A study of multiaccess computer communications

by P. E. JACKSON and CHARLES D. STUBBS
Bell Telephone Laboratories, Incorporated
Holmdel, New Jersey

INTRODUCTION

The communications characteristics of multiaccess* computing are generating new needs for communications. The results of a study of multiaccess computer communications are the topic of this paper. The analyses made are based on a model of the user-computer interactive process that is described and on data that were collected from operating computer systems. Insight into the performance of multiaccess computer systems can be gleaned from these analyses. In this paper emphasis is placed on communications considerations. For this reason, the conclusions presented deal with the characteristics of communications systems and services appropriate for multiaccess computer systems.

The problem

Digital computers requiring communications with remote terminals exhibit a set of communications needs which, in some respects, are different from those of both voice traffic and other record communications. It is important for the providers of data communications to have an understanding of the broad characteristics of this communication process so that new, more appropriate offerings can be designed to satisfy these needs.

Previous studies** by the manufacturers and providers of multiaccess computer systems have begun to characterize both the computer systems and their users. The principal interest of these studies, however, has been computer and/or user performance rather than data communications.

There are several reasons why the quantitative characterization of the communications process is timely but intricate. First, multiaccess computing is still in its infancy. Therefore, computer system design is going through a trial and error process with a high rate of change of system characteristics. Lacking a unified, well-tested body of technical knowledge applicable to the problems of multiaccess computing, systems designers have been led to heuristic solutions to system organization. Certain specific problems such as scheduling algorithms for single and multiple central processors have been studied in detail. No intensive, overall, general system studies, however, have been reported with the constraints of total cost minimization including the effects of system characteristics on communications costs and human factors such as reduction in efficiency due to long turn-around times.

Second, the rate of change of the size of the user community, the number of systems in operation, and the introduction of new equipment and operating systems is high. In fact, most systems are changing so rapidly that a detailed characterization of any one will probably be outdated before it is completed. The insight to be gained from such studies, however, far outweighs the drawback of obsolescence. Indeed, this situation calls for continued study and review.

Third, the applications of time-sharing are diverse. Where one of the parties in the transaction is a person, uses range from inquiry-response systems with short call durations of a minute or less, to scientific problem-solving and certain types of business information systems with call durations of 10 to 30 minutes, to computer aided learning with long call durations of one to two hours or more. Where the transaction involves an automatic terminal such as a telemetry device, call durations may be measured in milliseconds. Also, the volume of information exchanged in a computer-to-computer or computer-to-data-logger interaction varies widely from a small number of bits in polling, meter

* The word "multiaccess" is chosen to avoid confusion over the use of the word "time-shared" which is often used synonymously but which has a specialized meaning in some contexts.
** For example, see References 1, 2, 3 and 4.
reading and some banking and credit services, to a large number of bits in CRT displays, information retrieval and file manipulation. The speed of transmission is wide-ranging from the low bit rates of supervisory and control terminals to megabits per second for CRT displays.

Fourth, the data required for such studies are microscopic in nature. Unlike voice traffic, which can be characterized by measures of holding times, arrival rates and other parameters independent of a call's content, the characterization of calls to a computer requires some information about a call's content, e.g., timing information interrelating the transmission times of data characters is essential for the design of an efficient time division data multiplexer. An additional factor is that some of the desired statistics on these data have very skewed distributions. Thus, large data samples are required. The implications of these considerations upon our study are that:

a. new data gathering procedures and equipment are needed,
b. data analysis procedures must be capable of handling very large quantities of data,
c. legal, ethical, and business requirements related to communications and computing privacy must be satisfied.

The problem, then, is to provide communications services to a rapidly growing market of multiaccess computer systems and their terminals. These exhibit diverse and changing communications requirements. The study described below is directed at this problem.

The modus operandi for this study is an in-depth analysis of selected multiaccess computer communications systems. The subset of system types chosen for detailed study is composed of computer service providers whose systems are representative of multiaccess computer installations. Besides representativeness, additional prerequisites for the choice of a system to study were that the use of multiaccess computing be advanced, and that the provider of the particular system be knowledgeable in the communications area. By “advanced in multiaccess usage,” we mean that the system be fully operational on a daily basis with the initial break-in period accomplished. A final prerequisite for inclusion in the study is the willingness of the computer service provider to participate in the study.

*Part of the study reported herein involved the collection of data from three operating multiaccess computer systems. In every case these data were obtained on the premises of the computer service provider and with this full permission and cooperation. To ensure the privacy of the three systems under discussion, however, they are not identified by name.

To ensure that a cross section of on-line systems was included in the study, the characteristics of such systems were classified as shown in Table 1.

<table>
<thead>
<tr>
<th>Classification characteristics for multiaccess computer systems for communications study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Computer Type</td>
</tr>
<tr>
<td>2. I/O Device Type</td>
</tr>
<tr>
<td>3. Loading (Number of Simultaneous Users)</td>
</tr>
<tr>
<td>4. User's Applications</td>
</tr>
<tr>
<td>5. User Community (In-House or Utility)</td>
</tr>
<tr>
<td>6. Error Control (e.g., Echoplex)</td>
</tr>
<tr>
<td>7. Holding Time</td>
</tr>
</tbody>
</table>

In the table, by “computer type” we mean the manufacturer and model number of the central processor and the system configuration, i.e., whether or not a separate communications computer is used. Not all models of all manufacturers can be covered, but at least two large manufacturers were included for each application. I/O device types include teletypewriter-like terminals and TOUCH-TONE® telephones. Loading is the average percentage of ports that are active. User’s applications include scientific and business programming, inquiry-response systems, extended file retrieval and maintenance, message switching and mixtures of these. Both in-house and utility systems were included. Error control includes systems which retransmit each character back to the terminal (Echoplex) and those which do not. The systems selected for examination include short holding time systems with average call durations on the order of one or two minutes or less and long holding time systems with average holding times of 20 to 30 minutes.

From the systems selected for detailed analyses, measurements of three different categories were obtained. The first category included telephone facilities measurements such as occupancies and overflow counts on computer access lines (port hunting groups) and pen recordings of call durations from several terminal lines. The second category of measurements was made by the computer service providers within their computer systems by identifying the arrival and departure times of calls, the amount of central processor time used, the serving port and an identifier of call type. Distributions of call holding time, call interarrival time, CPU usage, and port loading can be obtained from these data. The third category of measurement was the collection of data at computer ports describing the characteristics of such microscopic statistics as intercharacter time. The first two categories of data are being used to formu-
late traffic and engineering practices. These will be used by telephone company personnel to provide appropriate computer communications by properly configuring existing telephone company equipments.

The third category of data is being employed in the analyses reported here. These data are required to investigate new systems and service characteristics such as the desirability of various transmission speeds or multiplexing methods, as they include detailed information on the timing relationships within a call.

An analytic model of the communications process between a multiaccess computer and a user at a remote console is the vehicle being used to conduct these analyses. The model describes the communications process in terms of random parameters which give the times between characters transmitted through the communications network. All of the parameters are measurable at the communications interface to the computer, i.e., none requires the gathering of data on internal computer processes such as the length of various queues.

The model is used to focus on the user-computer communications process and to exhibit how the characteristics of the computer and of the user affect communications requirements. It is also used to study the converse, i.e., how the constraints of the communications medium affect the user and the computer. The model does not directly represent the detailed characteristics of the computer system or its organization or the internal operation of the user's console. Rather, it reflects the effect of these on the characteristics of the communications signals entering and leaving the computer. From the characteristics of the communications process, however, it is possible to employ the model to predict the effects of changing system characteristics such as improving computer scheduling algorithms or increasing the computer's transmission rate. The following two sections further discuss this model.

The data stream model

The next two sections develop the data stream model, the analytical model used to describe the stochastic interactive communications process between user and computer. In this section, the basic parameters of the model are defined. In the next section, the relationships among the parameters are described and an expression for the holding time of the process is developed where holding time is the duration of a user-computer session.

Figure 1 illustrates the data stream model. A "call" (or a connect-disconnect time period) is represented as the summation of a sequence of time periods during which the user sends characters without receiving, inter-

[Diagram of the data stream model]

leaved with time periods during which he receives characters without sending. (This implies half-duplex operation. Simple modifications to the model would allow the accommodation of full-duplex operation.) The periods during which the user is sending characters to the computer are defined as user burst segments. The periods during which he is receiving characters sent from the computer are computer burst segments. A user burst segment, by definition, begins at the end of the last character of the previous computer burst segment. Similarly, a computer burst segment begins at the end of the last character sent by the user. The first burst segment of a call begins when the call is established and the last burst segment ends when the call is terminated as measured at the computer interface.

Within a given burst segment, there are periods of line activity and of line inactivity. The first inactive period of a user burst segment is defined as think time. That is, think time is the time that elapses from the end of the previous computer character until the beginning of the first user character in that burst segment. In most cases, think time is employed by the user to finish reading the previous computer output and to "think" about what to do next. The corresponding inactive period in a computer burst segment is called idle time. In some systems idle time represents time during which the user waits for the return of "line feed" after sending "carriage return"; in other systems, idle time represents time during which the user's program is being processed or is in queue. The remaining inactive periods within a burst segment are called inter-character times and interburst times. A prerequisite for their definition is the definition of a "burst."

Two consecutive characters are defined as belonging to the same burst if the period of inactivity between the characters is less than one-half character width. Thus, each "burst" is the longest string of consecutive characters where the period of inactivity between any two consecutive characters is less than one-half character width. All of the characters in a burst must, of course, be transmitted from the same party (user or computer).
For example, every character of an unbroken string of characters sent at line speed is in the same burst.

For characters within the same user burst, an inactive time between two consecutive characters is called a user intercharacter time. The corresponding parameter for computer bursts is computer intercharacter time. For bursts within the same user (computer) burst segment, the inactive time between two consecutive bursts is called a user (computer) interburst time.

Five final parameters of the data stream model are number of user bursts per burst segment, number of computer bursts per burst segment, number of characters per user burst, number of characters per computer burst, and temporal character width (time from start to end of one character).

For a given user-computer environment, a knowledge of the distributions of the parameters defined above allows the calculation of some interesting measures. Examples are distributions for (a) holding time, (b) percent of holding time during which the communication channel carries data, and (c) amount of delay introduced by the computer. The next section shows how some of these distributions can be calculated from the parameters.

**Relationships among data stream model parameters**

Let the following notation be introduced:

- \( \tau \) = holding time of call (seconds)
- \( S \) = number of burst segments in call
- \( T \) = think time (seconds)
- \( I \) = idle time (seconds)
- \( B \) = interburst time (seconds)
- \( N \) = number of bursts per burst segment
- \( M \) = number of characters per burst
- \( W \) = character width (seconds)
- \( C \) = intercharacter time (seconds)

The lower case letters "c" and "u" will be used as superscripts to B, N, M, W, and C to represent "computer" and "user" respectively. For example, \( N^c \) will represent the number of computer bursts per computer burst segment. The three indices of summation to be used are:

- \( i \) — to designate the \( i^{th} \) burst segment,
- \( j \) — to designate the \( j^{th} \) burst of a given burst segment,

and

- \( k \) — to designate the \( k^{th} \) character of a given burst.

In summing expressions over these indices, the primary index will be shown as a subscript and the secondary index (or indices), if any, will be enclosed in parentheses.

Using this notation, it is possible to construct an equation relating the holding time of a call to its component parts in the following manner:

a. In burst segment \( 2i \), in the \( j^{th} \) burst, the amount of time required by the \( k^{th} \) user character is \( W_{k}^{u}(2i, j) \). Summing over all \( k \) such characters in the burst, the time required is

\[
M_{j}^{u}(2i) = \sum_{k=1}^{N} W_{k}^{u}(2i, j) .
\]

Summing over all bursts in the burst segment gives

\[
\sum_{i=1}^{S} \sum_{j=1}^{M} M_{j}^{u}(2i) .
\]

If one defines all burst segments where \( i \) is even as user burst segments and assumes that the number of burst segments per call \( S \) is always even, the total contribution of user character times to total holding time is

\[
\frac{S}{2} \sum_{i=1}^{N} \sum_{j=1}^{M} W_{k}^{u}(2i, j) .
\]

Since \( S \) is usually large the error introduced by assuming \( S \) even, even when it is not, is small. Assuming the odd numbered burst segments are computer burst segments, the corresponding contribution of computer character times is

\[
\frac{S}{2} \sum_{i=1}^{N} \sum_{j=1}^{M} W_{k}^{c}(2i - 1, j) .
\]

b. A corresponding argument shows the total contributions of user intercharacter times are

\[
\sum_{i=1}^{S/2} \sum_{j=1}^{M} [M_{j}^{u}(2i) - 1] C_{k}^{u}(2i, j)
\]

and of computer intercharacter times are

\[
\sum_{i=1}^{S/2} \sum_{j=1}^{M} [M_{j}^{c}(2i - 1) - 1] C_{k}^{c}(2i - 1, j) .
\]
c. The total amount of user interburst time is
\[ \frac{S}{2} \left[ \sum_{i=1}^{N_u} \sum_{j=1}^{N_u} E_j(2i) \right] , \]
and the total computer interburst time is
\[ \frac{S}{2} \left[ \sum_{i=1}^{N_c} \sum_{j=1}^{N_c} E_j(2i - 1) \right] . \]

d. Total think time is
\[ \frac{S}{2} \sum_{i=1}^{N_u} T_{1i} , \]
and total idle time is
\[ \frac{S}{2} \sum_{i=1}^{N_u} I_{2i-1} . \]

The sum of these components is the holding time for a call. That is, the time of each burst segment summed over all burst segments is the holding time. The time of a burst segment equals the sum of the durations of all bursts, interburst times, and the think (or idle) time in that burst segment. The duration of a burst is equal to the sum of the character times and the intercharacter times contained therein. That is, the holding time of call \( \ell \), \( \tau_{\ell} \), is

\[ \tau_{\ell} = \frac{S}{2} \left[ \sum_{i=1}^{N_u} \left( T_{1i} + \sum_{j=1}^{N_u} E_j(2i) \right) + \sum_{i=1}^{N_u} \sum_{j=1}^{N_u} W_j(2i, j) \right] + \sum_{i=1}^{N_u} \left[ \sum_{j=1}^{N_u} E_j(2i - 1) \right] + \sum_{i=1}^{N_u} \left[ \sum_{j=1}^{N_u} C_j(2i - 1, j) \right] \] (1)

Knowing the distributions for the 12 parameters in Equation (1), it is theoretically possible to solve directly for the distribution of holding time. The mechanics of finding the solution are prohibitive, however, except for very restricted cases. One method of solving (1) is to find the moments of holding time rather than the complete distribution. This approach will be used here and, in fact, it will be sufficient for our purposes to solve merely for the mean value of holding time. In order to arrive at the solution, we assume that the random variables are stationary and mutually independent.*

Taking the expected value of both sides of (1), we obtain

\[ \tau = \frac{(S/2)}{T + B^*(N^* - 1) + N^*M^*W^*} \]
\[ + N^*C^*(M^* - 1) + I + B^*(N^* - 1) \]
\[ + N^*M^*W^* + N^*C^*(M^* - 1) \] (2)

where the symbol for each variable without a subscript implies its mean value. For further analysis, Equation (2) may be separated into four parts each having its own functional significance:

a. user send time (the total amount of time during which user characters are being transmitted)
\[ = \frac{(S/2)}{(N^*M^*W^*)} , \]

b. computer send time (the total amount of time during which computer characters are being transmitted)
\[ = \frac{(S/2)}{(N^*M^*W^*)} , \]

c. user delay (the sum of all inactive periods during user burst segments)
\[ = \frac{(S/2)}{[T + B^*(N^* - 1) + N^*C^*(M^* - 1)]} , \]

d. computer delay (the sum of all inactive periods during computer burst segments)
\[ = \frac{(S/2)}{[I + B^*(N^* - 1) + N^*C^*(M^* - 1)]} . \]

The sensitivity analysis performed in the next section is an investigation of the properties of these four parts and includes a discussion of how their interrelation affects holding time. If any of these parts can be

* Analyses of these assumptions have exposed their limitations. However, these assumptions have been shown to be reasonable for the analyses and conclusions of this paper.
reduced without increasing others, holding time can be reduced leading to possible cost savings.

Before discussing such analyses, it may be well to indicate what values have been observed for each of the 12 parameters. This will allow us to concentrate our attention on those parameters and those measures of parameters that promise to be the areas of greatest possible holding time reduction.

Collected data and sensitivity analyses

During the current study, data have been gathered on a large number of calls to each of several multaccess computer systems. For each system, the data have been partitioned into sets representing each of the 11 random parameters (the twelfth parameter, character width, is a constant). Probability density functions have been fitted to the data collected on each parameter from each system.

Data from three of the systems are discussed in this paper. These systems are labeled A, B, and C. Systems A and B have the same computer equipment and basically the same mix of user applications (programming—scientific). System C has computer equipment different from that of the other two systems and its mix of user applications is primarily business oriented. All three systems serve low-speed teletypewriter-like terminals. System B is rather heavily "loaded" compared to Systems A and C. Table II summarizes these characteristics for Systems A, B, and C.

Table II—Characteristics of systems studied

<table>
<thead>
<tr>
<th></th>
<th>System A</th>
<th>System B</th>
<th>System C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Type</td>
<td>Brand X</td>
<td>Brand X</td>
<td>Brand Y</td>
</tr>
<tr>
<td>Transmission Speed (Characters/sec)</td>
<td>10</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>Primary Application</td>
<td>Scientific</td>
<td>Scientific</td>
<td>Business</td>
</tr>
<tr>
<td>Load</td>
<td>Moderate</td>
<td>Heavy</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

Table III—Average parameter values

<table>
<thead>
<tr>
<th></th>
<th>(\bar{\mu})</th>
<th>(\sigma_{\bar{\mu}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>No. of Burst Segments</td>
<td>82.</td>
</tr>
<tr>
<td>T</td>
<td>Think Time (sec.)</td>
<td>4.3</td>
</tr>
<tr>
<td>I</td>
<td>Idle Time (sec.)</td>
<td>.05</td>
</tr>
<tr>
<td>B*</td>
<td>User Interburst Time (sec.)</td>
<td>1.6</td>
</tr>
<tr>
<td>B*</td>
<td>Computer Interburst Time (sec.)</td>
<td>16.</td>
</tr>
<tr>
<td>N*</td>
<td>No. of Bursts/User Burst Seg.</td>
<td>11.</td>
</tr>
<tr>
<td>N*</td>
<td>No. of Bursts/Computer Burst Seg.</td>
<td>3.3</td>
</tr>
<tr>
<td>M*</td>
<td>No. of Characters/User Burst</td>
<td>1.1</td>
</tr>
<tr>
<td>M*</td>
<td>No. of Characters/Computer Burst</td>
<td>47.</td>
</tr>
</tbody>
</table>

One characteristic of the data summarized in Table III deserves further comment. It is that measures which should be most sensitive to computer characteristics seem to just that. For example, the users of both Systems A and B have predominantly programming—scientific applications and the average numbers of characters per user burst segment \((N^*M^*)\) are 9.2 and 10.7 for Systems A and B, respectively, versus 13.8 for the primarily business oriented users of System C. Such relationships prevail in spite of widely different average computer delays. The average amount of time spent per computer burst segment in interburst delay, \((N^*-1)B^*\), is 1.4 seconds in System A and 35.8 seconds in System B.
Table IV—Mean values of holding time components

<table>
<thead>
<tr>
<th></th>
<th>System A</th>
<th>System B</th>
<th>System C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Holding Time, $\tau$</td>
<td>17</td>
<td>34</td>
<td>21</td>
</tr>
<tr>
<td>Minutes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average User Send Time, $(S/2)(N^uM^uW^u)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minutes</td>
<td>0.50</td>
<td>0.45</td>
<td>0.96</td>
</tr>
<tr>
<td>% of $\tau$</td>
<td>3%</td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>Average Computer Send Time, $(S/2)(N^cM^cW^c)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minutes</td>
<td>5.7</td>
<td>4.5</td>
<td>7.5</td>
</tr>
<tr>
<td>% of $\tau$</td>
<td>33%</td>
<td>13%</td>
<td>35%</td>
</tr>
<tr>
<td>Average User Delay, $(S/2)(T + B[N^u - 1] + N^cC^c[M^c - 1])$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minutes</td>
<td>10</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>% of $\tau$</td>
<td>58%</td>
<td>35%</td>
<td>53%</td>
</tr>
<tr>
<td>Average Computer Delay, $(S/2)(I + B[N^c - 1] + N^cC^c[M^c - 1])$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minutes</td>
<td>0.95</td>
<td>17</td>
<td>1.5</td>
</tr>
<tr>
<td>% of $\tau$</td>
<td>6%</td>
<td>51%</td>
<td>7%</td>
</tr>
</tbody>
</table>

Table IV summarizes the macroscopic characteristics of these data as they contribute to holding time. An inspection of the table leads to the following observations:

Observation 1: The average holding time for the heavily loaded system (System B) is considerably larger than for the lightly loaded Systems A and C (94 percent and 60 percent larger).

Observation 2: For the lightly loaded systems, A and C, Computer Delay is less than 10 percent of the total holding time but for System B Computer Delay accounts for over half of the total holding time.

Observation 3: User Delay is a significant component of holding time in all three systems, and in each case, is between 10 and 12 minutes.

Observation 4: User Send Time is less than 5 percent of total holding time in each system and is not a significant contributor to holding time.

Observation 5: Computer Send Time is smallest in both absolute value and in percent holding time for the heavily loaded System B.

These five observations lead to three broad areas of interest that are discussed in the next three sections. The first area is the relationships between holding time and computer delay (Observations 1 and 2). The second is the relationships between holding time and user characteristics (Observations 3 and 4). The third is the relationship between holding time and computer send time (Observation 5).

Relationships between holding time and computer delays

The Computer Delay times shown in Table IV indicate a large variability among computer systems. This section investigates this variability.

There are a number of convenient measures that can be used to describe “computer load” and “computer delays.” The “load” on a computer is a function of the number of simultaneous users who are “active” (in queue waiting for the computer to run their program or output to them), and the characteristics of user programs. For the purposes of the present discussion, data availability requires the use of “simultaneous users” as our measure of computer load. The manner in which a computer system reacts to a fluctuating load is a function of many additional variables, including characteristics of the scheduler.

Average computer delay may be calculated as the average total amount of computer delay per call, i.e., the sum of all the idle times, computer interburst times, and computer intercharacter times in a call.* This method was used in Table IV to demonstrate the effects of total computer delay on the average holding times of the three computer systems. It is also beneficial to examine the individual components of total computer delay. For example, it appears reasonable to divorce our measure of computer delay from the number of burst segments in a call by considering computer delay per burst segment. This new measure is reasonable because the number of burst segments

*Symbolically, average total computer delay per call is $(S/2) (I + B[N^c - 1] + N^cC^c[M^c - 1])$. 

From the collection of the Computer History Museum (www.computerhistory.org)
per call appears to be highly sensitive to user application type. Thus, calculating computer delay per burst segment reduces the dependence of our results on user application. We now consider contributions from average idle time (I), from average intercharacter times \((NcCc[M-1])\), and from average interburst times \((Bc[N-1])\). These three components of average computer delay per computer burst segment are shown in Table V.

As can be observed, the delays introduced by interburst times are the majority of all computer delay components in each system. The explanation for the relative sizes of these three components is as follows:

a. The characteristics of user programs are such that two or more quanta of execution time are required for a run to completion but output is generated by each quantum; and

b. The combination of system load, output buffer size, and characteristics of the scheduler preclude the immediate availability of additional output when the transmission of a computer burst is completed.

Because average InterBurst Time is the largest single contributor to average computer delay, it will be denoted by a special symbol, I-B-T. Figure 2 shows the relationships between holding time and I-B-T for each of the three computer systems. The three points in Figure 2 associated with computer systems are the observed values of holding time and I-B-T for those systems. The three lines are generated by changing I-B-T for each of the systems while holding all other parameters fixed. The slope of each line is equal to one-half the number of burst segments per call, i.e., \(S/2\), because when every other factor in Equation (2) is held constant, it becomes

\[
\text{Holding Time} = \text{Constant} + (S/2) (\text{I-B-T}).
\]

Note that the lines for Systems A and B are close together. As these two systems have similar configurations of hardware and software, support is given to the

<table>
<thead>
<tr>
<th>Table V—Components of average computer delay per computer burst segment (all times in seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average Computer Delay per Computer Burst Segment ((\Delta))</strong></td>
</tr>
<tr>
<td>Average Idle Time per Computer Burst Segment</td>
</tr>
<tr>
<td>% of (\Delta)</td>
</tr>
<tr>
<td>Average Interburst Time per Computer Burst Segment</td>
</tr>
<tr>
<td>% of (\Delta)</td>
</tr>
<tr>
<td>Average Intercharacter Time per Computer Burst Segment</td>
</tr>
<tr>
<td>% of (\Delta)</td>
</tr>
</tbody>
</table>

Figure 2—Average holding time versus average computer interburst time delay per computer burst segment
conjecture that increasing the load on System A would lead to values of I-B-T and holding time comparable to those of System B. Conversely, de-loading System B should result in these parameters having values comparable to those of System A.

Next, as it appears that holding time is a function of I-B-T and I-B-T is in turn a function of the loading on the computer and hence on the number of simultaneous users, an expression relating I-B-T and number of active users can be established.

To establish the relationship between I-B-T and number of simultaneous users, both quantities were measured on the systems as a function of time of day. The solid curve in Figure 3 indicates the average number of simultaneous users of System A for 15-minute periods of the day. The average I-B-T's were calculated for hourly periods on data from System A and a least squares fit of a variety of curve types was investigated. For these data the best fit is

\[ I-B-T = (0.18) \exp(0.13 u) \]  

(4)

where

\[ u = \text{number of simultaneous users}, \]

or

\[ u = (7.7)(1.7 + \ln I-B-T). \]  

(5)

The dashed curve in Figure 3 is a plot of \( u \) versus time of day by using Equation (4) and the actual measurements of I-B-T versus time of day. This fit seems to reflect the major characteristics of the data as shown by the solid line in the sense that at least the morning and afternoon busy periods are reflected along with the intervening noontime lull.

Using (4), a plot can also be made relating I-B-T and the number of simultaneous users. Figure 4 shows this relationship as well as three curves showing the effects of I-B-T on holding times in Systems A, B, and C. These latter curves are plotted from Equation (3) after substitution of (4). Figure 4 indicates that above some threshold (represented by the knees of the curves) the computer's grade of service deteriorates rapidly as additional users are accepted.*

Relationships between holding time and user characteristics

Tables VI and VII, below, are used to illustrate several relationships between holding time and user characteristics in this section and between holding time and computer send times in the next section.

---

Table VI—Send time information
(all times in minutes)

<table>
<thead>
<tr>
<th></th>
<th>System A</th>
<th>System B</th>
<th>System C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Holding Time (( \tau ))</td>
<td>17.</td>
<td>34.</td>
<td>21.</td>
</tr>
<tr>
<td>Average Total Send Time (( R )) = ( (S/2)(N^*M^<em>W^</em> + N^*M^<em>W^</em>) )</td>
<td>6.2</td>
<td>5.0</td>
<td>8.4</td>
</tr>
<tr>
<td>% of ( \tau )</td>
<td>36%</td>
<td>15%</td>
<td>40%</td>
</tr>
<tr>
<td>Average User Send Time ( (S/2)(N^*M^<em>W^</em>) )</td>
<td>.50</td>
<td>.45</td>
<td>.96</td>
</tr>
<tr>
<td>% of ( \tau )</td>
<td>3%</td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>% of ( R )</td>
<td>8%</td>
<td>9%</td>
<td>11%</td>
</tr>
<tr>
<td>Average Computer Send Time ( (S/2)(N^*M^<em>W^</em>) )</td>
<td>5.7</td>
<td>4.5</td>
<td>7.5</td>
</tr>
<tr>
<td>% of ( \tau )</td>
<td>33%</td>
<td>13%</td>
<td>35%</td>
</tr>
<tr>
<td>% of ( R )</td>
<td>92%</td>
<td>91%</td>
<td>89%</td>
</tr>
</tbody>
</table>

Table VI shows, for each system, the average holding time (\( \tau \)), the average total send time (\( R \)), and the average user and computer send times. These quantities are measured both in minutes and, for the latter three categories, as a percentage of holding time.

Table VII is constructed identically for delay quantities rather than for send time quantities.
Table VII—Delay information
(all times in minutes)

<table>
<thead>
<tr>
<th></th>
<th>System A</th>
<th>System B</th>
<th>System C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Holding Time (τ)</td>
<td>17</td>
<td>34</td>
<td>21</td>
</tr>
<tr>
<td>Average Total Delay (D)</td>
<td>11.5</td>
<td>29.2</td>
<td>13.7</td>
</tr>
<tr>
<td>% of τ</td>
<td>64%</td>
<td>85%</td>
<td>60%</td>
</tr>
<tr>
<td>Average Total User Delay</td>
<td>10.2</td>
<td>12.1</td>
<td>11.5</td>
</tr>
<tr>
<td>% of τ</td>
<td>58%</td>
<td>35%</td>
<td>53%</td>
</tr>
<tr>
<td>% of D</td>
<td>92%</td>
<td>40%</td>
<td>88%</td>
</tr>
<tr>
<td>Average Total Computer Delay</td>
<td>.95</td>
<td>17.3</td>
<td>1.5</td>
</tr>
<tr>
<td>% of τ</td>
<td>6%</td>
<td>51%</td>
<td>7%</td>
</tr>
<tr>
<td>% of D</td>
<td>8%</td>
<td>60%</td>
<td>12%</td>
</tr>
</tbody>
</table>

Table VI indicates that (a) in all three systems average user send time accounts for less than five percent of average holding time and less than 12 percent of average total send time, and (b) the users of System C inputted about three times as many characters as the users of the other two systems. A conclusion that can be drawn from (a) is that user send time is an insignificant contributor to holding time. Even if average user send time increased by a factor of three, the increase in holding time would be only one to two minutes, assuming that total user delay remains fixed. The reasons for (b) are probably a combination of, first, the business oriented applications of many of the system’s users and, second, the rather low computer delays experienced in System C. It is possible that this increase in user input volume in System C is encouraged by the small computer delays experienced in that system. The same factor could also be partially responsible for the greater degree of on-line interaction in System C which when compared to Systems A and B had about double the average number of burst segments per call.

Table VII indicates that average total delay D (the sum of average user and computer delays) accounts for more than half of average holding time. The lowest percentage is for System C where average total delay is...
60 percent of average holding time; the highest is System B with 85 percent. Of these delays, users contributed from 40 percent (System B) to 92 percent (System A). Another observation is that the absolute value of average user delay is between 10 and 12 minutes in all three systems. This consistency is rather remarkable when one considers the diversity of other parameters that affect user delay.

One might conjecture further about how user send time and delay characteristics would be affected by different transmission speeds, different program applications, different levels of user sophistication, and many other variables. For example, the user who inputs his prepared program from punched tape eliminates the user delay introduced by the user who performs this function by the hunt and peck method. An analogous "user" characteristic is computer send time which is determined almost entirely by the amount of computer output requested by the user. The next section discusses how this measure affects holding time.

Relationships between holding time and computer send time

Table VI shows that the system with the highest load, i.e., System B, has the smallest user send times and computer send times in both absolute value and in percent of holding time. This fact may be partially caused by the user's tendency to limit the amount of on-line I/O when he experiences long computer delays. The second, and more useful, observation that can be made from Table VI is that in systems which are not heavily loaded, average total send time may be on the order of 35 percent to 40 percent of average total holding time. Of this time, approximately 90 percent is computer send time.

We may infer from these data that, barring changes in user patterns and other influencing factors, holding time may be materially reduced by providing a high-speed channel from the computer to the user and a high-speed printer, or other display device at the user's location. This system redesign would enable a decrease in W*, computer character width, with a corresponding decrease in

\[
\text{Holding Time} = \text{Constant} + \left(\frac{S}{2}\right) (N \cdot M \cdot W^*). \quad (6)
\]

Figure 5 is a plot of average holding times versus computer-to-user channel speed assuming all other factors remain constant. Of the three systems, two of them, viz., Systems A and B, have computer channel speeds of 10 characters per second. System C transmits at 15 characters per second. Note that if System C transmitted at 10 characters per second, its average holding time would be expected to increase to about 25 minutes from 21 minutes.

If computer channel speed were infinite, the average holding times for Systems A, B, and C would be 11.7, 29.3, and 13.7 minutes, respectively. To approach these minima within ten percent would require computer channel speeds of about 60 ch/sec for System A, 20 ch/sec for System B and 100 ch/sec for System C. If a channel speed of 360 ch/sec were available, the computer send times would be less than one half minute and less than 2.5 percent of average holding time in each system.

At this point one might conjecture that there are at least two components of holding time that are likely to increase if computer transmission rates are increased. The first is think time because, in at least some instances, the user utilizes computer send time to read the output he receives. Hence, if the computer outputs the same number of characters in a much shorter time interval, the user may increase his think time in order to do the same amount of reading and thinking. This interplay of responses and transmission rates has effects on holding time in that it suggests the existence of some upper bound on computer transmission rate beyond which decreases in computer send time are matched by equal increases in user think time. The result is that average holding time cannot be further reduced by increasing computer transmission rates. In order to attach quantitative significance to this conjecture let us assume that the average user employs currently 10 percent of computer send time for reading and thinking.* This

* Ten percent is certainly a high figure for the user who is listing his program to provide a paper copy for filing purposes but is low for the user who is checking every comma and parenthesis in order to find a program bug.
assumption implies that a tenfold increase in computer transmission rate will result in minimum average holding time. For Systems A and B this rate is 100 characters per second (1000 words per minute) and yields holding time of about 12.5 minutes and 29.5 minutes, respectively. These average holding times are 28 percent and 13 percent less than the current holding times and seven percent and one percent greater than the theoretical minima shown in Figure 5 for Systems A and B. For System C, the corresponding transmission rate is 150 characters per second. This would reduce the average holding time in System C by 31 percent to about 14.5 minutes which is six percent greater than the theoretical minimum of 13.7 minutes. The minimum holding times implied by this assumption are shown as dashed lines in Figure 5.

The second variable that may naturally increase if computer transmission rates increase is the quantity of output requested by the user. For example, if the computer transmitted at 360 characters per second, computer send time would be less than 2.5 percent of average holding time in all three systems, and users may find it quite convenient to request two or three times as much output as they do presently. This increase in output will not severely affect average holding times, however, as even a tripling of computer send time would increase average holding time by less than a minute assuming a 360 character per second computer rate.

It should be noted that these figures are heavily dependent on such factors as user application. For example, if the application is not scientific programming but rather inquiry-response one might anticipate drastically shorter holding times. Consider the telephone directory assistance operator who makes a five second query of a computer that responds in one second with the beginning of a 1000 character transmission to be displayed with a video terminal. Assume that "holding time" is defined to be

\[ \tau = K + I + 1000 W_z. \]

For the numbers we have assumed,

\[ \tau = 6 \text{ seconds} + 1000 W_z. \]

his program to provide a paper copy for filing purposes but is low for the user who is checking every comma and parenthesis in order to find a program bug.

and at 10 characters per second, this results in a 106 second holding time of which 94 percent is computer send time. At 360 characters/second, \( \tau = 8.8 \) seconds but computer send time is still 32 percent of \( \tau \). At 4000 characters/second, \( \tau = 6.25 \) seconds of which four percent is computer send time. Figure 6 shows a graph of holding time versus computer channel speed for the short holding time example we have assumed.

### SUMMARY AND CONCLUSIONS

The results of a study of the communications considerations of serving multiaccess computer systems have been presented. A model of user-computer interaction, as observed at the communications interface, has been developed. Summary data from three “long holding time” computer systems have been given for the parameters of this model.

Examination of these data has revealed that

a. Computer introduced delays can be a large component of holding time and, above some threshold, are acutely sensitive to the number of simultaneous users. The largest component of computer delay occurs during those periods when the computer is outputting to remote users. The conclusions to be drawn from these findings stem from the consideration of holding time as being composed of periods of computer outputting activity, and periods of no computer outputting.
Of the inactive periods, some time is due to user-dependent delays and some to computer-dependent delays, some, such as execution time, may not be reducible and others, such as delays due to overhead, can be reduced. It should be noted that changes in the computer system such as changes to the scheduling algorithm, or changes to the communications control unit, can strongly influence computer delays. Thus, it is within the computing system that some reductions in holding time may be made resulting in communications economies. As not all computer delays can be eliminated in a heavily loaded system, the technical and economic feasibility of employing data multiplexers at the computer to decrease the number of access lines should be explored.

b. The average number of characters sent by the computer to the user is an order of magnitude greater than the number of characters sent by the user to the computer. If other parameters did not change drastically, the availability of higher transmission rates for computer outputting would effect significant reductions in average holding time. A computer transmission rate of 360 characters per second would reduce total computer send time below one half minute per call. To achieve this rate requires either high speed or asymmetric data sets and correspondingly higher speed output terminals.

c. Delays introduced by the user are a significant contributor to average holding time. The average user delays in the three systems reported are remarkably close in absolute values. As user delays are appreciable, the multiplexing of inputs from user terminals which are geographically clustered appears attractive and is being studied.

The data analyzed in this report are from systems having primarily scientific and business problem-solving applications. The users of such systems demonstrate a wide range of sophistication in their use of these systems. As users educate themselves in the efficient use of multiaccess computers and their terminals, the data traffic characteristics of these systems will change.

Studies of multiaccess computer communications systems are continuing. Data are being collected from systems with markedly different terminal types, average holding times, and user applications. Analyses of these data will allow the characterization of such systems in a manner analogous to that reported above for “long holding time” systems.

Implications of results

The implications of the results of this study extend into several aspects of computer communications. The study has produced quantitative indications of the degree to which computer operations can influence such communications parameters as holding time. The developers of computer systems, in turn, have noted that the provision of computer hardware and software to accommodate data communications is a major problem area. In order to jointly optimize the computation-communication solution to the problem, it is apparent that closer coordination between the computer and communications systems designers would be extremely fruitful in terms of economic and technological improvements to overall systems design.

One of the impediments to finding rapid and robust solutions to the problems of multiaccess computer communications has been the unavailability of data descriptive of the user-computer interaction process. The acquisition of the data reported here is a contribution toward the removal of that obstacle.

These data are currently being used in systems engineering studies at Bell Telephone Laboratories to further define the systems requirements for new systems and services to satisfy the needs of the multiaccess computer community.

The analyses made of these data support for the first time in more than qualitative terms some proposals proffered in the past as solutions to the data communications problems associated with multiaccess computer systems. The delays which are introduced by both user and computer suggest the possibilities for effective employment of multiplexing techniques. For example, it has been shown for the systems studied that the average total send time (the sum of user send time and computer send time) is as little as five to nine minutes. This corresponds to 15 to 40 percent of average holding time. For the other 60 to 85 percent of average holding time the communications channel is idle. One method of obtaining higher utilization of these facilities is by time division multiplexing. The following assumes a multiplexing technique in which the user channel is independent of the computer channel. Only one to five percent of average holding time (or three to eight percent of the average user burst segment) is user send time. Thus, for 95 to 99 percent of an average call, the user-to-computer channel is idle and could be made

* Multiplexing is not always economical, of course, despite the idle times. Other important considerations involve the geographical placement of terminals and computers and several statistical traffic characteristics other than average occupancy.
available to additional users. Average computer send times for the three systems were from 13 to 35 percent of average holding time (21 to 86 percent of the average computer burst segment) indicating that higher usage of the computer-to-user channel may be realized by the use of appropriate multiplexing techniques.

The asymmetric nature of the data flow in multi-access computer systems suggests that different transmission treatments may be appropriate for computer-to-user versus user-to-computer transmissions. The large volumes of computer-to-user data are an order of magnitude greater than volumes in the opposite direction. The provision of computer transmission rates of 100 to 200 characters per second could reduce average holding times up to 30 percent. As the user is capable of generating characters for transmission at a much slower rate, the application of asymmetric channels or data sets receives quantitative support and is now being studied. Provision for higher computer transmission rates would require, of course, user terminals with accordingly higher input rates.

The final conjecture receiving quantitative support from the above analyses is that users themselves contribute substantially to the communications costs of their real-time computer access calls by introducing delays. Some of these delays are likely to decrease as users gain proficiency. Others are due to the inveterate characteristics of human users. As they pertain to the use of communications and to the use of computers1,2 these characteristics are being intensively studied to enable the design of versatile and responsive computer/communications systems.

ACKNOWLEDGMENTS

Many people have contributed their efforts to various parts of this study. Data acquisition was accomplished with the considerable help of the American Telephone and Telegraph Company and the Bell System Operating Companies. Contributions to the model and the analyses and many helpful criticisms were made by Messrs. E. Fuchs, R. J. Price and R. J. Roddy, all of Bell Telephone Laboratories.

Our special thanks are extended to the companies whose computer systems are being studied. Without their full permission and very helpful cooperation these analyses would not be feasible.

REFERENCES

1 A L SCHERR
An analysis of time-shared computer systems
Massachusetts Institute of Technology Project MAC
Thesis MAC-TR-18 June 1965
2 T G RAYNAUD
Operational analysis of a computation center
Operations Research Center MIT July 1967
3 H SACKMAN
Experimental investigation of user performance in time shared computer systems retrospect, prospect, and the public interest
System Development Corporation May 1967
4 G E BRYAN
JOSS: 20,000 hours at the console, a statistical summary
Rand Corporation August 1967
5 P E JACKSON
A fourier series test of goodness of fit
To be published
6 ANON.
Switching systems
AT & T Co 1961
7 M GREENBERGER
The priority problem
MIT November 1965
8 E G COFFMAN JR
Analysis of two time sharing algorithms designed for limiting swapping
J A C M July 1968
9 E G COFFMAN JR, L. KLEINROCK
Computer scheduling methods and their countermeasures
Proc S J C C 1968
10 L J COHEN
Theory of the operating system
The Institute for Automation Research Inc 1966
11 W F BAUER, R H HILL
Economics of time shared computing systems—Part II
Datamation Vol 13 No 12 41–49 December 1967
12 W E SIMONSON
Data communications: The boiling pot
Datamation Vol 13 No 4 22–25 April 1967
13 W W CHU
A study of the technique of asynchronous time division multiplexing for time-sharing computer communications
Proc of the 2nd Hawaii International Conference on Systems Sciences January 1969