ANALYSIS OF AN INFINITE BUFFER SYSTEM WITH RANDOM SERVER INTERRUPTIONS†

HERWIG BRUNEL†
Department of Computer Science, Ghent State University, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium

Scope and purpose—In this paper a discrete buffered system is considered where the output line of the buffer—which is assumed to be of infinite size—is interrupted at random instants for a random period. The reason to study this kind of system was the observation that most existing analyses for this system model the interruption process through a Bernoulli sequence of independent random variables or through a first order Markov process, whereas in practical applications of the system, such as an integrated voice-data system (see e.g. [3–6]), the interruption process is of a more complicated nature. The model we propose in this paper is, in this sense, more realistic. Comparison of our results with earlier results clearly shows the importance of a precise model for the interruption process.

Abstract—An infinite buffer with general arrival process, synchronous transmission, one single output channel and random server interruptions is considered.

As opposed to previous analyses the interruption process of the output line is kept rather general, i.e. the server is assumed to be in one of two states, “available” or “blocked”, where the sojourn time of the blocked state is arbitrarily distributed and the sojourn time of the available state has a density function which is a mixture of a finite number of geometric densities. For this general case the probability generating function of the buffer occupancy at various time instants is derived.

The results of the study are applied to the case where the server interruptions are due to the presence of speech at the input of the transmission channel of an integrated voice-data system. Some considerable deviations from earlier results are found.

1. INTRODUCTION

The behavior of both finite[1–5] and infinite[6–11] buffers in (computer) communication systems has been analyzed quite extensively in the past several years. Not only has the case been studied where the servers are available for all times, e.g. [1–2], but a considerable research effort has also been spent in analyzing buffers where the servers are subjected to random interruptions in time [3–12]. Sherman[6] and Hsu[7] have considered this problem for the case of asynchronous transmission of data from an infinite buffer and a uniform input rate. The case of synchronous transmission of data has also been treated by many authors both for finite[3–5] and infinite[8–12] buffer sizes. A Poisson arrival process is considered in [3, 8–10], a mixed arrival process (Poisson and compound Poisson) is treated in [4, 5] whereas the arrival process is arbitrary in [11, 12]. In most of these analyses the random interruptions of the servers are modeled through a Bernoulli sequence of independent random variables, i.e. it is assumed that the probabilities of the servers being available (σ) or blocked (1 − σ) during an arbitrary clock time interval are constant and independent of the state of the servers in the previous clock time intervals. In other words, the interruption process of the servers is completely specified by only one parameter σ which equals the fraction of time in which the servers are available. In other analyses, e.g. in [5] and [11], the server interruptions are studied through a first order Markov process. In this case the “available periods” and “blocked periods” of the servers (expressed in clock time periods) are geometrically distributed random variables with independent parameters. Thus, in this case the interruption process is completely specified by two parameters.

†Work supported by the Belgian National Fund for Scientific Research.
†Herwig Brunel is currently working towards a Ph.D. degree in Computer Science at the "Interfacultair Centrum voor Informatica", Ghent State University, Gent, Belgium. He received the M.S. degree in Electrical Engineering and the degree of licentiate in Computer Science in 1978 and 1979 respectively, both at Ghent State University, Gent, Belgium. His main research interests include queueing theory, computer communications and performance evaluation.

This work was supported by the Belgian National Fund for Scientific Research.
The present paper investigates a discrete buffered system with general arrival process, synchronous data transmission (periodic service opportunities), infinite buffer size and one single server (output channel) which is subjected to random interruptions in time.

The interruption process is kept quite general: the output channel is considered to have two possible states ("available" or "blocked"), the sojourn times of the blocked state may have an arbitrary (discrete) distribution, whereas the "available periods" are distributed according to a probability distribution function that is a mixture of a finite number of geometric densities. The reason for choosing this particular form of probability density function for the "available periods" lies in the observation that: (1) the system is still mathematically tractable under this condition; and (2) a large class of discrete distributions can be written as (or at least approximated by) such a mixture.

Under these assumptions expressions are derived for the probability generating functions of the number of messages in the buffer after available periods, after blocked periods and after a randomly chosen clock time interval. The results show that our model is a generalisation of several prior models. It also becomes clear that buffer behavior for this kind of system is very dependent on the exact form of the distributions of blocked and available periods, even for the same value of the probability of having the output channel available (or blocked).

Finally it is shown how the model can be used in order to determine buffer behavior of an integrated speech-data communication system, where data is multiplexed in analog speech.

2. DESCRIPTION OF THE INVESTIGATED SYSTEM

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 service time of the messages is constant. The clock time period is defined as the time required to service one message. The data transmission is synchronous, i.e. the data is 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 interruption process of the output channel is characterized in the following way: as time goes by the output channel is available for a number of clock time periods, then is blocked for a number of clock time periods, then is available again, and so on. Let us denote the state where the output channel is available by $A$ and the state where it is blocked by $B$. Then the state of the server alternately equals $A$ or $B$ for some time. Let us denote by $X_d(v)$ and $X_b(v)$ the lengths of the $v$th interval where the server is in state $A$ or $B$ respectively, expressed in clock time periods. We assume that the $X_d(v)$'s are i.i.d. according to a probability mass function $a(n)$, that the $X_b(v)$'s are i.i.d. according to a probability mass function $b(n)$ and that the $X_d(v)$'s are independent of the $X_b(v)$'s:

\[
\begin{align*}
a(n) &= \text{Prob}[X_d(v) = n], \quad n = 1, 2, \ldots \\
b(n) &= \text{Prob}[X_b(v) = n], \quad n = 1, 2, \ldots
\end{align*}
\]

Under these conditions the server interruption process is completely specified by the two probability mass functions $a(n)$ and $b(n)$. Throughout this paper it will be assumed that $b(n)$ is an arbitrary probability density function, whereas $a(n)$ is a mixture of geometric

![Fig. 1. The investigated buffer system.](image)
density functions, i.e.

\[ a(n) = \sum_{i=1}^{m} r_i a_i^n, \]  

where \( 0 \leq a_i < 1 \) for \( i = 1, 2, \ldots, m \) and

\[ \sum_{i=1}^{m} \frac{a_i r_i}{1 - a_i} = 1, \]

for normalisation.

Here \( m \) denotes the number of geometric densities needed in the mixture (in order to obtain a good approximation of the actual density), the \( a_i \)'s are the parameters of the \( m \) densities and the \( r_i \)'s are measures for the relative importance of each of the \( m \) geometric densities in the mixture.

Let \( P_s(z) \) and \( P_b(z) \) indicate the probability generating functions of \( X_s(v) \) and \( X_b(v) \) respectively.

The total number of arriving messages during the \( k \)th clock time interval is denoted by \( e_k \). The \( e_k \)'s are assumed to be i.i.d. with probability mass function \( e(n) \), i.e.

\[ e(n) = \text{Prob}[n \text{ messages arrive during the } k \text{th clock time interval}] = \text{Prob}[e_k = n], \quad n = 0, 1, 2, \ldots \]

Let \( E(z) \) indicate the corresponding probability generating function.

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.

3. 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 "available period" or a "blocked period" respectively, and \( C_0(z) \) and \( D_0(z) \) the corresponding probability generating functions.

Let \( c_1, c_2, \ldots, c_n, \ldots \) be the random variables which equal the equilibrium number of messages in the buffer after the first, second, \ldots, \( k \)th, \ldots clock time period of an "available period", and \( d_1, d_2, \ldots, d_n, \ldots \) the random variables which equal the equilibrium number of messages after the first, second, \ldots, \( k \)th, \ldots clock time period of a "blocked period". The corresponding probability generating functions are denoted by \( C_1(z), C_2(z), \ldots, C_n(z), \ldots \) and \( D_1(z), D_2(z), \ldots, D_n(z), \ldots \) respectively.

Figure 2 illustrates these definitions.

Further we define for each value of \( x \) in the interval\([0, 1)\) the function \( f(x, z) \) by

\[ f(x, z) = \sum_{i=1}^{n} C_i(z)x^i. \]  

Here \( x \) denotes a dummy variable. The meaning of \( x \) and \( f(x, z) \) will become clear in the next section.

4. ANALYSIS OF BUFFER BEHAVIOR

Let us establish the state equations of the buffer system.

During "blocked periods" no departures can take place; thus we have

\[ d_k = d_{k-1} + e_k^p, \quad k = 1, 2, \ldots \]
where \( e_k^B \) is the number of arrivals during the \( k \)th clock time interval of a "blocked period". Since \( d_{k-1} \) and \( e_k^B \) are mutually independent, we have

\[
D_k(z) = E(z)D_{k-1}(z), \quad k = 1, 2, \ldots
\]

It follows that

\[
D_0(z) = [E(z)]^D D_0(z).
\]  

(3)

Now it is clear that \( C_0(z) \) can be seen as the probability generating function of the number of messages in the buffer after the \( X_k \)th clock time interval of a "blocked period" where \( X_k \) is the length of the "blocked period". Hence

\[
C_0(z) = E[z^c] \\
= E_{X_k}[E[z^c|X_k]] \\
= E_{X_k}[D_{X_k}(z)] \\
= D_0(z). E_{X_k}[E(z)]^X_k \\
= D_0(z). P_d(E(z)).
\]  

(4)

During "available periods" we have

\[
c_k = \begin{cases} 
  c_{k-1} + e_k^A - 1 & \text{if } c_{k-1} > 0 \\
  e_k^A & \text{if } c_{k-1} = 0 \end{cases}, \quad k = 1, 2, \ldots
\]

where \( e_k^A \) is the number of arrivals during the \( k \)th clock time interval of an "available period". Combining the two equations above, we obtain

\[
c_k = e_k^A + (c_{k-1} - 1)^+
\]

where the notation \( (\ldots)^+ \) is used to indicate \( \max(0, \ldots) \). Since \( e_k^A \) and \( c_{k-1} \) are mutually independent random variables this leads to

\[
C_k(z) = E(z) C_{k-1}(z) + (z - 1)C_{k-1}(0)/z, \quad k = 1, 2, \ldots
\]  

(5)

From the definition of \( D_0(z) \) it is clear that \( D_0(z) \) can be viewed as the probability generating function of the number of messages in the buffer after the \( X_k \)th clock time interval of an "available period" where \( X_k \) is the length of the "available period". Thus

\[
D_0(z) = E[z^c] \\
= E_{X_k}[E[z^c|X_k]] \\
= E_{X_k}[C_{X_k}(z)] \\
= \sum_{k=1}^{\infty} a(k)C_k(z)
\]
Analysis of an infinite buffer system

\[= \sum_{k=1}^{\infty} \sum_{i=1}^{m} r_i(x_i) \frac{C_i(z)}{z - \alpha_i E(z)} = \sum_{i=1}^{m} r_i \sum_{k=1}^{\infty} C_i(z) (x_i)^k = \sum_{i=1}^{m} r_i f(x_i, z).\]

Here the function \(f\) was defined in eqn (2). An expression for \(f(x, z)\) can be found from eqn (5) by multiplying this equation by \(x^k\) and summing over all values of \(k \geq 1\). This leads to

\[f(x, z) = \frac{x E(z)}{z - x E(z)} \left\{ C_0(z) + (z - 1)[C_0(0) + f(x, 0)] \right\},\]

and

\[D_0(z) = C_0(z) E(z) \sum_{i=1}^{m} \frac{r_i \alpha_i}{z - \alpha_i E(z)} + (z - 1) E(z) \sum_{i=1}^{m} \frac{m_i}{z - \alpha_i E(z)}\]

where the \(m_i\)'s are defined by

\[m_i = r_i \alpha_i \{C_0(0) + f(x_i, 0)\}.\]

Let us introduce

\[g(z) = \prod_{i=1}^{m} \left\{ z - \alpha_i E(z) \right\}\]

and

\[g(z) = \frac{g(z)}{z - \alpha E(z)},\]

then we obtain

\[D_0(z) g(z) = C_0(z) E(z) \sum_{i=1}^{m} r_i \alpha_i g(z) + (z - 1) E(z) \sum_{i=1}^{m} m_i g(z).\]

Combining (4) and (10) we finally obtain

\[C_0(z) = \frac{(z - 1) E(z) P_{\Delta} E(z)}{g(z) - E(z) P_{\Delta} E(z)} \sum_{i=1}^{m} m_i g(z)}{g(z) - E(z) P_{\Delta} E(z) \sum_{i=1}^{m} r_i \alpha_i g(z)}\]

and

\[D_0(z) = \frac{(z - 1) E(z) \sum_{i=1}^{m} m_i g(z)}{g(z) - E(z) P_{\Delta} E(z) \sum_{i=1}^{m} r_i \alpha_i g(z)}\]

The \(m\) remaining unknown parameters \(m_i\) can be determined as follows. Using Rouché's theorem (see e.g. [15]) it can be shown that, whenever the condition for a stochastic equilibrium is met, the denominator of \(C_0(z)\) and \(D_0(z)\) has exactly \(m\) zeros in the unit disk.
\[ \{ z : |z| \leq 1 \}, z_0 = 1, z_1, z_2, \ldots, z_{m-1}, \] which must be zeros of the numerator of \( C_0(z) \) and \( D_0(z) \)
as well since probability generating functions are analytic in the unit disk \( \{ z : |z| \leq 1 \} \). The resulting \( m-1 \) linear equations in the \( m_i \)'s (no equation is found for \( z_0 = 1 \)) together with the normalisation condition \( C_0(1) = 1 \) or \( D_0(1) = 1 \) yield the desired values of the \( m_i \)'s. Using these an expression for \( C_0(z) \) and \( D_0(z) \) in terms of known quantities only can be derived.

We can proceed with the derivation of an expression for the probability generating function \( N(z) \) of the number of messages in the buffer at random clock times.

Let \( t \) indicate the end of a randomly chosen clock time interval. The probability that during this clock time interval the output channel is available (blocked) is given by
\[
\sigma = \frac{E[X_d]}{E[X_d] + E[X_b]},
\]
(13)

Let us consider two cases (Fig. 3):

1. The time \( t \) lies "in" a "blocked" period. Let \( X^*_b \) denote the length of the particular "blocked" period which contains \( t \). Since \( t \) was chosen randomly, we have
\[
\text{Prob}[X^*_b = k] = \frac{kb(k)}{E[X_b]}, \quad k = 1, 2, \ldots
\]

Let \( l_b \) indicate the number of clock time intervals in \( X^*_b \) before \( t \), and \( L_b(z) \) its probability generating function. Then:
\[
\text{Prob}[l_b = l] = \sum_{k=l}^{\infty} \text{Prob}[l_b = l | X^*_b = k] \cdot \text{Prob}[X^*_b = k] = \sum_{k=l}^{\infty} \frac{b(k)}{E[X_b]}
\]
since \( t \) was chosen at random.

Furthermore
\[
L_b(z) = \sum_{i=1}^{\infty} \text{Prob}[l_b = l | z^l] = \frac{z}{z - 1} \frac{P_b(z) - 1}{E[X_d]},
\]
(14)

Let \( N_b(z) \) indicate the probability generating function of the number of messages in the buffer at time \( t \) (in a "blocked" period). Then:

![Fig. 3. Sample function of the server state (A or B) versus time, a long time after the first arrivals. Illustration of the two possible positions of a random clock time \( t \). Definitions of \( l_b \), \( X^*_A \), \( X^*_B \), \( N_A(z) \) and \( N_B(z) \).](image-url)
Analysis of an infinite buffer system

\[ N_d(z) = E_d[D_a(z)] \]
\[ = D_d(z) \cdot L_g(E(z)) \]
\[ = D_d(z) \cdot \frac{E(z)}{E(z) - 1} \cdot \frac{P_g(E(z)) - 1}{E[X_g]}, \quad (15) \]

where we have used eqns (3) and (14).

(2) The time \( t \) lies "in" an "available" period. Let \( I_a \) indicate the number of clock time intervals in this particular available period before \( t \). Then:

\[ \text{Prob}[I_a = l] = \sum_{i=1}^{m} \frac{a(k)}{E[X_a]} \]

where this result was derived as for "blocked" periods. Using equation (1) we obtain

\[ \text{Prob}[I_a = l] = \sum_{i=1}^{m} s_i(a_i) \]

(16)

where

\[ s_i = \frac{r_i}{(1 - \alpha_i)E[X_a]} \quad (17) \]

i.e. the probability distribution function of \( I_a \) is also a mixture of \( m \) geometric densities with parameters \( \alpha_1, \alpha_2, \ldots, \alpha_m \).

Let \( N_d(z) \) indicate the probability generating function of the number of messages in the buffer at time \( t \) (in an "available" period). Then:

\[ N_d(z) = E_d[C_{I_a}(z)] \]
\[ = \sum_{i=1}^{m} \text{Prob}[I_a = l] \cdot C(z); \]

or if we use eqn (16):

\[ N_d(z) = \sum_{i=1}^{m} s_i f(a_i, z). \]

Using the eqns (6)–(9) and (17) we obtain the following expression:

\[ N_d(z) = \frac{E(z)}{E[X_a]g(z)} \left[ C_d(z) \sum_{i=1}^{m} \frac{\alpha_i r_i}{1 - \alpha_i} g(z) + (z - 1) \sum_{i=1}^{m} \frac{m_i}{1 - \alpha_i} g(z) \right]. \]

Since

\[ \sum_{i=1}^{m} \frac{\alpha_i r_i}{1 - \alpha_i} = 1 \]

the sum

\[ \sum_{i=1}^{m} \frac{\alpha_i r_i}{1 - \alpha_i} g(z) \]

may be considered as an average of the \( g(z) \)'s. If we use the notation \( \bar{g}(z) \) for this sum we finally obtain:

\[ N_d(z) = \frac{E(z)}{E[X_a]g(z)} \left[ C_d(z)\bar{g}(z) + (z - 1) \sum_{i=1}^{m} \frac{m_i}{1 - \alpha_i} g(z) \right], \quad (18) \]
The probability generating function \( N(z) \) of the number of messages in the buffer at random clock times is given by

\[
N(z) = \sigma N_a(z) + (1 - \sigma)N_b(z)
\]

where \( \sigma \) is the quantity defined in equation (13).

Using the expressions we derived for \( N_a(z) \) and \( N_b(z) \) and substituting the functions \( C_b(z) \) and \( D_b(z) \) by the expressions given in eqns (11) and (12), we find:

\[
N(z) = \frac{(z - 1)E(z)}{S.g(z)} \sum_{i=1}^{m} \frac{m_i a_i}{1 - a_i} g(z) + \frac{(z - 1)E(z)}{S.g(z)[E(z) - 1]} \left[ \frac{(z - 1)E(z)P_b(E(z))}{g(z) - E(z)P_b(E(z))} \sum_{i=1}^{m} r_i a_i g(z) \right] - \sum_{i=1}^{m} m_i g_i(z).
\]

(19)

Here the quantity \( S \) is defined as

\[
S = E[X_a] + E[X_b].
\]

(20)

Equation (19) gives us the desired expression of the probability generating function of the buffer occupancy at random clock times as a function of known quantities only (the parameters \( m_i \) are determined as explained above). From this function the important performance measures such as mean value and variance of the buffer occupancy can be easily derived, using the moment generating property of the probability generating function.

5. AVAILABLE PERIODS WITH GEOMETRIC DENSITY FUNCTION

In this section we consider the special case where the available periods have a geometric density function, i.e. where the parameter \( m \) equals 1. In our model we then have:

\[
a(n) = r a^n
\]

where

\[
r = \frac{1 - a}{a},
\]

\[
g(z) = z - az E(z);
\]

\[
g_j(z) = 1;
\]

\[
\tilde{g}(z) = 1.
\]

From eqns (11) and (12) we obtain:

\[
C_b(z) = \frac{m_a(z - 1)E(z)P_b[E(z)]}{z - E(z)[a + (1 - a)P_b[E(z)]]}
\]

(21)

\[
D_b(z) = \frac{m_b(z - 1)E(z)}{z - E(z)[a + (1 - a)P_b[E(z)]]}
\]

(22)

where \( m_i \) follows from the normalisation condition \( C_b(1) = 1 \) or \( D_b(1) = 1 \):

\[
m_i = 1 - \bar{c} \cdot \{1 + (1 - a)E[X_b]\}
\]

where \( \bar{c} \) is the arrival rate, i.e. the average number of arrivals per clock time interval.
Furthermore, from eqn (19) we find:

\[
N(z) = \frac{(\sigma - \bar{\lambda})(2 - 1)E(z)}{1 - E(z)\left[\sigma + (1 - \alpha)P_y(E(z))\right]} \frac{1 - E(z)[\alpha + (1 - \alpha)P_y(E(z))]}{z - E(z)[\alpha + (1 - \alpha)P_y(E(z))]} \tag{23}
\]

where \(\sigma\) is the quantity defined in equation (13).

The expression in eqn (23) corresponds to the results found by Towsley[11] in his work on multiplexers operating in a two state Markovian environment.

We now focus attention on the special case where both “available periods” and “blocked periods” have geometric density functions with parameters \(\sigma\) and \(1 - \sigma\) respectively, as in the work of Hsu[8]. In our model we then have:

\[
\alpha = \sigma \quad \quad P_y(z) = \frac{\sigma z}{1 - (1 - \sigma)z^2}
\]

From eqns (21)–(23) we then obtain:

\[
C_0(z) = \frac{(\sigma - \bar{\lambda})(2 - 1)E^2(z)}{z - E(z)[\sigma + (1 - \alpha)P_y(E(z))]} \tag{24}
\]

\[
D_0(z) = \left(1 - \frac{\bar{\lambda}}{\bar{\lambda}}\right) \frac{(\sigma - \bar{\lambda})(2 - 1)E(z)[1 - (1 - \alpha)E(z)]}{z - E(z)[\sigma + (1 - \alpha)P_y(E(z))]} \tag{25}
\]

\[
N(z) = \frac{(\sigma - \bar{\lambda})(2 - 1)E(z)}{z - E(z)[\sigma + (1 - \alpha)P_y(E(z))]} \tag{26}
\]

For the special case of a Poisson arrival process

\[
E(z) = \exp[-\lambda(1 - z)]
\]

eqn (26) reduces exactly to the expression derived by Hsu[8].

The examples above show that our model is a generalisation of several previous models with more restrictive server interruption processes. The importance of the exact form of the probability mass function of “available periods” and “blocked periods” can be illustrated by the following comparison. Let us compare all the cases where both “available periods” and “blocked periods” have geometric distributions and where the (long time) probability of having an available output channel is given by the same value \(\sigma\), i.e. where

\[
a(n) = \left(\frac{1 - \sigma}{K}\right)\left(1 - \frac{\sigma}{K}\right)^{n-1};
\]

\[
b(n) = \left(\frac{\sigma}{K}\right)\left(1 - \frac{\sigma}{K}\right)^{n-1}; \quad n = 1, 2, \ldots
\]

Here the parameter \(K\) can take all values greater than \(\max(\sigma, 1 - \sigma)\). The probability generating function of the buffer occupancy at random clock times can be found from eqn (23) with:

\[
\alpha = 1 - \frac{1 - \sigma}{K}
\]

and

\[
P_y(z) = \frac{\sigma z}{K - (K - \sigma)z^2}.
\]
The corresponding mean queue length is given by

\[ \bar{N} = \frac{dN(z)}{dz} \bigg|_{z=1}, \]

As a result we obtain:

\[ \bar{N} = \bar{e} + \frac{E''(1)}{2(\sigma - \bar{e})} + K \frac{(1 - \sigma)\bar{e}}{\sigma - \bar{e}} \]

where \( E''(1) \) denotes the second derivative of \( E(z) \) evaluated at \( z = 1 \).

The expression for \( \bar{N} \) shows that the parameter \( K \)—which is a measure for the absolute lengths of “available” and “blocked” periods rather than \( \sigma \) which has to do with the ratio of these lengths—has a great and direct influence on the values of \( \bar{N} \) for variable values of the mean arrival rate \( \bar{e} \).

This means that analyses which characterize the interruption process by only one parameter \( \sigma \) (the fraction of time during which the output channel is available) considered as the probability of having an available channel during an arbitrary clock time interval, i.e. taking \( K = 1 \), may lead to results which deviate considerably from the actual buffer occupancies.

6. AN APPLICATION: SYNCHRONOUS MULTIPLEXING OF DATA IN SPEECH

We consider a communication system here where one transmission line is used both for the transmission of data (from a buffer) and for the transmission of a speech signal. This channel is available for the transmission of data, whenever a “silent period” occurs in the speech signal; the channel is blocked during so-called talkspurts. Similar voice-data systems are described for instance in[3–6]. Sherman[6] considers an infinite buffer size, continuous buffer states and asynchronous multiplexing of data in speech. In[3–5] a finite buffer size is considered, but the interruption process of the transmission line is modeled through a Bernoulli sequence of independent random variables[3, 4] or through a first order Markov process[5]. In both cases the talkspurts and the silent periods are considered as random variables with geometric density functions (discrete equivalent of continuous exponential distribution).

This may be a good approximation with respect to the talkspurts, but it certainly is not as far as the silent periods are concerned. It has been found that for prose reading two types of silence gaps occur: short gaps with a mean duration on the order of 10 ms on one hand and longer gaps of about 1–2 s on the other hand. For conversational speech a third type of silence gaps occurs corresponding to the silences when the speaker listens to the other person[13, 14]. If each type of silence gaps is modeled through a geometric random variable, and the relative occurrence of the different types of gaps is known, the silent periods themselves can be described through a mixture of geometric densities, resulting in a better approximation than a single geometric density, which does not take into account the fact that there are different types of silence gaps[6].

Under these circumstances the behavior of the data buffer of the voice-data system can be studied using the model we developed in this paper, if we let the talkspurts correspond to the “blocked periods” and the silent periods in speech to the “available periods” of the server. Let us consider the case where the speech signal comes from prose reading. We will use the following notations:

average talkspurt length = \( 1/\alpha \)

average silent periods = \( 1/b_1 \) (short gaps)

relative occurrence of
short silence gaps = \( A \)

long silence gaps = \( 1 - A \).
In our model then
\[ a(n) = Ab_1(1 - b_1)^{n-1} + (1 - A)b_2(1 - b_2)^{n-1} \]
\[ b(n) = a(1 - a)^{n-1} \]
and thus
\[ m = 2; \]
\[ r_1 = \frac{Ab_1}{1 - b_1} \]
\[ \alpha_1 = 1 - b_1 \]
\[ r_2 = \frac{(1 - A)b_2}{1 - b_2} \]
\[ \alpha_2 = 1 - b_2 \]
and
\[ P_s(z) = \frac{az}{1 - (1 - a)z} \]
The expressions for \( C_0(z) \), \( D_0(z) \) and \( N(z) \) may be found from the eqns (11), (12) and (19), using the above data. Let us concentrate on \( C_0(z) \) which is the most important when it comes to determine the size of the data buffer, since it has to do with the number of messages in the buffer just after a "blocked period".
From eqn (11):
\[ C_0(z) = \frac{a(z - 1)E^2(z)[m_1g_1(z) + m_2g_2(z)]}{[1 - (1 - a)E(z)]g(z) - aE^2(z)[Ab_1g_1(z) + (1 - A)b_2g_2(z)]] \]
The unknown parameters \( m_1 \) and \( m_2 \) must be determined as explained in Section 4. In general the zeros of the denominator \( d(z) \) of \( C_0(z) \) can only be found in a numerical way. For some particular arrival processes \( (E(z)) \) however these zeros can be found as explicit functions of the parameters of the system and a closed form expression for \( C_0(z) \) can be derived.
Let us consider a Bernoulli arrival process with parameter \( \lambda \), i.e.,
\[ E(z) = (1 - \lambda) + \lambda z \]
then it can be shown that
\[ C_0(z) = \frac{1 - \omega}{z - \omega} (1 - \lambda + \lambda z)^2 \]
where \( \omega \) is the only zero of \( d(z) \) outside the unit disk in the \( z \)-plane. The corresponding mean buffer occupancy is then given by
\[ N_{\text{max}} \Delta C_0(1) = 2\lambda + \frac{1}{\omega - 1}. \]
This can be shown to be given by
\[ N_{\text{max}} = \frac{2\lambda(1 - \lambda)[1 - \lambda + \sqrt{4\lambda d_0 + [(b_1 + b_2)\lambda - a(1 - \lambda)]^2}]}{2\lambda d_0 + (1 - \lambda)a(1 - \lambda) - \lambda(1 - \lambda) + \sqrt{4\lambda d_0 + [(b_1 + b_2)\lambda - a(1 - \lambda)]^2}} \]
where \( d_0 = a(1 - \lambda)[(1 - A)b_1 + Ab_2] - \lambda b_1 b_2. \)
Let us compare this result to the result which is found if the silence gaps are modeled through a random variable with geometric density, instead of a mixture of two geometric densities, as would be done with any of the prior models.

In this case we have

$$a(n) = b(1 - b)^{n-1}$$

where $1/b$ is the mean length of silent periods, i.e.,

$$\frac{1}{b} = \frac{A}{b_1} + \frac{(1 - A)}{b_2}.$$ \(C_d(z)\) for this case follows from eqn (21) where $a = 1 - b$ and

$$P_d(z) = \frac{az}{1 - (1 - a)z}.$$  

If we use the same Bernoulli input process as above this leads to the following expression:

$$C_p(z) = \frac{1 - \omega^*(1 - \lambda z)^2}{z - \omega^*(1 - \lambda + \lambda z)}$$

where

$$\omega^* = 1 + \frac{a - \lambda(a + b)}{\lambda b + (1 - a)\lambda(1 - \lambda)}.$$ 

The corresponding mean buffer occupancy is

$$\bar{N}_{geom} \triangleq C_p'(1) = \frac{(a + 1)(1 - \lambda) - \lambda^2 b}{a - \lambda(a + b)}.$$  

In Fig. 4 plots of $\bar{N}_{mix}$ and $\bar{N}_{geom}$ as functions of the mean arrival rate $\lambda$ are given for the following realistic values of the parameters[6]:

![Fig. 4. Mean buffer occupancy of an integrated voice-data system just after talkspurts, versus the mean arrival rate $\lambda$ per clock time interval, for the case of a Bernoulli arrival process with parameter $\lambda$: $\bar{N}_{geom}$ for geometrically distributed talkspurts and silent periods; $\bar{N}_{mix}$ for geometrically distributed talkspurts and silent periods which have a density function that is a mixture of two geometric densities.](image-url)
Analysis of an infinite buffer system

\[
\begin{align*}
1/a &= 500 \text{ clock time periods} \\
1/b_1 &= 100 \text{ clock time periods} \\
1/b_2 &= 10,000 \text{ clock time periods} \\
A &= 0.95.
\end{align*}
\]

These values are found in a system where the clock time frequency is 10 kHz, the mean talkspurt length 50 ms, the mean length of short gaps 10 ms, the mean length of long gaps 1 s and where 95% of the silent periods are short gaps.

The plots clearly show that for all permitted values of \( \lambda \), \( N_{\text{max}} \) takes considerably larger values than \( N_{\text{peoc}} \). This means that the mean buffer occupancy may be substantially underestimated if a model is used where both talkspurts and silent periods are treated as geometrically distributed random variables.

Thus it may be quite important to use a realistic model for the distribution of the silent periods. The model we developed in this paper provides a means to do so, since a wide class of discrete density functions can be approximated by a mixture of geometric densities. In this sense, our model for the interruption process of the output line is much more general than most of the previous models, which allow only a geometric density function for the “available period” length (and for the “blocked period” length).

Furthermore the model can be used with an arbitrary arrival process (\( E(\tau) \)), whereas in many models the arrivals are restricted to the Poisson or the compound Poisson type. However, in general this implies that the \( m \) unknown parameters \( m \) can only be found in a numerical way since the roots of a possibly transcendental equation are to be determined.

7. CONCLUSIONS

We have considered an infinite buffer system with general arrival process, synchronous transmission, one single output channel and random server interruptions, where the stochastic nature of the server interruption process has been modeled through two independent sequences of i.i.d. random variables: the \( X_\lambda(\tau) \)'s for the lengths of the periods during which the server is available (with a density function which is a mixture of a finite number of geometric densities) and the \( X_\delta(\tau) \)'s for the lengths of the periods during which the server is blocked (with arbitrary density function). Explicit formulas for the probability generating functions of the buffer occupancy after blocked periods, after available periods and after random clock time intervals are obtained, provided the roots of a (generally transcendental) equation are determined. From these functions the average buffer occupancy and if desired higher moments of the buffer occupancy can be derived. The results show, among other things, the importance of the exact form of the distributions of blocked and available periods. Finally, it is shown how the analysis can be applied to an integrated voice-data system in order to determine buffer behavior of the data buffer in this system.

REFERENCES