0(z) = \sum_{k=1}^{\infty} p_A(k)C_k(z), \]

which, upon substitution of (1), can be written as

\[ D_0(z) = \sum_{k=1}^{r_A} \sum_{i=1}^{n} (1-\alpha_i)(\alpha_i)^{k-1}C_i(z) \]

\[ = \sum_{i=1}^{r_A} a_i f_A(\alpha_i, z) \quad (6) \]

where $f_A(\alpha_i, z)$ was defined in (4).

An expression for the function $f_A(x, \alpha_i)$ can be derived as follows. Consider the evolution of the buffer occupancy during $A$-times; it is described by the equations ($k > 0$)

\[ c_k = \begin{cases} c_{k-1} + e_{A,k} - \xi_k & \text{if } c_{k-1} > 0 \\ e_{A,k} & \text{if } c_{k-1} = 0. \end{cases} \]

Here $e_{A,k}$ is the number of arrivals during the $k$th clock time interval of an $A$-time interval which has generating function $E_A(z)$, and $\xi_k$ is the number of available output channels during this clock time interval (which equals 1 or 0 with probabilities $\sigma_A$ or $1 - \sigma_A$, respectively). By introducing generating functions, the equations (7) can be transformed into a relationship.
between \( C_k(z) \), \( C_{k-1}(z) \), and \( C_{k-1}(0) \):

\[
C_k(z) = \frac{E_A(z)}{z} \left\{ \left[ (1 - \sigma_A)z + \sigma_A \right] C_{k-1}(z) + \sigma_A (z - 1) C_{k-1}(0) \right\}
\]

for all \( k > 0 \). By multiplying this equation by a factor \( (1 - x^k) \) and summing the result over all values of \( k > 0 \), a relationship between \( f_A(x, z) \) and \( f_A(x, 0) \) can be obtained, which upon solution for \( f_A(x, z) \) yields

\[
f_A(x, z) = \frac{E_A(z) \left\{ \left[ (1 - \sigma_A)z + \sigma_A \right] C_0(z) + \sigma_A (z - 1) \left[ (1 - x)C_0(0) + x f_A(x, 0) \right] \right\}}{z - x E_A(z) \left[ (1 - \sigma_A)z + \sigma_A \right]}.
\]

If we define

\[
E_A^*(z) = \left[ (1 - \sigma_A)z + \sigma_A \right] E_A(z)
\]

then the expression for \( f_A(x, z) \) can be rewritten as

\[
f_A(x, z) = \frac{(1 - x) E_A^*(z) C_0(z) + m_A(x)(z - 1) E_A(z)}{z - x E_A^*(z)}.
\]

An expression for \( D_0(z) \) can now be derived by combining (6) and (11); this leads to

\[
D_0(z) = E_A^*(z) C_0(z) \sum_{i=1}^{r_A} \frac{(1 - \alpha_i) a_i}{z - \alpha_i E_A^*(z)} + (z - 1) E_A(z) \sum_{i=1}^{r_A} \frac{m_A(\alpha_i) a_i}{z - \alpha_i E_A^*(z)}.
\]

If we introduce the new notations

\[
g_A(z) \triangleq \prod_{i=1}^{r_A} [z - \alpha_i E_A^*(z)]
\]

\[
g_{A,i}(z) \triangleq \frac{g_A(z)}{z - \alpha_i E_A^*(z)}
\]

\[
h_A(z) \triangleq \sum_{i=1}^{r_A} (1 - \alpha_i) a_i g_{A,i}(z)
\]

(12) leads to the following relationship between \( C_0(z) \) and \( D_0(z) \):

\[
g_A(z) D_0(z) = h_A(z) E_A^*(z) C_0(z) + (z - 1) E_A(z) \sum_{i=1}^{r_A} m_A(\alpha_i) a_i g_{A,i}(z).
\]

A second relationship between \( C_0(z) \) and \( D_0(z) \) can be derived in a similar manner, by considering the evolution of the buffer occupancy during \( B \)-times instead of during \( A \)-times; this relationship reads

\[
g_B(z) C_0(z) = h_B(z) E_B^*(z) D_0(z) + (z - 1) E_B(z) \sum_{i=1}^{r_B} m_B(\beta_i) b_i g_{B,i}(z)
\]

where the new notations are self-explanatory.

The relationships (16) and (17) provide us with two linear equations in \( C_0(z) \) and \( D_0(z) \), which can thus be solved for; this yields

\[
C_0(z) = (z - 1) \frac{E_A(z) E_B^*(z) h_B(z) \sum_{i=1}^{r_A} m_A(\alpha_i) a_i g_{A,i}(z) + g_A(z) E_B(z) \sum_{i=1}^{r_B} m_B(\beta_i) b_i g_{B,i}(z)}{g_A(z) g_B(z) - E_A(z) E_B^*(z) h_A(z) h_B(z)}
\]

\[
D_0(z) = (z - 1) \frac{E_B(z) E_A^*(z) h_A(z) \sum_{i=1}^{r_B} m_B(\beta_i) b_i g_{B,i}(z) + g_B(z) E_A(z) \sum_{i=1}^{r_A} m_A(\alpha_i) a_i g_{A,i}(z)}{g_A(z) g_B(z) - E_A(z) E_B^*(z) h_A(z) h_B(z)}
\]

Equations (18) and (19) contain expressions for \( C_0(z) \) and \( D_0(z) \) in terms of known quantities on one hand, and the \( r_A + r_B \) unknown quantities \( m_A(\alpha_i) (i = 1, 2, \ldots, r_A) \) and \( m_B(\beta_i) (i = 1, 2, \ldots, r_B) \) on the other hand. These unknown parameters can be determined as follows. Using Rouché’s theorem (see, e.g., [16]) as explained in the Appendix of [10], one can show that, whenever the condition for a stochastic equilibrium is met, i.e., whenever the average number of messages that arrive during one clock time interval is lower than the average number of messages that can possibly leave the buffer per clock time interval, the denominator of \( C_0(z) \) and \( D_0(z) \) has exactly \( r_A + r_B \) zeros inside the unit disk \( \{ z | |z| \leq 1 \} \) of the \( z \)-plane, one of which equals unity. These must be zeros of the numerator of \( C_0(z) \) or \( D_0(z) \) as well, since generating functions are analytic inside the unit disk of the complex plane. The resulting \( r_A + r_B - 1 \) linear equations in the unknown parameters (no equation is obtained for the zero \( \rho_A = 1 \), together with the normalization condition \( C_0(1) = 1 \) or \( D_0(1) = 1 \), yield the desired values of the \( m_A(\alpha_i) \)’s and the \( m_B(\beta_i) \)’s. Using these, expressions for \( C_0(z) \) and \( D_0(z) \) in terms of known quantities only can be derived.

B. Derivation of \( N_A(z), N_B(z), \) and \( N(z) \)

Let \( t \) indicate the moment just after a randomly chosen clock time interval. The probability that this clock time interval belongs to an \( A \)-time or \( B \)-time is given by \( \rho_A \) or \( \rho_B \), respectively, where

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

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

Let us first consider the case where the clock time interval under consideration lies in an \( A \)-time, and let \( L_A \) denote the
number of clock time intervals in this $A$-time before $t$ (Fig. 3). Then using the definitions of Section III, the number of messages in the buffer at time $t$ is given by $c_{L_A}$, and its generating function is $N_A(z)$. It follows that

$$N_A(z) = \sum_{j=1}^{\infty} \text{Prob } [L_A = j] C_j(z).$$

It can be shown (see, e.g., [15]) that, since the moment $t$ was chosen arbitrarily,

$$\text{Prob } [L_A = j] = \sum_{k=0}^{\infty} \frac{\rho_A(k)}{E[A\text{-time}]}.$$

Since the distribution of the $A$-times is a mixture of geometrics [see (11)], this can be expressed as

$$\text{Prob } [L_A = j] = \sum_{i=1}^{r_A} a_i \cdot \frac{1 - \alpha_i}{(1 - \alpha_i)E[A\text{-time}]}.$$

where the quantities $a_i$ are defined as

$$a_i = \frac{a_i}{(1 - \alpha_i)E[A\text{-time}]}.$$

This means that the density function of $L_A$ can be written as a mixture of the same set of geometric densities used in the expression of $p_A(t)$. It follows that $N_A(z)$, just as $D_B(z)$, earlier, can be written as a weighed sum of the $f_A(\alpha_i, z)$'s:

$$N_A(z) = \sum_{i=1}^{r_A} a_i f_A(\alpha_i, z).$$

Notice that at this point of the analysis, the functions $f_A(\alpha_i, z)$ are known functions of $z$; indeed, from (11) it follows that $f_A(\alpha_i, z)$ is given by

$$f_A(\alpha_i, z) = \frac{(1 - \alpha_i)E_A^\alpha(z)C_0(z) + m_A(\alpha_i)(z - 1)E_A(z)}{z - \alpha_i E_A(z)}$$

which contains known quantities only.

Next, consider the case where the clock time interval immediately before $t$ belongs to a $B$-time. The generating function of the buffer occupancy at such a time $t$ is $N_B(z)$, given by

$$N_B(z) = \sum_{i=1}^{r_B} b_i f_B(\beta_i, z)$$

where the parameters $b_i$ and the functions $f_B(\beta_i, z)$ are given by similar expressions as $a_i$ and $f_A(\alpha_i, z)$ earlier.

Finally, the generating function $N(z)$ of the buffer occupancy at a random clock time can be found from the equation

$$N(z) = \rho_A N_A(z) + \rho_B N_B(z).$$

This concludes the analysis of the buffer system.

V. DISCUSSION

The model we developed and analyzed in this paper is a direct extension of the model presented by Towsley [13] in his work on statistical multiplexers in a two-state Markovian environment. The most important differences between our study and Towsley’s are as follows.

1) The sojourn times of the two environment states (the $A$-times and $B$-times in our model) are considered as geometrically distributed random variables in Towsley’s study, whereas their distributions are allowed to be mixtures of arbitrary numbers ($r_A$ and $r_B$) of geometrics in our study, which is much more general.

2) The method of analysis presented in the previous section is different from Towsley’s, in that it departs explicitly from the distributions of the $A$-times and $B$-times, whereas these distributions are only implicitly present in Towsley’s method: their geometric nature is represented by the Markovian nature of the environment. The general framework of our method could therefore, in principle, be used for any type of distribution for the $A$-times and $B$-times, whereas Towsley’s method is an ad hoc solution restricted to geometric distributions.

3) Our study not only yields the generating function of the buffer occupancy at random clock times ($N(z)$); it also leads to results at the beginning of an $A$-time ($C_0(z)$) or a $B$-time ($D_B(z)$), which may be important in certain applications, as will become clear further.

Notice that the case of a Markovian environment can be obtained as a special case of our study; it corresponds to the following choice of parameters:

$$r_A = r_B = 1$$

$$a_i = b_i = 1$$

$$\alpha_i = \alpha$$

$$\beta_i = \beta.$$ 

Here $\alpha$ and $\beta$ denote the parameters of the geometric distributions of the $A$-times and $B$-times.

In this special case our results reduce to

$$C_0(z) = (z - 1) - \frac{m_A(\alpha)(1 - \beta)E_A(z)E_A^\alpha(z) + m_B(\beta)E_B(z)(z - \alpha E_A^\alpha(z))}{[z - \alpha E_A(z)](z - \beta E_B^\beta(z)) - (1 - \alpha)(1 - \beta)E_A^\alpha(z)E_B^\beta(z)}$$

$$D_B(z) = (z - 1) - \frac{m_A(\alpha)E_A(z)(z - \beta E_B^\beta(z)) + m_B(\beta)(1 - \alpha)E_B(z)E_A^\alpha(z)}{[z - \alpha E_A(z)](z - \beta E_B^\beta(z)) - (1 - \alpha)(1 - \beta)E_A^\alpha(z)E_B^\beta(z)}$$

$$N_A(z) = D_B(z)$$

$$N_B(z) = C_0(z)$$

$$N(z) = (z - 1) - \frac{m_A(\alpha) \rho_A E_A(z)(z + (1 - \alpha - \beta)E_B(z)) + m_B(\beta) \rho_B E_B(z)(z + (1 - \alpha - \beta)E_A(z))}{[z - \alpha E_A^\alpha(z)](z - \beta E_B^\beta(z)) - (1 - \alpha)(1 - \beta)E_A^\alpha(z)E_B^\beta(z)}.$$ 


The expression (29) for \( N(z) \) corresponds exactly to the result found by Towsley [13]; the expressions (27) and (28) for \( C_d(z) \) and \( D_0(z) \) are new results of the present study.

VI. APPLICATIONS OF THE MODEL

Discrete-time buffer systems of the type treated in this paper are encountered in a variety of computer and communication systems that involve the synchronous transmission of (some kind of) data units (messages, packets, digits, etc.). Examples are statistical multiplexers, computer terminals, loop systems, integrated voice-data systems, demultiplexing facilities, packet-switching networks, and many others. In all these applications a buffer system is used for the temporary storage of data units awaiting transmission via some kind of communication channel. This storage may be necessary for different reasons; the most important are:

1) an intermittent stream of data units must be transformed into a more regular data stream;

2) data units generated by many users are assembled prior to their transmission, so that the users can have access to the communication channel concurrently;

3) the communication channel is disabled periodically or is unreliable.

It is clear that the input and output characteristics of the buffer system depend on the specific application it is used in. The mathematical model presented in this paper can be used to describe a wide variety of buffer systems, depending on the particular choices we make for the quantities that describe the system, i.e., \( \sigma_A, \sigma_B, E_a(z), E_b(z), r_A, r_B, a_i, b_i (i = 1, 2, \ldots, r_A), b_i, \beta_i \). In this section we discuss some of these applications of our study.

A. Buffer System with Random Server Interruptions

Consider a buffer system where the output channel is interrupted at random points in time for a random time interval. This situation occurs for instance in voice-data systems [6]–[9], where data are transmitted only during the so-called silent periods of the speech signal, or in a communication system with two kinds of messages, where low-priority messages are transmitted at a constant rate.

The server interruptions correspond to the talkspurts of the speech signal in the first example, and to the servicing of high-priority messages in the second example. The behavior of the data buffer of a voice-data system or the low-priority buffer in a system with two kinds of messages can be analyzed by means of the model presented in this paper, provided the following choice of parameters is made:

\[
\begin{align*}
\sigma_A &= 1 \quad \text{(available periods of the output channel)} \\
\sigma_B &= 0 \quad \text{(blocked periods of the output channel)} \\
E_A(z) &= E_B(z) = E(z)
\end{align*}
\]

if \( E(z) \) denotes the generating function of the number of arrivals per unit time interval.

The choice of the parameters \( r_A, r_B, a_i, a_j, b_i, \beta_i \) depends on the precise nature of the available and the blocked periods of the output channel of the buffer system, and should be made such that the resulting mixtures of geometrics for the \( A \)-times and \( B \)-times correspond as well as possible to the distributions of the available and blocked periods of the output line. For instance, in the case of an integrated voice-data system, where the available periods of the output line (of the data buffer) correspond to the silent periods in the speech signal, the distribution of the periods of pauses in the speech signal (see, e.g., [6], [15], [17]) can be modeled by a geometrically distributed random variable with different mean values, so that the distribution of the silent periods is a mixture of geometrics.

The talkspurts of the speech signal, i.e., the blocked periods of the output line of the data buffer, can be conveniently described by means of a (single) geometric distribution [17].

It is clear that in this application our model gives a more realistic description of the silent periods than Towsley’s model, which would characterize various types of pauses in the speech signal by means of one single geometric density. Furthermore, our model has the additional advantage that it yields an expression for the generating function \( E(z) \) of the data buffer occupancy just after a talkspurt, when the data buffer typically contains most messages. Towsley’s study would only give an expression for \( N(z) \), which describes the buffer occupancy at random clock times.

B. Buffer System with Two Kinds of Input Traffic

In many communication systems, the arrival stream of data into the system is not stationary, i.e., the characteristics of the arrival process can vary in time. This situation occurs, for instance, in a statistical multiplexer whose arrival stream is built up of messages sent by a varying number of users, or in a node of a communication network, where messages from a varying number of network nodes are received. A varying arrival process also occurs in any buffer system where some of the input channels are subjected to random interruptions (due to failures) or where the traffic carried by (some of) the input channels consists of alternate routed traffic or overflow traffic from some other service system (see, e.g., [18], [19]).

If there are essentially two kinds of input traffic, the study presented in this paper can be used to determine the behavior of buffer systems with this kind of arrival process. It suffices to choose the generating functions \( E_a(z) \) and \( E_b(z) \) of the model such that they correspond to the two possible types of arrival stream. The parameters \( \sigma_A \) and \( \sigma_B \) can be used to model an unreliable output channel; choosing \( \sigma_A \neq \sigma_B \) allows the introduction of correlation between the reliability of the output channel and the type of input traffic. The distributions of the \( A \)-times and \( B \)-times can be adapted somewhat to the distributions of the durations of the two kinds of input traffic, by an appropriate choice of the parameters \( r_A, r_B, a_i, a_j, b_i, \beta_i \).

A special case of this kind of buffer system, namely, a buffer system with random arrival interruptions, will be treated in detail in Section VII.

VII. EXAMPLE: BUFFER WITH ARRIVAL INTERRUPTIONS

Consider a buffer system whose arrival stream is subject to random interruptions in time. Such a system can be analyzed by means of the study presented in this paper, provided we choose:

\[
\begin{align*}
E_A(z) &= 1 \quad \text{(no arrivals during \textit{A}-times)} \\
E_B(z) &= E(z) \quad \text{(arrivals characterized by generating function \( E(z) \) during \textit{B}-times)}
\end{align*}
\]

In order to keep the complexity of the computations within reasonable limits, we will assume that the distribution of the \( A \)-times is a mixture of two geometrics:

\[
\rho_A(n) = a(1 - a_1)(a_1)^{n-1} + (1 - a)(1 - a_2)(a_2)^{n-1}
\]

and that the \( B \)-times have a geometric distribution,

\[
\rho_B(n) = (1 - \beta) \beta^{n-1}.
\]

In terms of our general assumptions, introduced in Section II, this means: \( r_A = 2, r_B = 1, a_i = a, a_2 = 1 - a, b_1 = 1, \beta_i = \beta \).

Finally, assume that the arrivals during the \( B \)-times are governed by a geometric arrival process, i.e.,

\[
E(z) = \frac{1}{1 + \lambda - \lambda z}.
\]
where \( \lambda \) denotes the mean number of arrivals per clock time interval.

It is clear that in this buffer system, the buffer occupancy will take its highest values just at the periods when the arrival stream was available, i.e., just after the \( B \)-times, or equivalently, at the beginning of the \( A \)-times. Since the buffer occupancy at such time instants is described by the generating function \( C_d(z) \), we will now concentrate on \( C_0(z) \). A general formula for \( C_0(z) \) was derived in Section IV [see (18)]; for the special case considered here, this formula reduces to

\[
C_0(z) = \frac{(z-1)[m_1(z-\alpha_2 A(z)) + m_2(z-\alpha_1 A(z))B(z) + m_3(z-\alpha_1 A(z))(z-\alpha_2 A(z))]}{[z-\alpha_1 A(z)][z-\alpha_2 A(z)][zL(z)-\beta B(z)]-(1-\beta)A(z)B(z)h_A(z)}.
\]

(30)

Here the functions \( A(z) \), \( B(z) \), \( L(z) \), and \( h_A(z) \) are defined as

\[
A(z) = \sigma_4 + (1-\sigma_A)z
\]

\[
B(z) = \sigma_B + (1-\sigma_B)z
\]

\[
L(z) = 1 + \lambda - \lambda z
\]

\[
h_A(z) = a(1-\alpha_1)(z-\alpha_2 A(z)) + (1-a)(1-\alpha_2)(z-\alpha_1 A(z))
\]

and the unknown parameters \( m_1, m_2, \) and \( m_3 \) replace the original unknowns \( m_4(\alpha_1), m_4(\alpha_2) \) and \( m_B(\beta) \).

From (30) it follows that \( C_0(z) \) is a rational function of \( z \): the numerator is a polynomial of degree 3, whereas the denominator is a polynomial of degree 4. From Section IV we know that the denominator of (30) has exactly three zeros inside the unit disk of the complex plane, which must be zeros of the numerator of (30) as well, in order for \( C_0(z) \) to be an analytic function of \( z \) inside the unit disk. It thus follows that \( C_0(z) \) can be expressed as

\[
C_0(z) = \frac{1-z_0}{z-z_0}
\]

(31)

where \( z_0 \) denotes the only zero outside the unit disk of the polynomial

\[
[z-\alpha_1 A(z)][z-\alpha_2 A(z)][zL(z)-\beta B(z)]+(1-\beta)A(z)B(z)h_A(z)
\]

(32)

and we have used the normalization equation \( C_0(1) = 1 \) for the complete determination of \( C_0(z) \).

In general, the parameter \( z_0 \) can only be found in a numerical way by means of a digital computer, for each given set of system parameters \( (\alpha_1, \alpha_2, \beta, a, \sigma_A, \sigma_B, \lambda) \), e.g., by use of the Newton–Raphson scheme. However, once \( z_0 \) has been found, all the important performance measures of the buffer system can be derived from (31). For example, the mean buffer occupancy \( \bar{C} \) (at the beginning of an \( A \)-time) can be obtained from

\[
\bar{C} = \frac{dC_0(z)}{dz} \bigg|_{z=1} = \frac{1}{z_0-1}.
\]

(33)

A computer program was developed in order to obtain plots of the mean buffer occupancy \( \bar{C} \) versus the mean arrival intensity \( \lambda \), under various circumstances. Some results are given in Figs. 4 and 5.

In Fig. 4 we consider a buffer system with random arrival interruptions and an unreliable output channel, where the probability of correct transmission is independent of the arrival interruptions, i.e., \( \sigma_A = \sigma_B (= 0.8) \). The durations of the available periods of the arrival stream (the \( B \)-times in our model) are geometrically distributed with parameter \( \beta = 0.9 \), i.e., \( E[B \text{-time}] = 1/(1 - \beta) = 10 \). Two different cases are considered for the distribution of the arrival interruptions (the \( A \)-times in our model). First, it is assumed that the distribution of the \( A \)-times is a mixture of two geometrics with the following choice of parameters: \( \alpha_1 = 0.5, \alpha_2 = 0.99, a = 0.9 \). I.e., there are two kinds of arrival interruptions: 90 percent of the interruptions have mean duration \( 1/(1 - \alpha_1) = 2 \); the remaining 10 percent have mean duration \( 1/(1 - \alpha_2) = 100 \); the overall mean length of an \( A \)-time is then given by 11.8. Second, the \( A \)-times are modeled as geometrically distributed random variables, with the same mean value 11.8 as in the first case. The figure shows that for all possible values of \( \lambda \) the mean buffer occupancy \( \bar{C} \) is higher in the first case \( (C_1) \) than in the second case \( (C_2) \), which illustrates the importance of the precise distribution of the \( A \)-times: a simplified model which uses one single geometric distribution for the two kinds of arrival interruptions tends to underestimate the mean buffer occupancy. This observation alone justifies the generalization from a Markovian environment to a non-Markovian environment, carried through in this paper.

Fig. 5 illustrates the influence of correlation between the availability of the output channel and the availability of the arrival stream. More specifically, the figure contains three curves of the mean buffer occupancy \( \bar{C} \) versus the mean arrival intensity \( \lambda \), which correspond to the same arrival interruption process (the same distribution of \( A \)-times and \( B \)-times), but to different choices for the parameters \( \sigma_A \) and \( \sigma_B \), i.e., the availability of the output line. The arrival interruptions are characterized by the following choice of parameters: \( \alpha_1 = 0.98, \alpha_2 = 0.996, \beta = 0.99, a = 0.75 \), i.e., there are two kinds of \( A \)-times: 75 percent have mean value \( 1/(1 - \alpha_1) = 50 \), and the remaining 25 percent have mean value \( 1/(1 - \alpha_2) = 250 \), resulting in an overall mean length of 100 for the \( A \)-times; the mean length of the \( B \)-times is given by \( 1/(1 - \beta) = 100 \) as well. This means that the arrival stream is interrupted during 50 percent of the time. The availability of the output channel is different for the three curves: \( \sigma_A = 1, \sigma_B = 0.5 \) for curve \( C_1 \), \( \sigma_A = \sigma_B = 0.75 \) for curve \( C_2 \), and \( \sigma_A = 0.5, \sigma_B = 1 \) for curve \( C_3 \). Notice that the "mean availability" of the output channel is 75 percent in all three cases. The figure shows that the mean buffer occupancy \( \bar{C} \) is very dependent on the type of correlation between the availability of the output channel and the availability of the arrival stream. This correlation is negative in the case of curve \( C_1 \), zero in the case of curve \( C_2 \), and positive in the case of curve \( C_3 \). Our results show that a model in which correlation is allowed between input and output may lead to both higher or lower estimations of the mean buffer occupancy than a model without correlation: higher if the correlation is negative, lower if the correlation is positive.

VIII. CONCLUSIONS

We have developed and analyzed a mathematical model for a discrete-time queueing system with one single server, operating in a two-state environment. Arrivals to the system are described by a general but state-dependent arrival process. The server of the system is subject to random interruptions, which are modeled through a Bernoulli sequence of i.i.d. random variables with state-dependent density functions.

As opposed to Towsley’s study [13], the analysis is not restricted to a Markovian environment, i.e., an environment in which the two states have geometrically distributed sojourn times. In our model the sojourn times of the environment
states are considered as discrete stochastic variables with probability density functions that can be written as weighted sums of finite numbers of geometric densities, which is much more general.

As a result of the analysis we have found expressions for the probability generating function of the equilibrium number of messages in the system, at various time instants. These expressions depend on a finite number of unknown constants, the determination of which requires the solution of a possibly transcendental equation, which may be the most tedious part of the analysis.

A few applications of our model were indicated, and an example was worked out in detail.

---

Herwig Bruneel was born in Zottegem, Belgium, on September 7, 1954. He received the M.S. degree in electrical engineering, the degree of Licentiate in computer science, and the Ph.D. degree in computer science in 1978, 1979, and 1984, respectively, all from Ghent State University, Ghent, Belgium.

Since 1979 he has been with the Belgian National Fund for Scientific Research (FNWO) at the Department of Computer Science, Ghent State University. His main research interests include queueing theory and statistical modeling of computer communications systems.