THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 50, No. 9, November, 1971
Printed in U.S.A.

Traffic Analysis of a Ring Switched Data Transmission System

By J. F. HAYES and D. N. SHERMAN

(Manuscript received April 12, 1971)

This paper is concerned with a study of traffic and message delay in a ring switched data transmission system. The system, by asynchronous multiplexing and data storage, shares transmission facilities among many users. It is the random component of message delay due to buffering that is the focal point of our study.

The basic configuration of the system is a ring connecting stations where traffic enters or leaves the system. A mathematical model of the ring is developed which accommodates an arbitrary number of stations and any given pattern of traffic between stations. Studied in detail is the uniform traffic pattern in which each user is identical and communicates equally to all others. Intrinsic to the model is a recognition of the bursty nature of data sources. Other factors that are taken into account are line and source rates as well as the blocking of data into fixed size packets. Formulas are derived from which average message delay induced by traffic in the ring can be calculated.

The results of the study are presented in a set of curves where normalized delay due to traffic within specific system configurations is plotted as a function of the number of stations and source activity. The delay here is normalized to average message lengths. An important parameter of these curves is the ratio of source rate to line rate. The results show that, in certain quite reasonable circumstances, the delay is less than two average message lengths.

Rings of 10, 50, and 100 users have been simulated on a digital computer. Data obtained from these investigations are presented and compared to the theoretical estimates for line busy and idle periods and message delay. The results of simulated average message delay show that, for interstation link utilizations of about 60 percent, the difference between the theoretical estimates and experimental observations is small.

I. INTRODUCTION AND BACKGROUND

In a recent paper[1] J. R. Pierce has proposed a data communication network in which users are connected in a ring or loop topology. In this paper we study the behavior of this network. We examine the relationship between source and line utilizations and the message delay within the system. We propose mathematical models of station behavior and, by making use of reasonable input data traffic models, predict the average delay as a function of network parameters. A principal result of our study shows that, in many cases of interest, average delay in storage is less than two message lengths.

The network is *buffered*, operates on a *distributed control* philosophy, and user entry is gained by *asynchronous* multiplexing into the line bit stream. Other systems have used the ring topology. For example, IBM offers a synchronous, nonbuffered system[2] in which a central controller monitors the loop and allows access by the users. Buffered, centrally controlled systems have been proposed by W. D. Farmer and E. E. Newhall[3] and A. G. Fraser.[4] In the first a computer controls several peripheral devices, while in the second, several computers are interconnected.

The characteristics of data sources and the requirements of users make the technique of message switching applicable to data communications. One important characteristic is that data sources are often bursty; i.e., relatively short sequences of bits followed by long pauses. A basic need of many data customers is rapid response to data bursts. One way to meet this need is by devoting a line to a source for a long period of time. However, because of the long pauses between data bursts the line will be underutilized. On the other hand, dropping the line between data bursts may be inefficient because of long setup times. For the line switching techniques that are currently used, the setup time will often be longer than the holding time.

Message switching provides a way of efficiently using the line and obtaining rapid response. The idea is to asynchronously multiplex several sources onto the same line. For example, in the present system, data from each source are formed into fixed size packets and supplied with a header. The header contains source and destination addresses as well as bookkeeping information. Packets from all sources are buffered and fed onto the line according to some scheduling algorithm. (One such algorithm will be seen presently when the detailed operation of the ring is described.) Clearly as more packets seek access to the same line, more storage is required. The storage requirements and the attendant buffering delay are important characteristics of system operation. It is precisely these characteristics that are the focal point of our study.

A sketch of the basic ring system is shown in Fig. 1. As indicated, traffic flows in one direction around the ring from station to station. The stations are indicated as B-boxes in Fig. 1. For purposes of explanation let us begin with the operation of the B-box before considering the other components of the loop (A-box and C-box). The data source is connected to the B-box where its output is formed into fixed size blocks or packets and supplied with header. A message from a source may consist of several packets. If the line is free a packet is multiplexed on the line immediately. The packet is then passed from B-box to B-box to its destination. At each B-box on its itinerary the address of a packet is examined to determine whether the packet's destination is at that particular B-box. This examination entails a fixed delay for each B-box which can be calculated given the source and destination. It is, however, the random delay encountered by the last bit of a message before it gets on the loop that commands our

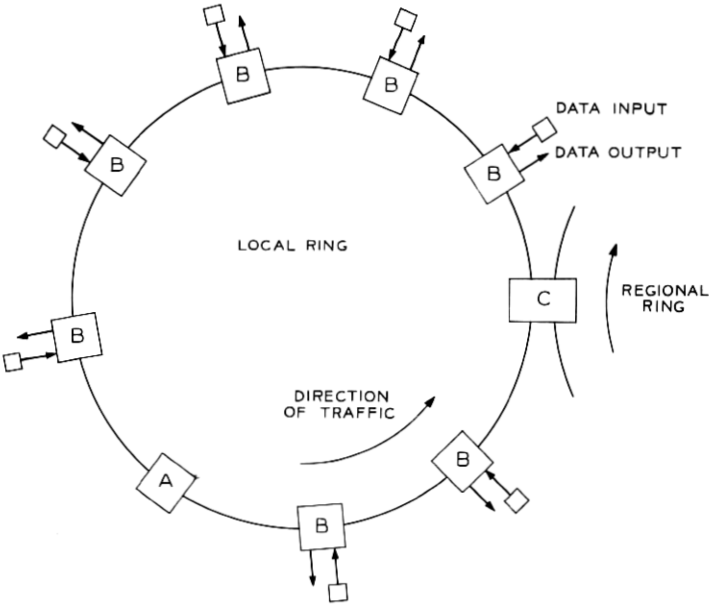

Fig. 1—Pierce ring.

attention. A fundamental property of the system is that traffic presently in transmission has priority over traffic seeking entrance to the ring. If the line at a B-station is busy with information packets passing through it, the packets produced at that station are buffered until the line is free. The reading onto the line of a message consisting of more than one packet will be interrupted when packets from another station pass through. The reading of the message is resumed from the point of interruption when the line is free again.

In order to explain the mechanism of multiplexing packets on and off the line, it is helpful to draw an analogy between the ring and a conveyer belt (see Fig. 2 for an illustration). Time slots into which packets may be placed circulate around the loop. The A-box insures that synchronism is maintained (see below). At the beginning of each time slot is a marker indicating whether the ensuring packet slot is empty or full. The B-box senses this marker and acts accordingly. In a full packet, address bits follow the occupancy marker. If a B-box senses its own address the packet is removed from the line and sent to its destination. This same B-box may take advantage of the empty slot to feed its own packet on the line. Of course if a packet slot is full and destined for a source beyond a particular B-box, then the line is momentarily blocked for that B-box.

While there are many B-boxes in a ring, there is only one A-box. The A-box has two basic functions. The first, as mentioned earlier,

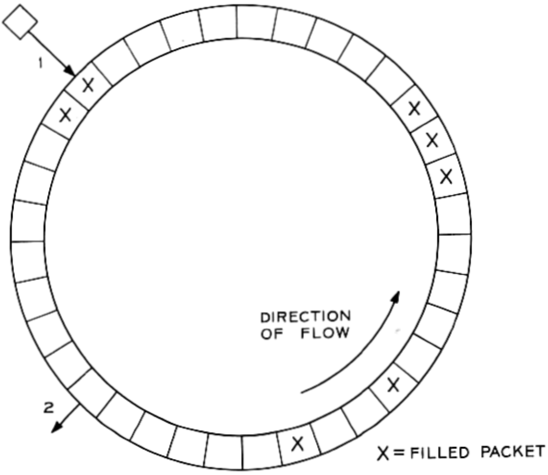

Fig. 2—Conveyor belt. Packets enter at 1 and leave at 2.

is synchronization of the ring. The second function is that of preventing the buildup of traffic in the ring due to undeliverable packets. The header of each packet passing through an A-box is marked. If a packet tries to pass through an A-box a second time it is either destroyed, creating an empty packet slot, or sent back to its destination. Sending a packet back to its destination is done simply by interchanging source and destination addresses. In this way a busy signal is provided.

The C-box shown in Fig. 1 provides interconnection of rings (see Fig. 3). Packets destined for a station outside a particular ring have addresses indicating this and are picked off by the C-box in exactly the same way that intraring traffic is picked off by B-boxes. This traffic is buffered and multiplexed onto the next ring in the same way that traffic from a local station is multiplexed on a ring. Since traffic already on a loop has priority, inter-ring traffic will suffer some delay. A likely realization of the C-box suggested by W. J. Kropfl[5] is the tandem connection of buffer and B-boxes shown in Fig. 4.

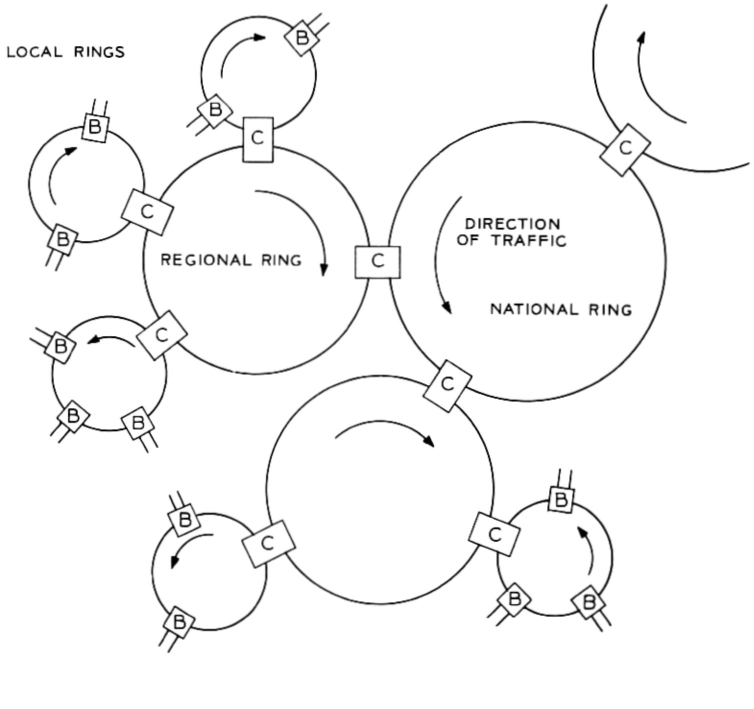

Fig. 3—Three-stage hierarchy of Pierce ring.

In the remainder of this paper a mathematical model of the ring will be developed and analyzed. The results of this study will be presented in the form of sets of curves which illustrate the behavior of normalized delay as a function of either the number of stations in the ring or the utilization of the source.

In order to carry this analysis forward it was necessary to make several assumptions and approximations. Thus in order to verify our results and to refine our model, a simulation program was developed. Simulation results have been obtained for rings ranging from 10 stations to 100 stations. These results are compared to the results of analysis.

II. GLOSSARY OF TERMS

C_b—Line rate in bits per second.

C_p—Line rate in packets per second.

b_i—Bit rate during the activity period of the ith source.

B_p—Number of information bits per packet.

H—Number of header bits per packet.

$1/\lambda_i$—Average duration of idle period of ith source in seconds.

$1/\mu_i$—Average duration of active period of ith source in seconds.

Q_i—Average number of packets per message at ith source.

γ_i—Ratio of source packet rate to the line packet rate.

u_i—Utilization of source i [see equation (3)].

θ_i—Intensity of source i [see equation (4)].

r_i—Average number of packets per second from source i.

N—Number of stations in ring.

P_{ij}—Portion of traffic from station i to station j.

R_k^*—Traffic passing through station k in packets per second.

R_k—Traffic out of station k in packets per second.

U_k^*—Line utilization as seen by station k [see equation (9a)].

Θ_k^*—Line intensity as seen by station k [see equation (10)].

U_k—Line utilization after station k [see equation (9b)].

Θ_k—Line intensity after station k [see equation (10b)].

$1/\Lambda_k^*$—Average duration of line idle period in seconds as seen by station k.

$1/M_k^*$—Average duration of line busy period in seconds as seen by station k.

III. SYSTEM MODEL

3.1 *Source Input Model*

In this section a mathematical model for a ring is presented thereby

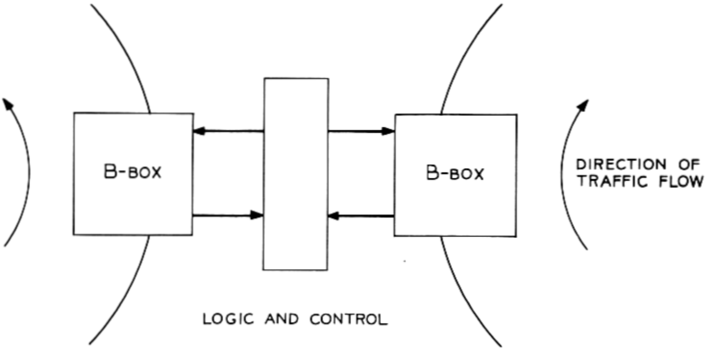

Fig. 4—Tandem connection of B-boxes for ring switching.

laying a foundation for succeeding sections where the model is analyzed. We are primarily interested in the traffic characteristics of the ring and the attendant delay. In order to focus on this aspect of ring operation, we assume that there are no equipment failures or transmission errors so that, once on the ring, a packet is ultimately delivered to its destination. Thus the action of the A-box in destroying packets or providing a busy signal is not part of the model. At present the study is confined to an analysis of a single ring. Finally the study was predicated on light to moderate loading of the system. In a heavily loaded system long queues of messages form and the response time of the system may be excessive for data applications.

Throughout the analysis the parameter N designates the number of B-boxes or stations that are on the ring. The capacity of the line connecting the stations to one another is designated as C_b bits/second.

To each B-box is connected a source, which as we have noted above, is bursty in nature. The output of the ith data source is modeled as consisting of alternate idle and active periods. During the latter, transmission is at a constant rate of b_i bits/second. The durations of the source activity and idle periods are assumed to be exponentially distributed and statistically independent of one another. Successive idle and busy periods are also independent of one another. The average durations in seconds of the active and idle periods of the source connected to the ith station in the loop are denoted as $1/\mu_i$ and $1/\lambda_i$ respectively. Studies of computer user statistics[6,7] indicate that the foregoing is a reasonable, if somewhat simplified, model of a data source.†

† An initial study of this model of the data stream is due to R. J. Pilc.[8]

Each message (bits generated in a source activity period) is bundled into an integral number of fixed size packets. The maximum number of information bits in each packet is B_p . The number of header bits that accompany each packet is denoted by the parameter H. In Appendix A it is shown that the number of packets in a message is geometrically distributed with mean

$$Q_i = \frac{1}{1 - \exp\left(-B_p\mu_i/b_i\right)}, \qquad i = 1, 2, \cdots, N. \tag{1}$$

During an active period of a source, the rate at which packets are produced is $\mu_i Q_i$. The rate at which packets can be transmitted on the line is $C_b/(B_p + H)$. An important quantity in our consideration is the ratio of source packet rate to line packet rate,

$$\gamma_i = \frac{\mu_i Q_i}{C_p}, \qquad i = 1, 2, \cdots, N. \tag{2a}$$

It may well be that in some applications a typical message consists of many packets, i.e., $b_i/\mu_i \gg B_p$. In this case we have

$$\gamma_i \cong \frac{b_i/B_p}{C_p}. \tag{2b}$$

Other source parameters that are important in our study can be derived. The utilization of source i or the fraction of time that source i is active is

$$u_i = \frac{\lambda_i}{\lambda_i + \mu_i}, \qquad i = 1, 2, \cdots, N. \tag{3}$$

The intensity of source i is defined as

$$\theta_i = \frac{\lambda_i}{\mu_i}, \qquad i = 1, 2, \cdots, N. \tag{4}$$

The average number of packets/second transmitted from source i is

$$r_i = \frac{Q_i}{1/\lambda_i + 1/\mu_i} = \mu_i Q_i u_i \, . \tag{5}$$

IV. LINE TRAFFIC

4.1 *Line Utilization*

From the way in which traffic is multiplexed onto the line it is clear that message delay is dependent upon the line traffic. As a prelude to the calculation of delay, the relevant characteristics of line traffic are considered in this section. A precise mathematical characterization of line traffic is extremely difficult. In fact, for reasons that will be explained presently, we encounter the most difficulty in this phase of the analysis.

We begin by calculating the average line utilization at each point in the ring. The basic assumption in this calculation is that a conservation law holds so that, over a sufficiently long time period, the average packet rate into a station is equal to the average packet rate out of a station. The traffic into and out of a station includes data to and from a customer connected to the station as well as line traffic. The implication is that there is no continuous buildup of packets in storage at a station, which is as it should be for normal ring operation.

Let the number of packets/second emanating from the source connected to station i ($i = 1, 2, \cdots, N$) be designated as r_i. In terms of parameters defined earlier $r_i = Q_i[1/\lambda_i + 1/\mu_i]^{-1}$. P_{ij} is defined as the portion of traffic originating at station i that is destined for station j with $P_{ii} = 0$. The average number of packets per second going from station i to station j is $P_{ij}r_i$. All of these packets pass through each station on the ring between stations i and j. The average number of packets per second from station i passing through station k is given by

$$R_{ik}^* = \begin{cases} r_i \sum_{1}^{i-1} P_{ij} + r_i \sum_{k+1}^{N} P_{ij} & \text{if} \quad 1 < i < k,\ k \neq N \\ r_i \sum_{1}^{i-1} P_{ij} & \text{if} \quad 1 < i < k,\ k = N \\ r_i \sum_{k+1}^{N} P_{ij} & \text{if} \quad i = 1,\ k \neq N \\ r_i \sum_{k+1}^{i-1} P_{ij} & \text{if} \quad k + 1 < i \leqq N \\ 0 \quad \text{otherwise.} \end{cases} \tag{6}$$

The total volume of traffic passing through station k is given by

$$R_k^* = \sum_{i=1}^{N} R_{ik}^* . \tag{7}$$

The total volume of traffic out of station k, including traffic from the local source, is

$$R_k = R_k^* + r_k , \quad k = 1, 2, \cdots, N. \tag{8}$$

Perhaps the distinction between R_k and R_k^* here can be emphasized by referring to Fig. 5. Here a B-box is shown as being split into its two functions, viz., taking data off the line and reading data onto the line. At point Z the line carries all of the traffic out of station k at an average rate of R_k packets/second. At point Y the average traffic rate seen by the local source as it attempts to multiplex data on the line is R_k^* packets/second. (Throughout the remainder of the analysis an asterisk on a line traffic parameter denotes a quantity as seen by the location station after message deletion.) Since the packet capacity of the line is C_p, the line utilization as seen by the source at station k is

$$U_k^* = R_k^*/C_p . \tag{9a}$$

The line intensity at this point is

$$\Theta_k^* = R_k^*/(C_p - R_k^*). \tag{10a}$$

The utilization and intensity on the line after station k are given by

$$U_k = R_k/C_p \tag{9b}$$

$$\Theta_k = R_k/(C_p - R_k). \tag{10b}$$

In the case where ring traffic is symmetric, i.e.,

$$P_{ij} = \begin{cases} 0 & i = j \\ 1/(N-1) & i \neq j \end{cases} \tag{11}$$

and $r_i = r$, $i = 1, 2, \cdots, N$, these expressions simplify greatly. We have

$$R_k^* = r\left(\frac{N}{2} - 1\right). \tag{12}$$

4.2 *Busy Period*

We turn now to the calculation of the average duration of idle and busy periods on the line. An exact calculation of either quantity is

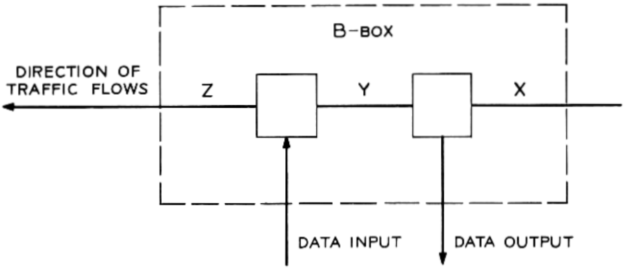

Fig. 5—Station decomposition.

difficult for all but very specific cases. The reason for this difficulty can be seen by examining the mechanism for multiplexing traffic on and off the line. Packets destined for a particular station may be interspersed in a sequence of contiguous packets. This sequence is broken up at random when the packets are delivered. On the other hand, the packets with the same destination may be concentrated at the beginning of a long sequence. In this case the long sequence will be scarcely affected when the packets in question are taken off the line.

In the remainder of this section are presented two approaches to the problem of calculating the average duration of the line busy period. While neither calculation is exact, both take into account the various factors involved, and, in those special cases where the answer is known, the same correct result is obtained. The basic difference between the two methods lies in the kind of line traffic that they are designed to model. As we proceed in the calculation this will be pointed out.

Both of these approaches draw on a result from storage theory concerning the data sequence out of a buffer. Suppose that n sources feed into a common buffer. The sources turn on and off at random transmitting at a constant rate during the active period, and data are read out of the buffer at a constant rate. The output will consist of successive idle and busy periods. Suppose that the durations of input idle periods are exponentially distributed with mean $1/\alpha_i$, $i = 1, 2, \cdots, n$. By definition, the probability of an input idle period terminating in any incremental interval dt is $\alpha_i\ dt$. An output idle period terminates when any one of the input idle periods terminate. Thus the probability of an output idle period terminating in any incremental interval dt is $\sum_{i=1}^{n} \alpha_i\ dt$; consequently the duration of the output idle period is exponentially distributed with mean $1/\sum_{i=1}^{n} \alpha_i$.

In both calculations the line traffic intensity [see equations (10a) and (10b)] is assumed to be equal to the ratio of the average durations of the line busy period to the line period. It is implicit in this assumption that the active and idle periods on the line are independent of one another and that they are stationary in the mean.

The first method that calculates the average duration of the busy period looks at each station individually. In equation (6) an expression for R_{ik}^*, the number of packets per second from station i passing through station k, is given. If this were the only source active, the line intensity at station k would be

$$\Theta_{ik}^* = \frac{R_{ik}^*}{C_p - R_{ik}^*}.$$

The average length of a message from source i is Q_i packets. We take the average duration of the idle period of the packet stream in seconds from the ith source as

$$1/\Lambda_{ik}^* = Q_i/C_p\Theta_{ik}^* = \frac{Q_i(C_p - R_{ik}^*)}{C_pR_{ik}^*}. \tag{13}$$

Now the assumption is made that the durations of the idle periods in the packet streams are exponentially distributed. To find the average duration of the idle period of the packet stream as seen by the source connected to station k, we call upon the result from storage theory quoted earlier. The duration of the idle period is exponentially distributed with mean

$$1/\Lambda_k^* = 1\Big/\sum_{\substack{i=1\\i\neq k}}^{N} \Lambda_{ik}^* . \tag{14}$$

The average duration of the busy period of the line in seconds as seen by the source at station k is given by

$$\frac{1}{M_k^*} = \frac{\Theta_k^*}{\Lambda_k^*}. \tag{15}$$

Combining (13), (14), and (15) we obtain

$$\frac{1}{M_k^*} = \frac{\Theta_k^*}{\displaystyle\sum_{\substack{i=1\\i\neq k}}^{N} \frac{C_pR_{ik}^*}{Q_i(C_p - R_{ik}^*)}}. \tag{16}$$

When the ring traffic is symmetric [see equation (11)], these results simplify. We have

$$\Lambda_k^* = \sum_{j=2}^{N-1} \frac{\lambda u((j-1)/(N-1))}{1 - \gamma u((j-1)/(N-1))}, \tag{17}$$

where $\mu = \mu_i$, $\Theta = \Theta_i$, and $u = u_i$, $i = 1, 2, \cdots, N$. In many cases of interest γ is small so that the term $\gamma u((j-1)/(N-1))$ in the denominator of (17) contributes little and may be ignored. In this case we may make a convenient approximation:

$$\Lambda_k^* \cong \mu u\left(\frac{N}{2} - 1\right), \qquad k = 1, 2, \cdots, N \tag{18}$$

and

$$\frac{1}{M_k^*} = \frac{\gamma}{\left[1 - \gamma u\left(\frac{N}{2} - 1\right)\right]}. \tag{19a}$$

This expression may be put in the alternate form

$$1/M_k^* = \gamma(1 + \Theta^*)/\mu. \tag{19b}$$

The key assumption in this calculation suggests the region of application of this model. In the packet stream from station i passing through station k, at average rate R_{ik}^*, it is assumed that the messages stay together. Thus the average busy period of this stream is Q_i packets long. The model is not valid when the messages from individual sources are broken up in the act of multiplexing. Thus the model should hold best when the sources emit short messages or when line utilization is low.

The foregoing method of calculating the duration (length) of the average busy period on the line essentially ignores the ring and treats each station as a separate entity. In contrast the second method makes explicit use of the ring structure. The algorithm begins with the assumption that the length of the line busy period at the input to a station is known. Based on this assumption the length of the busy period on the line out of the station is calculated. The process continues all the way around the ring until one returns to the starting point. The ring is closed by setting the duration of the final busy period equal to that of the initial busy period.

The algorithm for calculating the change in the busy period is best explained by referring to Fig. 5. The line intensities at points X and Y are Θ_{k-1} and Θ_k^* respectively. The assumption, fundamental to this approach, is that whole busy periods are deleted from the data stream. While there may be fewer busy periods at point Y, the average length of a busy period is the same from X to Y. The average durations of the idle periods are related by

$$\frac{1}{\Lambda_k^*} = \frac{\Theta_{k-1}}{\Theta_k^*} \frac{1}{\Lambda_{k-1}}. \tag{20}$$

The durations of the idle periods between Y and Z can be related by calling upon the previously quoted result from storage theory. If the length of the idle period at point Y is assumed to be exponentially distributed with mean $1/\Lambda_k^*$, and the length of the source idle period is exponentially distributed with mean $1/\lambda_k$, then the duration of the idle period at point Z is exponentially distributed with parameter

$$\Lambda_k = \Lambda_k^* + \lambda_k . \tag{21}$$

The average duration of the busy period at point X is given by

$$\frac{1}{M_k} = \frac{\Theta_k}{\Lambda_k}. \tag{22}$$

One can continue in this fashion until the starting point is reached.

In the case of a ring with a symmetric distribution of traffic the solution simplifies considerably since, by assumption, we have

$$\frac{1}{M_{k-1}} = \frac{1}{M_k}.$$

Substituting into equations (20), (21), and (22) we have that

$$\frac{1}{M_k} = \frac{\Theta - \Theta^*}{\lambda} = \frac{\Theta - \Theta^*}{\mu\theta}, \tag{23a}$$

where $\lambda = \lambda_i$, and $\theta = \theta_i$, $i = 1, 2, \cdots, N$. This result can be put into the form

$$\frac{1}{M_k} = \frac{\gamma}{(\lambda + \mu)[1 - \gamma u(N/2)][1 - \gamma u(N/2 - 1)]}. \tag{23b}$$

The key assumption in this calculation is that entire busy periods are destined for a single station. As the line traffic begins to build up, messages from different stations will tend to cluster in the same sequence and the validity of the model is weakened. The model should hold well when messages are multiplexed into light to moderate traffic.

V. DELAY CALCULATIONS

In this section we shall consider models for calculating message delay. As mentioned earlier we are primarily interested in calculation of message delay that is induced by traffic in the ring. Other delays such as propagation and processing delays are invariant with traffic intensity and are fixed for a given implementation of the ring.

Two separate approaches to the calculation of message delay have been considered. These approaches are complementary in the sense that they apply to different kinds of source traffic while neither approach is applicable to the whole range of source traffic. The exponential on-off model of the source as presented in the foregoing is difficult to handle analytically. The difference in the two approaches lies in the way this exponential on-off model is approximated.

The first approach that we shall consider uses a classical queueing theory model for the source.† Messages arriving at a terminal are viewed as customers arriving for service. An analogy is drawn between the length of the message (in bits or packets) and the service time of

† The queueing theory model of a source in message switched networks is widely used, e.g., Refs. 9 and 10.

a customer in queueing theory. Thus the exponential on-off source is approximated by messages with exponential length which arrive at a Poisson rate.

The queueing theory model is well suited to sources that are not very active. (However the line into which these sources are multiplexed may be very active if many sources are connected to the ring.) When the source is more active the queueing model has some inherent inaccuracies. The queueing model implies that the distribution of time between messages is an exponentially distributed random variable and there is a nonzero probability of successive messages overlapping. In contrast for the exponential on-off model, the distribution between beginnings of successive messages is the convolution of two exponential distributions and the probability of message overlap is zero.

The second approach to delay calculation uses a smoothed version of the traffic out of a source. Thus if the *average* number of bits per second out of a source is X bits/second, the calculation assumes that bits emanate from the source at a *constant* rate of X bits/second. This model of the source is meant to take up where the previous model leaves off, i.e., active sources. The effect of smoothing the bit flow will be less deleterious for more active sources.

The analysis of message delay based on a modified $M/G/1$ queue† is based on work by B. Avi-Itzhak and P. Naor.[11] The line into which a message is multiplexed is viewed as a server that is subject to random breakdown. As the line is either idle or busy the server is operating or under repair. Four assumptions on probability distributions are necessary in order to carry out the analysis: messages arrive at a Poisson rate, the interval between message arrivals is independent of the message size, the duration of the line idle period is exponentially distributed, and the lengths of line idle and busy periods are statistically independent. The distributions of the size of a message and of the duration of a line busy period are arbitrary.

We take the arrival rate of messages from source i as λ_i. The amount of time required to multiplex a message onto a completely free line at station i is denoted by the random variable S_i. It is assumed that the duration of the idle period on the line is exponentially distributed with parameter Λ_i^* [see equations (17) and (21)]. The random variable L_i denotes the duration of the line busy period as seen by the source at station i. In Appendix B it is shown that the Laplace-Stieltjes

† According to standard queueing theory notation an $M/G/1$ queue is one where a single server accommodates customers that arrive at a Poisson rate with an arbitrarily distributed service time per customer.

transform of T_i, the delay suffered by a message at station i, is given by

$$\mathcal{L}_{T_i}(v) = (1 - U_i^* - \gamma\theta_i)\,\frac{[\Lambda_i^*\mathcal{L}_{L_i}(v) - \Lambda_i^* - v]\mathcal{L}_{S_i}(\Lambda_i^* - \Lambda_i^*\mathcal{L}_{L_i}(v))}{\lambda_i - v - \lambda_i\mathcal{L}_{S_i}(v + \Lambda_i^* - \Lambda_i^*\mathcal{L}_{L_i}(v))}, \tag{24}$$

where $\mathcal{L}_{L_i}(v)$ and $\mathcal{L}_{S_i}(v)$ denote the $L - S$ transforms of L_i and S_i respectively. By differentiating $\mathcal{L}_{T_i}(v)$ with respect to v and allowing v to approach zero, one obtains the following expression for the expected value of T_i,

$$\begin{aligned} E[T_i] = {} & E[S_i]\,\Theta_i^* \\ & + E[S_i^2]\,\frac{\lambda_i(1 + \Theta_i^*)^2}{2[1 - \gamma_i\theta_i(1 + \Theta_i)]} \\ & + E[L_i^2]\,\frac{\Lambda_i^*}{2(1 + \Theta_i^*)[1 - \gamma_i\theta_i(1 + \Theta_i^*)]}. \end{aligned} \tag{25}$$

Now the mean number of packets per message is Q_i. Since the line packet rate is C_p packets per second, the average time that it takes to multiplex a message onto a clear line is

$$E[S_i] = Q_i/C_p = \frac{\gamma_i}{\mu_i}.$$

In the previous section we have made estimates of the line busy period which we have designated as $1/M_i^*$. These quantities can be substituted into (25) yielding

$$\begin{aligned} E[T_i] = {} & \frac{\gamma_i\Theta_i^*}{\mu_i} + \frac{\gamma_i^2}{\mu_i}\,\frac{\theta_i(1 + \Theta_i^*)^2(1 + \beta_{S_i})}{2[1 - \gamma_i\theta_i(1 + \Theta_i^*)]} \\ & + \frac{1}{M_i^*}\,\frac{U_i^*(1 + \beta_{L_i})}{2[1 - \gamma_i\theta_i(1 + \Theta_i^*)]}, \end{aligned} \tag{26}$$

where

$$\beta_{S_i} = \frac{\text{Var}(S_i)}{E^2(S_i)}$$

and

$$\beta_{L_i} = \frac{\text{Var}(L_i)}{E^2(L_i)}.$$

The quantities β_{S_i} and β_{L_i} indicate the sensitivity of the calculation of delay to assumptions about probability distributions. In the next section sample calculations are presented in which we assume that

$\beta_{L_i} = \beta_{S_i} = 1$ as would be the case if S_i and L_i were exponentially distributed.

By taking a second derivative of equation (24) and letting v approach zero, the second moment of delay can be found. For brevity we have omitted this expression; however, calculations based on it will be presented in Section VII.

$\mathcal{L}_{T_i}(v)$ can also be used to calculate the probability that a message has zero delay. Let us assume that $\lim_{u \to \infty} \mathcal{L}_{L_i}(v) = 0$ and that S is exponentially distributed. Then

$$\begin{aligned} \Pr\,[\text{zero message delay}] &= \lim_{v \to \infty} \mathcal{L}_{T_i}(v) \\ &= (1 - U_i^* - \gamma\theta_i)\left[\frac{\gamma\mu_i}{\gamma\mu_i + \Lambda_i^*}\right]. \end{aligned} \tag{27}$$

As mentioned previously, the second calculation of delay is predicated on the assumption that the data flow from the source is at a *constant* rate equal to the average rate from a source. Thus we assume that the source associated with station i generates data at a constant rate of r packets per second. Because the line is not continuously available to receive these packets buffering is required. It is the content of the buffer that is the key element in the calculation of delay.

As stated earlier, the line traffic consists of alternate busy and idle periods. In the following it is assumed that the durations of the busy and idle periods are independent and exponentially distributed. The mean values of these quantities are known from the calculations in the previous section. Under these assumptions, constant input rate and exponentially distributed durations of line idle and busy periods, it can be shown[12][†] that the probability density of the content of the buffer in packets of the ith station is given by

$$f(B_i) = K_i\delta(B_i) + (1 - K_i)\alpha_i \exp(-\alpha_i B_i), \tag{28}$$

where $\delta(\cdot)$ is the Dirac delta function and K_i and α_i are related to the parameters of the system by

$$K_i = \frac{C_p - r_i(1 + \Theta_i^*)}{(C_p - r_i)(1 + \Theta_i^*)} \tag{29a}$$

and

$$\alpha_i = \frac{M_i^*}{r_i} - \frac{\Lambda_i^*}{C_p - r_i}. \tag{29b}$$

[†] In the reference, the solution was obtained for a hyper-exponential density. In the present work the density is exponential, which can be obtained from the hyper-exponential density.

Notice that K_i is the probability that buffer i is empty. The average content of buffer i in packets is

$$E[B_i] = \frac{1 - K_i}{\alpha_i}. \tag{30}$$

The delay suffered by a packet is the amount of time that it spends in the buffer while waiting to be put on the line. From Little's Theorem[13] the average delay of a packet is the average content of the buffer divided by the arrival rate:

$$E[T_i] = E[B_i]/r_i\,, \quad i = 1, 2, \cdots, N. \tag{31}$$

Substituting (29) and (30) into (31) yields

$$E[T_i] = \frac{1}{M_i^*}\left[\frac{U_i^*}{1 - \gamma_i u_i(1 + \Theta_i^*)}\right]. \tag{32}$$

VI. RESULTS—AVERAGE DELAY

This section is devoted to numerical examples of the foregoing results. We shall look at two contrasting modes of system operation, complete symmetry and complete asymmetry. In the symmetric case all stations transmit equally to all other stations and the destination matrix is given by equation (11). All stations in this case are precisely alike in their traffic characteristics. This symmetric traffic pattern may be encountered on a national ring which connects regional rings. The presumption here is that each of the regional rings to which it is connected receives and transmits the same volume of traffic.

The asymmetric case models the situation where one station on the ring receives all of the output of the other stations. This singular station, in turn, distributes its traffic equally among the remaining stations. The asymmetric traffic pattern will be encountered in a ring which is composed of inquiry response users connected to a computer through a C-box.

Let us begin with the symmetric ring. Calculations were made for two separate models of delay. Model 1 is based on the queueing model of delay that resulted in equation (26). The model of line busy period here is that which led to equations (16)–(19). The combination of these two models is best suited to the situation where there are many lightly loaded sources on the line which send short messages. Model 2 is a combination of the smoothed approximation to the source that led to equation (32) and the calculation of line busy period lengths that resulted in equation (23).

The estimates of average delay yielded by Models 1 and 2 are given in terms of average message lengths. Presumably, in a particular application, the duration of an average message is known and the delay in seconds due to line traffic can be evaluated.

The results of the calculation for the symmetric ring are shown in Figs. 6–10. On Figs. 6 and 7 delay normalized to the average message duration is shown as a function of the number of stations for two different values of source utilization. The difference between these curves lies in the factor γ which is equal to 1 on Fig. 6 and 1/30 on Fig. 7. The value of 1/30 for γ corresponds roughly to sources with 50×10^3-bit-per-second active rate feeding into a 1.5×10^6-bit-per-second line. The results predict that, for $\gamma = 1/30$, the ring can accommodate many stations with delay less than one average message length.

Another view of system performance is shown on Figs. 8–10, where delay is shown as a function of source utilization for 10-, 50-, and 100-station rings when $\gamma = 1$. We see from these curves that, for moderate line loading, delays are not large. For example, when the line utilization on a 50-station ring (see Fig. 9) is 0.5, the delay is less than two average message lengths. Also shown on Figs. 8–10 are the results of simulations which will be discussed presently.

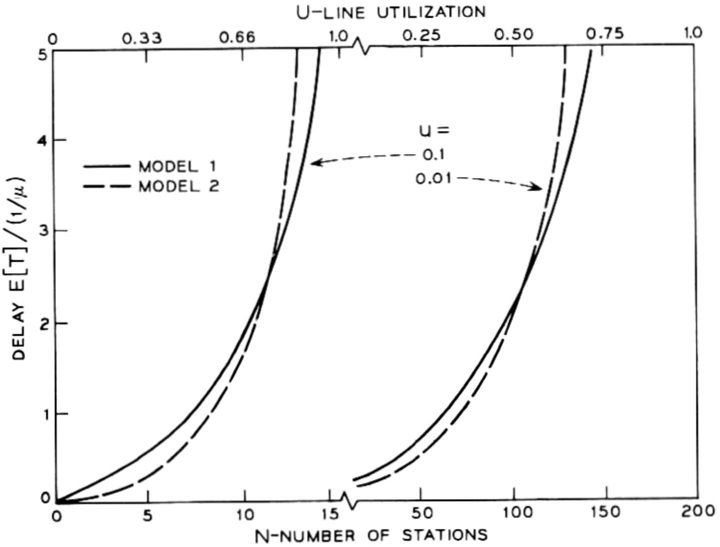

Fig. 6—Normalized delay for $u = 0.1$ and 0.01 (uniform inputs), $\gamma = 1$.

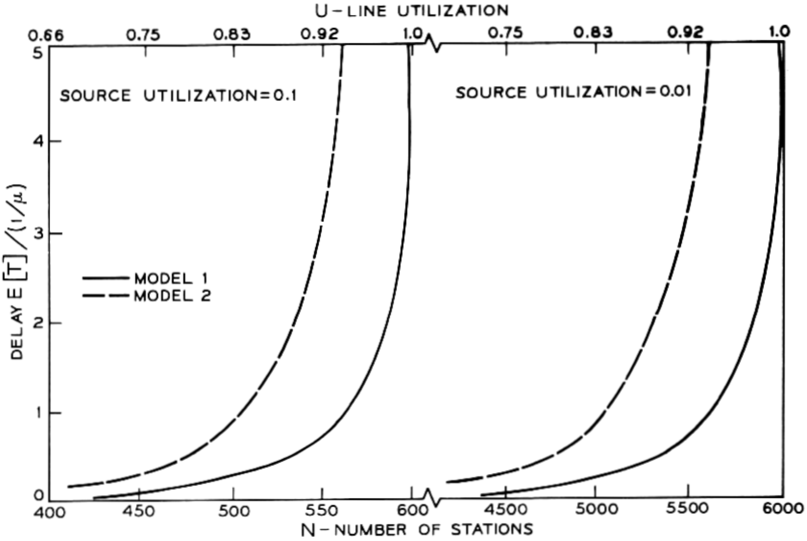

Fig. 7—Normalized delay for $u = 0.1$ and 0.01 (uniform inputs), $\gamma = 1/30$.

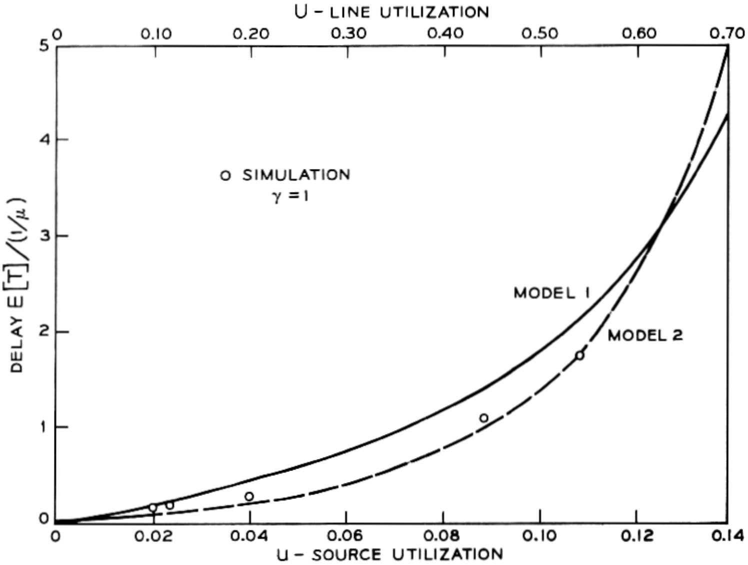

Fig. 8—Normalized delay for 10-station ring (uniform inputs).

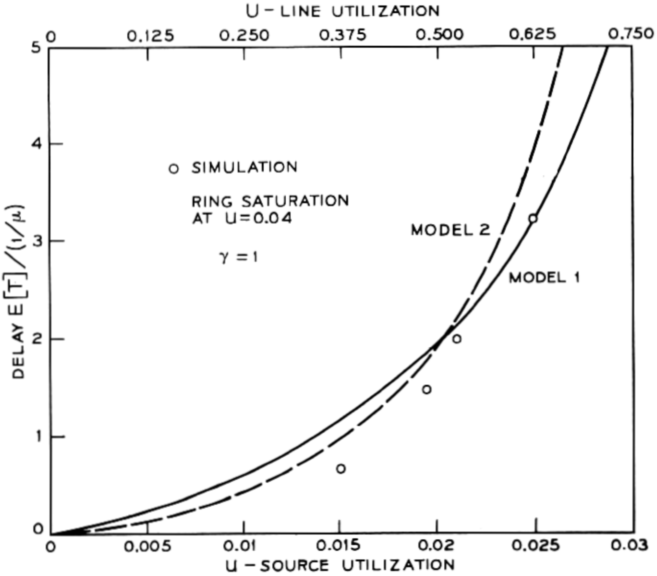

Fig. 9—Normalized delay for 50-station ring (uniform inputs).

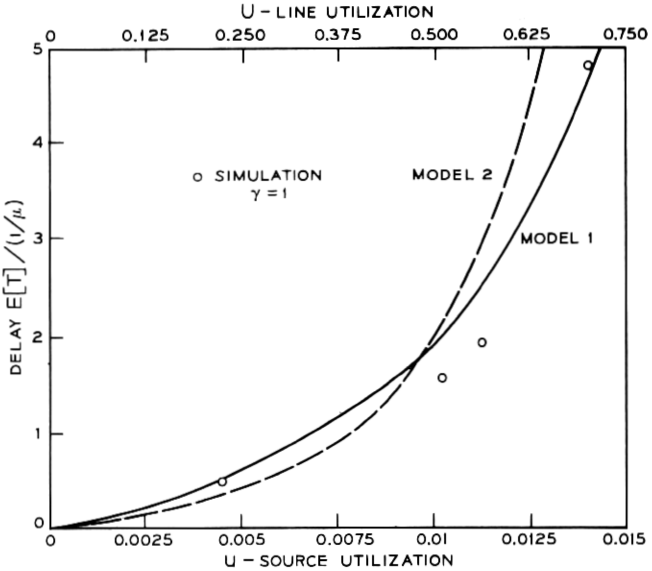

Fig. 10—Normalized delay for 100-station ring (uniform inputs).

A comparison of Figs. 10 and 11 shows the impact of different values of γ. On both figures delay is shown as a function of line utilization for 100-station rings. The difference between them is that $\gamma = 1$ on Fig. 10 and $\gamma = 1/30$ on Fig. 11. For $U = 0.5$ and $\gamma = 1$ (Fig. 10), we find that the delay is two message units. In contrast, for $\gamma = 1/30$ and the same line loading (Fig. 11), the delay is near zero. When $\gamma = 1/30$, the line has very large capacity compared to the source data rates, even when the line is moderately loaded.

It is convenient to use Model 2 to examine the asymmetric case of many users communicating with a single computer on the ring. Let u_c denote the utilization of the computer and u_s denote the utilization of the customer in each of the N stations connected to the computer. We assume that the computer distributes its traffic equally among all the other stations. Applying equations (7) and (9) the line utilization as seen by the ith station after the computer is

$$U_i^* = \gamma_c u_c \frac{(N - i)}{N} + \gamma_s u_s (i - 1), \qquad i = 1, 2, \cdots, N, \qquad (33)$$

where γ_c and γ_s are the ratios of source packet rate to line packet rate for the computer and the user, respectively [see equation (2)].

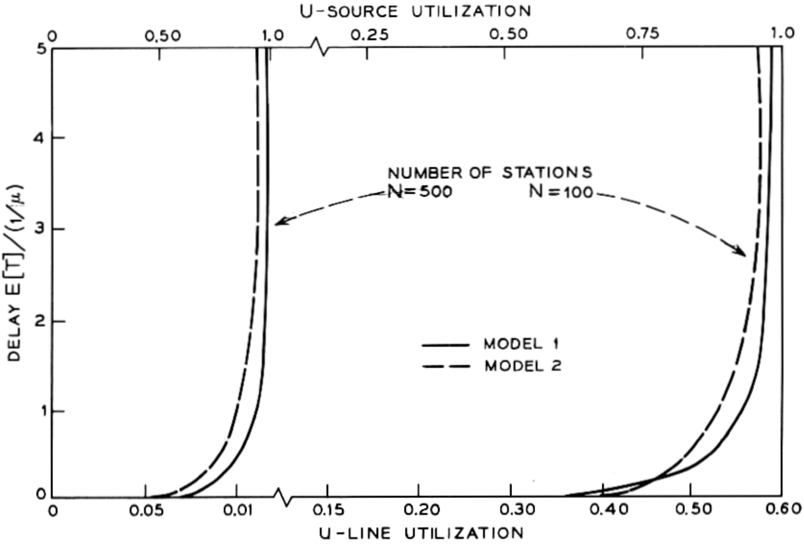

Fig. 11—Normalized delay for 100 and 500 stations (uniform inputs), $\gamma = 1/30$.

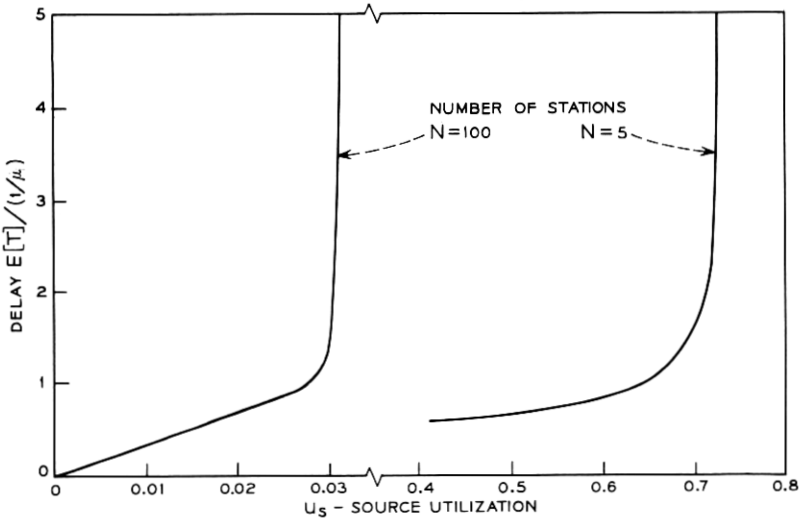

Fig. 12—Normalized delay at most critical station for $N = 5$ and 100 stations, $\gamma = 1/30$ (asymmetric case).

In our computations we assume that the computer-to-user traffic is ten times the user-to-computer traffic ($u_c\gamma_c = 10\ \gamma_s N u_s$).† Equation (33) leads us to conclude that under this assumption the most critical station in the ring is the one right after the computer ($i = 1$) since the line traffic at this point is heaviest.

Equation (33) may be used in conjunction with equation (32) to find the delay normalized to the line busy period at the most critical station. Because of the way the line busy period is calculated in Model 2, we may in turn normalize the line busy period to the activity time of the computer. The results are shown in Fig. 12 where delay, normalized to the computer busy period, is shown as a function of u_s and u_c and the number of stations as parameters. In Fig. 12 we have taken $\gamma_s = 1/30$ and $\gamma_c = 1$.

VII. SIMULATION RESULTS AND COMPARISON WITH THEORY

An investigation of the system was carried out by means of simulation as well as analysis. Single rings comprising the one A-box and from 10 to 100 B-boxes were simulated. All of the simulation results were for symmetric rings in which each station has identical traffic

† Studies of computer traffic support this assumption.[6]

characteristics and transmits equal portions of its traffic to the other stations.

In the simulation, as in the analysis, our attention was focused on average delay. Nevertheless, as we shall see, the simulation yielded information on other characteristics of delay which can be compared to analytical results.

The simulations attempted to mirror as closely as possible the actual operation of the ring. At each station a sequence of message and idle period lengths are chosen randomly from exponential distributions. As the messages are generated they are assembled into an integral number of packets. Bit stuffing is used to round out the last packet in a message. The destination of each message is found by a random selection from $N - 1$ equally probable choices.

After packets are generated they are given an initial time tag and sent to a buffer. The packets are stored in the buffer until they are multiplexed on the line. When a packet is multiplexed on the line, the time is noted and a difference in multiples of packets is taken with the initial time for each packet. It may happen, especially in a lightly loaded system, that a packet is multiplexed on the line immediately. In this case the time difference is zero.

These time differences indicate the delay suffered by a packet due to line traffic. By noting the time difference for the last packet in a message, we have a measurement of message delay. Histograms of message delay were compiled and results drawn from these histograms will be presented in the sequel.

A key step in the theoretical calculation was the estimation of line busy and idle period durations. Accordingly, in order to check the consistency of the models, measurements were made of line busy and idle period durations. These measurements will also be presented in the sequel.

In order to keep the simulation effort within reasonable bounds, it was necessary to fix some of the parameters of the system. Thus for most of the data that follow the average message length was fixed at 100 bits. The messages were quantized into 125-bit packets ($B_p = 125$). Header information was neglected ($H = 0$). Source utilization was varied by choosing appropriate idle durations. In this case 70 percent of the messages are of one packet duration. As a check, selected simulations were run with different ratios of message length to packet size.

As seen in the previous section, the theoretical models estimated average delays of one or two average message lengths for light to moderate line loadings. For $\gamma = 1/30$, estimates of average delay were small even for reasonably heavy line loadings. In Figs. 8, 9, and 10 average delays yielded by simulation are shown in comparison with theoretical results. These averages were taken over all stations for each of the line loadings shown. In Fig. 8 we see that for the 10-station ring the theoretical estimates given by Model 2 are quite close to simulation results, differing by substantially less than a message duration. The estimates produced by Model 1 for moderate line loads are also fairly good, overestimating delay by about 0.3 message duration. Theoretical estimates compare well with simulation results for the 50- and 100-station ring as well. In general both models overestimate the simulation delays somewhat. For line loadings below 0.5, Model 1's estimates are within 0.5 message duration above simulation values, while Model 2's estimates are somewhat closer. For line loadings above 0.5, Model 1's estimates are closer to simulation results.

It is significant that delay estimates obtained for line loadings below the knee of the curve ($U \cong 0.5$ for the 10-station ring) are well within an average message duration. Since the system would probably be operated in this region, delay estimates for these line loadings are important.

Each of the simulation points presented above is an average of between 7000 and 14,000 data points obtained through simulation. Along with average delay the standard deviation of delay was estimated. Estimates of the standard deviation of average delay were obtained by assuming that the standard deviation obtained from simulation was equal to the standard deviation of the underlying distribution of delay. The results showed that standard deviations of the averages are reasonably low. For example, for a 10-station ring at a line loading of 0.43, 10,000 data points were taken and the standard deviation was estimated to be 1.61 average message durations. The standard deviation of the mean is estimated to be 0.016 average message duration which is less than 2 percent of the average delay. The situation is the same on the 100-station ring. At a line loading of 0.52 where 7000 data points were taken, the standard deviation of the mean was estimated to be approximately 2 percent of the average delay.

Average delay is clearly not the only characteristic of delay that is important in judging the performance of a system. Both the simulation and the analysis provided results on characteristics of delay beyond averages. As mentioned earlier, histograms of message delay were compiled by the simulation program. On Fig. 13 the cumulative probability graph of message delay is shown. Also shown on Fig. 13 is the

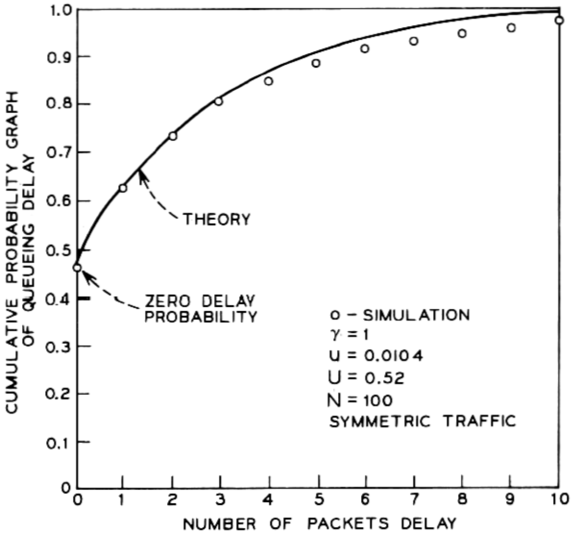

Fig. 13—Cumulative probability graph of queueing delay in a 100-station ring. Data compared with estimates from Model 2.

probability distribution predicted by Model 2 from the same line loading and system configuration [see equation (28)]. The probability distribution for Model 1 was not readily obtainable. The cumulative probability graph shown here is typical of other system configurations and line loadings. It shows a nonzero component at zero delay with diminishing probabilities of larger packet delays.

In order to summarize simulation results along these lines, three sets of numbers will be presented. They are probability of zero delay and a conditional mean and standard deviation of delay. The mean and standard deviation of message delay here are conditioned on the message having nonzero delay.

In Table I are shown the probability of zero delay given by simulation and by Model 1 [equation (27)] and Model 2 [equation (29a)]. For the smallest utilization in the 10-station ring the error between simulation and theory is within 15 percent. For the other line loadings the estimate obtained from Model 2 is within 10 percent of simulation values while Model 1 considerably underestimates zero message delay probability. For the highest loading considered, $U = 0.7$ on the 100-station ring, neither model predicts the simulation value very well.

Estimates of means and standard deviations of delay conditioned

Table I—Probability of Zero Message Delay

N–Number of Stations	U–Line Utilization	Simulation	Theory	
			Model 1	Model 2
10	0.10	0.906	0.833	0.918
10	0.20	0.80	0.688	0.833
10	0.43	0.565	0.418	0.624
10	0.54	0.464	0.311	0.514
100	0.235	0.773	0.622	0.769
100	0.52	0.454	0.318	0.485
100	0.54	0.418	0.264	0.420
100	0.70	0.252	0.178	0.304

on nonzero delay for both simulation and theory are presented in Table II. Comparison of the theoretical estimates of the mean shows that for the lightest loading in both the 10- and the 100-station rings the analysis estimates the delay within 3 percent. For the heavier loadings, excluding the $U = 0.7$ loading on the 100-station ring, the error is about 20 percent of the simulated results for Model 1 and about 25 percent for Model 2.

In summary, the data presented in Tables I and II indicate that a substantial portion of the messages (i.e., those that have no delay) are accounted for, and that the remaining messages have an average delay that is predictable, in the lighter loads, within 3 percent. For example, in the 10-station ring with a line loading of 0.1, 90.6 percent

Table II—Mean and Standard Deviation for Delayed Messages (In Average Message Lengths)

N–Number of Stations	U–Line Utilization	Simulation		Theory		
		Mean, $E[T \mid T > 0]$	Standard Deviation $\sigma[T \mid T > 0]$	Model 1		Model 2†
				Mean	Std Dev	Mean
10	0.10	1.18	0.79	1.19	1.22	1.18
10	0.20	1.34	0.99	1.44	1.46	1.44
10	0.43	2.06	1.89	2.33	2.48	2.58
10	0.54	2.66	2.75	3.12	3.34	3.78
100	0.235	1.57	1.59	1.61	1.61	1.69
100	0.52	2.95	3.20	3.13	3.16	4.26
100	0.54	3.29	3.55	3.78	3.81	4.83
100	0.70	6.48	7.89	5.60	5.65	10.72

† The density in this case is exponential, and thus the mean and standard deviation are equal.

of the messages have no delay and the remaining 9.4 percent have an average delay of 1.18 average message durations.

One observed difference between theory and simulation lies in the prediction of zero message delay by Model 1, which, for moderate line loadings, noticeably underestimates this quantity. There is evidence that the difficulty here may lie in the way that Model 1 treats quantization effects. For example, message lengths were taken to be exponentially distributed when, in fact, they are geometrically distributed. As pointed out in the beginning of this section, most of the simulation results are carried out for an average message length of 100 bits and a packet length of 125 bits. Investigations of the effect of quantizing messages into packets are continuing.

As mentioned earlier in this section, data were gathered on line busy and idle periods in order to check the consistency of our models. In Table III are shown the average durations of line busy periods for different loadings measured at one station in the ring. For comparison the average line busy periods predicted by Models 1 and 2 are also shown. Model 1 underestimates the length of the busy period by approximately 10 percent. The estimates given by Model 2 consistently overestimate the duration of the line busy period. For line loadings greater than 0.5, the estimate is poor.

The average durations of line idle periods as found in simulation are shown in Table IV. Again Model 1 underestimates while Model 2 overestimates. The error for Model 1 is higher than for the corresponding line busy periods. For line loadings greater than 0.5, Model 2 yields high estimates.

If the data on the durations of line busy periods obtained from

TABLE III—LINE BUSY PERIODS[†] (PACKETS)

		Simulation		Theory	
No. of Stations	U–Line Utilization	Mean	Standard Deviation	Mean (Model 1)	Mean (Model 2)
10	0.10	1.58	1.02	1.53	1.65
10	0.20	1.81	1.41	1.70	1.98
10	0.43	2.66	2.74	2.24	3.36
10	0.54	3.20	3.64	2.80	4.76
100	0.235	1.94	1.52	1.78	2.36
100	0.52	2.96	3.07	2.80	5.87
100	0.54	3.32	3.87	3.24	7.9
100	0.70	5.08	7.32	4.49	14.5

[†] The results shown here are averages for a single station. The line utilization is averaged over the entire ring.

TABLE IV—LINE IDLE PERIODS[†] (PACKETS)

No. of Stations	U–Line Utilization	Simulation		Theory	
		Mean	Standard Deviation	Mean (Model 1)	Mean (Model 2)
10	0.10	14.07	13.30	17.60	18.90
10	0.20	7.22	6.80	6.80	7.93
10	0.43	3.40	2.97	2.84	4.48
10	0.54	2.74	2.24	2.34	3.95
100	0.235	6.34	5.71	5.95	7.70
100	0.52	2.96	2.75	2.54	5.85
100	0.54	2.78	2.39	2.3	5.65
100	0.70	2.05	1.58	1.95	6.20

[†] The results shown here are averages for a single station. The line utilization is averaged over the entire ring.

simulation are used in the calculation of average delay the conclusions do not change substantially. Except for the $U = 0.7$ load point in the 100-station ring, the average delays predicted by Model 1 increase by less than 10 percent. Below line loadings of $U = 0.5$, the predictions of Model 2 decrease but are still fairly close to simulation points. Regions where the line loading is light to moderate are of greatest interest since it is most likely that systems would be operated in this region. For higher line loadings, delay in terms of average message lengths is large and small changes in loading lead to large changes in delay.

VIII. CONCLUSION

We conclude with a summary of our results. The analysis and the simulation of 10-, 50-, and 100-station rings show that for $\gamma = 1$, i.e., source rate and line rate equal, message delay is less than 2 average message durations for line loading up to 0.5 of capacity. For line loadings greater than 0.5, delay increases substantially. For $\gamma = 1/30$, i.e., a low-speed source feeding into a high-speed line, average delay in terms of average message durations is low for line loadings up to 0.8. Analytical and simulation results indicate that there is a nonzero probability of zero message delay at all line loadings. The probability of zero message delay is greater than 0.5 for line loadings less than 0.5. Analysis and simulation indicate that the probability of a specific message delay decreases monotonically with the value of delay. The rate of decrease is similar to that of an exponential distribution (see Fig. 13).

For light to moderate line loads the theoretical models of average message delay show good agreement with the simulation results. For high line loads some of the approximations used in the theoretical models are weakened and agreement is not as good. However, even for heavy loading, there is meaningful agreement between analysis and simulation since both predict large delay and large shifts in delay with small shifts in line loading.

Some of the discrepancy between analysis and simulation may be due to the fact that the analysis did not take into account quantizing effects. For example, line busy periods must be an integral number of packets in duration, whereas the analysis treated line busy periods as a continuous random variable. Investigation of these effects on message delay will be carried out in the future.

The simulation program is capable of simulating systems where the traffic pattern is not symmetric and work will continue in this direction.

IX. ACKNOWLEDGMENTS

The authors wish to thank M. J. Ferguson for many fruitful discussions. Acknowledged also is the assistance of B. Avi-Itzhak and M. Segal in deriving results on the server with breakdown queueing model.

The simulation programming was done by R. R. Anderson, who is also responsible for coaxing from the computer all of the simulation points that are shown here. The authors are deeply appreciative of his efforts.

APPENDIX A

Effect of Quantization

In the text we have assumed that the duration of a source active period is exponentially distributed with parameter μ_i (for source i). During the active period the source emits bits at a constant rate of b_i bits/second. The data is bundled into packets with B_p bits/packet. The probability that there are j packets in a message is given by

$$\begin{aligned} D_i &= \int_{(j-1)B_p/b_i}^{jB_p/b_i} \mu_i \exp(-\mu_i t)\, dt, \qquad i = 1, 2, \cdots, N \\ &= \exp(-(j-1)B_p\mu_i/b_i)[1 - \exp(-B_p\mu_i/b_i)]. \end{aligned}$$

Let

$$A \triangleq \exp[-B_p\mu_i/b_i];$$

then

$$D_j = A^{j-1}(1 - A).$$

The mean number of packets per message is

$$1/M_i = \sum_{j=1}^{\infty} jD_j = \frac{1}{1 - \exp(-B_p\mu_i/b_i)}.$$

APPENDIX B

In an earlier paper Avi-Itzhak and Naor[11] found the average delay of a customer arriving at a server that is subject to random breakdown. Recently, using a similar line of reasoning, Avi-Itzhak[14] found the Laplace–Stieltjes transform of the density of this delay T_i' to be:

$$\mathfrak{L}_{T_i'}(v) = (1 - U_i^* - \gamma\theta_i) \cdot \frac{[\Lambda_i^*\mathfrak{L}_{L_i}(v) - \Lambda_i^* - v]\mathfrak{L}_{S_i}(v + \Lambda_i^* - \Lambda_i^*\mathfrak{L}_{L_i}(v))}{\lambda_i - v - \lambda_i\mathfrak{L}_{S_i}(v + \Lambda_i^* - \Lambda_i^*\mathfrak{L}_{L_i}(v))}, \qquad (34)$$

where $\mathfrak{L}_{L_i}(v)$ and $\mathfrak{L}_{S_i}(v)$ are the Laplace–Stieltjes transforms of L_i and S_i respectively. (Recall that in the main body of the text L_i is defined as the duration of the line busy period at station i and S_i is defined as the time required to multiplex a message on a free line.) In this derivation it is assumed that the line idle periods are exponentially distributed with parameter Λ_i^*. It is also assumed that messages arrive at a Poisson rate λ_i.

In order for this result to be applicable to our problem, some modification of equation (34) is necessary. The expected value of T_i' yielded by (34) when the line is entirely free for a long period of time is $E[S_i]$. But this is a delay that is due to multiplexing alone and is not a function of line traffic. Since we are interested in delay that is dependent on line congestion we remove this multiplexing delay.

The delay described by equation (34) is the sum of two independent components, the waiting time and the residence time. The waiting time is the interval from when the message first arrives until it is first multiplexed on the line. The residence time is the interval in which the message is multiplexed on the line, including line busy periods during which the message is blocked. The L − S transform of the density of the residence time is given by

$$\mathfrak{L}_{S_i}(v + \Lambda_i^* - \Lambda_i^*\mathfrak{L}_{L_i}(v)) = \int_0^{\infty} dt\, e^{-vt} \sum_{k=0}^{\infty} \int_0^t dx\, \frac{(\Lambda_i^* x)^k e^{-x\Lambda_i{}^*}}{k!} f_{L_i}^{*k}(t - x) f_{S_i}(x), \qquad (35)$$

where $f_{S_i}(x)$ is the density of the message multiplexing time and $f_{L_i}^{*k}(x)$ is the k-fold convolution of the line busy period. This expression is obtained by adding the message multiplexing times and all of the intervening line busy periods. Now if we simply add together only the line busy periods, removing the line multiplexing time, we have

$$\mathfrak{L}_{S_i}(\Lambda_i^* - \Lambda_i^* \mathfrak{L}_{L_i}(v)) = \int_0^\infty dt\, e^{-vt} \sum_{k=0}^\infty \int_0^\infty dx \frac{(\Lambda_i^* x)^k}{k!} e^{-x\Lambda_i^*} f_{L_i}^{*k}(t) f_{S_i}(x). \tag{36}$$

Equation (34) becomes

$$\mathfrak{L}_{T_i}(v) = (1 - U_i^* - \gamma\theta_i) \cdot \frac{[\Lambda_i^* \mathfrak{L}_{L_i}(v) - \Lambda_i^* - v] S_i(\Lambda_i^* - \Lambda_i^* \mathfrak{L}_{L_i}(v))}{\lambda_i - v - \lambda_i \mathfrak{L}_{S_i}(v + \Lambda_i^* - \Lambda_i \mathfrak{L}_{L_i}(v))}. \tag{37}$$

REFERENCES

1. Pierce, J. R., Coker, C. H., and Kropfl, W. J., "Network for Block Switching of Data," IEEE Conv. Rec., New York, March 1971.
2. Steward, E. H., "A Loop Transmission System," 1970 ICC, vol. 2, pp. 36–1, 36–9.
3. Farmer, W. D., and Newhall, E. E., "An Experimental Distributed Switching System to Handle Bursty Computer Traffic," Proc. ACM Conf., Pine Mountain, Georgia, October 1969.
4. Fraser, A. G., "The Coordination of Communicating Processes," unpublished work.
5. Kropfl, W. J., "An Experimental Data Block Switching System," unpublished work.
6. Jackson, P. E., and Stubbs, C. D., "A Study of Multi-access Computer Communications," AFIPS, Conf. Proc., vol. 34, p. 491.
7. Jackson, P. E., and Fuchs, E., "Estimates of Distributions of Random Variables for Certain Computer Communications Traffic Models," Proc. ACM Conf., Pine Mountain, Georgia, October 1969.
8. Pilc, R. J., unpublished work.
9. Chu, W. W., "An Analysis of Buffer Behavior for Batch Poisson Arrivals and Single Server with Constant Output Rate," IEEE Trans. Commun. Tech., *COM-18*, No. 5 (October 1970), pp. 613–619.
10. Kleinrock, L., *Communications Nets–Stochastic Message Flow and Delay,* New York: McGraw-Hill, 1964.
11. Avi-Itzhak, B., and Naor, P., "Some Queueing Problems with the Service Station Subject to Breakdown," Oper. Res., *11*, No. 3, 1963, pp. 303–320.
12. Sherman, D. N., "Data Buffer Occupancy Statistics for Asynchronous Multiplexing of Data in Speech," Proc. ICC, San Francisco, California, June 1970.
13. Little, J. D. C., "A Proof of the Queueing Formula L = λW," Oper. Res., *9*, 1961, pp. 383–387.
14. Avi-Itzhak, B., unpublished work.