Performance of some Local Area Network Technologies

by

Samuel T. Chanson
Arun Kumar
Ashok V. Nadkarni

Technical Report 82-8

August 1982

Dept. of Computer Science,
University of British Columbia, Vancouver, B.C.,
Canada V6T 1W5

ABSTRACT

This paper classifies local area network (LAN) technologies according to their topology and access method. The characteristics of the popular LAN technologies (namely Ring/Token passing, Ring/Message slots and Bus/Contention) are discussed. Analytic models are developed to estimate the mean packet delay time of each technology as a function of the network loading for various packet sizes and number of active stations. It is found that in the case of slotted rings (but not the other two technologies) an optimal value of the number of active stations exists which minimizes the mean delay time at all load levels given a packet arrival rate. The LAN technologies are compared with regard to their performance, reliability, availability, maintainability, extensibility, fairness and complexity.

It is hoped that potential users may be able to select the appropriate technology for their intended applications based on their specific performance requirements and operation environment. As well, LAN designers may benefit from the insight provided with the analysis.

ACKNOWLEDGEMENT

The authors would like to thank Prem Sinha for useful discussions. This work was supported in part by the Canadian Natural Sciences and Engineering Research Council under grant No.3554.

1. Introduction

Local Area Network (LAN) technologies have been the subject of considerable research and development efforts in recent years[1-6]. By network technology, we mean the mechanism, both hardware and software, by which various computing facilities are interconnected for communication.

The advantages of interconnecting computing facilities are numerous, some of which are listed below:

1) it allows devices (such as line printers, tape and disk units) as well as files and software systems to be shared,
2) provides convenient access to facilities in different physical locations,
3) facilitates the co-operation and co-ordination of projects (chiefly due to points 1 and 2 above),
4) allows the workload to be shared among the computing facilities,
5) points 2 and 4 above mean the availability and reliability of the computing system are improved,
6) allows computing power to be expanded incrementally (through the addition of nodes (or stations) in the network),
7) facilitates the implementation of electronic mail systems,
8) performance may be improved through parallel computations at different nodes.

Though in principle each station may be connected to every other station in the network by separate high speed links, network technology is concerned with making such connections cost effective through sharing of the communications facilities.

A local area network is generally considered to be one which covers a "limited" geographical area such as a building or a group of buildings within a few kilometers of one another. LANs typically also exhibit certain attributes such as high data rates, a high degree of interconnection between devices on the network with each station having the potential of communicating with every other station. As well, each station generally listens to every transmission, whether addressed to it or not [5]. Because of the characteristics of LANs, the technologies used are quite different from those of long-haul networks [6]

Various LAN technologies have been developed. It is increasingly clear that no single technology is superior to the others in all respects. This is evidenced by the IEEE 802 Committee (which is considering the lower level protocol standards for LANs) proposing two incompatible systems to be standardized : contentions (CSMA/CD busses) and token passing (both for rings and busses) [7] .

This paper compares the characteristics and performance of the popular LAN technologies. It is hoped that potential users may be able to select the appropriate technology for their intended applications based on their specific performance requirements and operation environment. As well, network designers may benefit from the insight provided with the analysis. An understanding of the basic principles of LAN (such as those contained in [3]) is assumed.

2. Classification of LAN technologies

LAN technologies consist of at least the following aspects:

a) Physical data transport mechanism:

E.g., digital signalling using baseband (single channel) or broadband (modulated, multi-channel) techniques, over twisted pair, coaxial cable, fiber optics or radio .

b) Topology (the connectivity characteristics of the network nodes):

E.g., star, ring, bus which may be fully connected, hierarchical, cross-connected or irregular.

c) Sharing technique (the bandwidth allocation scheme for multiple users):

E.g., dedicated (non-shared), time division multiplexing, frequency division multiplexing, contention.

d) User services and protocols:

Each LAN invariably has its own terminal support functions. Despite the efforts of the IEEE 802 Committee, the American National Standards Institute (ANSI), the International Standards Organization (ISO) and other organizations, standards for higher-level protocols are not expected in the near future.

Given the physical environment the network is to operate in and the financial considerations, the choice of the physical data transport mechanism is usually straightforward. As well, there are just too numerous user services and protocols in existence. We shall, therefore, concentrate on b) and c) only. LAN technologies will thus be classified according to their topology and sharing technique (or access method, which determines which device can use the transmission line at any given time). Only the popular LAN technologies will be studied: namely Bus/Contention, Ring/Token passing, Ring/Message slots. For simplicity, hierarchical connections will not be considered and packet-switching is assumed.

2.1 Desirable Characteristics of local area network technologies

Only those characteristics which are influenced by the network topology and/or the access method are discussed.

1) Performance

Given that communications facilities are to be shared, it is clear that a station may not be able to access the channel the instant a packet is ready for transmission. The delay time in waiting to get onto the channel usually increases as the network loading increases. This delay time should be as low as possible.

Because of the typical applications and the high data rates of LANs (1-10 Mbits/sec.) the mean utilization of the channel is usually very low [11,21]. Thus mean channel efficiency or throughput is often not a design consideration [21]. However, a good technology would maximize the proportion of useful data carried through the network at all load levels.

2) Reliability

A station or a section of the channel failure should not affect the rest of the network. As well, hardware and software failure rate should be as low as possible. The IEEE 802 Committee, for example, recommends that the number of undetected error should not exceed once per year[7].

3) Availability

Rapid fault isolation and short mean-time-to-repair enhance the availability of the network. A guaranteed upper-bound on the waiting time to access the network further increases its availability.

4) Maintainability

Errors should be easy to detect and correct. It should be possible to remove a station from the network for repair without disturbing the rest of the network. A network which is easy to maintain usually is more available than one which is not.

5) Extensibility

It should be possible to add (or remove) stations from a network without disturbing it. The maximum length of the network should be reasonably long. As well, the network technology should allow easy adaptation to new transportation media (which can be orders of magnitude faster than is currently available).

6) Fairness

By fairness we mean no request from any station to access the channel should be discriminated against. In other words, the variance of the delay time to gain access to the channel should be low. In addition, the mean delay time should be nearly equal for all stations.

7) Complexity

The cost of the network as well as its reliability and maintainability is usually directly related to its complexity. Thus the simpler the technology the better.

Before the different LAN technologies are compared we first give a brief description of the salient features of each technology. The reader is referred to [3,8,10,12,17] for details.

3.1 Contention Busses

There are several variations of this technology. The nodes or stations are all attached to a passive transmission line (the bus). In the simplest scheme, the stations may transmit at any time. A collision occurs if more than one station transmit simultaneously, in which case the stations involved will retransmit after some random time (backoff). This unrestrained contention of the use of the channel results in very low useful throughput rate , with the maximum value at approximately 0.186 of the transmission rate. This is the classical Aloha scheme [9].

If time is divided into equal length slots and the stations are synchronized so that they may transmit only at the beginning of a slot, the maximum useful throughput rate of this so called Slotted Aloha scheme increases to twice that of the basic scheme.

In their 1976 paper [10], Metcalfe and Boggs present an improvement to the scheme which has since been referred to as carrier sense multiple access with collision detection (CSMA/CD). This scheme is used in the Ethernet. Each station monitors the channel continuously whether it is transmitting or not. It transmits only if it senses the channel to be idle. Due to propagation delay, collisions may still occur, however. A station that has just transmitted will abort the transmission immediately if a collision is detected, and retransmit after a random backoff time. This scheme has been shown to yield channel utilization in excess of 90% [11].

Except for the slotted varieties, packet sizes need not be fixed.

3.2 Basic Ring Topology

As its name implies, the channel is in the form of a loop. Stations are usually connected to the channel through active repeaters (Figure 1).

Although the topology provides for two links connected to every station, most rings have operated with unidirectional data flow. An important consequence is that unlike the bus topology, a packet on the channel may only be picked up by one station at a time if the network is operating correctly. Also a packet is sent by a station on only one link connected to a single station (the one down stream).

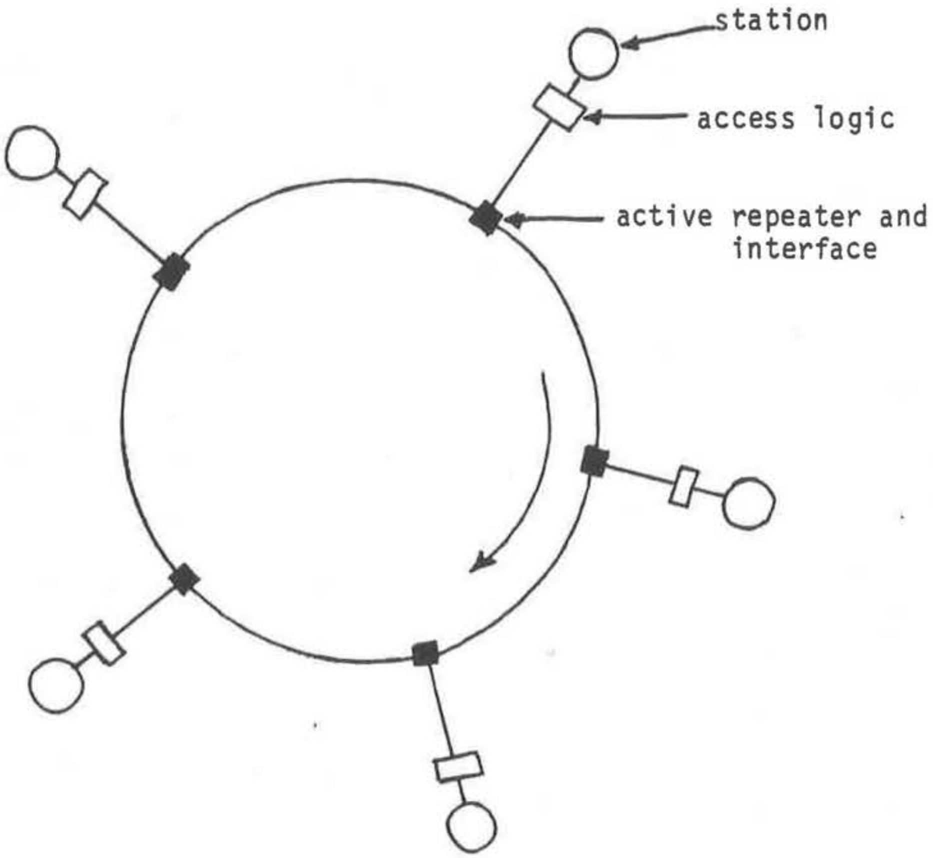


Figure 1. Basic Ring Topology

3.2.1 Slotted Rings

This access method requires one of the stations on the ring to be designated the "monitor station". It initially transmits a number of empty packets (slots), which usually are of fixed size. If any station wishes to send data, it marks a passing packet which is empty and puts the data and the destination's address into the packet. The packet circulates around the ring and is checked by each station. If the destination's address matches with the station address, the data is read from the packet and the packet is marked as empty. If the packet completes a full circuit of the ring and is detected by the sending station without having been emptied, the sending station empties the packet and returns an error code to its associated machine's operating system. The monitor station also checks each packet for errors and clears a marked packet if it has completed two circuits.

Note that this scheme allows more than one station to transmit and receive packets simultaneously if the number of slots exceeds one. The Cambridge Ring [12,17] is an example of the slotted ring.

3.2.2 Token Rings

There is no "monitor" station in this scheme. A special bit pattern known as the token circulates around the ring. There is only one token. A station may only transmit a packet onto the ring if it is in possession of the token. It does this by replacing the token by a special bit pattern often called the "connector", append the data after it and the token after the data. Unlike the slotted rings, the packet can be of variable size.

Every station on the ring "sees" the packet and may verify its integrity and its address. However, only the sending station may remove the packet (simply by destroying the connector). Thus the token provides decentralized control. (see, for example, [13])

4. Comparison of Technologies

We shall now compare the various LAN technologies with respect to each of the characteristics listed in section 2.1. If only the topology (and not the access method) has an impact on a characteristic, the two ring technologies will be discussed together under the heading of 'rings'.

4.1 Performance

The performance index considered relates to the responsiveness of the network and shall be called the delay time t. It is the time interval between the instant a packet is ready for transmission and its arrival at its destination (sometimes also known as the response time). The time t consists of the delay time tw to access the channel, the propagation delay tp and the time required for the entire package to move into the buffer at the destination station,tm.

The last two components are functions of the transmission rate of the channel, the distance between the source and destination stations and the packet size. tw is a function of the network loading as well as the efficiency of its protocols . We shall derive the relationship between t and the normalized load (defined as the ratio of packet arrival rate and the channel transmission rate) for each of the technologies. (See

Appendix for the analytic models). The normalized load is used so as to remove the channel transmission rate as an explicit independent parameter in the relationship. It also has more intuitive meaning than the actual load . The normalized load varies between zero (no transmission) and one (network is saturated and no more bandwidth is available). Since t depends on the packet size and increases as the packet size increases (everything else remains constant), the absolute delay time may be misleading in some cases. It is customary to use the delay time normalized by a base packet size B. Thus the normalized delay time with respect to packet size x is simply its absolute delay time multiplied by B and divided by x. The normalized delay time represents the time required to transmit B bits of information through the network. The absolute as well as the normalized delay is plotted against the normalized load for various values of packet size (Figures 1,2,4,6). To study how well the technologies respond to increase in the number of stations, the absolute delay is also plotted against the normalized load for various values of the number of active stations in the network (Figures 3,4,7).

For the slotted ring technology, there are two additional system parameters which affect performance - slot size and the number of slots in the ring. For simplicity, we shall assume the slot size to be the same as the packet size. This eliminates a parameter and simplifies the model. Figure 5 shows the relationship of t as the normalized load changes for different number of slots, keeping the slot size constant at 64 bits. In Figure 7, as the number of slots varies, the slot size is also changed to maintain the same number of bit delay in the ring.

In all the plots, the channel has a transmission rate of 3 Mbits/sec. and is 1 km long.

a) Contention Bus (Figures 1-3)

The model is actually that of the Ethernet (Appendix I), which is the most popular contention bus technology. It is a modification of the one by Almes and Lazowska [22]. Instead of an infinite population model, our model allows the mean delay time to be expressed in terms of the number of active stations in the network.

To better utilize the bandwidth of the channel, Ethernet recommends the minimum packet size be 512 bits for a 10 Mbits/sec. line ?8? (or 154 bits for a 3 Mbits/sec. line). The 64-bit packet size in the figures are there to show that performance will degrade drastically when smaller packet sizes are used.

Observe that the mean delay time remains very low when the packet size exceeds the recommended minimum value until the load is close to the maximum channel capacity. Also, the larger the packet size, the less time is required to transmit each bit of information and the channel can handle a larger loading before reaching saturation (Figure 2). (In this paper, saturation is loosely defined as the state when mean delay time approaches infinity.) Because the tm component of the delay time is linearly proportional to the packet size, the absolute delay time per packet decreases as the packet size decreases when the load is light. At higher loads, however, the contention interference increases and tw dominates t which, for the same load level, is smaller the larger the packet size due to the lower packet arrival rate (Figure 1).

It is interesting to observe that for a given load and particularly for large packet size, the mean delay time is not very sensitive to the number of active stations on the bus (Figure 3). Even with only 4 active stations, the result is very similar to that of 128 stations. This is an advantage of the contention bus technology. It implies that one can add stations to the network without having to worry about performance degradation.

b) Token Ring (Figure 4)

The model (Appendix II) is a refinement of the one by Tanenbaum [3]. Queuing delay of packets at the stations is considered.

Under the assumption that a station is allowed to put all waiting packets onto the ring when the token arrives, the mean absolute delay time for a given normalized load is independent of the packet size (see Appendix II). Thus it is not meaningful to talk about normalized delay time.

Figure 4 shows that as the number of active stations increases, the delay characteristics get poorer. For example, when the number of active stations increases to 512 from 128 the mean delay time goes up by approximately a factor of 2 at all load levels. This is because an additional station adds to the walk time delay (i.e., the time required for a bit to go once around the idle ring) thus increasing the propagation delay tp. Contention bus technologies do not suffer from this disadvantage.

c) Slotted Ring (Figures 5-7)

Very few work exist on the analytic modelling of the slotted ring technology. Our model (Appendix III) makes use of several results of queuing theory to obtain the mean steady state value of t. For simplicity, it is assumed that the packet size coincides with the slot size and that the gaps between slots are negligible.

In Figure 5, the slot size is kept constant while the number of slots n is varied. Thus, the more the number of slots, the longer the walk time. This explains the increase in absolute delay time as n increases when the load is light. However, for a given load, the probability of finding an empty slot increases as n is increased. At higher loads, this wait time tw dominates the delay time and the mean delay time t actually decreases as n goes up. As well, more load can be accommodated before the system saturates.

In Figure 6, the walk time is kept constant as the number of slots is varied. Thus the slot size decreases as n increases. The normalized delay (with respect to slot size of 256 bits) is better for larger slot (and thus packet) sizes.

Figure 7 shows that for a given slot size and number of slots in the ring there is an optimal number of active stations N which minimizes the mean delay time and maximizes the saturation load. For a given load, the larger the value of N the less the mean number of packets waiting at each station. Thus the mean queue wait time monotonically decreases as N increases. However, because the number of stations ready to transmit has increased, the mean wait time to acquire an empty slot is monotonically up. Furthermore, the walk time also increases monotonically with N. The opposing effects on the mean delay time as N varies imply that there is a value of N which minimizes the mean delay time for a given load. Figure 7 shows that this value remains constant for all load levels. In our case, the number lies between 2 and 128. It may be possible to compute the number mathematically. Knowledge of this information allows us to estimate how close an existing ring is operating from its theoretical optimum. It also allows an estimate of how many stations should be attached to the ring.

4.2 Reliability

a) Rings

The basic ring topology has often been criticized as being unreliable on the ground that an open circuit anywhere or the failure of any repeater will disrupt the entire network. This is certainly a problem with a large number of repeaters strung together. Current ring designs, however, often use redundant paths or fail-safe bypassing to avoid this problem.

A common design is the so-called "star-shaped ring" [14]. Here, the ring is actually implemented in a wire centre at the hub of a star-configured network (Figure 2).

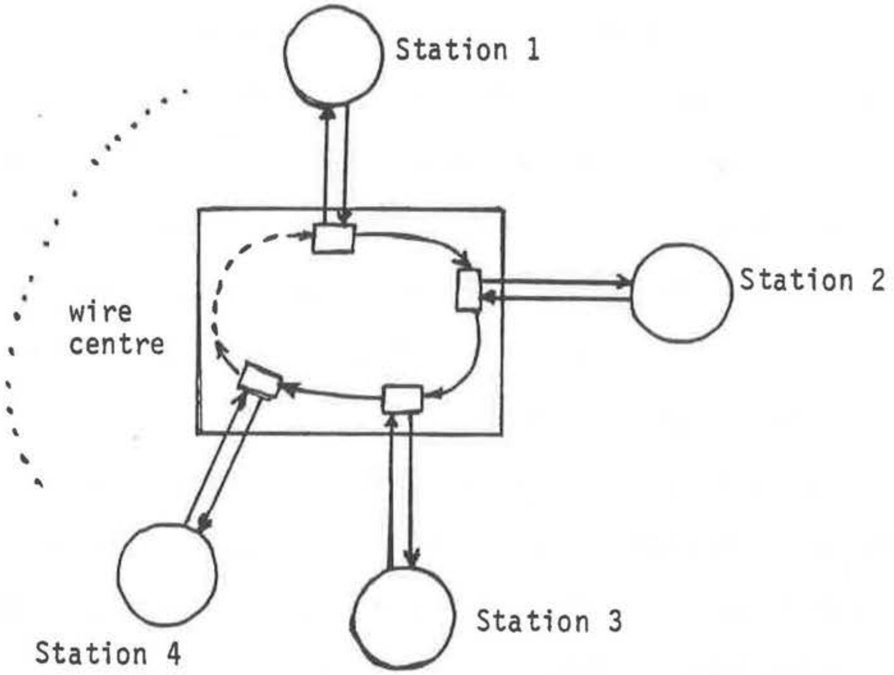


Figure 2. Star-shaped ring

The wire centre is a passive device which allows stations to be connected onto the ring, often through the use of relays remotely energized by the stations. If a station fails (such as due to power failure) or if any of the cabling between the station and the wire centre is broken, the relay falls back to its normal state which bypasses the station at its attachment point. As well, the centralized location of the wire centre facilitates the maintenance and reconfiguration of the network, thus enhancing its availability. Furthermore, the circulating control signal of the ring topology lends itself to fully synchronous operation. At high data rates (such as those in excess of 10 Mb/sec.) synchronous operation provides much more reliable performance than the burst mode operation of the contention bus systems.

We shall assume this or similar ring design in subsequent discussions.

a.1 Token ring vs slotted ring

The decentralized control scheme of the token ring technology means that there is no single station failure which may bring the entire network down. The monitor station of the slotted ring, however, is a critical component whose failure may disrupt the whole system.

b) Busses

The bus is essentially a passive device thus its reliability is much higher than that of the basic ring design. The wire centre concept of the star-shaped ring, however, has reduced the advantage offered by this aspect of the bus.

While there are no central control points in the bus technology, a short circuit or a transceiver which fails to stop transmitting will disrupt the bus. Thus we feel that the reliability attributable to topology is comparable for the bus and the star-shaped ring. The reliability attributable to the access method has much to do with the corresponding complexity and will be discussed under that heading.

4.3 Availability

4.3.1 Rings

The wire centre concept allows rapid isolation of a misbehaving station. As well, the centralized location facilitates maintenance and reconfiguration of the network, thus enhancing its availability.

Ring technologies, particularly the token-passing scheme and also the slot message design with appropriate protocols, provide a guaranteed maximum in the queue wait time to access the channel. This is primarily due to the synchronous control operation and the unidirectional data flow in ring technologies.

4.3.2 Contention Busses

A basic weakness of the contention scheme is that an unlucky station may have to wait a long time to gain control of the bus, though it seldom happens in practice. This is especially so when the network loading is heavy [15].

4.4 Maintainability

4.4.1 Rings

Again, the wire centre concept allows the system to fail gracefully. This reduces the maintenance burden, since not all failures require instant attention. Stations may be removed for repair without disrupting the network. The centralized location also allows failures to be rapidly detected.

4.4.2 Contention Busses

It is easy to remove a faulty station for repair without disturbing the rest of the network. Detection of the source of hardware error is not as convenient as that provided by the wire centre environment[16]. However it is felt that the contention scheme simplifies software error detection and correction.

To obtain maximum transceiver performance, current bus designs often require direct coupling of an active component to the cable. To suppress transient surge of current (such as from lightning), the coaxial cable ground shield should not be grounded at more than one point. This requirement and in order to divide a long cable into sections for trouble shooting result in the Ethernet specification [8] of no ground for the cable. Such a floating conductive system provides extreme hazard for the maintenance crew should it be accidentally shorted to an electric power conductor.

Ring technologies do not suffer from this problem.

4.5 Extensibility

The basic ring design, with the active elements forming part of the data path, does not allow easy addition of stations without disturbing the rest of the network. The wire centre design overcomes this problem to a large extent.

It is very easy to attach devices to the passive bus (usually a coaxial cable) at any point simply by using clamp-on connectors. This however could cause the transmission medium to become unbalanced, making the problem of electromagnetic compatibility between the net and other electrical equipment closeby more difficult to engineer [16]. Nevertheless the bus technology allows at least as easy an incremental growth as that offered by the current best ring technology.

In the case of CSMA/CD technology, high channel efficiency is dependent on the ability to detect collision while the packet is being transmitted. This requirement means the end-to-end propagation delay of the bus must not exceed the packet transmission time. Given the typical packet length, the product of the maximum speed of the transmission medium and the maximum length of the network is fixed. The Ethernet specifications, for example, limit the maximum line length to 2.5 km with a transmission speed of 10 Mbits/sec.[8]

It is also difficult to take advantage of the higher speed transmission media such as optical fibers in contention buses. In addition to the transmission speed issue, it is difficult to turn optical fiber into a broadcast medium. Ring technologies have no such restrictions. The ring size may be extended far beyond that of the basic Ethernet by the use of repeaters. However, as repeaters amplify both signal and noise indiscriminately, the ring cannot be extended indefinitely without using expensive regeneration devices. As ring technologies usually provide one-way, point-to-point transmission, it is straightforward to replace wire links with fiber optic links. An experimental slotted ring network using this transmission medium has been constructed.[17]

4.6 Fairness

4.6.1 Token Rings

Because the token passes through each of the stations on the ring one at a time in a round-robin fashion, it is easy to see that if an upper limit is placed on the length and number of packets that can be sent at a time, a station is guaranteed to receive the token within a fixed time. It is thus easy to design a fair protocol for this technology.

4.6.2 Slotted Rings

Under normal working conditions, fairness is rarely a problem with this technology. However, with some protocols, it is possible for a station to be locked out indefinitely. Take the case of the Cambridge Ring (which is a slotted ring) using the Basic Block protocol [17] for example. Consider the situation where three stations are trying to send packets to a fourth station at the same time. If for two of these stations, the data available are so fast that there are no delays between blocks and that they are insistent, then the third station may be frozen out for an indefinite period [18]. A proper protocol will alleviate this problem.

4.6.3 Contention Busses

Basic contention schemes provide no upper bound for the maximum number of interferences of packet. Thus theoretically, it is possible for a request to wait indefinitely to be successfully transmitted. Furthermore, because of the backoff formula used in most schemes, collided packets are discriminated by having to wait even longer. Thus a packet that has just arrived actually has a higher probability of getting to its destination before a packet that has arrived earlier and which has suffered one or more collisions. The variations of waiting times are therefore high. There are research activities (such as Slotted Ethernet [19]) which aim at overcoming this problem.

4.7 Complexity

The reader is referred to Saltzer et al. [16] for details which contains an excellent discussion of the complexity issue of the token ring and the contention bus technologies.

4.7.1 Rings

Generally, the ring net is mostly a digital design and thus can take advantage of the rapid advancement in digital technology and VLSI. One complexity arises from the fact that in a ring, the repeaters must use a common clock rate which must result in an integral number of bit times of delay when traversing the ring. There are known solutions (such as using phase-locked-loops [20]) to overcome this problem.

Since in a ring, a station has only one output link which is connected to exactly one other station, the analog portion of the ring repeater is much easier to engineer.

Another intrinsic property of ring interfaces is that it must be able to selectively remove a message from the ring or pass it on. This makes ring interfaces more complex than the interfaces to the passive bus. To increase the reliability, availability and maintainability, the simple basic ring topology is replaced by a more complex design such as the star-shaped ring with a wire centre. This requires considerably more cabling as well as devices to bypass faulty stations. Unless the building(s) is already wired in this way for other reasons, this would increase the complexity and cost of installing the network. As well, token ring design is more complex than slotted ring design because of the decentralized control scheme. The designer of the token ring technology must solve the problems of distributed initialization and recovery. Thus more complex protocols must be established to handle such problems as:

-how to initialize the first token
-what if the device holding the token crashes,
-what if two or more devices have picked up a token.

4.7.2 Contention Busses

Clocks at different stations need only to agree on a common frequency. The slotted variations, however, require synchronization of the clocks as well.

The complexity associated with broadcast, contention bus technology is largely due to its analog component. A transmitter's signal must be received by all receivers on the cable. These receivers are at different distances from the transmitter. Thus they will experience different attenuations and echoes. Also, receivers must be able to hear every transmitter. As well, to detect collision while transmitting, the receiver of an active transceiver must be capable of detecting the weakest other transmitter and distinguishing it from its own echoes.

In order to allow a station to send a packet to itself (for testing, for example) and for the receiver and transmitter to be started independently, full duplex mode of transmission is required. This is, for example, recommended by the Ethernet specifications [8]. Most experimental Ethernets, however, use only half-duplex [23].

It would appear, however, from the chart of the approximate price per connection of a number of existing networks [13], that the cost of contention bus connection with current technology is slightly lower than that for ring networks.

5. Conclusions

Local area network technologies are classified according to their topology and access method. The characteristics and performance of the popular technologies are discussed. It is clear that no single LAN technology outperforms the others in all respects under all load conditions.

In general, the CSMA/CD bus technology should have lower mean delay time than the ring technologies when the load is light (i.e., the probability of collision small). This is because unlike the ring technologies where a station must wait for the arrival of the token or an empty slot, a station can transmit the packet at once. It is interesting to note that the acknowledgement of the arrival of a packet in the ring technologies is essentially the original packet itself. Thus the packet cannot be removed from the channel until it has completed at least one round trip. If the packet is very large, this is wasteful of channel bandwidth. In the contention bus technology, a small acknowledgement packet can be sent.

Another advantage of the contention bus technology is that its performance is relatively insensitive to the number of active stations on the network particularly when the mean packet size is large. Thus large number of users may be supported. In the ring technologies, mainly because of the increased walk time as the number of stations increases, delay time suffers.y

Except for the slotted bus varieties [19], contention bus allows variable packet size. Furthermore, the larger the packet size , the smaller the delay time per bit of data transmitted [11]. In fact because of the need to detect collision during transmission to improve channel efficiency, there is a lower bound on the size of packet (512 bits for Ethernet with a 10 Mbits/sec. bus [8]). Thus if the application calls for relatively infrequent transmission of fairly long, variable size packets from a large number of stations, the contention bus technology is better than the two ring technologies discussed from a performance standpoint.

Broadcast contention bus technology suffers from the fact that the upper bound of delay time is non-deterministic. Thus it is not appropriate for real time applications. Furthermore, the theoretical maximum network length is generally shorter than the ring technologies (e.g., it is 500 meters per line, 2.5 km between any two stations in a hierarchical network for Ethernet [8]). As well, it seems difficult for broadcast type technology to use fiber optics as the transport medium. Also the predominantly analog design of the contention bus may not be able to take advantage of advances in digital technology and the corresponding decrease in cost in the future.

The wire centre concept of star-shaped rings has removed much of the maintainability problems once considered a major disadvantage of ring technologies. However this increases the installation cost. It is generally easier to implement a fair protocol for rings which are more suitable for smaller packet sizes. The slotted ring requires the slot size to be fixed which may result in wasted bandwidth if the slot is not filled. Thus the slot is usually quite small. (The Cambridge Ring Basic Block allows only 16 bits of data to be carried in a packet of 38 bits [17]). Thus it is more suitable for small but fixed size transmissions. Its performance could be better than token ring when the number of slots is large enough as it allows more than one station to transmit and receive simultaneously.

The token ring allows transmission of packets of different sizes efficiently. The decentralized control scheme also means that the failure of a single station need not disrupt the rest of the network. The scheme, however, is more complex which generally implies lower reliability and maintainability (all else being equal).

It is evident that no technology is "best" for all applications. More quantitative measurements are needed to enhance our understanding of the performance of LAN technologies. As well, more work needs to be done to improve the accuracy of analytic models and to make them more realistic.

Figure 1 . Absolute Mean Delay vs. Normalized Load for a CSMA/CD Bus

N=128 PS=64,256,512,1024 bits

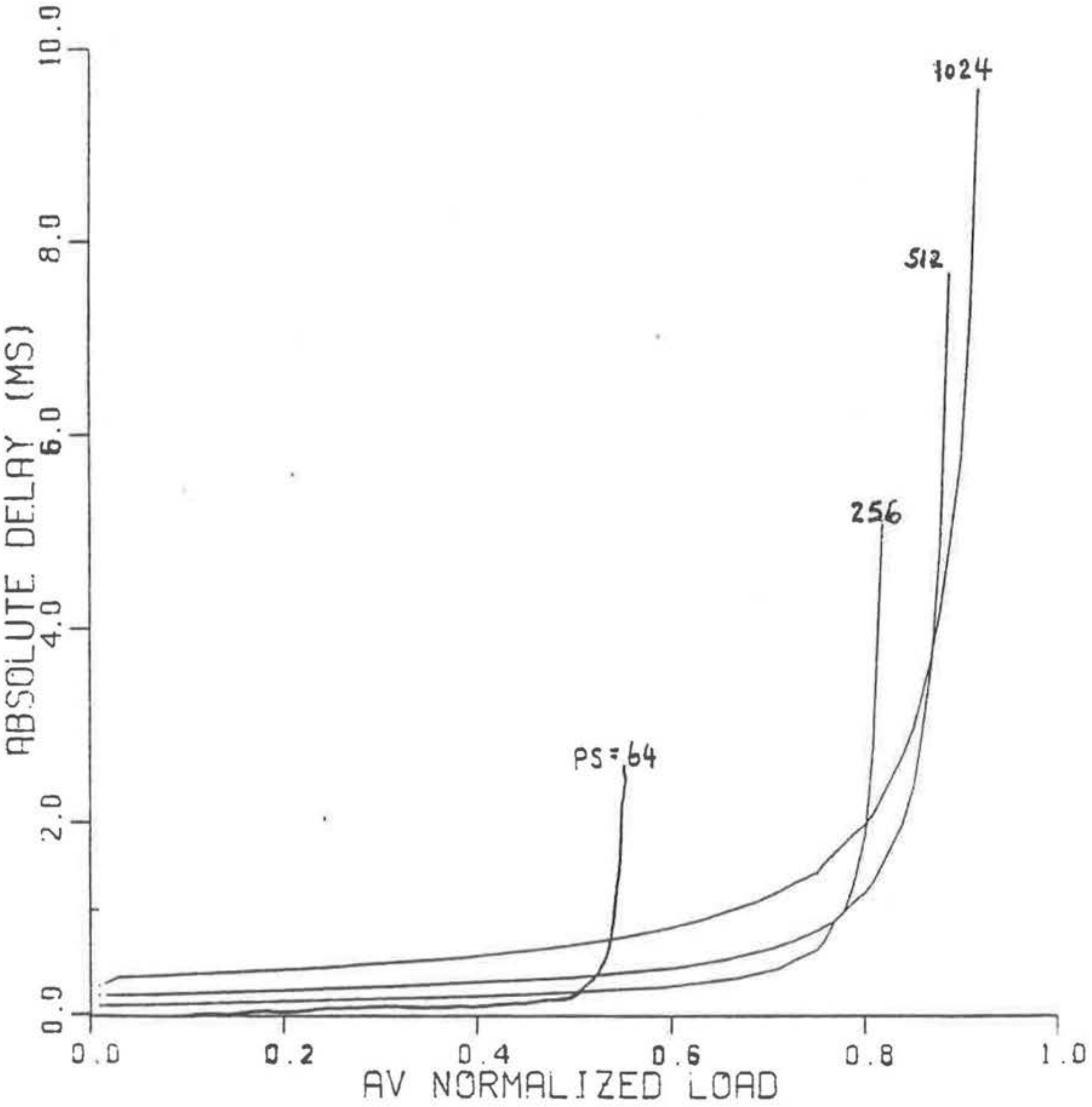


Figure 2. Normalized Mean Delay vs. Normalized Load for a CSMA/CD Bus

N=128 PS=64,256,512,1024 bits

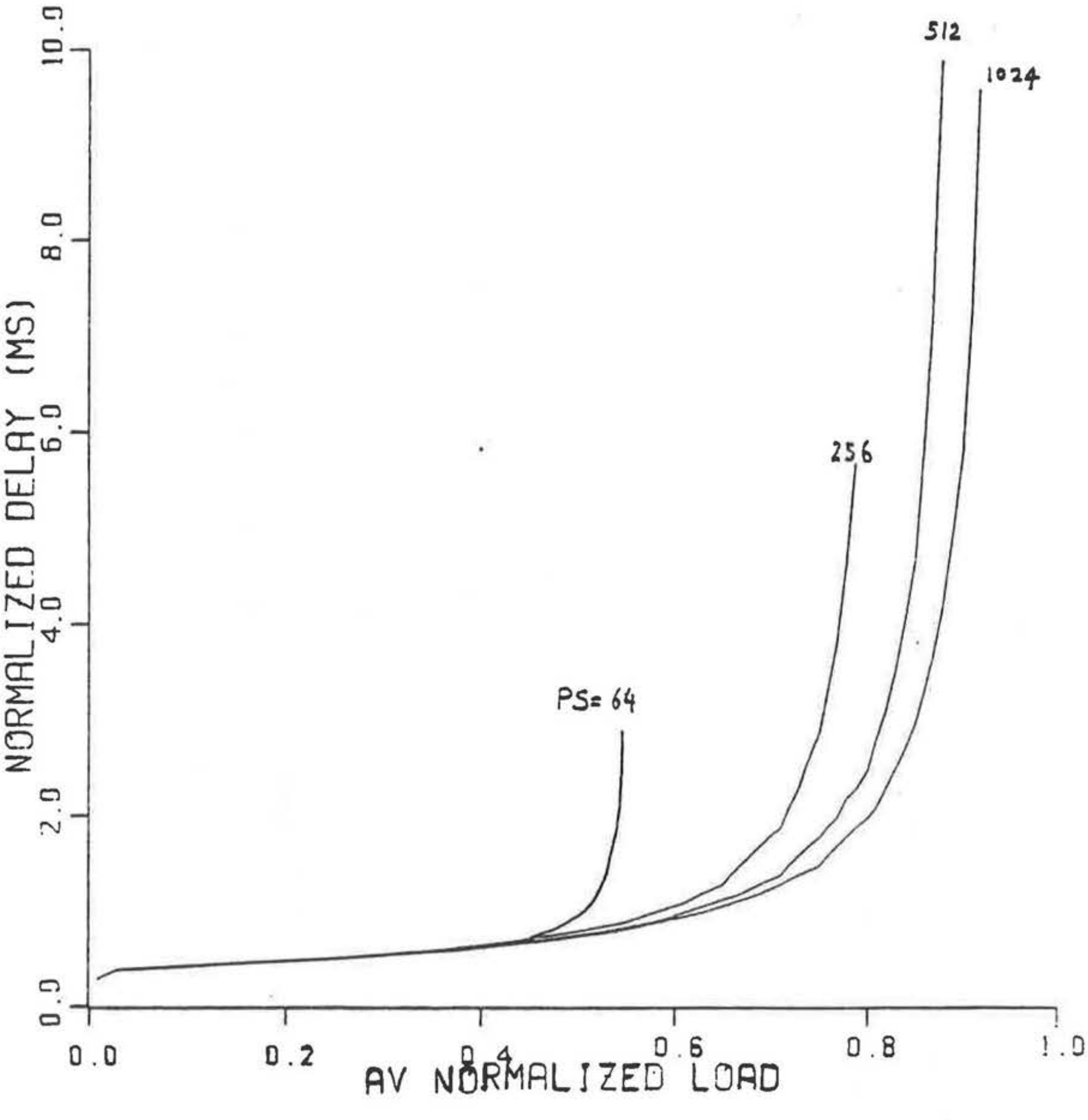


Figure 3 : Absolute Mean Delay vs. Normalized Load

for a CSMA/CD Bus

N=2,4,8,128 PS=1024 bits

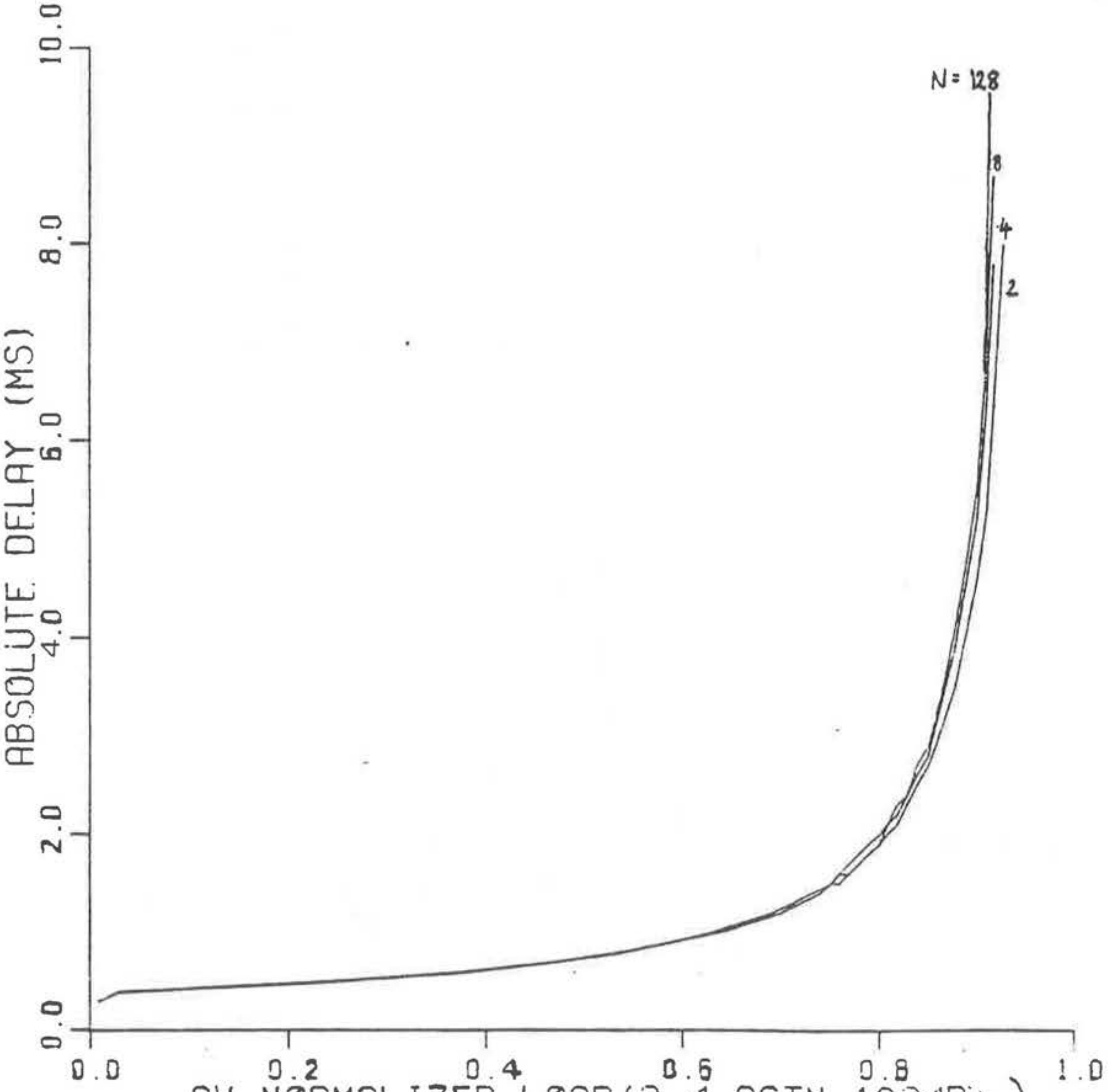


Figure 4 : Absolute Mean Delay vs. Normalized Load for a Token Ring

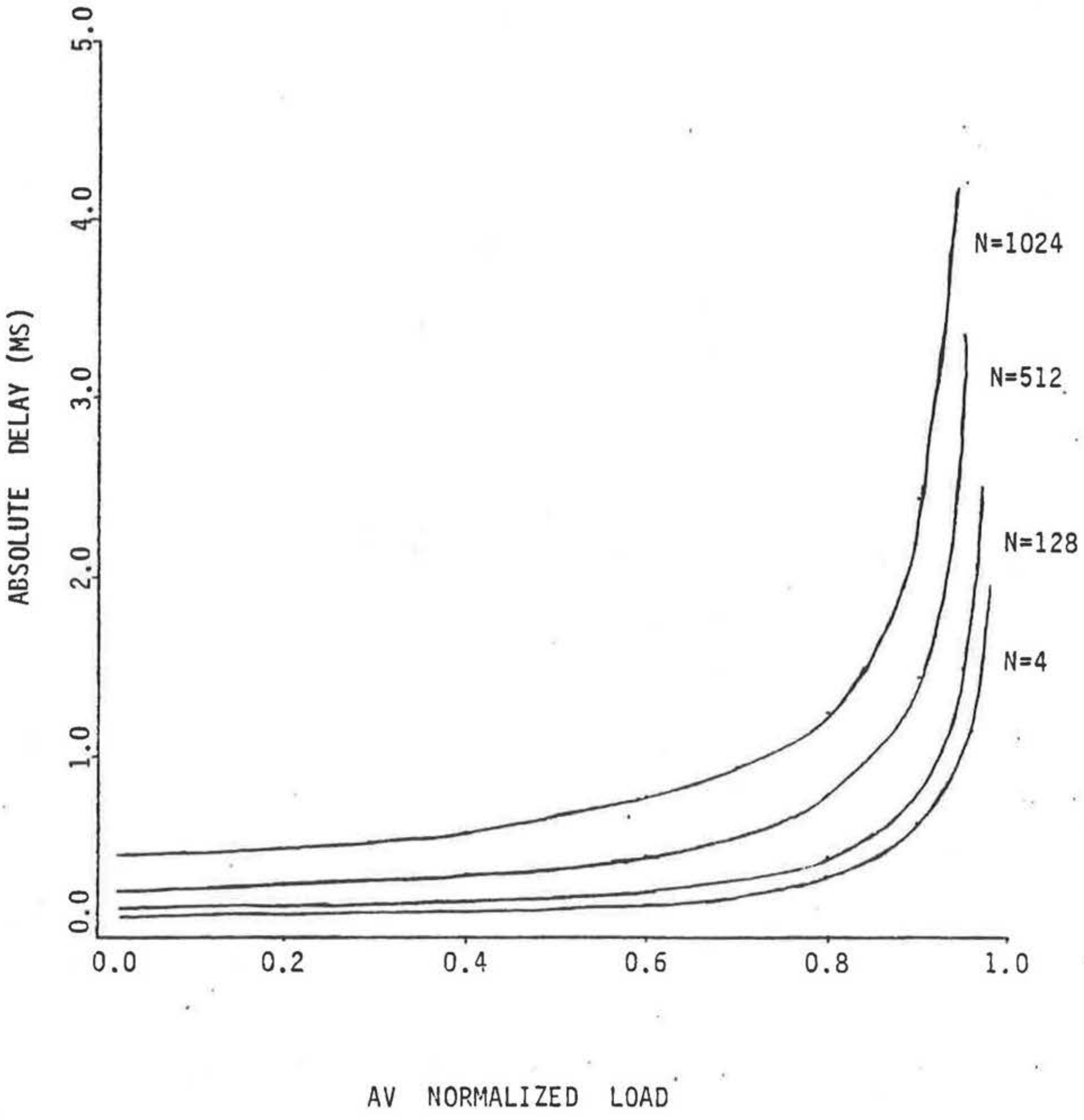


Figure 5 : Absolute Mean Delay vs. Normalized Load for a Slotted Ring

N=128, PS=64 bits

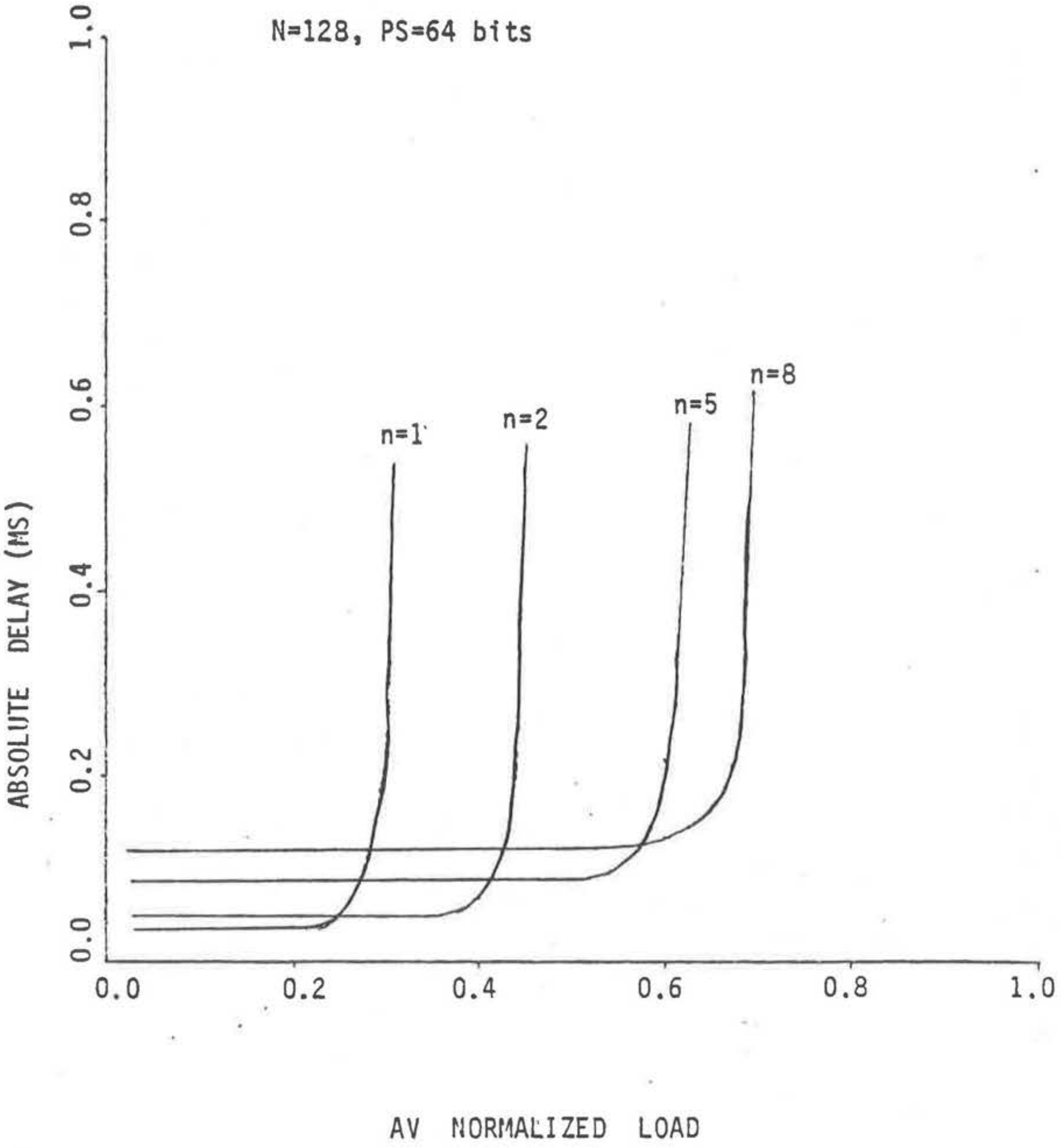


Figure 6 . Normalized Delay vs. Normalized Load for a Slotted Ring

N=128

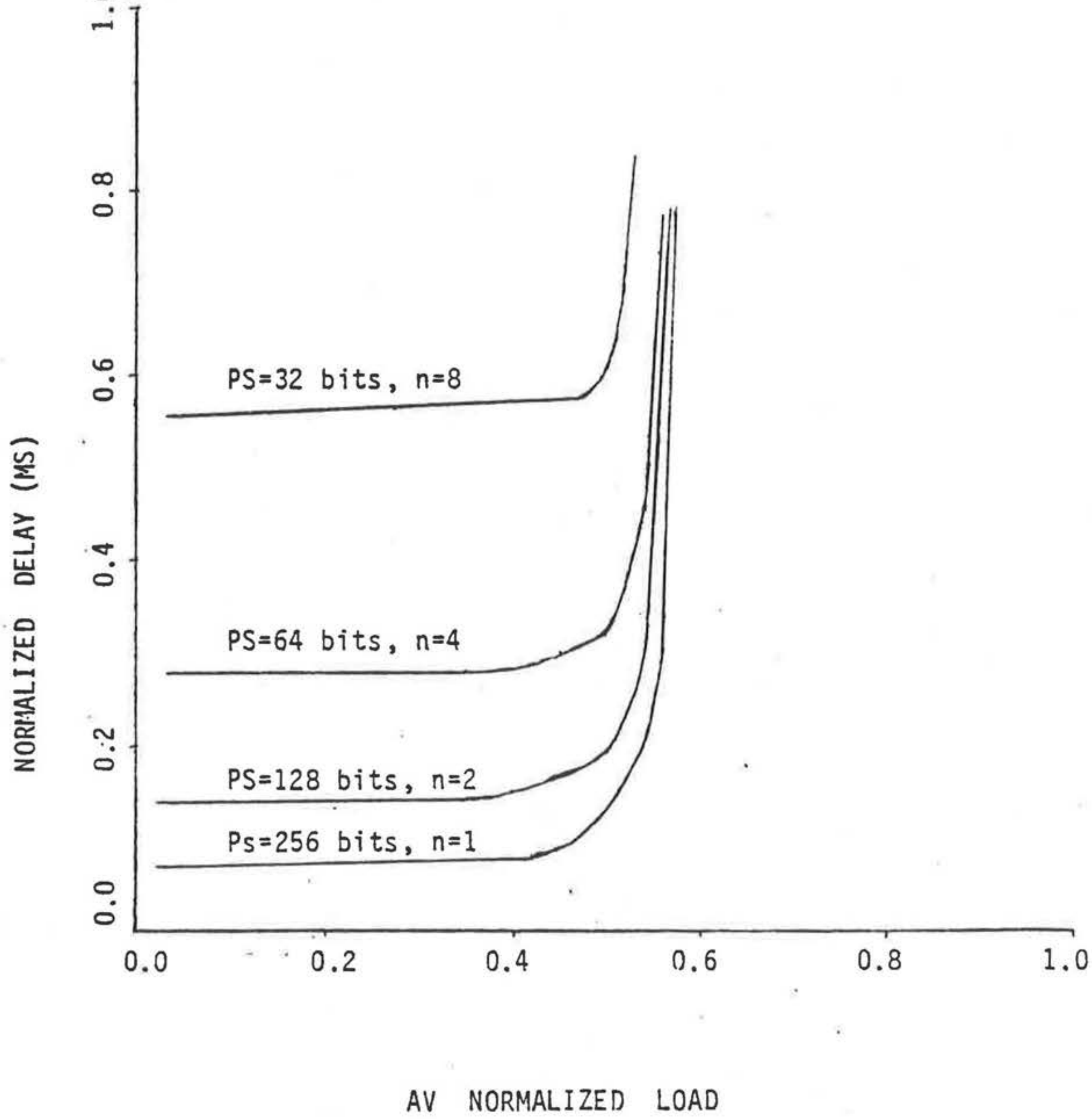
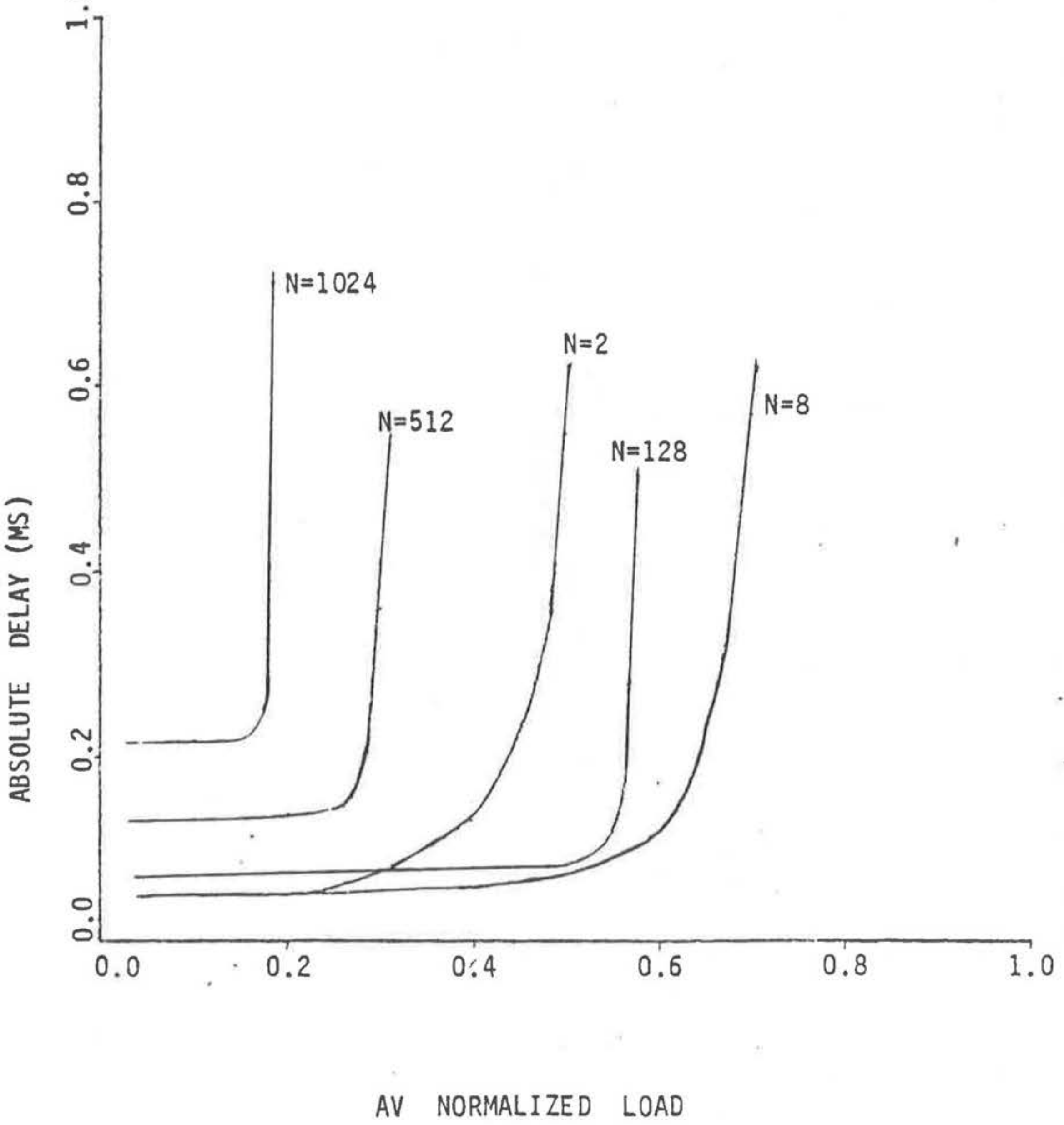


Figure 7 : Absolute Mean Delay vs. Normalized Load for a Slotted Ring

n=4, PS=64 bits



REFERENCES

[1] Clark,D.D.,Pogran,K.T. and Reed,D.P.,"An Introduction to Local Area Networks", Proc.IEEE, Vol.66, No.11,pp.1497-1517,Nov.1978.

[2] Cotton,I.W., "Technologies for Local Area Computer Networks", Computer Networks, Vol.4,pp. 197-208, 1980.

[3] Tanenbaum,A.S.,Computer Networks,Ch.7, Prentice Hall,Englewood Cliffs,N.J., 1981.

[4] Hamacher,V.C.,"Local Area Networks",Proc. of CIPS, pp.243-254, May, 1982.

[5] Kryskow,J.M. and Miller,C.K.,"Local Area Networks Overview - Part 1: Definitions and Attributes",Computer Design, Vol.20, No.2, pp.22-35, Feb.1981.

[6] McQuillan,J.M.,"Local Networks Architectures",Computer Design, Vol. 18, No. 5, pp.18-26, May 1979.

[7] Graube,M.,"Local Area Nets: a pair of standards",IEEE Spectrum, Vol. 19, No. 6, pp.60-64, June 1982.

[8] The Ethernet:Data Link Layer and Physical Layer Specifications, Version 1.0, D.E.C., Intel Corp. and Xerox Corp., Computer Communication Review, Vol.11, No.3, pp.20-66, July 1981.

[9] Abramson,N., "The Aloha System", Proc., Fall Joint Computer Conference, Vol.37, AFIPS Press, pp.281-285, 1970.

[10] Metcalfe,R.M., and Boggs,D.R.,"Ethernet : Distributed Packet Switching for Local Computer Networks", Comm. ACM, Vol. 19 No. 7, pp.395-404, July 1976.

[11] Shoch,J.F., and Hupp,J.A.,"Measured Performance of an Ethernet Local Network", Comm. ACM , Vol.23, No.12, pp711-720, Dec 1980.

[12] Needham,R.M.,"System Aspects of the Cambridge Ring", Proc. 7th Symposium on Operating Systems Principles, pp.82-85, Dec. 1979.

[13] Sterry,R.E.,"Ring Nets : Passing the token in local network circles",Data Communications, pp.97-133, Dec. 1981.

[14] Saltzer,J.H., and Pogran,K.,"A star-shaped ring network with high maintainability", Proc. Local Area Communications Network Symposium, Mitre Corp., pp.179-190, May 1979.

[15] Kleinrock.L., and Tobagi, F.A., "Packet Switching in radio channels : Part I - carrier sense multiple-access modes and their throughput-delay characteristics", IEEE Trans. Comm., Vol. 23, No. 12, pp.1400-1416, Dec. 1975.

[16] Saltzer, J.H., Clark, D.D., and Pogran,K.T., "Why A Ring?", SIGCOMM Computer Comm. Review, Vol.11,No.4,pp211-217,Oct. 1981.

[17] Wilkes, M.V., and Wheeler, D.J., " The Cambridge Digital Communication Ring", Proc. Local Area Comm. Network Symposium, Mitre Corp., pp.47-61, May 1979.

[18] Kennington,C.J.,"Problems inherent in current ring protocols", INDRA Note 1040, Dept. of Computer Science, University College, London, Jan. 1981.

[19] Spaniol,O.,"Modelling of local computer networks", Computer Networks, Vol.3, pp.315-326, 1979.

[20] Leslie,I.M.,"Frequency Stability in a Unidirectional Ring of Phase Locked Loops", unpublished working paper of University of Cambridge Computer Laboratory, ca. 1979.

[21] Blair,G.S., " A Performance Study of the Cambridge Ring", Computer Networks, Vol.6, pp.13-20, 1982.

[22] Almes,G.T., and Lazowska,E.D., "The behaviour of Ethernet-like Computer Communications Networks", Proc. 7th Symposium on Operating Systems Principles, pp.66-81, Dec. 1979.

[23] Shoch,J.F. et al., "Evolution of the Ethernet Local Computer Network", Computer, Vol.15, No.8, pp.10-28, August 1982.

Appendix I Analytic Model For Ethernet

(The model is a modification of the one by Almes & Lazowska[22])

Consider an Ethernet System consisting of N active stations. Each station generates packets at a rate of λ packets/sec. Let PS be the length of packets in bits, C be the channel capacity in bits per seconds and S be the round trip propagation delay in seconds.

In order to simplify the analysis, we assume that time is divided into slots equal to the round trip propagation delay. If q stations desire to transmit at some given slot, and if each transmits with probability 1/q (the value that maximizes throughput), then the probability that some station acquires the bus is:

$$A_q = (1-1/q)^{q-1}$$

The mean number of slots wasted per contention is:

$$Z_q = \sum_{i=0}^{\infty} i*A_q*(1-A_q)^i = (1-A_q)/A_q$$

Thus, the instantaneous channel efficiency of the system is:

$$E_q = (PS/C)/(PS/C+S*Z_q)$$

We now construct a Markov's model, whose states are denoted by the number of stations ready to transmit. Let the system be in some state q>0. The system moves to state (q+1) at the rate $(N-q)*\lambda$, where (N-q) are the number of idle stations. The system goes to state (q-1) at the rate $(E_q*C/PS-\lambda*q)$.

The equilibrium probability Xi that the system is in state i is:

$$Xi = ((N-i)*\lambda/(E_i *C/PS-\lambda*i))*Xi-1$$

$$= \{\prod_{j=1}^{i}(N-j)*\lambda/(E_j *C/PS-\lambda*j)\}*X0$$

Since $\sum_{j=0}^{N} Xi=1$

We have

$$X0 = 1/(1+\sum_{i=1}^{N}\{\prod_{j=1}^{i}(N-j)*\lambda/(E_j *C/PS-\lambda*j)\})$$

Using Little's Law, the mean delay t can be calculated as:

$$t = (\text{Mean queue length}|\text{system in state } i)*Xi/(N*\lambda)$$

or, $t = (Ui/(1-Ui))/(N*\lambda)$

where $Ui = N*\lambda/(E_j *C/PS)$

Note the delay time t is the total time the packet stays in the network, including any queue wait time (which is zero in Almes and Lazowska's model).

Appendix II Analytic Model for the Slotted Ring

Let N = number of active stations,

n = number of slots on the ring,

λ = packet arrival rate at each of the stations,

w = walk time (i.e., time required by a packet to traverse the ring once),

p = probability a slot is non-empty.

We assume the gap between consecutive slots is negligible and a message is always correctly received when it arrives at its destination. Thus the results are optimistic and form the upper bound of performance. We also assume the protocol requires a message slot to be emptied by its original sender (i.e., it needs to go round the ring at least once before it may be used again).

If the system is not saturated, then

$$p = \frac{N*\lambda*w}{n}$$

since w is the mean length of time a packet stays in the ring. Now the mean time a station waits to find an empty slot

$$= \frac{w}{2n} + \sum_{K=1}^{\infty} p^{k}(1-p)*\frac{w}{n}*k$$

$$= \frac{w}{2n} + \frac{w}{n}*\frac{p}{(1-p)}$$

Thus the mean time for a station to put the packet at the top of its queue onto the ring

$$= \frac{w}{2n} + \frac{w}{n} * \frac{p}{(1-p)} + \frac{w}{n}$$

$$= \frac{w}{2n} + \frac{w}{n(1-p)} \qquad \underline{\qquad\qquad\qquad} \quad (1)$$

The mean delay time t = mean time a packet spends waiting at a station before it is completely injected into the ring +mean propagation delay time. ____________ (2)

Mean propagation delay = w/2

Assume each station may be modelled as an M/M/1 queuing system, the mean time spent by a packet in the system is simply 1/(mean service rate-mean arrival rate) where the mean service rate is given by the reciprocal of equation (1) and the mean arrival rate is $1/\lambda$. Thus the mean delay time t can be computed from Equation (2).

Appendix III Analytic Model for the Token Ring

The derivation follows the framework of Tanenbaum [3].

Let w = walk time (sec),

s = scan time (i.e., the mean interval between successive arrivals of the token at a station)(sec),

N = number of active stations,

λ = packet arrival rate at each station (packets/sec),

q = mean number of packets waiting at a station,

PS = mean packet size (bits),

C = mean service rate at each station (bits/sec) (=transmission rate of the channel.)

Assuming a station is allowed to transmit all of the waiting packets when it is in possession of the token, then

$$s = w + \frac{N*q*PS}{C} \qquad \text{———————} \quad (1)$$

$$q = s*\lambda \qquad \text{———————} \quad (2)$$

Substituting (2) into (1), we get

$$s = \frac{w}{1 - \frac{N*\lambda*PS}{C}} .$$

Now mean delay time = mean time to wait for the token + mean time a packet has to wait after the token has arrived before it is placed onto the ring + mean propagation delay time.

$$= \frac{S}{2} + \frac{q*PS}{2*C} + \frac{W}{2}$$

$$= \frac{S}{2} + \frac{S*\lambda*PS}{2*C} + \frac{W}{2}$$