ON A FINITE BUFFER PROBLEM

Bih-Hwang Lee
and
Arne A. Nilsson

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

December 1988

CCSP-TR-88/30

1. INTRODUCTION

In data communication, there are three major components, the source system, the transmission medium, and the destination system. Generally, the source system contains an input device and a transmitter and the destination system contains an output device and a receiver (Figure 1). In this paper, we will focus only on the source system.

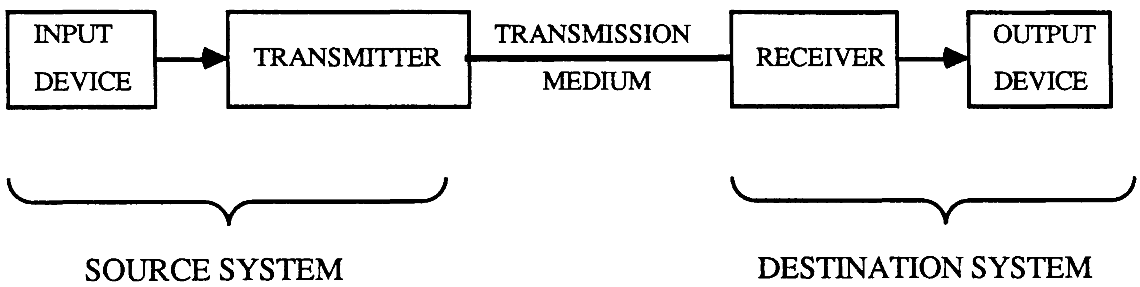

Figure 1 communication model

The input device can be any data generator or 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 from the coder. The data can immediately be transmitted by the transmitter if the transmission rate of the transmitter is greater than the rate of the coder. Unforturnately, the transmission rate of a transmitter is sometimes less than that of a coder, resulting in loss of data.

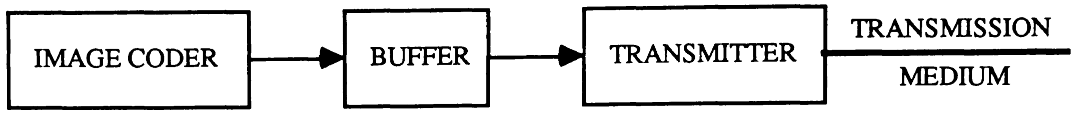

Figure 2 a buffer in the source system

In order to prevent the data loss, we need a buffer in which the excess input to the transmitter is temporarily stored (Figure 2). An interesting question is that of how large a buffer we need. Of course, there will be no loss if we have an infinite buffer, but this is very inefficient and also totally unrealistic. We also know that if the size of the buffer is small, the loss can be considerable, unless we adapt a smaller rate of a coder; however, this implies a degraded performance. Consequently, we have to face a trade-off problem, which is to decide the optimum buffer size to achieve the required performance.

In this paper, we assume that the transmission rate of the coder is much greater than that of the transmitter, so the transmitter receives the entire input data instantaneously. Thus, the data has to be stored in a buffer before being transmitted, otherwise, the loss of data is inevitable.

2. ANALYSIS

For a finite-capacity resource system, a queueing model is usually a good tool to describe the system (Figure 3). We assume that (1) this is a FCFS system, (2) the inter-arrrival time between each input is a constant time period, T, and (3) the amount of data entering the buffer at each input time is a generally distributed random variable whose probability distribution function is denoted by B(x). According to this, the queueing system can be said to be a D/G/1 with variations, ie., a D/G/1 queue with finite buffer.

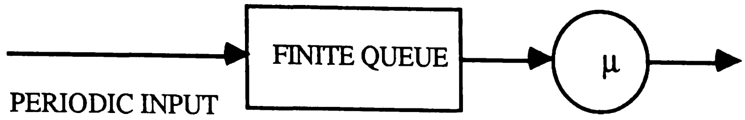

Figure 3 queueing model

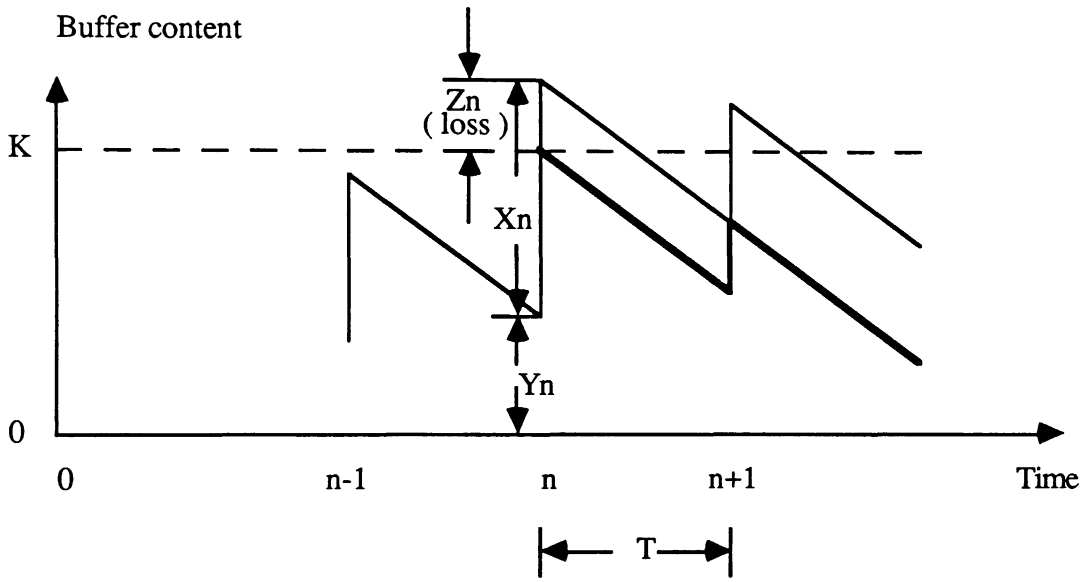

Figure 4 an embedded Markov chain description

The system can be analyzed by an embedded Markov chain (Figure 4). Let Y_n and X_n represent the buffer content, in bits, at time n and the message, in bits, entering the buffer at time n respectively. Any excess of the buffer's capacity, which is K bits, is lost. The amount of loss at time n is denoted by Z_n , where n = 0,1,2, ... For simplicity, let us assume that the transmission rate of the transmitter is one bit per second, we find

$$Y_{n+1} = [\, \min (Y_n + X_n , K) - T \,]^+ \; ; \tag{1}$$

$$Z_n = [\, Y_n + X_n - K \,]^+ . \qquad \text{where} \quad [\, x \,]^+ = \max [\, 0, x \,] . \tag{2}$$

Let $W_{n+1}(y)$ represent the probability distribution function of Y_{n+1}. Then, using some probability properties, we get $W_{n+1}(y)$ as follows.

$$
\begin{aligned}
W_{n+1}(y) &= \Pr\{ Y_{n+1} \le y \} \\
&= \Pr\{ [\min(Y_n + X_n, K) - T]^+ \le y \} \\
&= \Pr\{ \min(Y_n + X_n, K) - T \le y \} \\
&= \Pr\{ Y_n + X_n \le y+T < K \text{ or } K-T \le y \} \\
&= \Pr\{ Y_n + X_n \le y+T < K \} + \Pr\{ K-T \le y \} \\
&= \Pr\{ Y_n + X_n \le y+T \text{ and } y+T < K \} + \Pr\{ K-T \le y \} \\
&= \Pr\{ Y_n + X_n \le y+T \} * \Pr\{ y+T < K \} + \Pr\{ K-T \le y \} \\
&= \Pr\{ Y_n \le y+T-x \mid X_n = x \} * \Pr\{ y < K-T \} + \Pr\{ K-T \le y \} \\
&= \left\{ \int_0^{y+T} W_n(y+T-x)\, dB(x) \right\} * \Pr\{ y < K-T \} + \Pr\{ K-T \le y \}
\end{aligned}
$$

$$
= \begin{cases} \int_0^{y+T} W_n(y+T-x)\, dB(x) & 0 \le y < K-T \\ 1 & y \ge K-T \end{cases} \qquad (3)
$$

Let us represent the stationary distribution for Y_{n+1} by

$$
W(y) = \lim_{n \to \infty} W_{n+1}(y) \qquad (4)
$$

where we assume that the limit exists. W(y) will be the stationary distribution for the buffer content, and also this limiting distribution is independent of the initial state Y_0. Thus we have

$$
W(y) = \begin{cases} \int_0^{y+T} W(y+T-x)\, dB(x) & 0 \le y < K-T \\ 1 & y \ge K-T \end{cases} \qquad (5)
$$

Let $L_n(z)$ denote the probability distribution function of the amount of loss at time n. Using some probability properties, we have

$$
\begin{aligned}
L_n &= \Pr\{ Z_n \le z \} \\
&= \Pr\{ [Y_n + X_n - K]^+ \le z \} \\
&= \Pr\{ Y_n + X_n - K \le z \} \\
&= \int_0^\infty \Pr\{ Y_n \le z + K - x \mid X_n = x \} \; dB(x) \\
&= \int_0^\infty W_n(z+K-x) \; dB(x) \\
&= \int_0^{z+K} W_n(z+K-x) \; dB(x) \; . \qquad (6)
\end{aligned}
$$

Similarly, let us represent the stationary distribution for Z_n by

$$L(z) = \lim_{n \to \infty} L_n(z) \qquad (7)$$

and the limit must exist. L(z) will be the stationary distribution for the amount of loss, and this limiting distribution is independent of the initial state Z_0. Thus we have

$$L(z) = \int_0^{z+K} W(z+K-x) \; dB(x) \; . \qquad (8)$$

Until now, we have obtained W(y) and L(z) for the system. In the next section, we will select a specific probability distribution function for B(x) and find a solution to such a particular system.

3. SOLUTION FOR THE ERLANG DISTRIBUTED DATA INPUT

In many applications, like telephone traffic theory and data manipulation, the Erlang distributed random service time has been widely used. For such a service demand, we have obtained the solution to the integral equation of the probability distribution function of the buffer content. If the probability distribution function $B(x)$ and the probability density function $b(x)$ are given as :

$$B(x) = 1 - e^{-k\mu x} \sum_{i=0}^{k-1} \frac{(k\mu x)^i}{i\,!} \quad ; \tag{9}$$

$$b(x) = \frac{d}{dx} B(x) = \frac{k\mu\,(k\mu x)^{k-1}}{(k-1)\,!}\, e^{-k\mu x} \quad , \tag{10}$$

then, based on the structure of the probability density function, we guess a solution of the form to be :

$$W(y) = \sum_{i=0}^{k} \alpha_i\, e^{-\beta_i y} \qquad \text{where} \quad \beta_0 = 0\,. \tag{11}$$

Using the integral equation, we derive it as follows.

$$\begin{aligned} W(y) &= \int_0^{y+T} W(y+T-x)\; dB(x) \\ &= \int_0^{y+T} \sum_{i=0}^{k} \alpha_i\, e^{-\beta_i (y+T-x)}\, \frac{k\mu\,(k\mu x)^{k-1}}{(k-1)\,!}\, e^{-k\mu x}\, dx \\ &= \sum_{i=0}^{k} \alpha_i\, e^{-\beta_i (y+T)}\, (k\mu)^k \frac{1}{(k-1)\,!} \int_0^{y+T} x^{k-1}\, e^{-(k\mu-\beta_i)\,x}\, dx \\ &= \sum_{i=0}^{k} \alpha_i\, e^{-\beta_i y} e^{-\beta_i T} \frac{(k\mu)^k}{(k\mu-\beta_i)^k} - \sum_{i=0}^{k} \alpha_i\, e^{-\beta_i (y+T)}\, (k\mu)^k e^{-(k\mu-\beta_i)(y+T)} \sum_{n=1}^{k} \frac{(y+T)^{k-n}}{(k-n)!\,(k\mu-\beta_i)^n} \end{aligned}$$

$$= \sum_{i=0}^{k} \alpha_i \, e^{-\beta_i y} e^{-\beta_i T} \frac{(k\mu)^k}{(k\mu-\beta_i)^k} - \sum_{n=1}^{k} e^{-k\mu(y+T)} (k\mu)^k \frac{(y+T)^{k-n}}{(k-n)\,!} \sum_{i=0}^{k} \frac{\alpha_i}{(k\mu-\beta_i)^n}$$

$$= \sum_{i=0}^{k} \alpha_i \, e^{-\beta_i y} \qquad (12)$$

Fortunately, the solution is obvious if the following two equations hold.

$$\left(\frac{k\mu}{k\mu-\beta_i} \right)^k = e^{\beta_i T} \qquad (13)$$

and

$$\sum_{i=0}^{k} \frac{\alpha_i}{(\,k\mu-\beta_i\,)^n} = 0 \qquad n = 1, 2, \ldots, k \qquad (14)$$

Let us first consider the transcendental equation, which is

$$\left(\frac{k\mu}{k\mu-\beta_i} \right)^k = e^{\beta_i T}$$

This is a k^{th}-root equation and β_i can be found as below.

$$\frac{k\mu}{k\mu-\beta_i} = (\,e^{\beta_i T} \; e^{j2\pi i}\,)^{1/k} = e^{(\,\beta_i T + j2\pi i\,)/k} \qquad i = 1, 2, \ldots, k \qquad (15)$$

$$\beta_i = k\mu \,[\, 1 - e^{-(\,\beta_i T + j2\pi i\,)/k} \,] \qquad i = 1, 2, \ldots, k \qquad (16)$$

where $j = \sqrt{-1}$.

In order to obtain α_i , We need to solve k+1 linear equations of the form

$$\sum_{i=0}^{k} \frac{\alpha_i}{(k\mu-\beta_i)^n} = 0 \qquad n = 1, 2, \ldots, k \tag{17}$$

and make sure that the boundary condition $W(K-T) = 1$ is satisfied. Thus, the probability distribution function of the buffer content, W(y), has been obtained.

Now, the probability distribution function of loss $L(z)$ can easily be found by solving the integral equation below.

$$\begin{aligned} L(z) &= \int_0^{z+K} W(z+K-x)\, dB(x) \\ &= \int_0^{z+T} dB(x) + \int_{z+T}^{z+K} W(z+K-x)\, dB(x) \\ &= B(z+T) + \sum_{i=0}^{k} \alpha_i\, e^{-k\mu(Z+T)} \sum_{n=1}^{k} \frac{[k\mu(Z+T)]^{k-n}}{(k-n)!}\, e^{\beta_i [(n/k+1)T-K]+j2\pi ni/k} \end{aligned} \tag{18}$$

4. EXAMPLE

For specified parameters, the probability distribution function of buffer content, W(y), and of loss, L(z), have been found. For the parameters of k=2, μ= 1.1, and K=2 (bits), the curve of the probability distribution function of buffer content, W(y) , is shown in Figure 5 and the probability of no loss is 94% in this case. Furthermore, we know that if the buffer size is increased, the loss is decreased. If the buffer size is increased immensely, i.e., an infinite buffer system, the probability distribution function of buffer content has also been found, and of course, there is no loss. The curve of the probability distribution function of buffer content in this case is also shown in Figure 5.

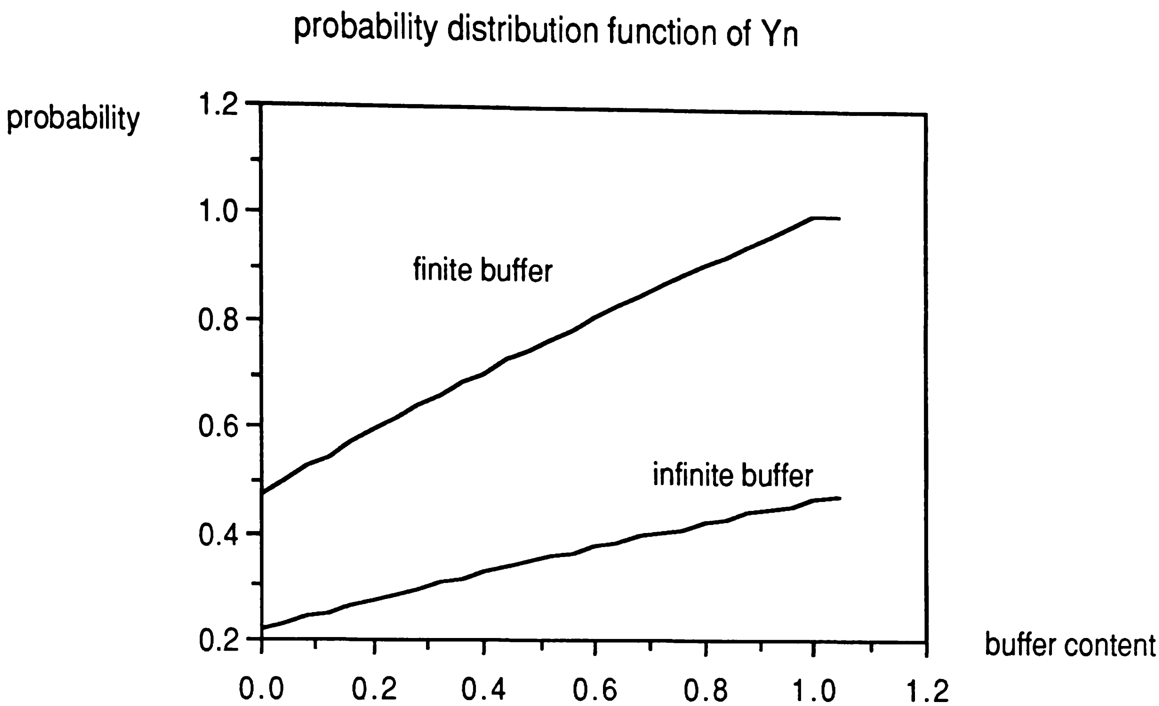

Figure 5 probability distribution functions for finite and infinite buffer

5. CONCLUSION

The buffer problem is an important factor in a queueing system. In other words, if the capacity of a system is finite, the loss has to be considered. Although a small buffer can minimize cost and a smaller transmission rate of a coder can reduce the loss of data, these characteristics will degenerate the performance of a system. A good performance can be obtained by using a bigger buffer, but the cost may be prohibitive. However, we can use the model derived in this paper to examine a system if a trade-off is acceptable.

6. REFERENCES

[1] L. Kleinrock, Queueing system Vol. 1 : theory, Wiley, 1976.

[2] D. J. Daley, "Single-server Queueing System with Uniformly Limited Queueing Time", J. Austral. Math. Society 4 (1964) pp. 489-505.

[3] J. W. Cohen, The Single server queue, North Holland publishing Co., Amsterdam, 1969.

[4] W. Stallings, Data and Computer Communications, Macmillan Publishing Co., 1985.