ABSTRACT

An analysis of the discrete-time queuing problem: binomial input, constant message length, infinite queue length, one server, is carried out. The state probabilities are calculated using the method of imbedded Markov chains, and the results are applied to the problem of determining characteristic features in the performance of traffic concentrators. The results are compared to the direct numerical solution of the equations of state for a finite buffer queue concentrator. The validity of the assumptions made for the distributions of the arrival process and message length, are discussed, and the use of simplified analytical models for solving complex queuing problems is analyzed.

1. INTRODUCTION

In the last years a growing interest in data communication networks based on message switching has been observed. This is primarily due to the fact that a message switching network will lead to:
- high utilization of the transmission lines
- no restriction on speed compatibility between different terminals or between terminals and high-speed transmission lines
- the terminals do not have to wait for a free channel in order to send messages.

One obvious disadvantage is that each message must be equipped with an address and certain administrative information, due to the asynchronous nature of the message transmission.

Traffic concentrators are vital elements in message-switched data communication systems. A concentrator receives messages from a finite number of sources (terminals or other concentrators), and transfers the data packets in order of their arrival, or following a preselected priority scheme, to a limited number of outgoing channels. To prevent message loss, the concentrator can be equipped with a buffer.

In recent years, many authors have been interested in the problem of dimensioning and analyzing traffic concentrators. The results so far achieved, clearly depend upon the assumptions made about the input process, the organization of the buffer queue and the service mechanism. To the author's knowledge, very little information of these properties in a data communication system is available, although some measurements on time-sharing systems have been analyzed [1].

Calculations of loss and delay in systems with one buffer queue have been carried out by Smith & Smith [2], for Poisson input and negative exponential holding time distribution. For Poisson input and constant message length, the buffer behaviour has been analyzed by Birdsall et al [3], Dor [4] and Chu [5],[6]. Maritsas and Martley [7] have studied Erlang input, and Chu [8] Batch Poisson arrivals, in the case of constant message length. In a paper published recently [9], Chu and Liang have analyzed buffer behaviour for mixed (Poisson and compound Poisson) input traffic.

The last three papers describe input processes which can take care of the bursty nature of the data traffic, as described by Fuchs and Jackson [1].

It is well known that both Poisson and Batch-Poisson input processes are based on the assumption of an infinite number of sources or terminals. When the number of terminals is small, Rudin [10] found it convenient to assume that the message arrivals in a specified time interval were binomially distributed. He made numerical calculations of loss probabilities and
average waiting times for multiple output channels.

In the following sections, the discrete-time queuing problem: binomial input, constant message length, one server and unlimited queue length, will be analyzed. Numerical results will be given when this model is applied to a traffic concentrator with finite buffer queue. A comparison is made with the direct numerical solution of the equations of state for the finite queue problem.

The model examined in this paper is based on asynchronous output of the buffer queue, as distinct from the works by the authors mentioned above. Synchronous output at fixed time marks separated by the constant service time, will always give conservative estimates of the buffer performance compared to asynchronous output.

2. MODEL DESCRIPTION

A traffic concentrator receives messages from N sources. The message length is constant, and the incoming transmission speed is $r_2$ (messages per time unit). The speed of the single output trunk is $r_1$, and to assure high utilization of this highspeed channel, it is assumed that $N\cdot r_1 > r_2$.

The buffer is able to store K messages simultaneously, but in the analytical study of this model it will be assumed that $K = \infty$. The messages will be served in order of their arrival.

The time required to transmit one message of constant length from the concentrator is

$$t_0 = \frac{1}{r_2}$$ (1)

The maximum number of message arrivals in $t_0$ is

$$n = \frac{Nt_0}{r_2}$$ (2)

With binomial input, the probability of j arrivals in time $t_0$ is

$$k_j = \binom{n}{j} (1-\beta)^{n-j} j, j = 0, 1, 2, \ldots, n$$

$$k_j = 0, j < 0, j > n$$ (3)

The arrival process, as described by equation (3), is strictly correct only when messages are received in the following way:

- the buffer accepts messages at discrete time marks separated by time intervals
- not more than one message may arrive at each time mark
- the arrival of a message is an event which is statistically independent of previous arrivals. The probability of message arrival at a given time is therefore a constant, $\beta$.

The mean number of arrivals during $t_0$, or the traffic offered, is

$$\rho = E(k_j) = n\beta$$

3. EQUATIONS OF STATE

The probability of j messages waiting in the buffer queue at time $t$ is $Q_j(t)$. The stochastic process $(Q_j(t))$ is non-Markovian, but a Markov chain can be constructed using the imbedding technique [11].

The state variables $(Q_j(t))$ are measured immediately after a message transmission,

$$(t_k+0), \text{ and the sequence } \{Q_j\}, j = 0, 1, 2, \ldots$$

where $Q_j = Q_j(t_k)$, will then form a Markov chain.

In statistical equilibrium, the equations of state can be written:

$$Q_j = \sum_{i=0}^{K-1} Q_i P_{ij}, j = 0, 1, 2, \ldots, K-1$$ (4)

Immediately after a message from the queue has been transmitted (the Markovian regeneration point), $Q_K = 0$. This is due to the fact that an incoming message does not have access to a specific place in the queue before the transmission is finished.

Introducing the fictitious probabilities $Q_K, Q_{K+1}, \ldots$, the equations of state will be:

$$Q_j = \sum_{i=0}^{+\infty} Q_i P_{ij}, j = 0, 1, 2, \ldots$$ (5)

Following Meisling [11], the transition probabilities are:

$$P_{ij} = k_{j-i+1} \quad i \geq 1$$

$$P_{oj} = kj \quad i = 0$$

and

$$Q_j = Q_0 k_j + \sum_{i=1}^{j-1} Q_i k_{j-i+1}, j = 0, 1, 2, \ldots$$ (6)

The normalizing condition is as usual:

$$\sum_{j=0}^{+\infty} Q_j = 1$$

4. OVERFLOW PROBABILITIES

At a regeneration point the queue length will be $j = K+i$ with probability $Q_{K+i}$. This is the state probability immediately after the completion of a message transmission, and the queue length just before this completion was $j = K+i$. For all $i \geq 0$ this implies that arriving messages would have been rejected if the queue had a finite length $K$.

The probability of a message being rejected may therefore be defined as

$$P_r = \sum_{j=K}^{+\infty} Q_j - \sum_{j=0}^{K-1} Q_j, K \geq 1$$ (7)

It is assumed that each message arrival must be completed before the transmission can begin, and consequently the model is not valid for $K = 0$.

An analytical expression for $P_r$ can be established using the generating functions:

$$\psi(z) = \sum_{j=0}^{+\infty} Q_j z^j \quad |z| \leq 1$$

$$K(z) = \sum_{j=0}^{+\infty} k_j z^j \quad |z| \leq 1$$

The following expression can be calculated, using equation (6) (see reference [11])

$$\psi(z) = Q_0 \frac{1-\beta z}{K(z)-z}$$ (8)

where

$$Q_0 = 1 - n\beta$$ (9)

Applying equation (3) one obtains:

$$K(z) = (\beta z + (1-\beta))^n$$ (10)
Generally, the following relationship exists between $\psi(z)$ and $\sum_{j=0}^{\infty} Q_j$ (see ref. [12], page 103)

$$\psi(z) = \sum_{k=0}^{\infty} z^k \sum_{j=0}^{k} Q_j$$

and consequently (using equation (8))

$$\sum_{k=0}^{\infty} z^k \sum_{j=0}^{k} Q_j = (1-n\beta) \frac{K(z)}{K(z) - z}$$

Provided $\left| \frac{z}{K(z)} \right| < 1$, this expression can be transformed into:

$$\sum_{k=0}^{\infty} z^k \sum_{j=0}^{k} Q_j - (1-n\beta) \frac{K(z)}{K(z) - z} = 0$$

The restriction $\left| \frac{z}{K(z)} \right| < 1$, will be analyzed in Appendix A.

The last term in equation (12) can, observing equation (10), be written

$$\sum_{k=0}^{\infty} z^k \sum_{j=0}^{k} Q_j - (1-n\beta) \frac{K(z)}{K(z) - z} = 0$$

With the abbreviation

$$A_{i,nk} = (-1)^i (nk+1-i) \left( \frac{\beta}{1-\beta} \right)^i$$

$$A_{0,0} = 1$$

$$A_{1,0} = 0$$

equation (12) will read

$$\sum_{k=0}^{\infty} z^k \sum_{j=0}^{k} Q_j - (1-n\beta) \sum_{i=0}^{\infty} A_{i,nk} z^i = 0$$

By forming the appropriate sum in the last term one obtains

$$\sum_{k=0}^{\infty} z^k \sum_{j=0}^{k} Q_j - (1-n\beta) \sum_{i=0}^{\infty} A_{i,nk} z^i = 0$$

$$Q_0 = (1-n\beta)$$

With $k = K-1$ in equation (16), equation (7) yields:

$$P_r = 1 - (1-n\beta) \sum_{i=0}^{K-2} \frac{A_{i,nk}(i+1)n}{(1-\beta)(K-1)-i}$$

or, observing $A_{i,0,nk} = 1$:

$$Q_k = \frac{1-(n\beta)}{(1-\beta)^{i+1}} \sum_{i=0}^{k-1} A_{i,0,nk}(i+1)n$$

$$k \geq 2$$

(18)

$Q_1$ can be calculated from equation (15):

$$Q_1 = \frac{1-n\beta}{(1-\beta)^{n-1}}$$

$$Q_1 = \frac{1-n\beta}{(1-\beta)^n}$$

The limiting process $n \to \infty$, $\beta \to 0$ will lead to results identical with the Poisson input case, as shown in Appendix B.

The mean queue length can be found directly from equations (8) and (9):

$$E(k) = \sum_{k=0}^{\infty} kQ_k = \frac{d\psi(z)}{dz} \Bigg|_{z=1} = \psi'(1)$$

Here $\psi'(1)$ is an expression of the form $Q_j$, and both numerator and denominator are evaluated in Taylor series about $z = 1$, which yields:

$$\psi'(1) = n\beta + \frac{K''(1)}{2(1-n\beta)}$$

(19)

or

$$E(k) = n\beta + \frac{n(n-1)\beta^2}{2(1-n\beta)}$$

(20)

6. MEAN WAITING TIME

The expression for the mean queue length, equation (20), will be approximately correct in the finite buffer queue case for small values of the overflow probability. From this expression, the mean waiting time can be obtained according to reference [10]. The mean number of arrivals during the mean waiting time, $E(w)$, plus the service time, $t_0$, must be equal to the mean queue length, $E(k)$:

$$(E(w) + t_0)a = E(k)$$

where $a$ = the mean number of arrivals per unit time

From chapter 2 it is known that the probability of a message arrival at given time marks, separated by $\Delta t = t_0/n$ is $\beta$. Thus:

$$a = \frac{\beta}{\Delta t} = n\beta$$

(21)

In terms of service time units, the mean waiting time can be written

$$E(w) = \frac{E(k)}{t_0} - 1 = \frac{E(k)}{\Delta t} - 1$$

(22)

7. NUMERICAL RESULTS

Numerical results have been calculated for a variety of the input parameters $p = n\beta$, $n$ and $K$. This is compared to the recursive solution of the equations of state in the case of finite buffer queue. The method is outlined as follows: The equations of state (equation (4)), can be written:

$$P_r = 1 - (1-n\beta) \sum_{i=0}^{K-2} \frac{A_{i,nk}(i+1)n}{(1-\beta)(K-1)-i}$$

or, using equation (20):

$$E(w) = \frac{(n-1)\beta}{(1-n\beta)^2}$$
Figure 1: Probability of message rejection as a function of buffer queue length

Figure 2: State probability as a function of place in buffer queue

Figure 3: Mean queue length as a function of traffic offered

Figure 4: Relative mean waiting time as a function of traffic offered
Probability of message rejection as a function of the maximum number of arrivals in a service interval

$P_r = (1-k_1)Q_{K-1} - Q_o k_{K-1} \sum_{i=1}^{K-2} Q_i k_i \cdot K \geq 2$ (25)

For $K=1$

$Q_o = 1$

$Q_0 = k_o$

and

$P_r = 1 - k_o$ (26)

In figure 1, $P_r$ is expressed as a function of the queue length $K$. It can be seen that the probability of message rejection becomes small even for moderate values of $K$. The agreement between the two methods of solution is good for small values of the offered traffic, $p = n\beta$, but the deviation is essential for $p > 1$ and for small values of $K$.

This is mainly due to the disagreement between the state probabilities for small $K$. Figure 2 shows this effect most clearly for $K=3$ and $p = 0.8$.

Figure 3 and 4, which give the mean queue length and the mean waiting time, confirm that the analytical solution does not give correct answers for a small number of queue places, and that the discrepancies increase for increasing values of $\rho$.

Generally, a uniform traffic pattern, with low variance ("smooth traffic"), leads to lower values of $P_r$ than random (Poisson) traffic.

From the same diagram (figure 5), the effect of variation in the speed ratio $r_1/r_2$, or the absolute values $r_1$ and $r_2$, can be extracted.

8. DISCUSSION

The model examined in this paper describes a queuing system without losses. The technique applied when solving the equations of state is well known, but in this case the calculation of the parameters $\{Q_k\}$ and $P_r$ was not strictly straightforward. The analytical expressions for $Q_k$ and $P_r$ are mostly of theoretical interest, as long as the recursive numerical solution of the equations of state is easily carried out in the case of finite buffer queue.

A comparison between these two approaches shows that the infinite buffer model describes the behaviour of the concentrator adequately in most cases.

This indicates that other infinite queue models can be used to examine practical finite queue problems, which, in many cases, simplifies the problem considerably. Merker's theory compared to Erlang's theory in telephone traffic theory, represents a similar way of approach for loss systems.

But it has to be stressed that only limited conclusions should be made from such simplified models.

A vital question is the following: to what extent can this or similar models predict the traffic flow through a concentrator in a data communication system? This clearly depends upon...
the reliability of the assumptions made for the
input process, the organization of the buffer
queue and the service mechanism.

In a theoretical analysis, the picture of the concentrator must necessarily be simple, and
some assumptions with regard to the input
traffic must be adopted. The complete traffic
pattern originating from the various needs of
the users and computers forming the data network
can not be included in a theoretical analysis.

The binomial input process takes care of
- the relative speed difference between
  in- and outgoing
  circuits
- variation in the number of terminals
- the fact that the messages have a
  finite minimum length.

On the other hand this input process will result
in a smoother traffic than can be expected from
experimental results [1]. Qualitatively, a
bursty traffic pattern will lead to higher
rejection probabilities than smooth traffic.

An analysis of the influence of variations in
the message length distribution is also of
great interest. As long as a certain minimum
length is required to transmit address and
administrative information, an exponential
distribution of the message length can not be
correct. However, the assumption of constant
message length can clearly be just as incorrect
in many cases.

Regardless of the detailed specification of the
concentrator and the input traffic, certain
qualitative features of the concentrator
performance can be described by this and
similar models.

Furthermore, these models describe the influence
of different parameters on the performance of
the concentrator. This information can be of
great importance in technical-economical
optimizations of data communication networks.

9. CONCLUSION

The main objectives of this paper is
- the analytical examination of the
  discrete-time queuing problem:
  binomial input, constant message
  length unlimited queue, one server
  - the application of this model to a
  specific finite queue problem.

Analytical expressions are obtained for the
parameters of main interest, and numerical
results are calculated.

This investigation shows that a simplified
analytical model can give a satisfactory
picture of the performance of a rather complex
queuing system.

However, the main problem in the dimensioning
of data communication systems, is the lack of
relevant measurements. Increased effort in
this direction is necessary if theoretical
studies shall give reliable contributions to the
design of data networks.

APPENDIX A

In chapter 4, the restriction \(|\frac{z}{K(z)}| < 1\) was a
necessary condition for the development of
equation (11).

From equation (10) it follows that

\[ K'(1) = p < 1 \]  

This means (see for example Prabhup [13], p 63)
that \( z = K(z) \) has no solution \( \xi \) for \( 0 < \xi < 1 \).

Therefore \( \frac{z}{K(z)} < 1 \) or \( \frac{z}{K(z)} > 1 \) for \( 0 < z < 1 \).

Introducing the function

\[ f(z) = \frac{K(z)}{z} = \frac{k_0}{z} + \frac{k_1}{z^2} + \frac{k_2}{z^2} + \cdots \]  

it can be seen that

\[ f(0^+) = \infty \]  
\[ f(1) = 1 \]  
\[ f'(0^+) = \infty \]  
\[ f'(1) = p - 1 < 0 \]  

The derivate of \( f(z) \) is a monotonously
increasing function in the domain \( 0 < z < 1 \),
and from equations (5) and (6) it follows that
\( f'(z) < 0 \) in this domain.

\[ f(z) > 1 \]  
for \( 0 < z < 1 \)

or \( \frac{z}{K(z)} < 1 \)  
for \( 0 < z < 1 \)

APPENDIX B

The state probabilities \( Q_k \) (equation (18)) must,
in the limit \( \beta \rightarrow 0 \) (and \( n\beta \rightarrow \beta \)), be identical with
the result for the Poisson arrival process.

The parameter \( A_{nk} \) will then be transformed as follows:

\[ \lim_{n \rightarrow \infty} A_{nk} = \lim_{n \rightarrow \infty} \left( \frac{n}{1-n} \right)^i \left( \frac{\beta}{1-\beta} \right)^i \]  

\[ n\beta \rightarrow \beta \]  

Using

\[ \lim_{m \rightarrow \infty} \frac{m^i}{(m-r)!m^i} = 1 \]  

one obtains

\[ \lim_{n \rightarrow \infty} A_{nk} = \lim_{n \rightarrow \infty} \left( \frac{1}{1-n} \right)^i \left( \frac{\beta}{1-\beta} \right)^i \]  

Introducing

\[ \beta = \frac{0}{n} \]  

\[ \lim_{n \rightarrow \infty} A_{nk} = (-1)^i \left( \frac{k\beta}{1-\beta} \right)^i \]  

Further \( \lim_{m \rightarrow \infty} (1-x)^m = e^{-x} \)
leads to

\[ \lim_{n \rightarrow \infty} (1-n)^{nk} = \lim_{n \rightarrow \infty} (1-\frac{0}{n})^{nk} = e^{-\frac{nk}{n}} \]  

Applying equations (4) and (6) to equation (18),
the queue length will be:
or, using $j = k - i$

\[ Q_k^p = (1-p) \left\{ e^{k_0 + \sum_{i=1}^{k-1} (-1)^i (k-i)p \frac{k(i-1)^i}{(k-1)^i} + \frac{(k-1)p^{i-1}}{(i-1)!}} \right\} \quad (7) \]

\[ Q_k^p = (1-p) \left\{ e^{k_0 + \sum_{j=1}^{k-j-1} (-1)^j k-j-1p \frac{(j-1)!}{(k-j-1)!}} \right\} \quad (8) \]

This expression is identical with known results from queuing theory, see for example Saaty [14], p 155.

**REFERENCES**

[1] E.Fuchs, P.E.Jackson

"Estimates of Distributions of Random Variables for Certain Computer Communications Traffic Models".


"Loss and Delay in Telephone Call Queuing Systems".


"Analysis of Asynchronous Time Multiplexing of Speech Sources".


[4] N.M.Dor

"Guide to the length of buffer storage required for random (Poisson) input and constant output rates".


"A study of asynchronous time division multiplexing for time-sharing computer systems".


AFIPS Press, Montvale N.J.


"Buffer Behaviour for Poisson Arrivals and Multiple Synchronous Constant Outputs".


"Buffer Length for Erlang Input and Constant Removal Rate".


"Buffer Behaviour for Batch Poisson Arrivals and Single Constant Output".


"Buffer Behaviour for Mixed Input Traffic and Single Constant Output".


"Performance of Simple Multiplexer-concentrators for Data Communication".


"Discrete-time queuing theory".


"Introduction to Congestion Theory in Telephone Systems".

Oliver & Boyd, 1960.


"Queues and Inventories".


[14] T.L.Saaty

"Elements of Queuing Theory".