Instantaneous Multiple Input Finite Buffer System

Bih-Wang Lee

Arne A. Nilsson

Center for Communications and Signal Procesing
Department of Electrical and Computer Engineering
North Carolina State University

TR-90/6

July 1990

INSTANTANEOUS MULTIPLE INPUT FINITE BUFFER SYSTEM

Bih-Hwang Lee

Arne A. Nilsson

Center for Communications and Signal Processing
Department of Electrical and Computer Engineering
North Carolina State University
Raleigh, NC 27695-7914
(Phone): 919-737-3015
(FAX): 919-737-7382

ABSTRACT

A buffer is needed to store the excess information if the transmission rate of a transmitter is less than its input rate. In this paper, a finite buffer is analyzed with instantaneous multiple input, which has a deterministic interarrival time and a generally distributed message size; the main objectives are to minimize the loss of data, and to maximize the utilization of the system. Three performance measures are used to achieve those purposes: the probability distribution function of the buffer occupancy is used to characterize the usage of the buffer; the probability distribution function of the amount of loss is used to determine if the loss is acceptable; and the utilization of the system refers to the need to make efficient use of transmission facilities. Analytical expressions are obtained for these performance measures. An approximation is obtained and compared with results given by a numerical method and a simulation experiment.

1. INTRODUCTION

Recently, computer communication networks have been used very widely to exchange information between two systems. Figure 1 shows a communication environment in which many stations are connected to a communication network. If a local station wants to send a message to a remote station, then a certain procedure is followed: the message must first go to the source node, then the message travels through zero or more intermediate nodes and arrives at the destination node; finally, the remote station receives the message.

Basically, the communication environment consists of three major systems: the source system, the transmission medium, and the destination system. Depending upon the functions of each element, we summarize the systems in Figure 2, e.g., a local station functions as an input device, a source node functions as a transmitter, and the intermediate nodes function as a transmission medium, etc.

The input device can be a computer, a data generator, or any data handling machine. If the input device is an image coder, the image information is manipulated into a series of digital data by the coder, and then the transmitter sends the data to the transmission medium upon receiving the data. The transmitter can immediately send data if its transmission rate is greater than its input rate. Unfortunately, the transmission rate of a transmitter is often less than the output rate of an input device, potentially resulting in a loss of data.

In order to prevent this data loss, we need a buffer in which the excess input to the transmitter is temporarily stored, Figure 3. An interesting question is how large a buffer we need. Of course, there is no loss if we have an infinite buffer, but this is very inefficient and totally unrealistic. We also know that if the buffer is small, the loss can be considerable unless we accept a reduction in the output rate of the input device; however, this reduction implies a degraded performance. Generally speaking, if the buffer size increases, then the cost increases, but the loss decreases. The main objectives of this research are to minimize the loss and to maximize the utilization of the system.

The performance of the buffered system can be evaluated by obtaining the probability distribution function of the buffer occupancy, the probability distribution function of the amount of loss, and the utilization of the system. From the probability distribution function of the buffer occupancy, we know how the buffer has been used. Furthermore, we use the probability distribution function of the amount of loss to determine if the loss is acceptable. The utilization of the system refers to the need to make efficient use of transmission facilities. We certainly hope that the transmitter has high utilization and that the amount of loss is relatively small or acceptable, i.e., the probability that the buffer is empty should be small, and the probability of no loss should be relatively high.

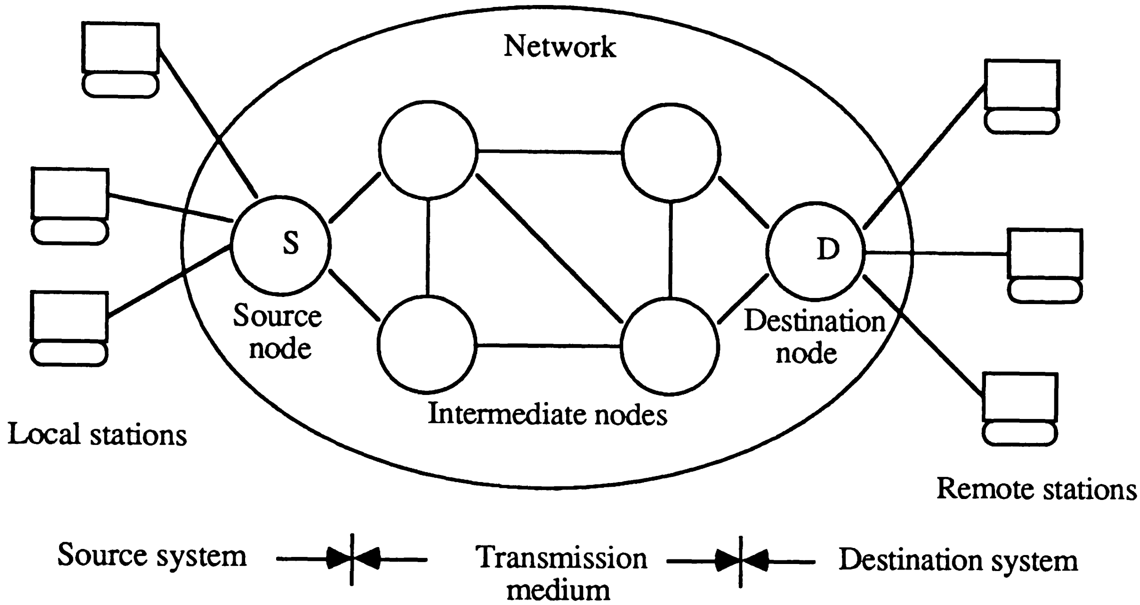

Figure 1 A communication environment.

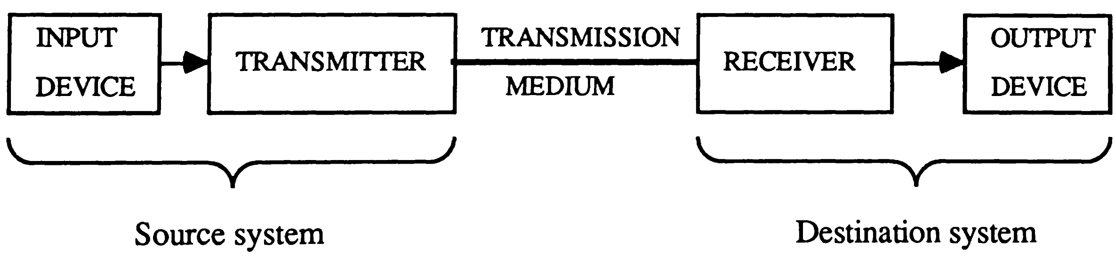

Figure 2 The communication model.

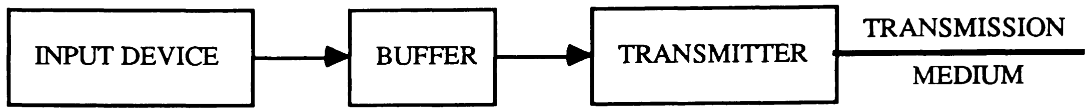

Figure 3 A buffered source system.

2. Description

We assume that there is a total of N input devices connected to a common buffer, Figure 4; these input devices are scheduled alternately to send a message within a time slot, which is a fixed time interval T. We also divide a time slot into N pieces, called time shift τ ($\tau = T/N$). After a time shift, another input line starts its own time slot to begin sending its message, i.e., the time slot of the next line lags by time τ. In other words, if the time slot of line 0 begins with time 0, then, in turn in round robin fashion, the time slot of the next line begins with time τ; the others begin with time 2τ, 3τ, ..., $(N-1)\tau$, etc. At time T, line 0 again obtains its own time slot to send the remaining part of the message; however, any message can be sent only at the beginning of its own time slot. We assume that the message size which can be brought into the buffer in a time slot is generally distributed; the probability distribution function of the message size is denoted by $B_i(x)$, and its density function is $b_i(x)$, where i=0, 1, ..., N-1. We also assume that the propagation delay from each input device to the buffer is the same as the others; otherwise, the time shift in our analysis is no longer a constant. Any input line may use several consecutive time slots, a busy-period, to send an entire message; then it remains in idle status for some time, an idle-period. We assume that the busy-period and the idle-period are geometrically distributed.

If the buffer is empty at the time when any message enters, then the message is transmitted through the medium immediately; however, if the input rate of the buffer is greater than its output rate, the excess message has to be stored in the buffer. In this paper, we assume that the output rate of the input devices is much greater than the transmission rate of the transmitter, i.e., an instantaneous input. When no message arrives and the buffer is not empty, the messages are still transmitted until the buffer becomes empty; that is, the system is entirely idle. The queueing model for this system is shown in Figure 5.

Since the buffer size is finite, any message in excess of the capacity of the buffer, K, is lost. However, in this system, not only is the excess message destroyed, but also the entire message brought into the buffer at the time slot is destroyed as well. Because a message should be recognized by carrying some information contained in the header and the trailer of the message; hence, the header and the trailer cannot be ignored. In data communication, a message must contain several kinds of information, such as routing and error controls, for transmitting data from a source system to another destination system. The buffer occupancy with timing for each line is shown in Figure 6. Note that the lost message could be re-sent, depending on the protocol of the source system, which we do not discuss here. However, if a lost message could be re-sent, we would still count the lost message as a loss in our system. On the whole, this is a very complicated, but realistic, system.

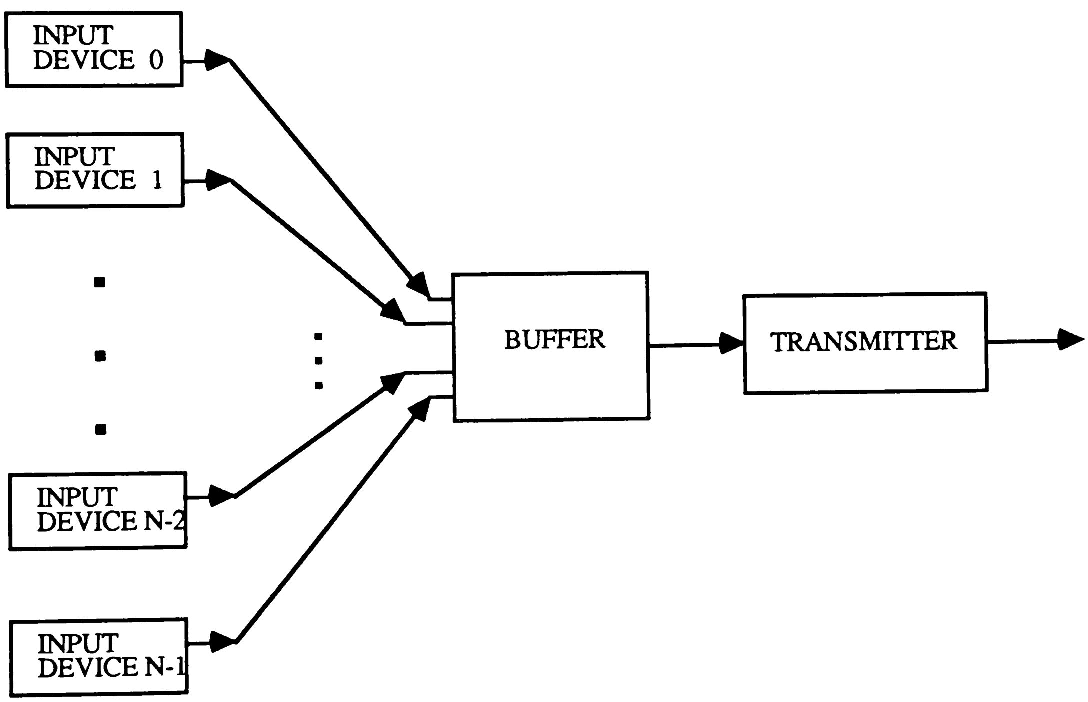

Figure 4 The buffered system with multiple input lines.

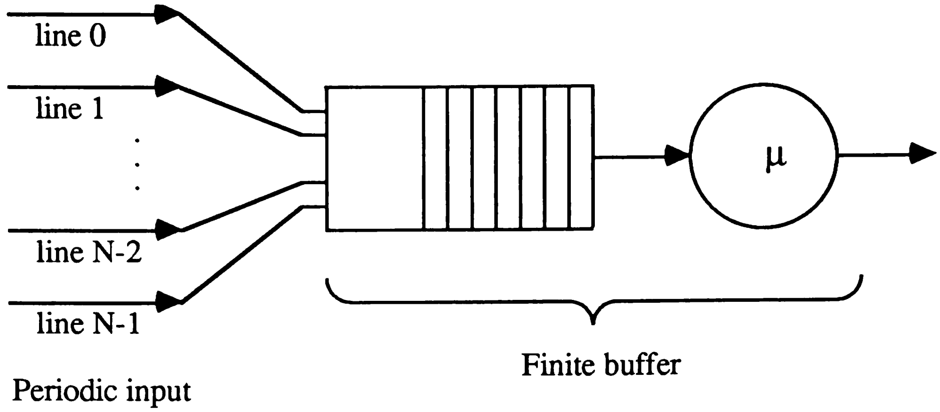

Figure 5 A finite buffer queueing model with multiple inputs.

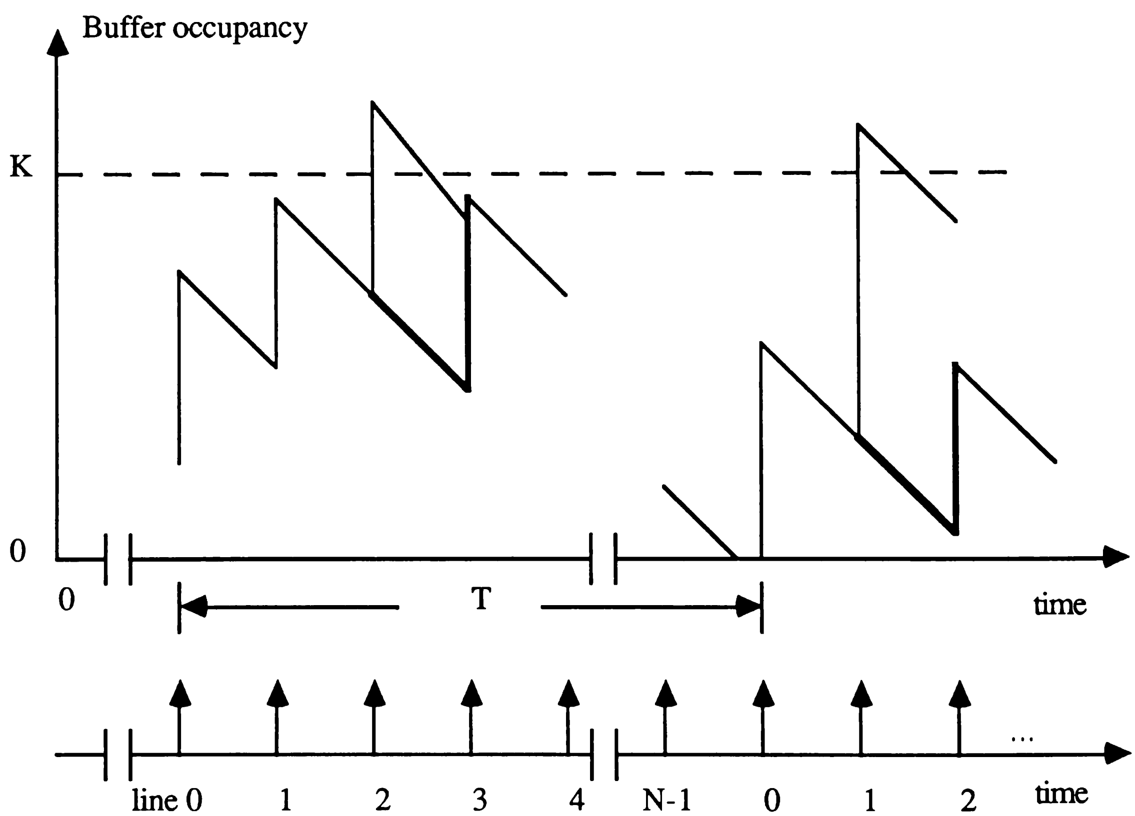

Figure 6 The buffer occupancy with the timing for each line.

3. Analysis

In order to analyze the system, we use an embedded Markov chain analysis. The embedded points are chosen to be the input time at which a new message has not been entered. We need to consider not only the probability distribution functions of the busy and idle periods, but also the probability distribution function of the message size sent at the beginning of the time slot. A detailed Markov chain description is shown in Figure 7.

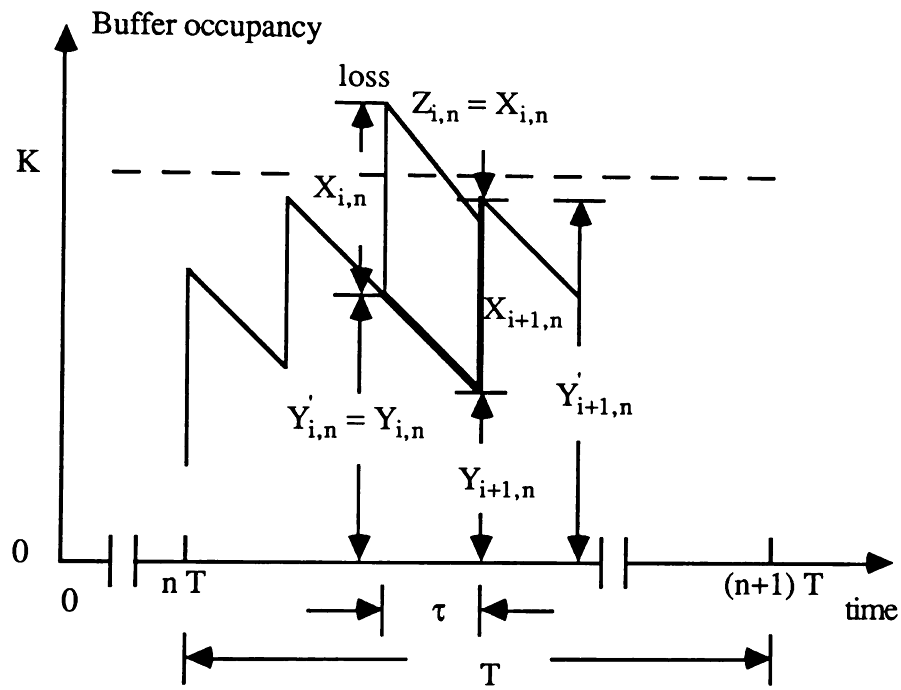

Figure 7 The embedded Markov chain description for multiple input lines.

Let $Y_{i,n}$ and $X_{i,n}$ represent the buffer occupancy before a new message enters at time slot n of line i, and the message size entering from line i to the buffer at time slot n of line i, respectively. A loss occurs if any part of a message exceeds the capacity of the buffer. The amount of loss at time slot n of line i is denoted by $Z_{i,n}$, where n = 0,1,2, ... , and i = 0, 1, ..., N-1. The units are in bits for all variables mentioned above; both time slot T and time shift τ are in seconds. For simplicity, we assume that the transmission rate of the transmitter is one bit per second; then, we find two equations to represent $Y_{i+1,n}$ and $Z_{i,n}$, respectively:

$$Y_{i+1,n} = [Y'_{i,n} - \tau]^+ , \tag{1}$$

where $[x]^+ = \max[0, x]$ and

$$Y'_{i,n} = \begin{cases} Y_{i,n} & \text{line i is idle at time slot n} \\ Y_{i,n} & \text{line i is busy at time slot n, and overflow occurs} \\ Y_{i,n} + X_{i,n} & \text{line i is busy at time slot n, and no overflow occurs.} \end{cases}$$

and

$$Z_{i,n} = \begin{cases} X_{i,n} & \text{if } Y_{i,n} + X_{i,n} > K \\ 0 & \text{elsewhere} . \end{cases} \tag{2}$$

Based upon the embedded Markov chain description, we analytically, by definition, obtain an integral equation for the probability distribution function of the buffer occupancy and equations for the probability distribution function of the amount of loss and for the utilization of the system.

3.1. The Probability Distribution Function of The Buffer Occupancy

According to the definition of $Y_{i,n}$, we know that the value of $Y_{i,n}$ is never more than K-τ; hence, the probability distribution function of the buffer occupancy must be equal to one for $y \geq K-\tau$. We let $W_{i,n}(y)$ represent the probability distribution function of $Y_{i,n}$. Consequently, we only need to consider the case of $y < K-\tau$; by definition, we have

$$\begin{aligned} W_{i+1,n}(y) &= \Pr\{Y_{i+1,n} \leq y\} \\ &= \Pr\{Y'_{i,n} \leq y + \tau\} \\ &= \Pr\{Y'_{i,n} \leq y+\tau \mid \text{line i is busy}\} \Pr\{\text{line i is busy}\} + \\ &\quad \Pr\{Y'_{i,n} \leq y+\tau \mid \text{line i is idle}\} \Pr\{\text{line i is idle}\}. \end{aligned} \tag{3}$$

However, we are only interested in the stationary distribution function of $W_{i,n}(y)$, which is denoted by $W_i(y)$; by definition, we have

$$W_i(y) = \lim_{n \to \infty} W_{i,n}(y) . \tag{4}$$

We assume that the limit exists and that the limiting distribution is independent of the initial state $Y_{i,0}$; hence, $W_i(y)$ is the stationary probability distribution function of the buffer occupancy at the time slot of line i, and its density function is $w_i(y)$. Similarly, for long-term behavior, we use Y_i for the buffer occupancy at the beginning of the time slot of line i and X_i for the message size to be brought into the buffer by line i. Furthermore, let BP_i and IP_i be the proportions of the expected values of the busy-period and the idle-period of line i, respectively. If the busy-period is geometrically distributed with parameter p_i, then its expected value is $1/p_i$; similarly,with parameter q_i for the idle-period , the expected value is $1/q_i$. Hence, we have

$$BP_i = \frac{1/p_i}{1/p_i + 1/q_i} \quad \text{and} \tag{5}$$

$$IP_i = \frac{1/q_i}{1/p_i + 1/q_i} \; . \tag{6}$$

Since we are considering the long-term behavior of the system, for convenience, we use BP_i for Pr{line i is busy} and IP_i for Pr{line i is idle} through this paper. Therefore, we obtain $W_{i+1}(y)$ as follows:

$$W_{i+1}(y) = \Pr\{Y'_i \le y+\tau \mid \text{line i is busy}\}\, BP_i + \Pr\{Y'_i \le y+\tau \mid \text{line i is idle}\}\, IP_i \tag{7}$$

We assume that the stationary probability distribution function of the buffer occupancy is continuous in [0, K-τ], and that $W_i(0) \ge 0$; hence, we have $W_i(0^+) \approx W_i(0)$ and $w_i(0) = W_i(0)\, \delta(y)$. We consider two cases ($0 \le y < K-2\tau$ and $K-2\tau \le y < K-\tau$) to obtain the probability distribution function of the buffer occupancy.

Case 1: $\mathbf{0 \le y < K\text{-}2\tau}$

$$\begin{aligned} W_{i+1}(y) &= \Pr\{Y'_i \le y+\tau \mid \text{line i is busy}\}\, BP_i + \Pr\{Y'_i \le y+\tau \mid \text{line i is idle}\}\, IP_i \\ &= \left\{ \int_0^{y+\tau} w_i(x)\, B_i(y+\tau-x)\, dx + \int_0^{y+\tau} w_i(x)\, [1-B_i(K-x)]\, dx \right\} BP_i + W_i(y+\tau)\, IP_i \\ &= \left\{ \int_0^{y+\tau} W_i(x)\, b_i(y+\tau-x)\, dx - W_i(y+\tau)\, B_i(K-y-\tau) - \int_0^{y+\tau} W_i(x)\, b_i(K-x)\, dx \right\} BP_i \\ &\quad + W_i(y+\tau) \end{aligned} \tag{8}$$

Case 2: **$K-2\tau \le y < K-\tau$**

$$W_{i+1}(y) = \Pr\{Y'_i \le y+\tau \mid \text{line i is busy}\}\, BP_i + \Pr\{Y'_i \le y+\tau \mid \text{line i is idle}\}\, IP_i$$

$$= \left\{ \int_0^{K-\tau} w_i(x)\, B_i(y+\tau-x)\, dx + \int_0^{K-\tau} w_i(x)\, [1-B_i(K-x)]\, dx \right\} BP_i + IP_i$$

$$= \left\{ B_i(y+2\tau-K) + \int_0^{K-\tau} W_i(x)\, b_i(y+\tau-x)\, dx - B_i(\tau) - \int_0^{K-\tau} W_i(x)\, b_i(K-x)\, dx \right\} BP_i + 1 \tag{9}$$

3.2. The Probability Distribution Function of The Amount of Loss

Let $L_{i,n}(z)$ denote the probability distribution function of the amount of loss at time slot n of line i; by definition $L_{i,n}(z) = \Pr\{Z_{i,n} \le z\}$. Again, we are interested in the stationary distribution function of $L_{i,n}(z)$, which is denoted by $L_i(z)$; by definition,

$$L_i(z) = \lim_{n \to \infty} L_{i,n}(z)\,. \tag{10}$$

This limiting distribution must exist and is independent of the initial state $Z_{i,0}$; hence, $L_i(z)$ is the stationary probability distribution function of the amount of loss at the time slot of line i. Similarly, for long-term behavior, we use Z_i for the amount of loss at the time slot of line i. In order to obtain $L_i(z)$, let us first determine its probability density function in the busy period, which is denoted by $l_i(x)$. According to the criteria of the loss measure when we only accept the whole message, the amount of loss is, obviously, either equal to zero or greater than τ; hence, we summarize $l_i(x)$ as follows:

$$l_i(x) = \Pr\{\text{loss} = x \mid \text{line i is busy}\}$$

$$= \begin{cases} B_i(\tau) + \int_\tau^{K} b_i(x)\, W_i(K-x)\, dx & \text{for } x = 0 \\ 0 & \text{for } 0 < x \le \tau \\ b_i(x)\, [1-W_i(K-x)] & \text{for } \tau < x \le K \\ b_i(x) & \text{for } x > K \end{cases} \tag{11}$$

Finally, we have $L_i(z)$ as:

$$L_i(z) = \Pr\{Z_i \le z\}$$

$$= \int_0^z l_i(x)\,dx\ BP_i + IP_i$$

$$= \begin{cases} \left[B_i(\tau) + \int_\tau^K b_i(x)\, W_i(K-x)\, dx \right] BP_i + IP_i & \text{for } 0 \le z \le \tau \\ \left[B_i(z) + \int_z^K b_i(x)\, W_i(K-x)\, dx \right] BP_i + IP_i & \text{for } \tau < z \le K \\ B_i(z)\, BP_i + IP_i & \text{for } z > K \end{cases} \tag{12}$$

3.3. The Utilization of The System, ρ

By definition, we have the utilization at the time slot of line i as follows:

$$\begin{aligned} \rho_i = BP_i \Bigg[& \int\!\!\int \Pr\{X_i = x \text{ and } Y_i = y \mid x+y \le K \text{ and } x+y \ge \tau\}\, dy\, dx \\ & + \int\!\!\int \Pr\{X_i = x \text{ and } Y_i = y \mid x+y \le K \text{ and } x+y < \tau\}\, \frac{x+y}{\tau}\, dy\, dx \\ & + \int\!\!\int \Pr\{X_i = x \text{ and } Y_i = y \mid x+y > K \text{ and } y \ge \tau\}\, dy\, dx \\ & + \int\!\!\int \Pr\{X_i = x \text{ and } Y_i = y \mid x+y > K \text{ and } y < \tau\}\, \frac{y}{\tau}\, dy\, dx \Bigg] \\ & + IP_i \left[\int \Pr\{Y_i = y \mid y \ge \tau\}\, dy + \int \Pr\{Y_i = y \mid y < \tau\}\, \frac{y}{\tau}\, dy \right] \end{aligned} \tag{13}$$

and the utilization of the system, ρ, is determined as:

$$\rho = \frac{1}{N} \sum_{i=0}^{N-1} \rho_i \tag{14}$$

We consider two cases ($K-\tau \ge \tau$ and $K-\tau < \tau$) to obtain ρ_i.

Case 1: **$K-\tau \ge \tau$**

$$\begin{aligned} \rho_i = 1 - \frac{1}{\tau} \Bigg\{ & \int_0^\tau W_i(y)\, dy + BP_i \Bigg[(K-\tau) \int_{K-\tau}^K W_i(K-x)\, b_i(x)\, dx + \int_0^\tau b_i(x) \int_0^{\tau-x} W_i(y)\, dy\, dx \\ & - \int_{K-\tau}^K x\, W_i(K-x)\, b_i(x)\, dx + \int_{K-\tau}^K b_i(x) \int_{K-x}^\tau W_i(y)\, dy\, dx - B_i(K) \int_0^\tau W_i(y)\, dy \Bigg] \Bigg\} \end{aligned}$$

$$= 1 - \frac{1}{\tau} \text{ E[idle period of line i]} \tag{15}$$

Furthermore, E[idle period of line i] can also be, by definition, obtained, i.e.,

E[idle period of line i]

$$= BP_i \Big[\int\int \Pr\{X_i = x \text{ and } Y_i = y \mid x+y \le K \text{ and } x+y < \tau\} (\tau - x - y)\, dy\, dx$$
$$+ \int\int \Pr\{X_i = x \text{ and } Y_i = y \mid x+y > K \text{ and } y < \tau\} (\tau - y)\, dy\, dx \Big]$$
$$+ IP_i \int \Pr\{Y_i = y \mid y < \tau\} (\tau - y)\, dy \tag{16}$$

Case2: **K-τ < τ**

$$\rho_i = 1 - \frac{1}{\tau} \Big\{ \tau - \int_0^{K-\tau} y\, w_i(y)\, dy + BP_i \Big[\int_0^{2\tau-K} B_i(x)\, dx + [B_i(2\tau-K) - B_i(K)] \int_0^{K-\tau} W_i(y)\, dy$$
$$+ \int_{2\tau-K}^{\tau} b_i(x) \int_0^{\tau-x} W_i(y)\, dy\, dx - (2\tau-K)\, B_i(\tau) + (K-\tau) \int_{\tau}^{K} W_i(K-x)\, b_i(x)\, dx$$
$$- \int_{\tau}^{K} x\, W_i(K-x)\, b_i(x)\, dx + \int_{\tau}^{K} b_i(x) \int_{K-x}^{K-\tau} W_i(y)\, dy\, dx \Big] \Big\}$$
$$= 1 - \frac{1}{\tau} \text{ E[idle period of line i]} \tag{17}$$

4. Approximation for Erlang Distributed Data Input

We have obtained the integral equation for the probability distribution function of the buffer occupancy, and the equations for the probability distribution function of the amount of loss and for the utilization of the system; however, we cannot solve for them because the message size is generally distributed. In this section, we choose an Erlang distribution for the message size and attempt to solve the equations.

It is very difficult to find an analytical solution for the integral equations of (8) and (9). We have, however, obtained a polynomial approximation as follows:

$$W_i(y) = C_{i,0} + \sum_{n=1}^{N} C_{i,n}\, y^n \qquad \text{for } N = 3, 4, \text{ or } 5 \;, \tag{18}$$

where $C_{i,n}$ is the unknown coefficients to be determined.

The power series expansion for approximately solving the integral equation which we have used here normally proceeds as we try to match coefficients for the power of, in this case, y. Due to the complexity of the functions in the derivations, this approach is not feasible for our problem. We have instead adopted the least mean square error method to obtain the expression for the unknown coefficients $C_{i,n}$, n = 0, 1, ... , N. By this method, $\varepsilon_{L,i+1}$ is the mean square error for solving $W_i(y)$, where L is the number of points used to make a higher accurate operation. Certainly a good approximation can be achieved by using a large number for L. We find that it is sufficient to use the number 100 for L. In order to minimize the error, we take the derivative of $\varepsilon_{L,i+1}$ with respect to $C_{i+1,k}$ for k = 0, 1, ..., N, and set that derivative to zero. Lengthy manipulations give us an equation for the coefficients ($C_{i+1,0}, C_{i+1,1}, \ldots, C_{i+1,N}$). Note that we have the boundary condition to make N+1 linear independent equations to solve for the N+1 $C_{i,n}$'s for each line. After finding $W_i(y)$, then $L_i(z)$ and ρ_i can be analytically obtained. Note also that we have considered $W_i(y)$ for two cases; hence, this approximation should also be discussed in this manner.

5. Example

In this section we show through some results that the approximate solutions for $W_i(y)$ and $L_i(z)$ are very accurate, which is demonstrated by comparing the results obtained from the approximate analysis, a numerical method, and a simulation experiment. The case which we have used for comparison has an Erlang-2 distributed message size, three input lines, and the additional parameters $\mu = 1.1$, $T = 1$, and $K = 2$. Figures 8 to 15 show the results and the relative errors for $W_i(y)$ and $L_i(z)$, where i = 0 and 1. We omit the results for line 2 because they are very similar to the other lines. The results obtained from these three approaches are very close; therefore, we use a relative error to show how close the results obtained are from the approximation and the numerical method. The maximum relative errors are less than 0.2% and 0.11% for W(y) and L(z), respectively. We have found that if N = 4, then the approximate results are very close to the results obtained from the other two approaches. Clearly, increasing N increases the accuracy of the polynomial approximation; however, we can only use N = 3, 4, and 5 in this system because a large N increases the complexity of the iterative operation.

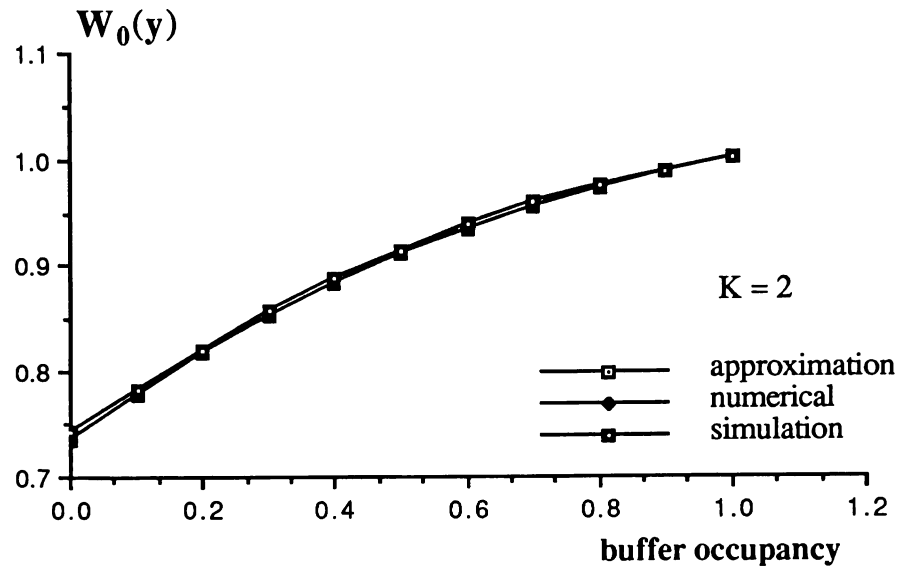

Figure 8 $W_0(y)$ obtained from the three approaches as K = 2.

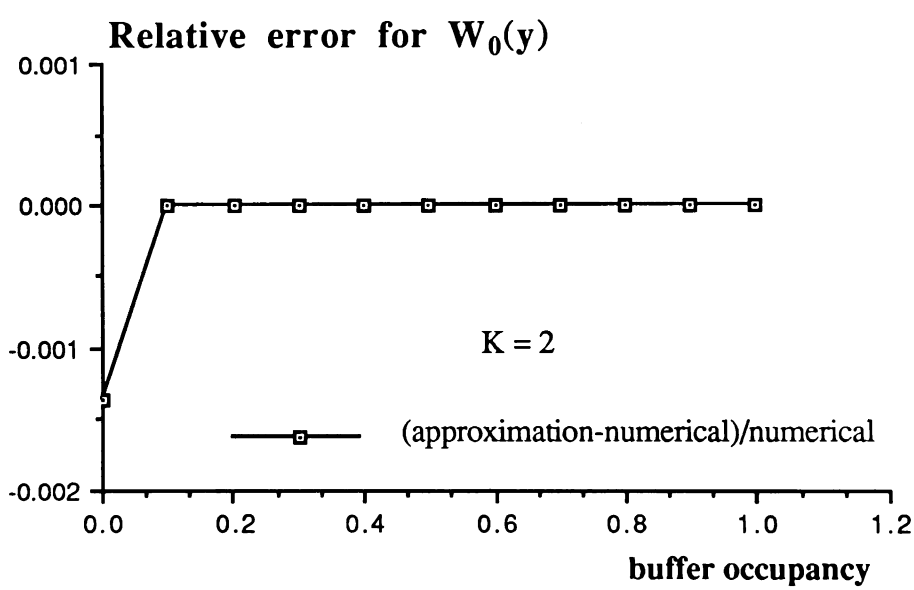

Figure 9 Relative error for $W_0(y)$ as K = 2.

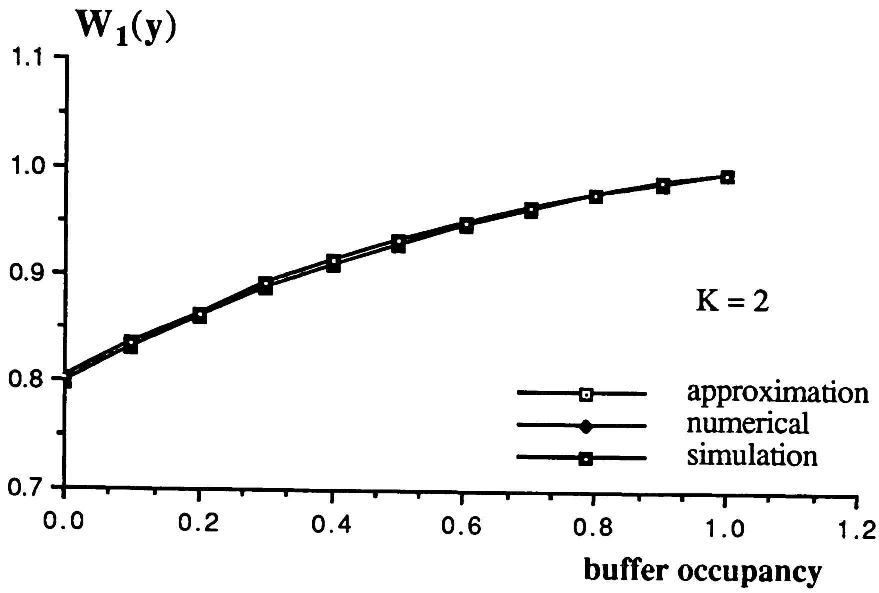

Figure 10 $W_1(y)$ obtained from the three approaches as K = 2.

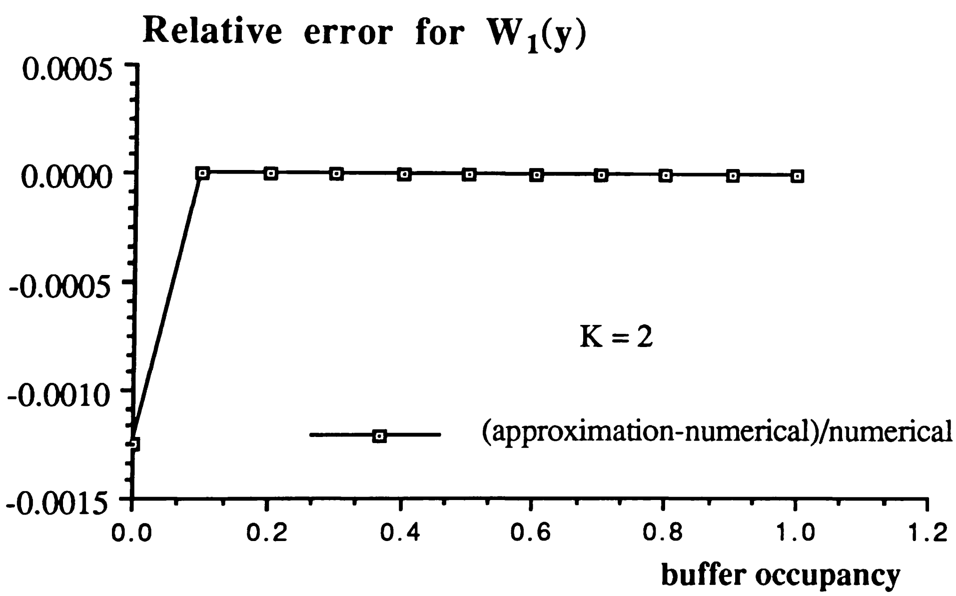

Figure 11 Relative error for $W_1(y)$ as K = 2.

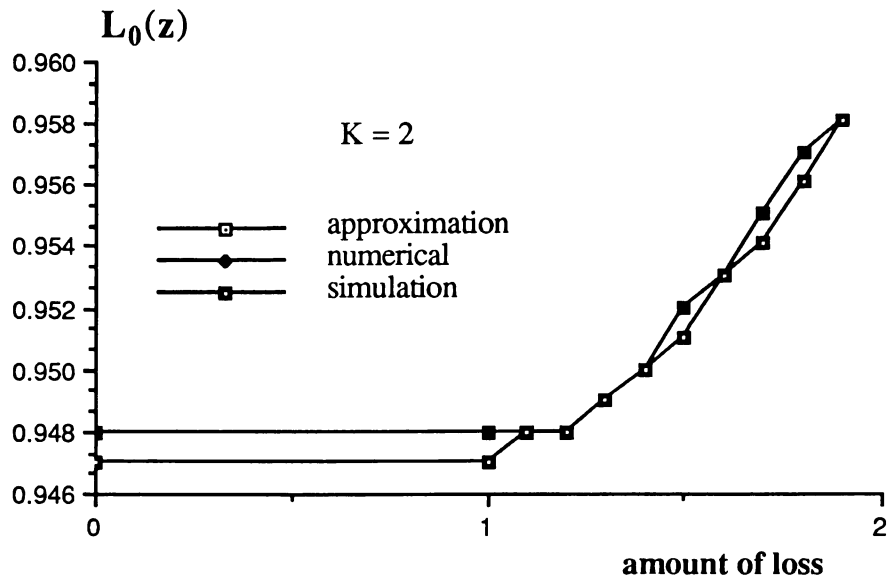

Figure 12 $L_0(z)$ obtained from the three approaches as K = 2.

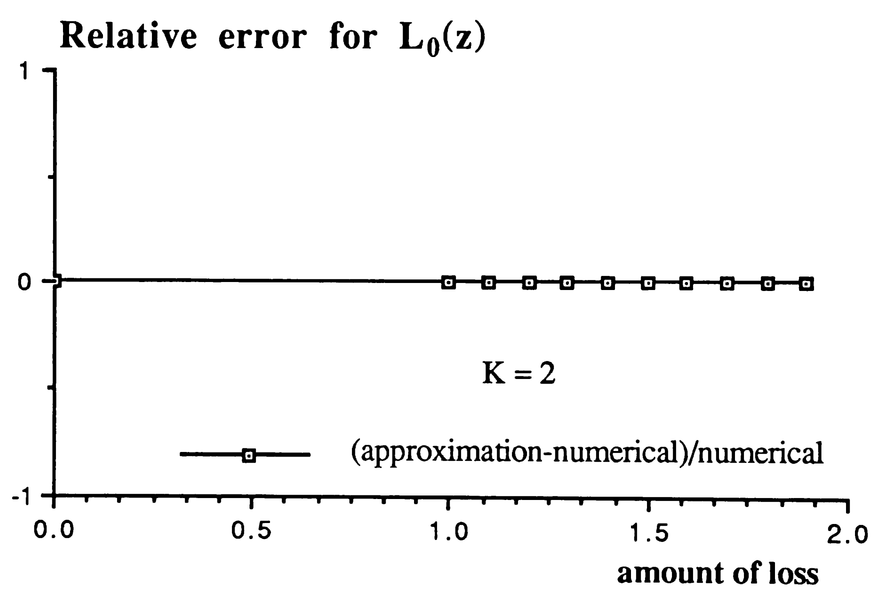

Figure 13 Relative error for $L_0(z)$ as K = 2.

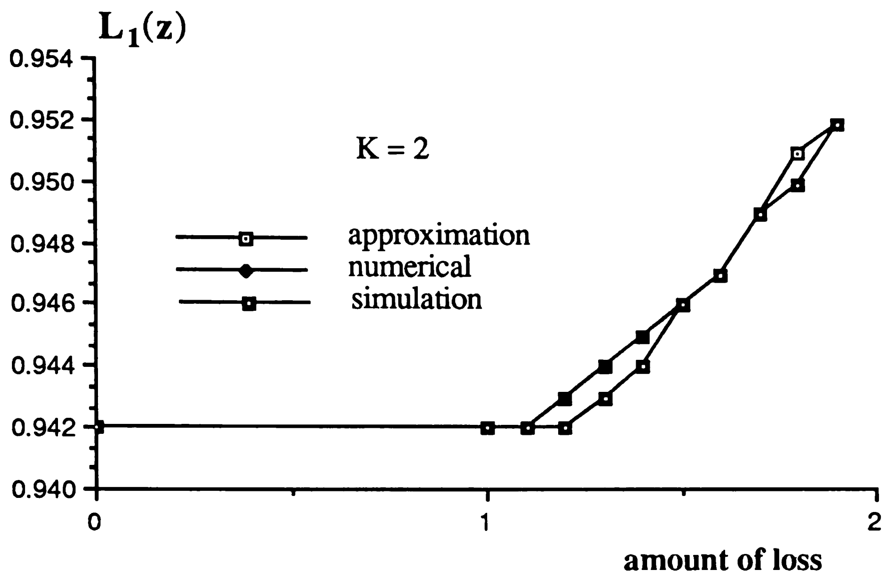

Figure 14 $L_1(z)$ obtained from the three approaches as K = 2.

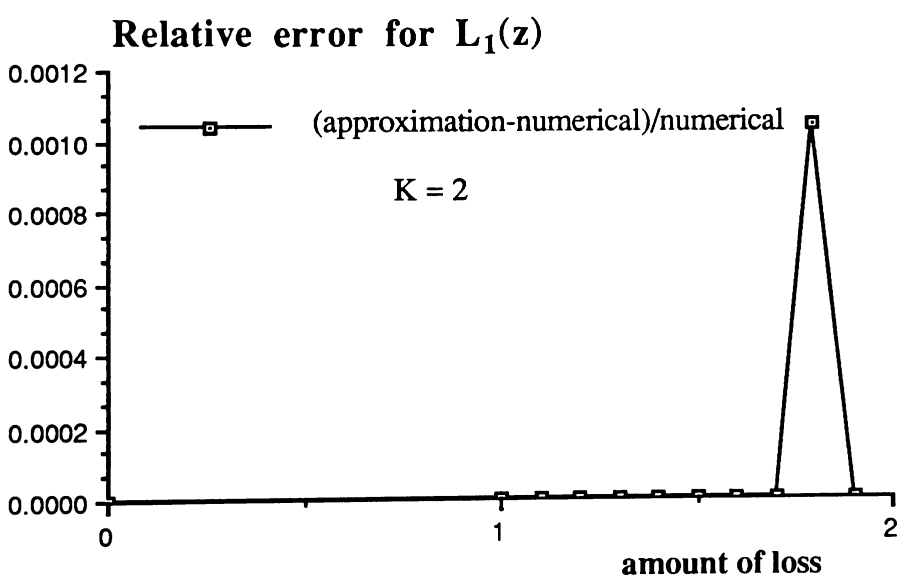

Figure 15 Relative error for $L_1(z)$ as K = 2.

6. Conclusion

In this paper we present solutions to the performance analysis of the finite buffer system with instantaneous multiple input. An integral equation is derived for the probability distribution function of the buffer occupancy at the beginning of an input message generation; equations are also derived for the utilization of the system and the probability distribution function of the amount of loss due to the overflow of the finite buffer. It is shown that a very accurate polynomial approximation does exist. The accuracy of this approximation is compared with the result obtained from a numerical solution of the integral equation, as well as the result obtained from the simulation experiment. The main advantages of the approximation are the accuracy and the reduction in computational complexity.