# DELAY-THROUGHPUT AND BUFFER UTILIZATION CHARACTERISTICS OF SOME STATISTICAL MULTIPLEXERS

**M. R. Karim**



***M. R. Karim*** *is a member of technical staff in the Services Access Engineering Department at AT&T Bell Laboratories in Holmdel, New Jersey, where he works on transmission systems for Universal Information Services. Mr. Karim joined the company in 1972. He holds B.S., M.S., and Ph.D. degrees in electrical engineering from, respectively, the University of Engineering and Technology at Bangladesh, Manchester University, and Stevens Institute of Technology.*

This paper presents the delay-throughput and buffer utilization characteristics of a few statistical multiplexers. Two of the various schemes that may be used to access the multiplexer are considered: the point-to-point access scheme where the traffic is carried between each user and the multiplexer over a separate link in a star topology, and the carrier sense multiple access protocol with collision detection (CSMA/CD) that is used over a shared communication link. Results from computer simulations show that for the $n \times 1$ statistical multiplexer using the star topology, the delay and buffer size utilized increase exponentially with the channel utilization factor $\rho$. The delay-throughput characteristics of a well designed multiplexer using the CSMA/CD access protocol are comparable to those of the $n \times 1$ multiplexer based on the star topology. With the CSMA/CD protocol, as the packet size is increased, the delay normalized to the packet size, while a function of $\rho$, first decreases and eventually tends to a limit.

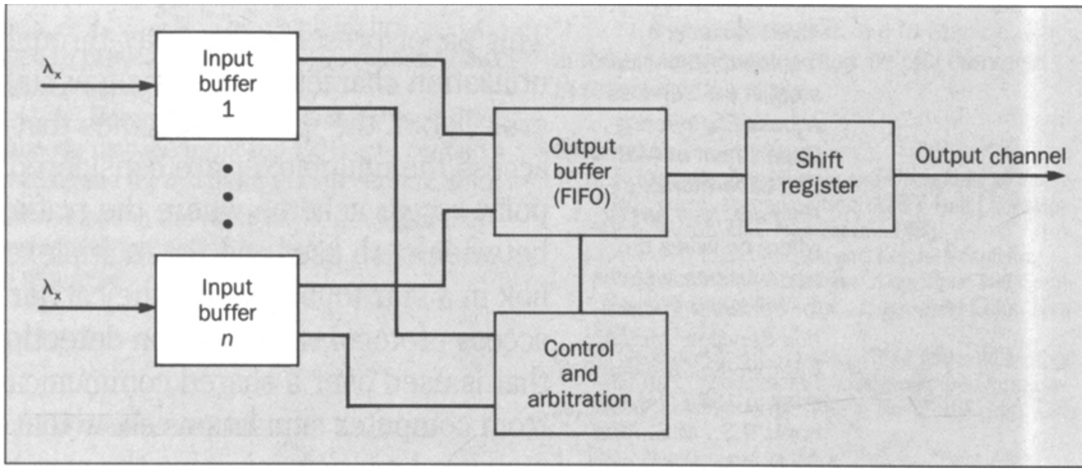


## Introduction

Currently, there appears to be a trend toward packet switching in telecommunications networks. Packet switching offer customers an end-to-end digital connectivity by integrating voice, data, image, and video economically and conveniently on a single physical link. In addition, by exploiting the bursty nature of the traffic, a packet switch can provide variable bandwidths in a more graceful way than a circuit-switched system.

Consider, for example, a normal voice conversation. Statistically, talk spurts appear only 40 percent of the time.[1,2] Similarly, a user

**Figure 1. An $n \times 1$ statistical multiplexer based on the point-to-point access scheme.**






terminal connected to a host computer sends out only a few characters per second, while the capacity of the computer may be several orders of magnitude higher.[3] Consequently, to use the available bandwidth effectively, the output of a user device is packetized and then statistically multiplexed with similar outputs of other devices before being transported to a packet switch.

User devices may be connected to a statistical multiplexer in several different ways. For example, they may access the multiplexer over dedicated, separate wire pairs as in a point-to-point distribution system.[4,5] Or, they may do so over a shared communication link, as in Xerox Corporation's Ethernet™ local area network, and contend for the resources at the same time (see below). This type of multiple access scheme has been widely used not only in local area networks[4,6] but also in packet switching systems with radio channels providing remote terminal access to host computers.[7-10]

Broadly speaking, there are two types of multiple access schemes. In one of them, a central controller polls each user sequentially or a group of users all at once. In this case, ready users are passive in the sense that they can transmit only in response to a query. In the other, the control is distributed rather than centralized, and one or more ready users may attempt to transmit at the same time using some algorithm. In this case, the users may be considered active.

A number of variations of the distributed control multiple access scheme are possible: the "Aloha,"[7,9] the "slotted Aloha,"[8,9] the non-persistent carrier sense multiple access (CSMA), and the 1-persistent and p-persistent CSMA[9] protocols. In the non-persistent CSMA protocol, each user monitors the channel before transmitting. If the channel is sensed idle, the ready user transmits its packet. Otherwise, it reschedules its transmission for a later time according to some retransmission delay distribution, and repeats the process all over again. Obviously, because of the non-zero propagation delay between any two users of the system, collisions are still possible.

In the CSMA/collision detection (CSMA/CD) protocol,[12,13] as in the CSMA scheme, each user monitors the channel before transmitting. If the channel is busy, the user does not transmit. If the channel is sensed idle, it transmits its packet, while at the same time listening to the channel to see if there is a collision. If there is one, the contenders, using the so-called collision consensus enforcement procedure, send a few random bytes to jam the channel to ensure that the collision on the channel is

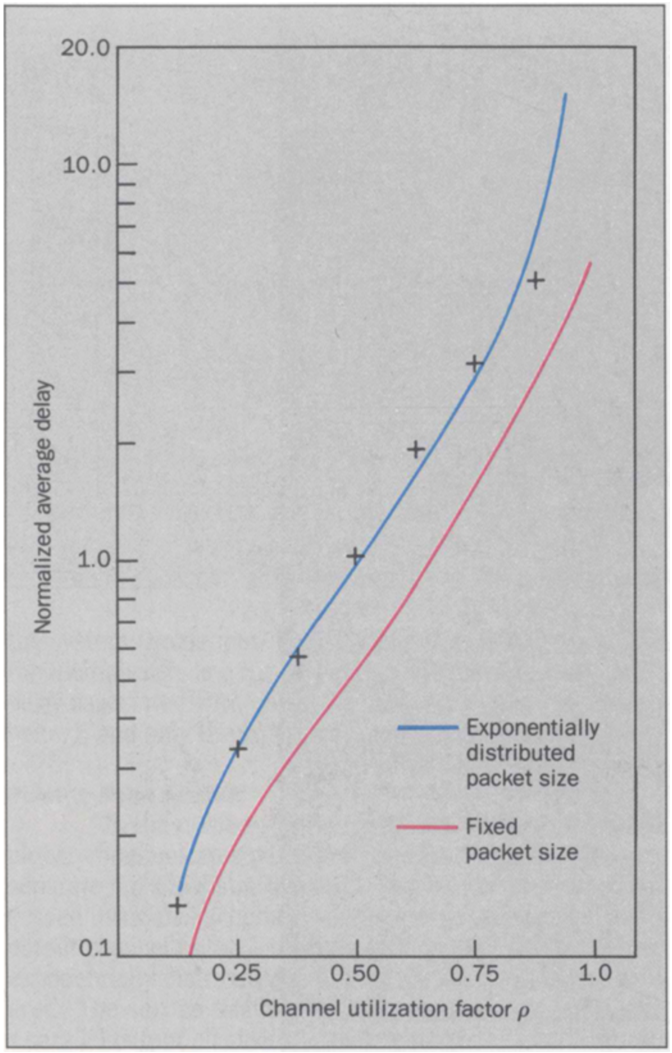



**Figure 2. The expected delay encountered by a packet in the statistical multiplexer of Figure 1 as a function of the channel utilization, ρ.**

detected by all contending users. Each user then stops, waits for a random period of time and then attempts to retransmit on the channel. Ethernet local area networks use the binary exponential back-off algorithm or some variation of it that allows the range of retransmission delays to be increased under heavy load conditions, thereby reducing the probability of repeated collisions.

This paper gives performance of some statistical multiplexers using different access schemes. Specifically, it presents the following parameters: the expected delay as a function of the channel utilization factor, the cumulative delay distribution, the expected value of the buffer size required as a function of the channel utilization, and its cumulative distribution.

### General Assumptions

For the remainder of this paper, several general assumptions apply. First of all, we suppose that packets originate from infinitely many sources. The packet arrival on each link is an independent, identically distributed random variable with exponential distribution; when all links are taken together, these random variables form a Poisson process. We will assume that the aggregate rate at which the packets arrive in the system is $\lambda$ per second, and that if there are $n$ input links, the packet arrival rate over each of these links is $\lambda_x = \lambda/n$.

In any packet communication system, some data communications protocol is used to insure the data integrity at the receiving end. One of the important functions of such a protocol is to overcome the effects of channel impairments that inevitably corrupt the transmission. Error-detecting codes are used to find errors in the message stream; then, if they detect errors, the transmitting end is asked to repeat the transmission, introducing an additional end-to-end delay. In this article, since no specific protocol is used, that additional delay is ignored. Similarly, since the line error rates and the exact error recovery procedures are among the factors that determine

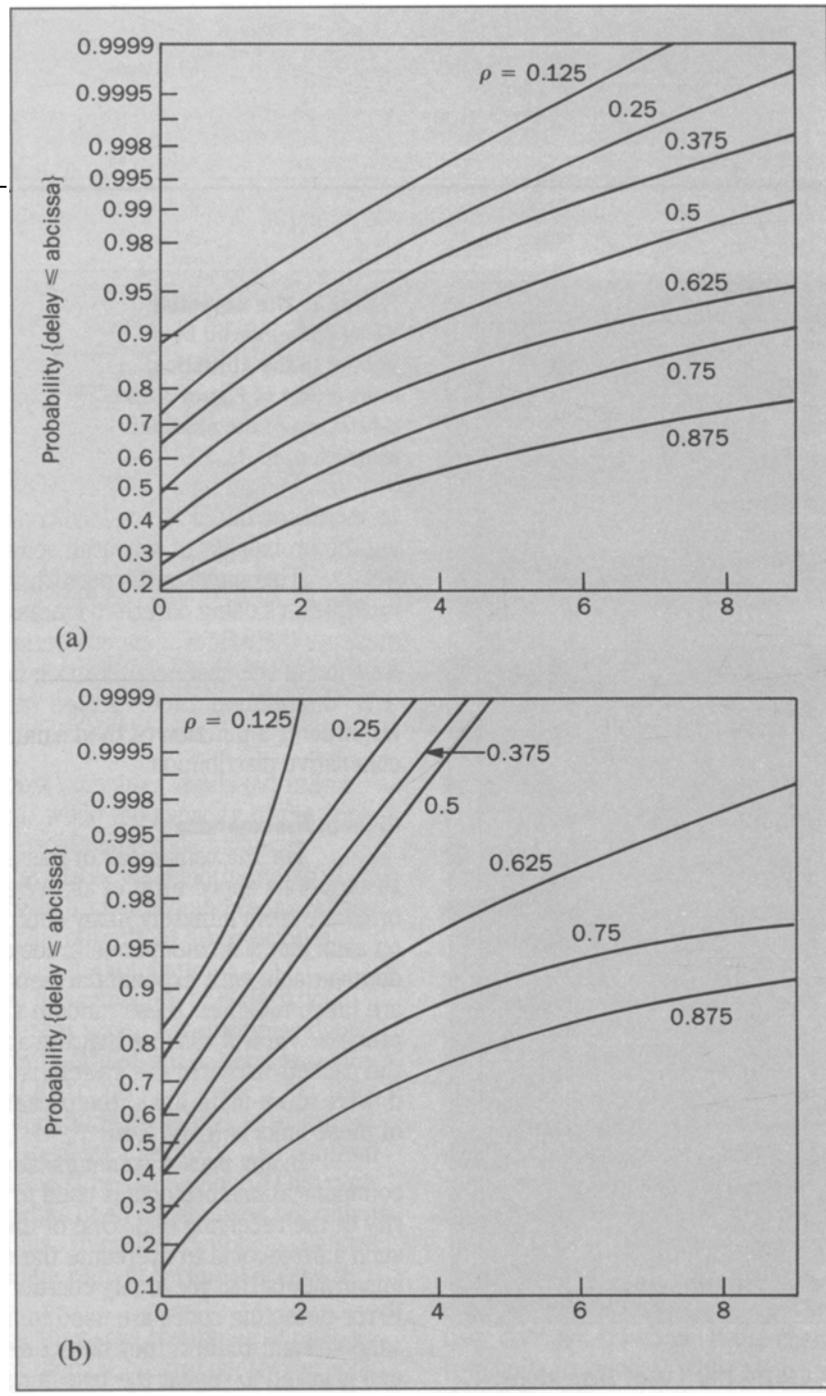



**Figure 3. The cumulative distribution for the statistical multiplexer of Figure 1 with ρ as a parameter. (a) Packet sizes are exponentially distributed. (b) Packet sizes are all fixed at unity.**

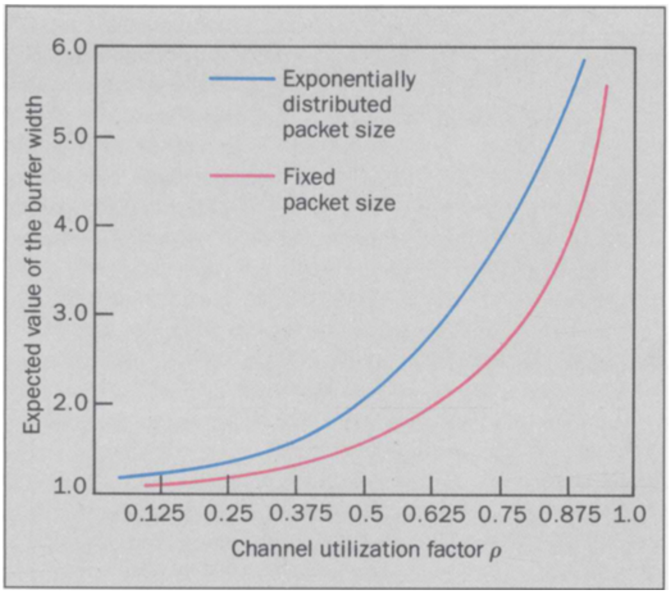



**Figure 4. The expected value of the buffer width utilized at the instant of a packet arrival in the statistical multiplexer of Figure 1 as a function of the channel utilization, ρ.**

the system throughput, the throughput characteristics of the multiplexers are not presented. Furthermore, the delay associated with polling or processing is ignored (see below), and only the buffering delay is considered.

**Point-to-Point Access**

In the point-to-point access mechanism the multiplexer communicates with each user on the input side over separate links in a star topology. Two multiplexers are discussed using this scheme. The first of them has a single output channel ($n \times 1$ multiplexer). In this case both exponentially distributed and fixed packet sizes are considered. The second multiplexer consists of $n$ input links and $c$ parallel output channels ($n \times c$ multiplexer), and uses a fixed packet size for any utilization factor. For each type of multiplexer, the delay-throughput, the buffer size requirement, and the probability distribution are given.

**$n \times 1$ Multiplexer with Point-to-Point Access.** Figure 1 shows a block diagram of the $n \times 1$ statistical multiplexer. Various polling protocols can be used to search the input buffers for any packet arrival.[11] In one such simple protocol, a controller sequentially polls each link. When it finds a packet in an input buffer, as indicated by a flag, it reads the packet, transfers it to the output buffer, resets the packet arrival flag and goes to the next input link. Obviously, the polling and the data transfer process will add to the overall delay that a packet undergoes in the system. This additional delay depends on the exact search mechanism, the number of input links being searched for packet arrivals and the number of ready links at the polling instant.[11] It is possible, however, to design the circuit such that the time required to search the ready links and transfer the packets to the output buffer is negligibly small compared to the interarrival times of the packets or the delay due to queueing in the buffers. Hence, in this section, this delay will be ignored.

If the output buffer contains no packet, the control element of the statistical multiplexer waits until one arrives. If the buffer contains a packet but the output channel is busy, it waits until the channel is idle. In this case, when the output channel becomes idle, the control element of the multiplexer takes the packet that is at the head of the queue, and sends it out on the output channel. We can visualize the output buffer as a first-in, first-out (FIFO) memory with a mechanism to convert the parallel data into a serial form.

**Exponentially distributed packet size.** We will assume here that the size of a packet is a random variable with the exponential distribution.

Results are obtained from computer simulations.[14] The blue curve of Figure 2 is the delay, averaged over all packets, as a function of the utilization factor ρ. Since the average delay increases linearly with the average value of

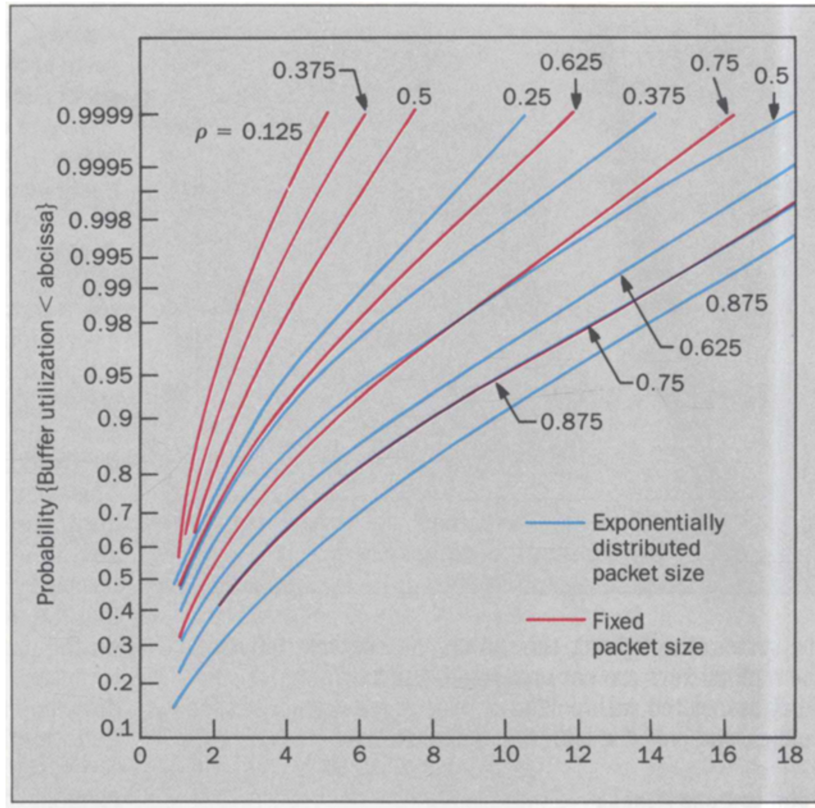



**Figure 5. The cumulative distribution of the buffer width utilized at the instant of a packet arrival in the statistical multiplexer of Figure 1 with channel utilization ρ as a parameter.**

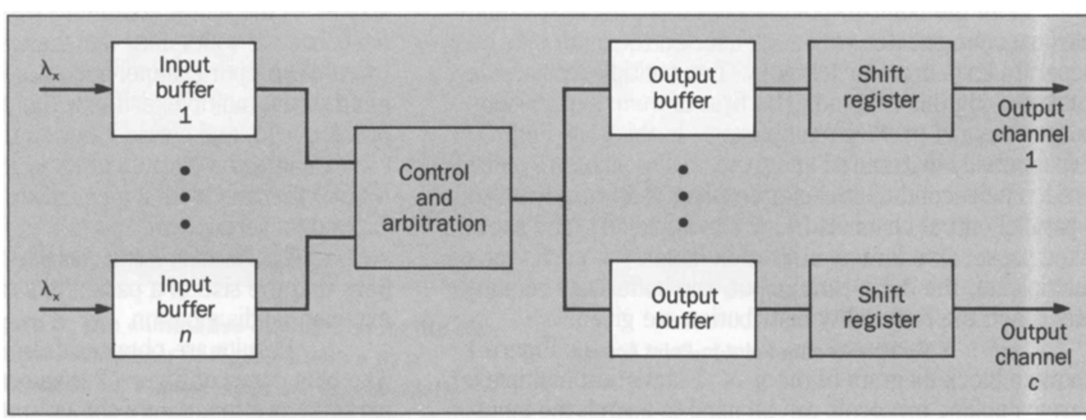



**Figure 6. An *n* × *c* statistical multiplexer with *n* input links and *c* parallel output channels. Each input buffer is capable of buffering one or more packets. Each arriving packet is constant in size.**

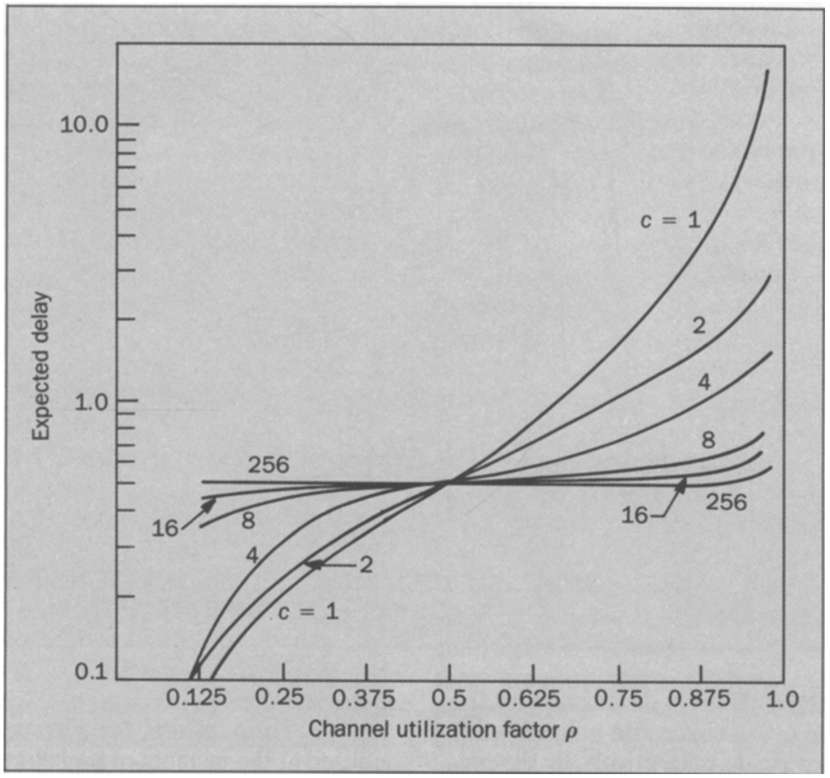



**Figure 7. The expected delay encountered by packets in the multiplexer of Figure 6 as a function of ρ with c as a parameter. Each packet size is fixed and is equal to unity.**

the packet size, the curves are normalized with respect to the latter.

Figure 2 can also be obtained analytically by considering the M/M/1 queue for the exponential packet size (and the M/D/1 queue for constant packets). The analytic results that are available are summarized in Appendix A. Values of the average delay as predicted analytically are shown alongside the blue curve of Figure 2; for $\rho \leqslant 0.75$, the agreement between the theoretical values and the simulation results is excellent. Having thus established the accuracy of our simulation, we will use it to obtain other parameters of interest for which theoretical results are not available.

Figure 3(a) gives the cumulative distribution for the delay with ρ as a parameter. The expected value of the buffer size utilized is given by the blue curve of Figure 4. Strictly speaking, the buffer size given in this figure is not the time average. Instead, it is the value of the buffer size utilized at the instant a packet arrives, averaged over all such discrete instants. It is done this way because it gives an upper bound on the buffer size required in order that the probability of buffer overflow does not exceed any given value (see the example below). The blue curves of Figure 5 are the cumulative distribution of the buffer size utilized.

**Constant packet size.** If the packet size does not vary randomly, but remains constant for all packets, the expected value of the delay is as given by the red curve of

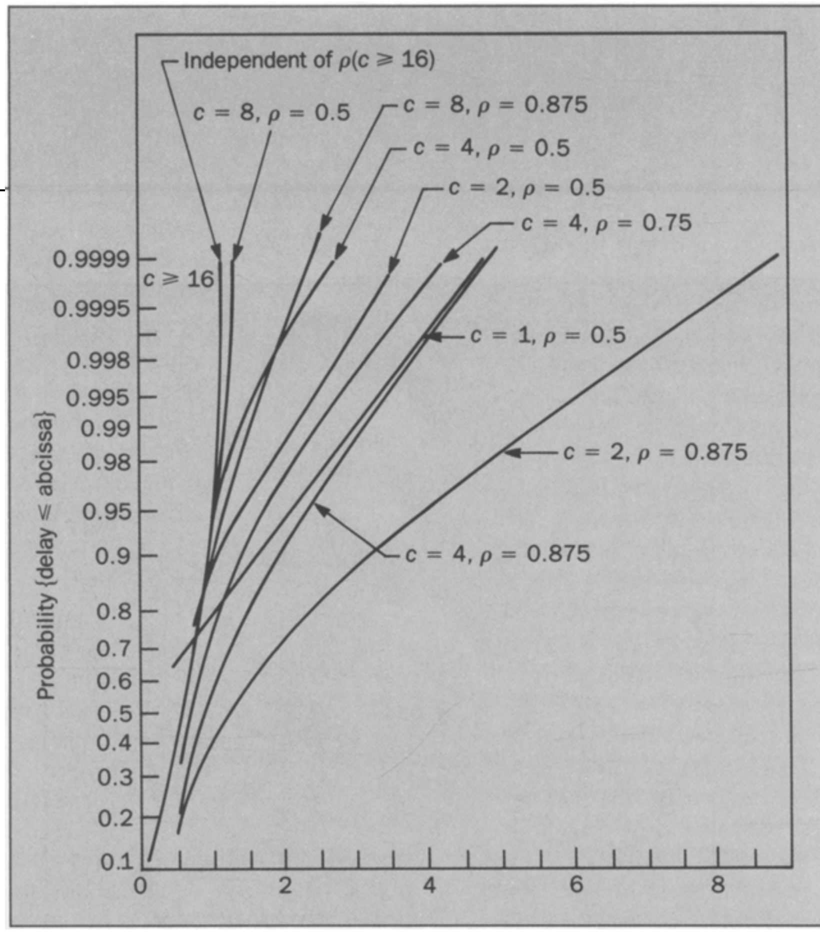



**Figure 8. The cumulative display distributions for the multiplexer of Figure 6 for different values of ρ and c. Each packet size is fixed and is equal to unity.**



Figure 2. Figure 3(b) is the cumulative distribution for the delay. The expected value of the buffer size utilized and its cumulative distribution are given, respectively, by the red curves of Figures 4 and 5.

**Discussion of results.** Suppose, for convenience, that the expected value of the exponentially distributed packet size is unity. Hence, in this case, packets are serviced at an average rate of $\mu = 1$. For example, assume that $\rho = 0.5$. Then, since $\mu = 1$, $\lambda = \mu\rho = 0.5$. So, if the unit of time is taken to be, say, a millisecond, packets, which are on the average one millisecond long, arrive at a rate of 500 per second and are serviced at a rate of 1000 per second. These packets, then, as seen from the curve, are each delayed on the average approximately 1 ms. For the same value of ρ, the probability that packets encounter delays longer than 1 ms is 0.32, as seen from Figure 3(a). About 1 percent of the packets will be delayed more than 8 ms. If the traffic is reduced so that the utilization factor is, say, 0.25 then only 0.05 percent of the packets are delayed 8 ms or more.

For $\rho = 0.5$, the average value of the buffer size utilized at the instants of packet arrivals is seen from Figure 4 to be about 2 ms, or twice the average packet size. In other words, if the data rate is, say, 1 Mb/s, and if the average packet size is 1 ms or 1000 bits, the average buffer utilized is 2000 bits long. The buffer utilization has a large variance. For example, about 15 percent of the time, it exceeds 4 ms and about 1 percent of the time, it is more than 10 ms. Thus, assuming that the statistical multiplexer has an output buffer equal to ten times the average size of the packet, the probability that the buffer will overflow is 1 percent for this value of ρ. In this case, the carried load for the multiplexers will be less than the offered load by the same amount, whereas if the buffer were unlimited, the two would be exactly the same.

Refer to Figure 2 again. Suppose that the expected packet size is doubled but that the utilization factor is maintained at a value of 0.5 by decreasing the packet

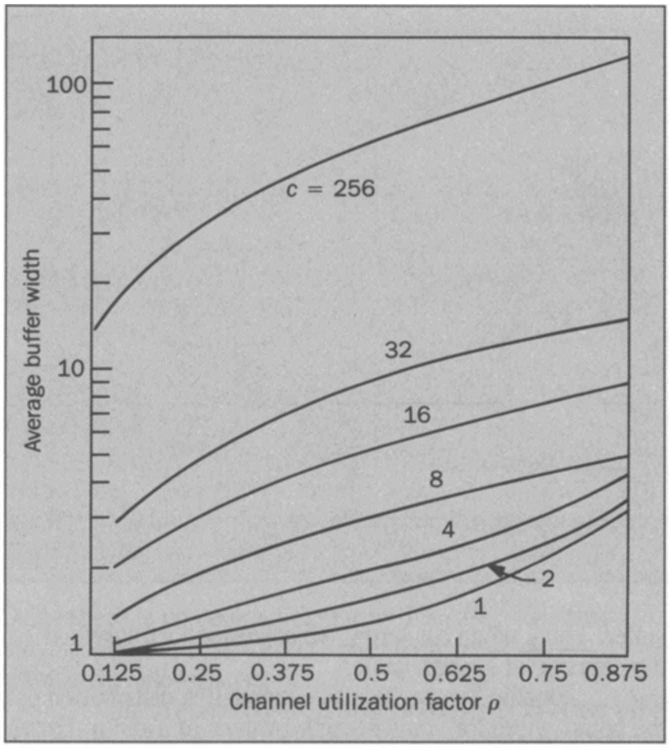



**Figure 9. The average buffer utilization at the instant of a packet arrival for the multiplexer of Figure 6. All packets are size unity.**

arrival rate. The average delay, then, is seen to have increased to 2 ms.

As for the case of the constant packet size, suppose that all packets are 1 ms long. From Figure 2, for $\rho = 0.5$, the expected value of the delay is about 0.5 or half the value for the previous case where the packet size is exponentially distributed. From Figure 3(b), for $\rho = 0.5$, about 22 percent of the packets are delayed longer than 1 ms and only 0.07 percent longer than 4 ms. Similarly, the average buffer utilization for $\rho = 0.5$ is, from Figure 4, about 1.5 ms. Figure 5 indicates that the buffer utilization exceeds 4 ms with a probability of 2 percent and exceeds 7.5 ms with a probability of only 0.01 percent. Thus, the delay and buffer size requirements significantly improve if packets are fixed in size and do not vary randomly.

**$n \times c$ Multiplexer with Point-to-Point Access.** We now consider a statistical multiplexer with n input links and $c$ parallel output channels.[15] The packet size for each link and channel is identical and constant. It is assumed for simplicity that $\mu = 1$, thus making the packet size unity. The results, however, are valid for all $\mu$.

Refer to Figure 6. Each input link is associated with an input buffer that is capable of buffering one or more packets. Each output channel consists of a single buffer that is the length of a packet with some mechanism for converting the data from the buffer into a serial form.



At the end of each service period, which is the same for all packets, the arbitration and control logic determines if there are any packets in the input buffers. If there are none, the multiplexer waits until one or more packets arrive. As soon as they arrive, they are serviced immediately. If there are $c$ or fewer packets in the input buffers at the end of the service period, they are sent out over the output channels. In this case, packets are delayed only a fraction of a service period. We assume that during a service period, no new packet arrivals are sent to the idle output channels. If there are more than $c$ packets in the input buffers, some of them must wait until the next service period when these ready packets plus any others that might have arrived during the last service period, up to a maximum of $c$, are sent out over the output links. Obviously, for equilibrium $c > \lambda$. We have chosen this service mechanism because it leads to some interesting characteristics, as we shall show shortly. Reference 16 analyzes some special cases of this queueing model and gives approximate values of some parameters of the queue. There are, however, no closed form expressions in the literature for the parameters of concern to us except when $c = 1$ (see below).

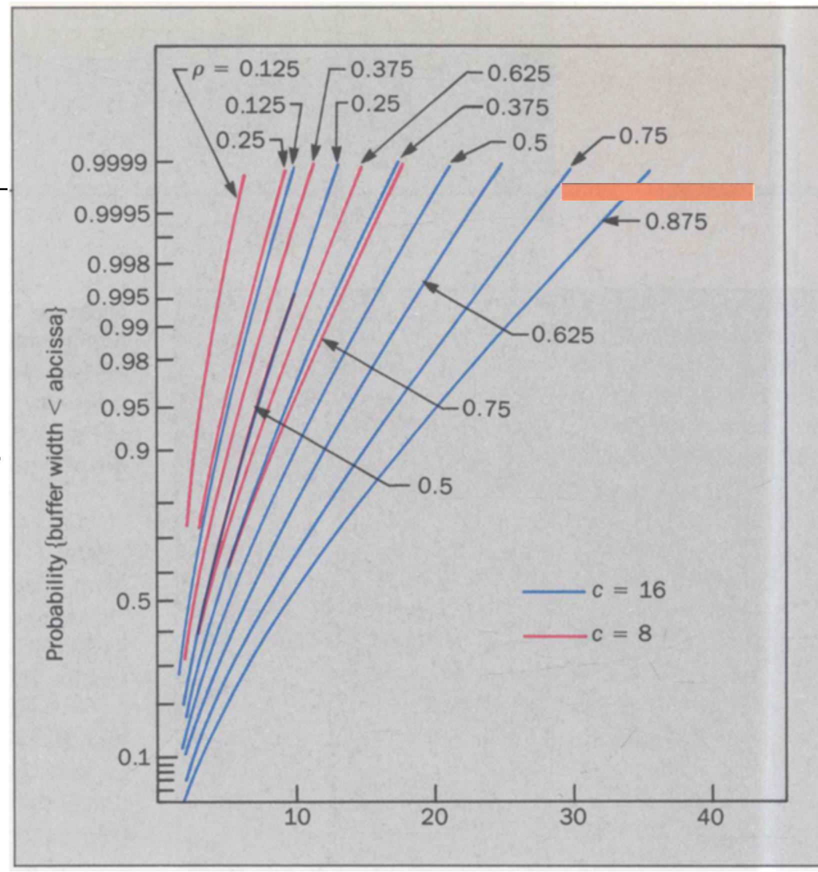



**Figure 10. The cumulative distributions of the buffer utilization for the multiplexer of Figure 6 for different values of ρ and c. Each packet size is fixed and is equal to unity.**

**Discussion of results.** Figure 7 gives the expected delay as a function of $\rho$ with $c$ as a parameter. Notice that when $c = 1$ the situation becomes exactly the same as for the $n \times 1$ multiplexer with constant packet size. Assuming that each packet is 1 s long, the throughput of the multiplexer is given by $\lambda = c\rho$. Hence, for any given $\rho$, the throughput increases linearly with $c$. At the same time, from Figure 7, as $c$ increases, the average delay decreases for $\rho > 0.5$ and increases for $\rho < 0.5$. (However, for any value of $c$, the delay $\rightarrow 0$ as $\rho \rightarrow 0$.) Interestingly, when $c = 8$ or more, the delay becomes virtually constant at about 0.5 for a large range of $\rho$. This asymptotic value of the delay for $c \geqslant 8$ and $\rho < 1.0$ is the same as for a multiplexer with $c = 1$ operating at $\rho = 0.5$. Thus, it is possible to operate an $n \times 8$ multiplexer at a high utilization factor and hence achieve a high throughput without a long delay.

Figure 8 shows the cumulative delay distribution for a few values of $\rho$ and $c$. For smaller values of $c$ ( $c < 8$ ), the distribution greatly depends not only on $c$ but on $\rho$ as well. For larger values of $c$, however, this dependence on $\rho$ diminishes. For example, assume that each packet is 1 ms long. Then, with $c = 8$, the delay exceeds 1.2 ms for $\rho = 0.5$ and 2.4 ms for $\rho = 0.875$ with a probability of 0.01 percent.

Figure 9 is a plot of the expected buffer width as a function of $\rho$ with $c$ as a parameter. Observe that the buffer requirement increases with $c$. It is important to note here that the buffer width shown in this figure is the aggregate buffer for all links taken together and not the per link buffer. Thus, when $c$ is large, we achieve short delays at the expense of a larger buffer.

Finally, Figure 10 is the cumulative distribution of the buffer width for some selected values of $c$ and $\rho$. For example, assuming the packet size to be 1 ms, when $c = 16$ and $\rho = 0.5$, the expected value of the buffer size utilized is seen from Figure 9 to be about 5.2 ms, whereas, as seen from Figure 10, about 50 percent of the time the buffer utilization will exceed this value and 10 percent of the time it will exceed even 10 ms.

### $n \times 1$ Multiplexer with CSMA/CD Access Scheme

The second access mechanism considered is the carrier sense multiple access protocol with collision detection. This protocol has been extensively used in many local area networks including Ethernet, and is suitable for a variety of broadcast channels such as radio, cables, twisted pairs, etc. The same access mechanism can also be used in a statistical multiplexer.

The access scheme is described in the Introduction. Our discussion is restricted to an active system.

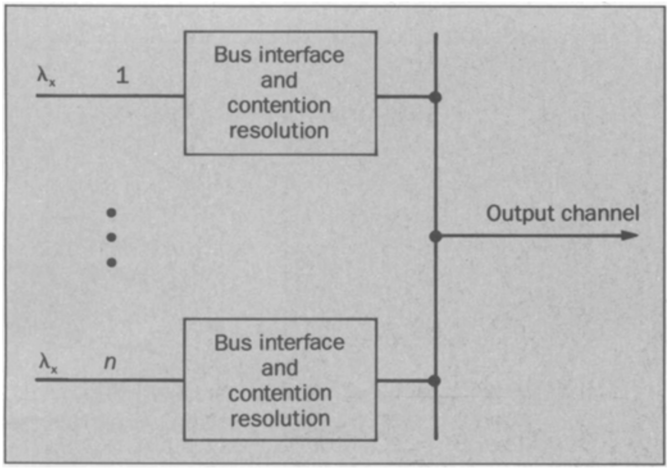



**Figure 11. An *n* × 1 multiplexer using the CSMA/CD access scheme.**

Since there is no common buffer in this case, our results only include the delay-throughput characteristics and cumulative delay distribution.

The average delay in this protocol, as in other protocols, depends on the channel utilization factor and the packet size. In addition, there are two other factors that contribute to the delay: the so-called carrier sense time or the propagation delay, $\alpha$, between two farthest stations connected to the shared channel, and the back-off or retransmission delay used by each party involved in a collision. Since the probability of a collision increases as $\alpha$ becomes larger, the average delay also increases with $\alpha$. Similarly, it increases as the range, $D$, of the back-off delay is increased, more significantly so for larger values of $\rho$ than for the smaller ones.

A local area network based on this protocol generally uses the binary exponential back-off algorithm whereby the parameter $D$ is increased in steps of the round-trip delay as the load on the system increases. In our study, for simplicity, $D$ is assumed to be constant throughout. Secondly, in the event of a collision, a station retransmits on the channel as many times as necessary until the transmission is successful.

A multiplexer with this access mechanism is shown in Figure 11. Figure 12 shows how the expected delay with this protocol varies as a function of $\rho$ with $\alpha$ as a parameter. In this protocol, the ratio of the average delay to the average packet size (the delay normalized to the packet size) is not constant but varies with the average packet size itself. Consequently, in Figure 12, the packet size is held constant at 663 μs (see the example below for this particular choice of packet size). Similarly, the carrier sense time, $\alpha$, and the range, $D$, of the back-off delay are also given in seconds. Furthermore, in this and Figures 13 and 14, $D$ is held constant at 10 ms. Thus, the back-off delay is assumed to be uniformly distributed between 0 and 10 ms. To see how the parameter $\alpha$ relates to the length of the physical channel, refer to the example below. In Figure 12, $\alpha$ assumes only three values: 100 μs, 10 μs and 0. Thus, $D/\alpha \geqslant 100$. This is a reasonable value since in many Ethernet applications this ratio is about 256.



Figure 13 shows the effect of varying the packet size on the average delay, which is expressed in multiples of the packet size. The parameter $\alpha$ is held at zero. This value of $\alpha$ is chosen because in a statistical multiplexer, as opposed to a local area network, the shared bus would very often be short enough to result in a negligibly small propagation delay. This implies that there are no collisions at all on the bus. For a given value of $\rho$, the delay normalized to the packet size decreases as the packet size increases. (The absolute value of the delay increases, though, with larger packets.) For a given offered load corresponding to a given value of $\rho$, fewer transmissions are required with larger packets. The result, therefore, is fewer collisions and consequently a higher system throughput.

Comparing Figures 2 and 13, we conclude that over a large range of packet size and channel utilization, the average delay for the CSMA/CD protocol is comparable to that of an $n \times 1$ multiplexer with the star topology. And for larger packets ($\geqslant$ 10 ms), it is uniformly smaller for all values of $\rho$.

A similar set of delay-throughput characteristics can be obtained for non-zero values of $\alpha$. We will not present those characteristics here but simply state our

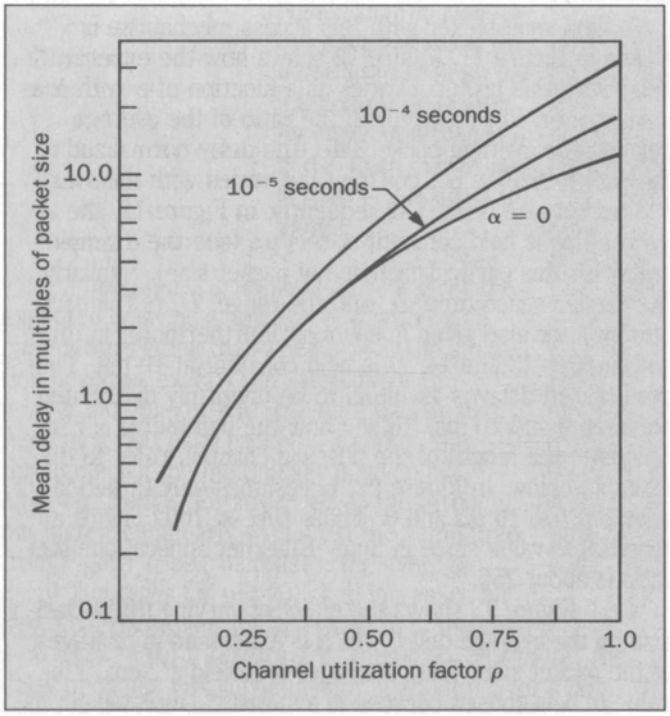



**Figure 12. The average delay in the CSMA/CD protocol as a function of the channel utilization for a few values of the carrier sense time, α, which is a measure of the round-trip propagation delay and hence the length of the shared medium. The packet sizes are exponentially distributed with a mean value of 663 μs.**

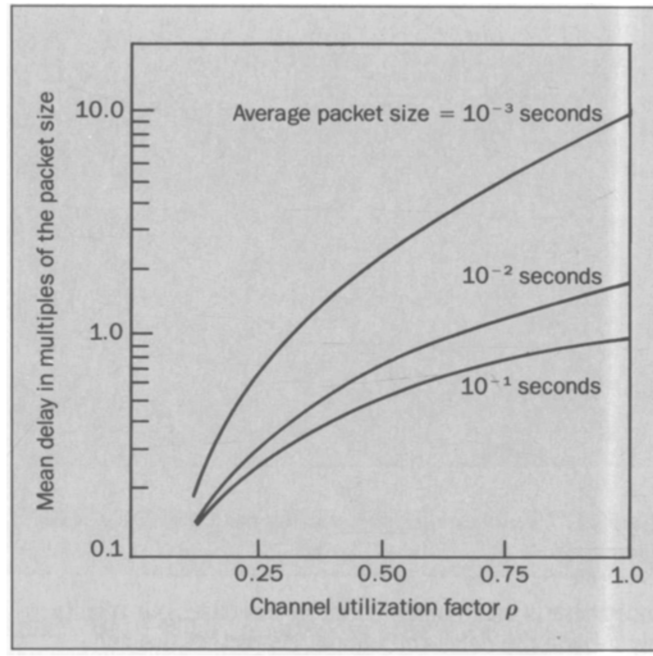



**Figure 13. The average delay for the CSMA/CD protocol as a function of the channel utilization with the packet size as a parameter. The packet sizes are exponentially distributed. The carrier sense time, α, is fixed at zero, indicating a bus of negligible length.**



observation that for any packet size, they are similar to those of Figure 12.

Although not explicitly shown in this article, the delay tends to a limit as the packet size increases. This is somewhat noticeable in Figure 13. For a well-designed system, this limit depends mostly upon the packet arrival and service rates, and not much on the carrier sense time or back-off delay. It can, therefore, be looked upon as the average queueing delay experienced by a packet, and hence is a function of the channel utilization.

Figure 14 is the cumulative delay distribution as a function of ρ for exponentially distributed packet sizes with a mean value of 663 μs.

**Example**

The example below illustrates the use of our results. Consider a point-to-point private network where the customer premises equipment are in two geographically separate locations and are connected together by packet links operating at 1.544 Mb/s, the standard T1 carrier rate. Assume that at either end there are various types of data equipment, which are broadly classified into three categories: slow speed terminals (generally asyn-

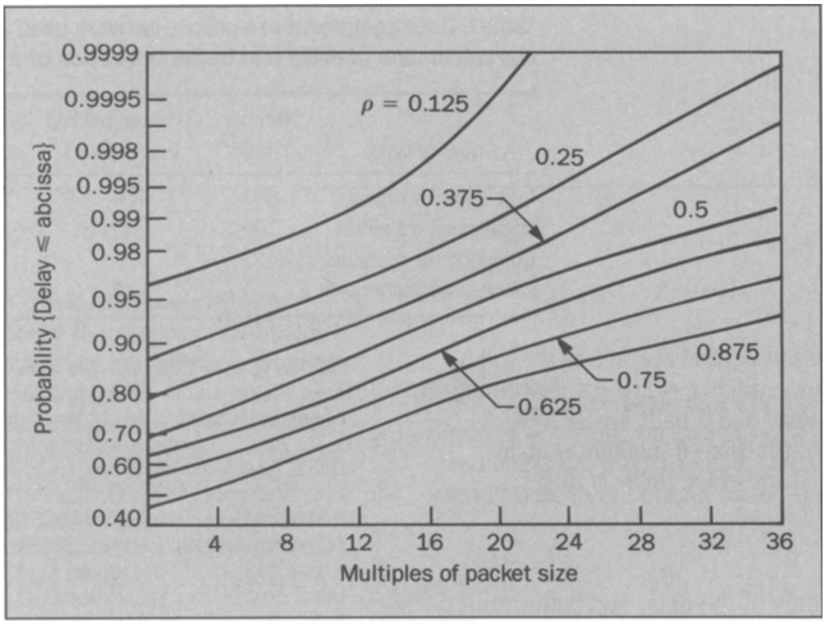



**Figure 14. The cumulative delay distribution for the CSMA/CD protocol as a function of the channel utilization. The packet sizes are exponentially distributed with a mean value of 663 μs. The carrier sense time, α, is assumed to be zero.**

chronous); moderate or high-speed terminals (generally synchronous); and high-speed data ports for computer to computer communications. The number of terminals in the network is large enough to justify the assumption that the packet arrivals form a Poisson process.

In the first case, it is assumed that the terminals and computer ports are unbuffered so that they all connect to the network at their own characteristic rates. Furthermore, the data equipment is sufficiently varied so that it is reasonable to assume the packet sizes to be exponentially distributed. Table I gives the number of devices in the network, the average circuit holding time and the associated data rate for each, and the total traffic (in bits/hour) generated in a busy-hour period.

The total number of data bits generated during the busy hour period as shown in the table is $2.76 \times 10^9$. Assuming a 25 percent overhead for the packet transmission, the utilization factor for a single output link is given by

$$\rho = \frac{1.25 \times 2.76 \times 10^9}{1.544 \times 10^6 \times 3600} \approx 0.62$$

Assume that the average packet size is 128 bytes. This packet size, at 1.544-Mb/s data rate, is 663.21 μs. From Figure 2, with $\rho = 0.62$, the expected value of the normalized delay for the statistical multiplexer of Figure 1 is 1.6. In absolute values, this delay is $1.6 \times 663.21$ μs $= 1061$ μs. From Figure 3(a), about 27 percent of all arriving packets are subjected to delays exceeding this value. Figure 4 gives the average normalized output buffer to be 2.75. Its absolute value is $2.75 \times 128 = 352$ bytes. But suppose we want the output buffer overflow probability to be 0.05 percent. Then from Figure 5 the required buffer, normalized to the packet size, is to be 18. Thus the output buffer should be 2304 bytes.



In the second case, suppose that all terminals and computer ports are buffered, and that all packets have a fixed size and are 128 bytes long. Referring to the above figures, one finds that the average delay in this case is 557 μs, that 36 percent of all packets experience delays greater than 557 μs, that the average output buffer utilized is 250 bytes and that for the same buffer overflow probability of 0.05 percent, we require a buffer that is approximately 1280 bytes long.

Consider now the multiplexer of Figure 11 with the CSMA/CD access scheme. Each data source enters the multiplexer at its own characteristic rate and is connected to the multiplexer at a bus of negligibly small length ($\alpha = 0$). The average packet size is again taken to be 663.21 μs. Referring to Figure 12 and using $\alpha = 0$ and $\rho = 0.62$, we find the average delay, expressed in multiples of the packet size, to be 5.2 with an absolute value of 3.45 ms. Referring to Figure 14, the probability that a packet experiences a longer delay is about 0.29.

**Table I. Data equipment in a private network used as an example to illustrate the calculation of delay and buffer utilization of a statistical multiplexer.**

| Source type | Number of units | Average holding time (mins.) | Average data rate (bytes/s) | Number of bits in a busy hour |
|---|---|---|---|---|
| Low-speed terminals | 600 | 54 | 2 | $3.11 \times 10^7$ |
| High-speed terminals | 100 | 40 | 300 | $5.76 \times 10^8$ |
| Computer-to-computer communication ports | 25 | 45 | 4000 | $2.16 \times 10^9$ |

For nonzero values of $\alpha$, this average delay will be higher. Suppose, for instance, that $\alpha = 10$ μs. Assuming a round-trip propagation delay of 3.4 ns/ft for 24 AWG twisted pairs, the length of the shared medium is about 2,941 ft. From Figure 12, the average delay is then 3.91 ms.

### Conclusion

This paper is a study of the delay and buffer utilization characteristics of statistical multiplexers using a few different access protocols. Wherever meaningful, we have presented the average values and cumulative distributions of the delay and buffer utilization. Two of the multiplexers use a central controller and are accessed by each individual data source over a physically separate link. The third multiplexer has no central controller, and uses the CSMA/CD access protocol. Our study shows that for the $n \times 1$ multiplexer based on the point-to-point access mechanism, the delay and buffer utilization increase with $\rho$. In this case, as the packet size is increased, the delay normalized to the packet size remains constant for any given $\rho$. Furthermore, both the delay and buffer utilization improve if all packets are of the same size. The $n \times c$ multiplexer with the same access scheme but using a fixed packet size exhibits the following property—if $c \geq 8$, the delay is only 0.5 (normalized to the packet size) for almost all values of $\rho < 1.0$. The delay-throughput characteristics of a well designed multiplexer using the CSMA/CD access protocol are comparable to those of the $n \times 1$ multiplexer using the point-to-point access scheme. With the CSMA/CD protocol, if the packet size is increased, the delay normalized to the packet size decreases and approaches a limit.



### References


1. P. T. Brady, "A Statistical Analysis of On-Off Patterns in 16 Conversations," *The Bell System Technical Journal*, Vol. 47, No. 1, January 1968, pp. 73-91.
2. J. G. Gruber, "Delay Related Issues in Integrated Voice and Data Networks," *IEEE Transactions on Communications*, Vol. COM-29, No. 6, June 1981, pp. 786-800.
3. R. D. Rosner and B. Springer, "Circuit and Packet Switching," *Computer Networks*, Vol. 1, No. 1, January 1976, pp. 7-26.
4. C. D. Tsao, "A Local Area Network Architecture Overview," *IEEE Communications Magazine*, August 1984, pp. 7-11.
5. A. S. Acampora and M. G. Hluchyj, " A New Local Area Network Architecture Using a Centralized Bus," *IEEE Communications Magazine*, August 1984, pp. 11-23.
6. D. D. Clark, K. P. Pogram and D. P. Reed, "An Introduction to Local Area Networks," *Proceedings IEEE*, Vol. 66, No. 11, November 1978, pp. 1497-1517.
7. N. Abramson, "The Throughput of Packet Broadcasting Channels," *IEEE Transactions on Communications*, Vol. COM-25, No. 1, January 1977, pp. 117-128.
8. L. Kleinrock and S. S. Lam, "Packet Switching in a Slotted Satellite Channel," *AFIPS Conference Proceedings*, Vol. 42, June 1973, pp. 703-710.
9. L. Kleinrock and F. A. Tobagi, "Packet Switching in Radio Channels: Part I—Carrier Sense Multiple Access Modes and Their Throughput-Delay Characteristics," *IEEE Transactions on Communications*, Vol. COM-23, No. 12, December 1975, pp. 1400-1416.
10. L. Kleinrock and F. A. Tobagi, "Packet Switching in Radio Channels: Part II—The Hidden Terminal Problem in Carrier Sense Multiple Access and the Busy Tone Solution," *IEEE Transactions on Communications*, Vol. COM-23, No. 12, December 1975, pp. 1417-1433.
11. S. S. Lam, "A Carrier Sense Multiple Access Protocol for Local Networks," *Computer Networks*, Vol. 4, No. 1, February 1980, pp. 21-32.
12. R. M. Metcalfe and D. R. Boggs, "Ethernet: Distributed Packet Switching for Local Computer Networks," *Communications ACM*, Vol. 19, No. 7, July 1976, pp. 395-404.
13. J. F. Shoch, Y. K. Dalal and D. D. Redel, "Evolution of the Ethernet Local Computer Networks," *Computer*, Vol. 15, No. 8, August 1982, pp. 10-26.
14. F. A. Tobagi, M. Gerla, R. W. Peebles and E. G. Manning, "Modeling and Measurement Techniques in Packet Communication Networks," *Proceedings IEEE*, Vol. 66, No. 11, November 1978, pp. 1423-1447.
15. W. W. Chu, "A Study of Asynchronous Time-Division Multiplexing for Time-Sharing Computer Systems," *AFIPS Conference Proceedings*, Vol. 35, November 1969, pp. 669-768.
16. T. L. Saaty, *Elements of Queueing Theory with Applications* McGraw-Hill, New York, 1961, pp. 153-162.

**Appendix A. Some Relevant Queueing Theory Results**

Results from queueing theory are available for a few parameters of some statistical multiplexers studied in this article.

**Exponentially Distributed Packet Size.** The statistical multiplexer of Figure 1 with an exponentially distributed packet size can be modeled by the well-known M/M/1 queue. Let $\lambda$ be the packet arrival rate and $\mu$ the service rate, that is, the rate at which packets are transmitted over the output channel. Thus, if the packet length is $k$ bits, $\mu = 1/k$. If $C$ bits per second is the capacity of the channel, we can define the channel utilization factor $\rho = k\lambda/C = \lambda/\mu C$. Normalizing with respect to $C$, we have $\rho = \lambda/\mu$.

Under steady state conditions of the queue, the probability that there are $n$ packets in the system (either waiting to be transmitted or in the process of transmission), is given by

$$p_n = \rho^n (1 - \rho)$$

The average waiting time in the output buffer is

$$W_q = \frac{\rho}{\mu (1 - \rho)}$$

The cumulative distribution, $P(\leq t)$, that a packet waits no more than $t$ in the output buffer is

$$P(\leq t) = 1 - \rho e^{-(1-\rho)\mu t}$$

The expected number of packets in the output buffer is

$$L_q = \frac{\rho^2}{1 - \rho}$$

The expected number of packets either in the output buffer or in the process of transmission over the output link is given by

$$L = \frac{\rho}{1 - \rho}$$

**Constant Packet Size** The statistical multiplexer of Figure 1 operating on packets of a fixed size can be analyzed using the M/D/1 queue. In this case, the average waiting time in the output buffer is given by

$$W_q = \frac{\rho}{2\mu (1 - \rho)}$$

The average number of packets in the system is

$$L = \frac{\lambda^2 - 2\lambda}{2(\lambda - 1)}$$

where the parameter $\mu$ has been replaced by 1, assuming that all packets are of size unity.

*(Manuscript received September 26, 1986)*