NBS
PUBLICATIONS

NAT'L INST. OF STAND & TECH R.I.C.
A11104 076574

NBSIR 84-2966

An IEEE 802.4 Token Bus Network Simulation

Jean-Luc Archambault

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Institute for Computer Sciences and Technology
Systems and Network Architecture Division
Gaithersburg, MD 20899

October 1984

U.S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

QC
100
.U56
84-2966
1984
c. 2

NATIONAL BUREAU
OF STANDARDS
LIBRARY
Lwc
QC100
U56
no. 84-2966
1984
c. 2

NBSIR 84-2966

AN IEEE 802.4 TOKEN BUS NETWORK SIMULATION

Jean-Luc Archambault

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Institute for Computer Sciences and Technology
Systems and Network Architecture Division
Gaithersburg, MD 20899

October 1984

U.S. DEPARTMENT OF COMMERCE, Malcolm Baldrige, *Secretary*
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Director*

AN IEEE 802.4 TOKEN BUS NETWORK SIMULATION

Table of contents

I) Introduction

II) The simulation model

III) The experiments

IV) The results

V) Conclusion

A discrete event simulation of token bus networks has been designed and implemented at the National Bureau of Standards, and used to analyze the performances of local area networks compliant with the IEEE 802.4 specifications. The model measures the utilization of the network, the rotation time, the waiting time of the data packets, and the queue lengths in the Medium Access Control sublayer. This paper presents the model, and studies the sensitivity of these variables to the offered load, the packet length, the target rotation times, and the number of stations.

Key words: Local Area Network; Modelling; Performance Analysis; Simulation; Standards; Token Bus.

I) INTRODUCTION

The National Bureau of Standards' Institute for Computer Science and Technology (ICST) has an ongoing standards program to develop protocol specifications for local area networks. ICST participates in the voluntary standards arena, assists other federal agencies deploy local networking technology and conducts research necessary to develop technical specifications for proposed Federal Information Processing Standards (FIPS). The Local Area Network Group in ICST has placed high priority on the development of a Token Bus local network FIPS.

Office automation networks may employ contention systems such as CSMA/CD that arbitrate use of the physical medium with statistical multiplexing techniques. But these are not appropriate for process control applications requiring prioritized and deterministic access with robust error detection and recovery mechanisms. Token passing networks are expected to exhibit these attributes.

That is why NBS and General Motors have commenced a joint project to be able to perform compliance tests and performance measurements on local area networks implementing the IEEE 802.4 specification. This project utilizes the skills and expertise of many companies, both users and manufacturers of token bus equipments.

One of its functional areas is modelling, with the goal of understanding the token bus access mechanism and evaluating the performances of such networks in various situations, including some that may not be practical in a measurement laboratory, such as a large number of stations. The simulation that has been designed is a preliminary modelling effort to gain some knowledge about the behavior of token bus networks and to help define the experiments to be conducted in the laboratory.

The purpose of this paper is to describe this simulation model and then to present and analyse some results of experiments conducted with it, to know the sensitivity of the measured variables to the network parameters.

II) THE SIMULATION MODEL

II.1. The 802.4 Protocol

The model is designed according to draft F of the IEEE 820.4 standard, but this protocol has been simplified, following the hypotheses:

there is a fixed number of stations in the ring, no noise on the channel, and no fault in the operations of the stations. Simply, everything happens as normally and smoothly as possible. So, the only functionalities of the IEEE 802.4 standard simulated here are sending data frames, token passing, soliciting new stations (without success), and token claiming (only when there is a single station).

Here are some details about these main functionalities:

a) sending data packets:

upon reception of the right to transmit (the token), a station sends the data packets pending in its highest access class (or "class 6") as long as data is in the queue and time (called "high priority token holding time") is available. Then the station comes to the next access class (number 4) and sends its packets; but here, the maximum time is computed differently: when the access class is allowed to transmit, it measures the difference between the current time and the time of the last entry into it, which is the time of the last rotation of the token, as seen by this class. The difference between the "target rotation time" which is assigned to the class and the actual rotation time is this maximum time available. If it is negative, no packet can be transmitted and the station goes to its next classes (number 2 and 0). Then, before passing the token, the station performs "ring maintenance".

b) maintaining the ring:

this is the solicistion of a new succwssor, which must be located between the soliciting station and its present successor. Here, "located between two stations" means to have an address between both others. But it is allowed to do so only at two conditions: a counter, loaded with its maximum value ("max_inter_solicit_count") at each solicitation and decremented upon every rotation must have come down to zero, and there is also a target rotation time, allowing the ring maintenance only if the token has rotated quickly enough before. This function consists of sending a solicit_successor frame on the bus, then waiting for any answer during one slot_time (two if the station has the lowest address); in this model, no answer comes, because there is no station not being in ring and wanting to be.

c) passing the token:

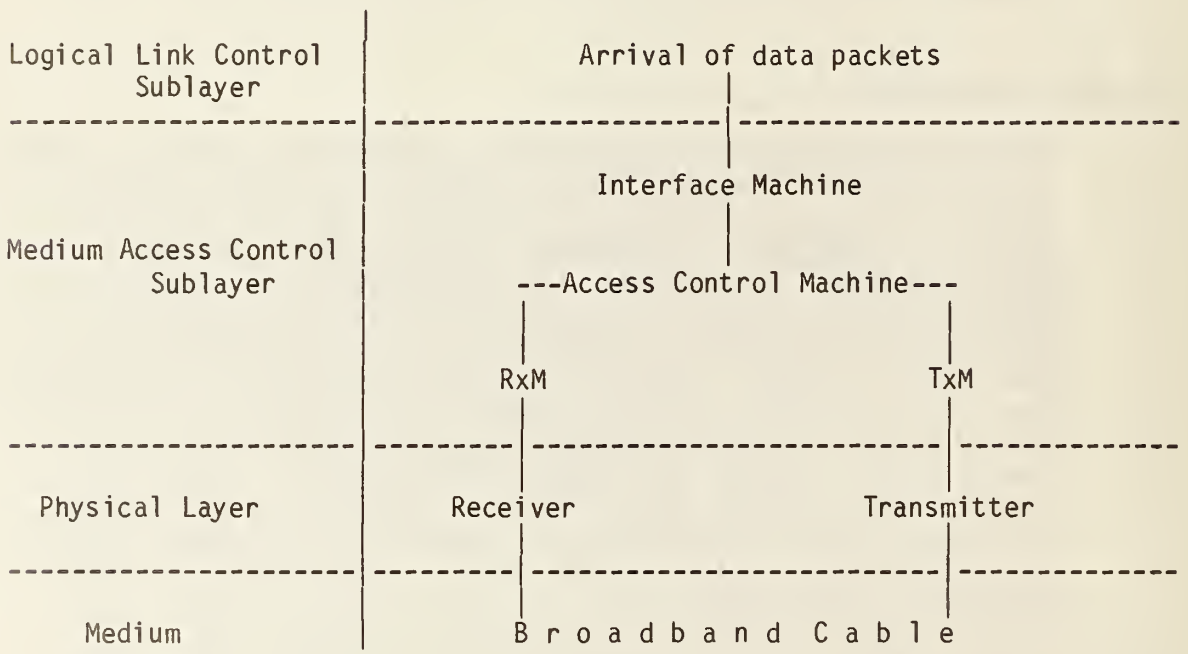
this is the next function performed by a station. A token frame is sent to the successor of the station, with the next lower address. The delay created here is just the time for the token to reach its destination, depending on the length of the cable and the positions of the stations. The station with the lowest address sends the token to the one with the highest address. The case of a single station in the logical ring is different: if it has pending packets when it is no longer allowed to send some (because of the maximum time available), it performs a "silent pass to self", and then may go on processing them. Else, it remains idle, sensing the channel for a period of time called "max_bus_idle", after which either it has still no pending data and starts another idle period, or else starts the token claiming process to re-initialize the logical ring, getting the right to transmit at the end.

For more details about these functions, see draft F of the 802.4 standard, especially section 5.

II.2. The model

The model is a discrete event simulation, based on draft F of the IEEE standard, with the restrictive hypotheses described above. Up to now, only the physical and medium access control (MAC) sublayer has been implemented, and the network operates with no connection, a packet is considered successfully transmitted as soon as its last bit has been sent on the medium. The network is made up of a certain number of stations connected to a broadband cable with a head-end remodulator.

Each station has the following structure, according to the specifications of the standard (Cf. draft F, page 1-5):



In this model, the Logical Link Control (LLC) sublayer generates the data packets to be transmitted. The Interface Machine (IFM) handles four queues (one per access class) and makes their packets available to the Access control Machine (ACM). The ACM is limited to the functions that have been described above. The Receive Machine (RxM) is not represented in this model because all its functions are irrelevant here (the stations need not know whether the bus is idle, for instance, because they are "black boxes" with a behavior described by the model, but without taking care of how things happen in details) and the Transmit Machine (TxM) just adds the necessary bytes to transform a packet into a frame. The physical layer is simply represented by the time required to transmit a frame at the given data rate. The medium and the remodulator create the delays between the stations.

II.3. The statistical functions

Statistical functions are at both ends of the model. At the "top" are random number generators, which are used in the simulation of the arrival of data packets. At the bottom are all the functions gathering and computing the statistical outputs of the model, and those deciding when to stop the run of the simulation. The method for stopping the simulation is the batch means method, with the ability to use either the mean rotation time or the waiting time of data packets in the highest access class or the waiting time of all the data packets as the parameters monitored. The basic idea of this method is to divide the simulation into batches of a fixed number of rotations (even when the stopping variable is not the rotation time). Then, the mean value of the chosen statistics is computed for each batch and, as an approximation, the central limit theorem is used to know whether the accuracy the user wishes is met or not. To improve the precision of the approximations made in this method, the first batches are ignored, to eliminate the "warming up" of the system. Of course, when a waiting time is chosen instead of the rotation time, the calculations are different because the number of data packets changes from one batch to another.

II.4. The parameters

Here is a list of all the inputs to the model that the user specifies:

* The data rate: according to the standard, it may be 1, 5 or 10 Mb/s.
* The slot time: it is the round trip propagation delay, as defined in page 6-2 of draft F.
* The latency at the head-end repeater.
* The number of stations connected to the cable: it may be between 1 and 1000.
* The propagation delays between the stations and the repeater: with the possibility to ask for a uniform distribution of them, computed from the value of the slot time.
* The number of packets a queue of the IFM can contain: it is the same for all the queues of all the stations.
* The offered load: as a percentage of the capacity of the channel.
* The arrival process for each access class of each station: the length of the packets is uniformly distributed between two bounds specified by the user; the interarrival times can be constant, uniformly or exponentially distributed, with a mean value computed from the offered load. It is also possible to have no packet arriving at a queue.
* The token timers: for each station, hi_pri_token_holding_time and the target rotation times of the three lower classes and of the "ring maintenance".
* The address of each station: only the 2 byte addresses are implemented.
* The "max_inter_solicit_count": its value is not randomized, although the 802.4 standard requires this (cf. p. 7-2 of draft F).
* The number of rotations in a batch: its default value is 50.
* The number of batches ignored at the beginning of the simulation: its default value is 40.
* The variable chosen to stop the simulation (see above).
* The value of the "criterion", which is the half-width of the confidence interval as a percentage of the mean value, for the variable monitored (Cf. II.3).

II.5. The architecture of the simulation

II.5.1. Overview

II.5. The Architecture of the simulation

II.5.1 Overview

The program implementing the model is classically organized in blocks and sub-blocks. There are four main blocks: NETWORK, which modelizes the machines of the network; INPUTS; OUTPUTS; STATISTICS. We now indicate all the sub-blocks of the program, and for each of them the main functions performed.

II.5.2. The NETWORK block

a) ARRIVAL sub-block: It is divided in two sub-units, RANDOM and ARRIVAL PROCESS.

*RANDOM:

it contains randon number generators which can operate according to an exponential or uniform distribution.

*ARRIVAL PROCESS:

it computes the interarrival time, the length of the data packets, and the length of the data unit of the claim_token frames. It uses these results to create new packets and to fill the queues of the access classes of the stations.

*Remark: in this model, the queues are always full; they are filled at the initialization and then, every time a packet is sent, a new one is put at the bottom of the queue to replace it. Of course, it may happen that a queue is not full, really; in this case, some of the packets have an arrival time bigger than the current time and cannot be sent before their arrival time. Also, this time is computed as the maximum of the arrival time normally given by the chosen process and of the current time; this is necessary because the queues are finite and so, packets could want to enter them when they are full.

b) INTERFACE MACHINE sub-block:

tells the NETWORK sub-block whether there are frames pending in a queue or a station and picks a packet out of them on request of this sub-block.

c) NETWORK sub-block:

This sub-block is the core of the simulation. It performs all the functions of the Medium Access Control sublayer and of the physical layer and the medium. It also asks for the updating of the various statistics and decides whether the simulation must be stopped. It calls, directly or not, all the functions of the NETWORK block.

Its logical organization may be represented thus:

```
while another batch is necessary
        |
        | for all the rotations of the batch
        |    |
        |    | for all the stations
        |    |    |
        |    |    | for all the access classes
        |    |    |    |
        |    |    |    | compute the rotation time
        |    |    |    | while it is permitted
        |    |    |    |    | send a data packet
        |    |    |    |    | update the packet statistics
        |    |    |    |    |
        |    |    |    | update the class statitics
        |    |    |    |
        |    |    | manage the ring
        |    |    |
        | compute the new relative half-width of the  confidence
        |         interval
        |
```

II.5.3. INPUTS block

a) INPUTS sub-block:

It reads from the user's answers to the questions asked by the INITIALIZATION sub-block from the terminal.

b) INITIALIZATION sub-block:

It asks the user for the values of the parameters of the simulation, including those related to the stopping criterion. A predefined set of inputs is available to save time, but it is also possible to specify on each run all the inputs. This block also initializes the simulation, creating all the required variables.

II.5.4. STATISTICS block

a) INTERMEDIATE sub-block: Performs the updating of the statistical variables necessary after every transmission of a packet (data or not), and after a pass of the right to transmit in an access class of a station.

b) FINAL sub-block: Makes all the calculations necessary at the end of the simulation to know the values of all the outputs.

c) CRITERION sub-block:

Gives the current estimation of the accuracy of the results for the variable chosen to know when to stop the simulation run (either the mean rotation time or the waiting time of the data packets in the highest access class or of all the packets).

*Remarks:

*The simulation must go on as long as the requested accuracy is not reached and the number of batches is smaller than a minimum (the ignored batches plus 50) and the number of data packets used in the measures is smaller than a minimum (2000).
*The number of batches is also limited to 1000.
*The code of the network() function being not very long, it often happens that there is no particular procedure corresponding to a given line of the synopsis described above. This also reduces the number of calls to functions and so improves the speed of the program.

II.5.5. OUTPUTS block

Prints all the values of the parameters of the simulation (at its beginning) and of the results (at its end) on the output file of the program.

II.5.6. Remark

At the beginning of the program are all the declarations of the constants, structures and variables (except some used very locally by I/O functions), and also the main procedure. Here is the list of the structures defined in the simulation: PACKET, CLASS, STATION, DATAKIND, containing both the variables required by the model of the network and those for the statistics about the packets, the access classes, the stations and the different kinds of data packets.

II.6. OUTPUTS

II.6.1. Global statistics

*duration of the simulation.
*number of batches performed, including the ignored ones.
*number of recorded packets and data packets, not including those in the ignored batches.
*waiting time of the data packets:
the maximum, the mean value and the coefficient of variation.
*the total utilization, i.e. the share of time used by the transmission of bits of any kind.
*the information utilization, i.e. the share of time used by the transmission of data bits, not including the headers of the data or the protocol frames.
*the mean rotation time, as observed by station number 0 (but it happens to be the same in all the stations, as a very good approximation).
*the time between two successor solicitations:
the maximum, the mean value, the mean of the coefficients of variation in all the stations, and the total number of such solicitations.

II.6.2. The global statistics for each access class

*the number of data packets sent.
*the waiting time of the data packets:
the maximum, the mean value, and the coefficient of variation.
*the information utilization.
*the queue length when the token enters this access class:
mean value and mean of the coefficients of variation in all the stations.

II.6.3. The statistics for each station

*the number of packets and data packets sent.
*the waiting time of the data packets:
mean value and coefficient of variation.
*the total and information utilizations.
*the statistics for each access class of this station:
_the number of data packets sent.
_the waiting time of the data packets:
_the maximum, the mean value, and the coefficient of variation.
_the time between two successor solicitations:
_mean value, coefficient of variation, number of such solicitations.
_the information utilization.
_the coefficient of variation of the rotation time as seen from this
access class.
the queue length when the token enters this access class:
mean value and coefficient of variation.

II.7. VALIDATION

The main and best way to validate the model and the program implementing it is to study the correlations between the results of the simulation and of the laboratory. This task has not been completed yet because the laboratory is not operational.

However, some validation has been made. In a first step, there was no other stopping criterion than a limit to the number of rotations the model was allowed to perform. It was used in some particularly simple configurations (such as a few stations, without the priority scheme) where the results were easily predicted, the access method being deterministic. The means to get accuracy was to have the program perform thousands and thousands of rotations. The convergence toward the expected results was observed. Then the statistical stopping criteria were tested by comparing the outputs they provided to previous ones. It appeared that discarding the results of, say, the first 2000 rotations allowed the model to reach the same accuracy in shorter runs. This value of 2000 was chosen after studying the "warming up" of the system, observing its behavior for 1000, 2000, 3000... rotations. The internal consistency of the results has also been verified, by checking that some relations which should exist between, for instance, the offered load and the mean rotation time are true. This was made also in situations for which it does not seem possible to predict all the outputs, and an example of this can be found in the comments on the experiments, below (Cf. section IV.1).

A remark about the accuracy:
The model is supposed to yield a value of the stopping variable such that the true value is with a 95% chance within a x% wide interval around it, where x is given at the beginning of the run. But one must be aware that several approximations were made, mainly when asserting that the variables considered on the batches are independent and identically distributed. Also, the random number generators create some uncertainty. In fact, it does not look reasonable to expect an accuracy better than 5 to 10% for all the outputs of the model. A way to improve it could be to run the same experiments several times with different "seeds" for the random number generators, and to compute the mean results. But in most of the applications, this should not be necessary.

III) THE EXPERIMENTS

Several sets of experiments have been conducted with the model, to study the influence of the offered load, the packet length distribution, the target rotation times, and the number of stations in the network.

All the other parameters of the model have kept a constant value throughout all the experiments. Here are their values:

* The data rate = 5 Mbits/s.
* The slot time = 45.6 microseconds.
* The latency at the head-end repeater = 10 microseconds.
* The propagation delays between the stations and the repeater: their distribution was uniform. The signals were supposed to need 6.4 microseconds to run along the cable (one way). This value comes from the previous two. For N stations on the bus, they are distant from their neighbor by 1/(N+1) times the length of the cable.
* The number of packets a queue of the IFM can contain = 10.
* The arrival process at each access class of each station: the length of the packets is uniformly distributed between two bounds specified by the user; the interarrival times are exponentially distributed.
* The token timers: hi_pri_token_holding_time = 409.6 microseconds.
* The address of each station: only the 2 byte addresses are implemented. and the addresses are equal to the number of the station, the closest to the remodulator being station number 0.
* The "max_inter_solicit_count": its value is not randomized, although the 802.4 standard requires this (cf. p. 7-2 of draft F). It is 255.
* The number of rotations in a batch = 50.
* The number of batches ignored at the beginning of the simulation: it is equal to 40 most of the time, except for loads from 90% with 20 stations or more because it would have made runs too long to keep 40, and in every batch, there are many packets (10000 for instance).
* The variable chosen to stop the simulation is the mean waiting time of the data packets in the high priority class.
* The value of the criterion = 7%.

In addition, the minimum number of measured batches has often been limited to 10 instead of the usual value of 50, when this gave enough data packets. Of course, the statistical criterion had always to be reached before the simulation stops.

The different sets of experiments are the following:

set #1: the number of stations is 4, the data packet length is constant and equal to 100 bytes, the target rotation time in the lower access classes is 10 milliseconds, and the offered load varies from 10% to 150% by steps of 10%, and then to 500% by steps of 50%.

set #2: same conditions as in the first one, except that the packet length is successively equal to 50, 150 ,and 205 bytes, plus another configuration where it is uniformly distributed between 10 and 400 bytes.

set #3: the number of stations is 4, the data packet length is constant and equal to 100 bytes, and the target rotation time in the lower access classes is 1.5 and then 3 milliseconds. The offered load has the usual behavior.

set #4: the number of stations takes the values 20, 40, 100, the packet length is constant and equal to 100 bytes; the target rotation time in the lower access classes is respectively equal to 50, 100, and 250 milliseconds. The offered load is limited to 200%, because most of the interesting phenomena happen in this range (but not all).

Remark: in the figures accompanying the following comments, the time unit is 1 microsecond, unless otherwise specified. The abbreviation "ms" stands for "millisecond".

IV.1. Relationships to the offered load

The results commented here come from the first set of experiments, as described in the previous section. Below are studies about the influence of the offered load (as a percentage of the channel capacity) on the time, queue lengths.

IV.1.1. Channel utilizations versus load

Figure 1 displays the information utilization (proportion of time spent sending data bits) and the total utilization (for all bits except the preambles) as percentages of the channel capacity. The maxima are 87.1% and 97.4% and are reached for loads around 90%. In fact, for offered loads below the maximum data utilization, this variable is equal to the load. When it attains the maximum, it remains at this value, without decreasing. From now on, "utilization" is used for "data utilization". The relative difference between the utilization and the offered load is less than 5% for loads up to 90%.

It is possible to give a formula for the maximum utilization, as a function of the parameters of the network: let token_time be the time to pass the token around the logical ring, and maxrt the maximum rotation time, which we may expect to be close to the target rotation time (and it is, see the next section). Let also lr be the ratio between the length of the data unit of a data frame and its total length (preamble included). Then, the maximum utilization is:

$$maxutil = lr*((maxrt - token_time)/maxrt)$$

because (maxrt - token_time) is the time used by the data frames. Here, the time to solicit successors has not been considered; so, this formula is only an approximation of the reality and of the results of the simulation model, but a good one, since this event does not occur often and has a duration of the same magnitude than the time to send one data frame. The value of token_time for all the experiments is 0.15 ms. Here, the values computed by the simulation model and this formula are equal to 87.1% and 87.2% respectively. But the actual maximum rotation time is smaller than the target value (see next section), and using the real value gives 87.1%. As the formal proof of the formula is independent of the use of the simulation model, the agreement between both values is a positive element for the validation of the simulation.

The values of the utilization in the different access classes will be studied later in the section on the influence of the target rotation time (IV.4.3).

It can also be noticed that, in this configuration, the access method is fair at all loads. The data utilization varies by a few percent of the total from one station to another, and no significant difference can be isolated from the unavoidable statistical variations (the offered load is only a mean value).
The maximum utilization is a crucial point for the studies about the dependence on the offered load, as it can be seen right below.

The total utilization, which does not count the preambles of the frames, has the same behavior, although it starts at a much higher value, which comes from the token passing (here, a little above 47%). This variable can be analytically predicted, but as the calculations need some results about the rotation time, they will be performed in section IV.3.1.1, for various packet lengths.

IV.1.2. Rotation time versus load

Figure 2 displays the mean rotation time (as seen by the station with the lowest address) as a function of the offered load. This variable remains small for loads up to 70% of the channel capacity. At this point, it is below 1/10 of the target rotation time. Then it increases very quickly, and when the load reaches the channel capacity, it is at its limit, which is here 95% of the target time. It is not very surprising to get the entire target rotation time greater than the actual limit value, because, no rotation being larger, their mean must be smaller. In fact, slow rotations where the lower access classes sent many packets may be followed by shorter rotations where they are not allowed to do so. However, it should be noticed that the actual rotation time can be larger than the target value by the length of a data packet, because it is enough to have time to start a transmission to be allowed to perform it fully. But here, the difference is small: a data frame lasts 0.18 ms while the target rotation time is 10 ms. The coefficent of variation is between 5% and 10% for loads from 90%: the differences from one rotation to another remain small. Notice that for loads up to 80%, this coefficient increases from 40% to 100%.

It is possible to give a formula for the mean rotation time when the load is not too large, which means not close to the maximum utilization. Always ignoring the time to solicit successors, the rotation time minus token_time (see previous section) is the time used sending data frames. So, the utilization is lr*((rt - token_time)/rt), where rt is the actual rotation time. But we have just seen that when the load is below the maximum utilization, it is equal to the utilization. This gives then:

$$\text{rotation time} = \text{token_time}/(1 - \text{load}/\text{lr})$$

The relative difference between the values predicted by the model and this formula is at most 2% for loads up to 80%. Once more, this a step toward a validation of the simulation. It can be noticed that in the above formula, the target rotation time and the high priority token holding time are absent. In fact, they are irrelevant to the actual rotation time only for loads below the maximum

utilization, which depends strongly on them. However, one can say, for a rough estimate, that for low loads, the rotation time is not sensitive to these parameters.

Notice also that the maximum rotation time is close to the target rotation time only because the high priority token holding time is small enough. More precisely, the target time is here larger than the maximum time that can be used by the high priority classes and the token passing. So, it can determine the limit of the rotation time. Further discussions of this kind of questioV.4.

IV.1.3. Waiting time versus load

Figure 3 shows the waiting time as a function of the offered load. Waiting time is the time spent by the data packets in the finite queue of the Interface Machine of the stations. The behavior is similar to the rotation time, although the convergence toward the limit is slower. Once more, three areas can be distinguished: for loads below the maximum utilization, it increases gradually, remaining under 6% of its value for very high loads when the load is at most 80%. Then a very sharp increase, and then from 100%, it goes up more and more gently and looks to have a limit of 28 ms.

The fact that the rotation time and the waiting time have the same kind of behavior is not very surprising. When a packet enters a queue, it has some packets before it and, for instance in the highest class, a fixed number of frames may be sent on every rotation (the packet length is constant). So, a data packet has to wait a certain number of rotations before being sent, whence the correlation between both variables. The waiting time of packets remains bounded, even for loads above 100% because it is not really computed from the user's point of view; it is the time spent in the queues of the IFM, which are finite, and so is the waiting time. Here, the limit value is around 3 times the maximum rotation time, and the highest classes may send at most 3 packets per rotation out of a queue of 10. But the value of the target rotation time is also relevant here (for more details, see section IV.4).

The coefficient of variation of the waiting time is abount 1 for low loads and decreases for loads between 90% and 150%, remaining below 1/3, and then increases again up to 1.

The maximum waiting time of the data packets can be an important issue for some applications. The results obtained here cannot provide an accurate information, because it is not a mean value. This time is less than 10 ms for loads up to 70%, then jumps to more than 30 ms and remains finally around 70 ms (the mean waiting time being around 30 ms). Of course, conducting the same experiments with different values of the seed of the random number generators could have given noticeably different values.

IV.1.4. Queue length versus load

Figure 4 shows the mean queue length when the right to transmit arrives in an access class, for class 6 and the mean of classes 4 to 0 (the three variables being always very close). Once more, the curve has the same shape than the two previous ones. The length remains below half a packet for loads up to 80%; in this range, it is approximately the same in all the access classes. The limit value is of course 10, the size of the queue, and in-between, it is smaller for the lower access classes. Thus we can suppose they may send more packets than the others. This can be checked by computing the maximum time used by the highest classes and the token passing , which gives by difference with the target rotation time what remains for the lower classes; here, the value is 0.65 ms per class, which makes 3.6 frames (preamble included), while class 6 is only allowed to 3. So, it is normal that the queue is smaller in the lower classes; but this comes from the values of the target times.

Once again, when the load is under the maximum utilization, things happen very smoothly: the queue length is less than 1 most of the time and independent from the target timers. Notice that when the load is under 100%, the offered load in each access class is under 1/(4*4) of 100%, which is 6.25%.

IV.2. Relationships to the utilization

All the variables that have been studied above have the same behavior as functions of the information utilization. That is why only one of them is represented in this section. Figure 5 shows the mean waiting time of the data packets as measured in the first set of experiments. The main character that can be observed is that the values remain small and slowly growing for loads say, 10 to 15% below the maximum utilization, and then grow to much higher values. Even for variables having a finite limit when the utilization becomes maximal, the slope of the curve remains very steep because the utilization has the fastest convergence.

IV.3. Sensitivity to the packet length

The results commented here come from the second set of experiments, as described in section III. Below is a study of the influence of the data packet length (mean and also distribution) on the same variables as in section IV.1.

IV.3.1. With a constant packet length

IV.3.1.1. Utilization versus packet length

Figure 6 shows the information utilization as a function of the offered load, for four different packet lengths. Something very simple can be observed: the line y = x , cut by horizontals gives the maximum utilization. This maximum increases with the packet length, because less tokens are required for the same number of data bits. The formula demonstrated in section IV.1.1 is yet true, and gives the following results: 78.2%, 87.1%, 90.6%, 92.6% , while the maximum utilizations predicted by the simulation are: 78.1%, 87.1%, 90.6%, 92.6%. The correlation is very good for all the packet lengths. For loads below the maximum utilization, the value is equal to the load with no signficant difference, for all packet lengths.

The total utilization has a behavior displayed by figure 7. It appears that it is larger for shorter packets when the offered load is below the smallest of the maximum utilizations, but its limit value is smaller. As said in section IV.1.1, the total utilization is predictable. Here are the calculations for the total utilization:

The rotation time is divided into two parts, one for the token passing, which is independent of the offered load, the balance for the data frames (not considering the successor solicitation). The four tokens last 4*11*1.6 = 70.4 microseconds (11 octets, 1.6 microsecond for each), and the data frames last the balance of the time, to correct by a multiplicative factor , called lr2, which is the ratio between the length of a frame without its preamble and with it. Dividing by the rotation time gives the required percentage:

$$tut = (70.4 + (rt - token_time)*lr2)/rt$$

But the rotation time is known as a function of the offered load (section IV.1.2), and this yields:

$$tut = tut0 + ld*(lr2 - lr*tut0)$$

where tut0 is the value for a null load, tut0 = 70.4/token_time, 70.4 being the duration of four token frames, token_time adding the time

for their transmission. Here, tut0 = 47.3% and the formulas are the following:

50 bytes:	tut = 47.3 + ld*0.624
100 bytes:	tut = 47.3 + ld*0.575
150 bytes:	tut = 47.3 + ld*0.559
205 bytes:	tut = 47.3 + ld*0.551

For all packet lengths and all loads up to 80%, the relative difference between both predictions is under 1.5%. At 90%, it is 8% for 50 bytes and 2% for 100 bytes, less than 1% for the others. The maximum total utilization can be found by replacing the load by the maximum data utilization in the formula. This gives 96.1%, 97.4%, 98%, 98.3%, while the values from the simulation are 96%, 97.4%, 98%, 98.3%. The correlation is quite good. It should be noticed that all the formulas that have been given up to now do not care about the distribution of the offered load. They only used the constant packet length and the assumption that all the processes out of data transmission and token passing take very little time. To include the preambles, add two octets to all the frames.

IV.3.1.2. Rotation time versus packet length

Figure 8 shows the mean rotation time as a function of the offered load for four packet lengths. Approximately, the rotation time is not sensitive on the packet length, as the curves show. For loads below the maximum utilization, the rotation time is a decreasing function of the packet length, and remains much smaller than the target value. But the relative differences can be rather large: for loads below 50%, it is at most 20%, but for loads between 60% and 90%, it is around 100%. The reason why is the difference in the maximum utilizations (see above). Indeed, the sharp increase in the rotation time begins before 78% for the 50 byte packets, while it is still slowly growing for the 205 byte packets, because their maximum utilization is 92.6%. For instance, the values for these two lengths and an offered load of 80% have a ratio equal to 8.

For low loads, the formula given in section IV.1.2 may be used and checked. It also shows a decreasing function of the packet length. For the 50 byte packets, the relative difference between the values given by both predictions is less than 6% for loads up to 70%. For the other packet sizes, it is less than 4% for loads up to 80%. This makes an additional verification of the formula.

For loads of 100% and above, the relative differences remain below 3%: the maximum rotation time is almost the same. The values given by the simulation are : 9.43 ms for 50 bytes, 9.50 ms for 100 bytes, 9.46 ms for 150 bytes, 9.7 ms for 205 bytes. The results for 100 and 150 bytes look to be in the wrong order, compared to the others. In fact, this cannot be regarded as significant here, because a 1% increase in the result for 150 bytes would reverse the order, and the accuracy of the rotation time as predicted by the simulation is not well known; it

depends on the "seed" of the random number generators (which is changed on each run) and on the assumptions of the batch means method. So, it might be around 1%, and no conclusion can be drawn from this strange fact. However, the limit rotation time looks to be an increasing function of the packet length (a 3% increase when the length goes from 50 to 205 bytes).

IV.3.1.3. Waiting time versus packet length

Figure 9 displays the waiting time of the data packets in the IFM queues as a function of the offered load for four different packet lenghts. The shape of the curves is the same than for the rotation times, but the maximum values differ much with the packet size. Once more, three areas: for loads up to 60%, the maximum relative difference between the results is less than 30%, and the waiting time increases with the packet length. Then from 70 to 90%, the difference changes signs and becomes much larger, up to 300%. For offered loads greater than 100%, the initial order comes back and the limit values are: 15.8 ms for 50 bytes, 28.4 ms for 100 bytes, 41.2 for 150 bytes, 55.0 ms for 205 bytes.

It is possible to roughly compute these values in the following way: there are 40 packets pending in a station at any time when the offered load is very high. The mean number n of data packets sent on a rotation in every class can be computed, as it was described in section IV.1.3 for 100 byte packets. Then the maximum mean waiting time is approximately 40*maxrt/n. This formula gives: 15.7 ms, 29.4 ms, 40.3 ms, 51.7 ms for the four packet sizes. The fit is not excellent, but gives a good estimate of the true value (the relative difference is at most 6%, and the accuracy of the predicted waiting time is probably not much better).

So, except for loads around the maximum utilization of the network, the waiting time of the packets increases with their length.

IV.3.1.4. Queue length versus packet length

Figure 10 presents the queue length for the highest access class and the four packet lengths. At any load, there is more queueing of packets when they are shorter. Their mean number is less than 1 for loads up to 70%, that is below all the maximum utilizations. The sudden increase begins sooner for shorter packets, as for the other variables. The fact that the curves for 100 and 150 bytes are close while the others are not comes from the number of packets a high priority class may send: at most 5 for 50 bytes, 3 for 100 bytes, 2 for 150 and 205 bytes. So, the 100 byte packets are sent 1.5 times faster than the 150 byte, but they arrive 1.5 times faster for the same offered load because they are 1.5 times smaller; thus, the two curves are do not differ much. On the other hand, the queue length is much smaller for 205 byte packets because they arrive less often but exit at the same rate. In the lower classes, the behavior is quite the same, although the packet length has less influen-

IV.3.2. With a variable packet length

The same experiments described above have been conducted with a packet length uniformly distributed between 10 and 400 bytes. Here is the comparison of their results with those for a constant 205 byte length. Since all the results are very close, no curve for this new set of outputs is appended. The network (or at least the variables that are measured here) is not sensitive to the packet length distribution. Here is what has been observed:

*channel utilizations: The information utilizations are similar, the relative differences being below 5% at all loads, generally around 1%. Up to a 90% offered load, the utilization is a little smaller for a variable packet length. The limit values are the same. Same thing for the total utilizations, which differ by at most 2% of their value, mostly less than 0.5%. The limit is 0.3% larger for a variable length, but this figure is not very significant. So, the formulas used with a fixed packet length look to give good approximations when this assumption is no longer true.

*mean rotation time: It is generally smaller for a variable packet size, but by at most 5%, except a 14% difference for a 90% offered load. The predicted limit value is also smaller, but by only 0.5%, which is not very significant. Once more, the formula giving the rotation time as a function of the load gives a good (although weaker) approximation, even with a variable packet length.

*mean waiting time: Mostly larger for a variable packet size, but by not more than 10%, except a 21% peak around an 80% offered load. The difference remains small, but noticeably larger than for both previous variables. The limit value and the the coefficient of variation of the waiting time are approximately the same.

*queue lengths: At the time the experiments with a uniformly distributed packet length were conducted, the measure of this variable was not implemented in the simulation program. But Little's formula, asserting that the mean number of "customers" in a queueing system is equal to the arrival rate times the mean time spent by them in the system can be used here, because the arrival rate is the same (as the mean length), and so is the mean time spent in the system, since the waiting time is roughly the same and the mean transmission time is the same, as the mean packet length. Thus it is possible to say that the mean queue lengths are not very different in both sets of experiments.

IV.4. Sensitivity to the target rotation times

All the results described here come from the set of experiments number 3, depicted in section III. Section IV.1 suggested that the target rotation time in the lower access classes may be an important parameter. That is why experiments were made in the same conditions as the first set, except the value of the TRT (target rotation time), which was given the values 1.5 ms and then 3 ms, instead of 10 ms. Why these values? The time to pass the token around the ring, plus the time the highest access classes may hold it is 2.35 ms. So, with a TRT of 1.5 ms , the lower access classes will not be allowed to send data at high offered loads, and with 3 ms, this will not happen but probably they will not be allowed to send many, because 3 ms is close to 2.35 ms. With 10 ms, they should have no problem.

Since there is much to say in each situation, they will be studied one by one.

IV.4.1. With a 1.5 ms target rotation time

IV.4.1.1. The utilizations

The curves that help understand what happens to the measured variables are on figure 11: it shows the data utilization in each access class as a function of the offered load. When it is below 80%, nothing special happens and all the offered load is sent. But then the utilization in the lower classes decreases and eventually becomes zero, while for the highest class it keeps on growing and then reaches a limit.

It is possible to perform some analytical predictions about this situation and thus to understand its main features. For low offered loads, there is no problem, every access class may send all the data it wants, because the utilization is far from its maximum and the rotation time far from its target value. So, the protocol limitations do not have to work and the utilization is equal to the load for all the classes. Things change when the rotation time reaches the TRT: the lower access classes will no longer be able to make the rotation time grow and so will not be allowed to send more packets. As class 6 does not have this restriction, it may go on sending frames and thus the share of the others must decrease. The load for which this happens can be computed with the formula about the rotation time given in section IV.1.2: the TRT being 1.5 ms, the load is 79.6%. And figure 11 shows that the curves for the lower classes separate from the other around 80%. When the load reaches 400%, the upper classes offer 100% of the channel capacity, and as they cannot be hindered by the others, they use all they can. But then the lower classes may not add any data, and so send nothing. The

rotation time can be calculated: it is token_time + 4*3*0.1808 ms because there are 4 stations and each class 6 may send at most 3 packets per rotation, each lasting 180.8 microseconds. This makes 2.32 ms , and the value given by the simulation is the same. Now, the maximum utilization is also known: using the formula of section IV.1.1 gives 82.9% , the measured value being 82.8%.

When the load is between 79.6% and 4*82.9 = 331.7% , the utilization of the three lower classes decreases. An upper bound for the mean of these three variables can be calculated:

when the rotation time is equal to the TRT, each access class of each station uses (TRT - token_time)/(4*4) because there are 4 stations and as many classes in each. This makes 84.4 microseconds, which is less than the maximum time for the highest class, which is the time to send 3 packets as already seen, i.e. 542.4 microseconds. So, this class may go on sending data without restriction, the others are not able to make the rotation time grow, and the utilization in class 6 remains equal to the load it offers. A consequence is that classes number 4, 2, and 0 have at most (maxutil - load/4) for them. And this value is zero when the load is 4 times the maximum utilization. All this determines a zone where the mean of the utilization in the three lower classes must remain. Figure 12 shows in dashed lines the expected behavior of the class 6 utilization and of the mean of the others. The solid lines represent the values given by the simulation model. The correlation between both kinds of prediction is good. But the equation

$$y = (\text{maxutil} - ld/4)/3$$

gives only a maximum value, the actual one being found by replacing maxutil by the real utilization, which is a little smaller.

Remark: if the offered load is not equally distributed among the four classes, the same argument can be used, adding when necessary the relevant multiplicative factors.

According to figure 11, the utilization in the lower classes is not the same in all of them. Class #4 sends more data, and then class #2, and last class #0. This is a consequence of the priority scheme (class #4 sends packets, and then if there is some time left class #2 does, and then class #0) and from the target rotation time which is the same for all three. Were it different, the share of each class would change.

Figure 13 shows the total and information utilizations as a function of the total offered load, this time for the whole network. When the load is below 79.6% (which is the utilization when the rotation time reaches its target value), both curves are like those of the previous situations. Then they become flat, but go on increasing (slowly). Eventually, they reach a limit when the load is 4 times the maximum data utilization. This can be easily explained from what has just been said about the access classes: the utilization is 79.6% when the lower classes become restricted, and its maximum value is 82.8%. Meanwhile, it goes up from the first value to the second, and slowly because their difference

is small compared to (4*82.8 - 79.6) = 251.6%. Similar calculations can be made for the total utilization.

IV.4.1.2. The mean rotation time

Figure 14 displays the mean rotation time versus the offered load. As for the utilization, it is divided into three areas: in the first one, the behavior is the usual one and the values are approximately the same than with a 10 ms TRT (Cf. section IV.1.2). The relative difference is less than 5% for loads up to 70% and increasing when the rotation time comes closer to 1.5 ms. It also comes from the variations created by the model itself, which is not absolutely accurate. Then the rotation time increases much more slowly from 1.5 ms to its maximum which is 2.32 ms as seen in the previous section.

IV.4.1.3. The mean waiting time

Figure 15 shows this variable versus the offered load. The usual three ranges of the load determine three parts of the curve. In the first one, the results differ of those for a 10 ms TRT by at most 0.5% for loads up to 60%, and by 3% at 70%. In the second zone, it grows slowly and in the third falls down suddenly and then finds a limit. The waiting time of a packet as it is computed by the simulation model is the time it spends queueing until it is sent on the medium. But if it is never sent, its waiting time is infinite and cannot be recorded. That is why, in the second area, the mean waiting time grows slowly: in the lower classes, it increases very much, but less and less packets are sent. The big fall happens when these classes send no more packet; then the waiting time is only the waiting time in class #6, and is around 7 ms, which is approximately 10/3 times the maximum rotation time, each class sending 3 packets per rotation and storing 10 (Cf. section IV.1.3).

Here, it may be useful to know the mean waiting time per access class, the global one having a limited meaning because of the big discrepancies from one to the other. Figures 16 and 17 give the corresponding curves. In the lower classes (fig. 16), the waiting times remain below 80 ms for loads up to 150%, but then increase very much, up to 3 to 5 seconds, and eventually fall down to 0 because no packet is sent. Figure 16bis focuses on the loads up to 200%. The differences between the three classes appear from an 80% offered load. In the highest access class, there is a first jump when the load is such that the rotation time equals the TRT (i.e. when the utilization stops growing linearly), and a second and much bigger one when the load offered by the class #6 of the stations reaches the maximum utilization of the network, as usual, because there things happen as if there was no lower access class.

IV.4.1.4. The queue lengths

Figure 18 shows the mean queue length when the right to transmit arrives in an access class of a station. The lower classes behave as usual (compare to figure 4), having a big increase around 80%, i.e. at the first critical point. Since the length is already of 10 packets for such loads, the second big increase in the waiting time cannot be seen on these curves. In the highest class, the behavior is almost classical, if it is measured as a function of the load offered by class #6 only: the big increase is when the load they give is around the maximum utilization of the network. The first critical point has a very slight influence.

IV.4.2. With a 3 ms target rotation time

IV.4.2.1. The utilizations

In this situation, there are differences from the previous one, but the same kind of arguments can be used to help understand what happens and to do some analytical predictions. Figure 19 is the equivalent of figure 11 and shows the utilization in the access classes versus the offered load. The main difference with the case where the TRT is 1.5 ms is that the limit final utilization for the lower classes is not zero. For low offered loads, the same remarks as before can be made (see section IV.4.1.1). When the rotation time becomes close to its target value, the lower classes cannot make it grow. But class 6 uses at this time, in each station, (TRT - token_time)/(4*4) because there are 4 stations and as many access classes per station. This value is here 0.178 ms, which is less than the authorized maximum of 0.542 ms. So, this class may go on sending frames and the total share of the others must decrease, which can be checked on figure 19. The load on which this happens can be computed by the same means than before, and the value is now 84%.

Studying the other end of the curves shows that the maximum time the upper classes plus the token passing may use is 2.32 ms, as before. The TRT being 3 ms, all the lower classes have at most 0.68 ms on each rotation, and so their share is 6.7%, counting only the data units of the frames. The simulation gives a maximum utilization of 6.6% per lower class. Then, all this yields the maximal utilization for the whole network, which is 3*6.7 + 64 = 84.1%, the value 64% being the maximum utilization of class 6, known from the TRT and their maximum token holding time. The simulation predicts a 64.3% value for this class, and an 84.05% value for the whole network. These values are reached when class 6 offers the maximum it may send, that is when the total offered load is 4 times this value, which is 256%.

Between 84% and 256%, the lower classes may at most use what is left by class 6, which is (84.1 - load/4)/3 for each of them. Thus the maximum utilization in the access classes is known, and figure 20 displays the results from both the simulation and the analytical model.

The comments about the differences in the utilizations of classes 4 to 0 are the same as in the previous section. However, one more thing should be added: why are the limit utilizations the same in the three lower classes, while the values differ for loads in the intermediate range? The reason why is that, when the transmissions of the highest classes are uniform (the same time used by every station on every rotation), then they must be so in the lower classes (if they have the same TRT). This can be seen by choosing a very simple situation with two stations, assuming each class 6 sends 1 frame and the TRT is 3 frames. There, the lower classes are one by one allowed to send 1 packet. On the other hand, if in the same situation it is assumed that from time to time the upper classes do not use their share of time, this creates a privilege for class 4, which may use the time left by class 6, and the situation remains the same for the last two. If class 4 does not use this advantage, class 2 may do so, and so on. In the average, class 4 has more opportunities than class 2, which has more than class 0.

Figure 21 shows the data and total utilizations of the whole network. Here, nothing peculiar is to be noticed, because, as has been seen above, the utilization reaches its maximum from loads equal to its value: the utilization for which the rotation time takes its target valueculations can be made for the total utilization.

IV.4.2.2. The mean rotation time

Figure 22 displays the mean rotation time versus the offered load in this configuration. It is once more possible to divide the curve in three areas corresponding to the three of figure 19 or 20, but the variation in the second area is very small; the curve looks like the one on figure 2 (for TRT = 10 ms). As it has been said above, the big variations in the utilization of the lower classes does not influence this variable very much.

For low loads, the value of the TRT changes almost nothing: the relative difference of the rotation times when the target value is 3 or 10 ms is at most 5% for loads up to 80%. And the difference with the results prdicted by the formula of section IV.1.2 is less than 6% for the same loads, mostly below 2%.

IV.4.2.3. The mean waiting time

Figure 23 shows its behavior as a function of the offered load. The same comment as for the rotation time can be made, except that the variations of the access class utilizations have bigger consequences. The values differ by at most 9% of those for a 10 ms TRT when the load is up to 70%. Figures 24 and 25 exhibit the waiting time in every access class. In all cases, there is a big increase when the offered load is around the maximum utilization, but much bigger in the lower classes, class 6 being yet allowed to send all it has to. As usual, the waiting time is the largest in class 0. It reaches a limit in the lower classes when class #6 offers its maximum utilization, approximately. The limit is the same in the three lower classes for the reasons given in section IV.4.2.1 about the utilizations.

IV.4.2.4. The queue lengths

Figure 26 shows the mean number of packets when the right to transmit enters an access class. It is not very different from figure 18 (for a TRT of 1.5 ms), although the length in class 6 is larger with TRT = 3 ms, which is normal, since the other queues send more packets.

IV.4.3. Back to a 10 ms target rotation time

This situation has been studied in section IV.1, but not extensively, for nothing was said about the utilization in each access class. Figure 27 shows this, providing the mean of the values in the three lower classes, because they are very close at any load. The behavior is rather different from what was observed before. All the utilizations grow linearly and suddenly, it falls down in class 6 (the so called "high priority class"), and then limits are reached.

Here, the point where the rotation time attains its maximum value can be computed, but this gives to class 6 more time than it is allowed. So, what happens first is that this class uses all the time it may. Indeed, the rotation time is at its target value when the utilization is equal to $1r*(TRT - token_time)/TRT$, according to section IV.4.1.2, which is here 87.2%. This would give to class 6 a time equal to $(TRT - token_time)/4 = 2.4$ ms , while the maximum time it may use is $4*3*0.1808 = 2.\overline{2}$ ms (4 stations, 3 frames per rotation, each lasting 180.8 microseconds). So, class 6 reaches its limit before the rotation time does. This happens when the upper class uses 2.2 ms; the other classes do the same (they are not yet restrained by the TRT), and so the data utilization of the whole network can be computed and is 87%, giving 21.7% for each class. At this point, the curves diverge, as can be seen on figure 27, because the same phenomenon as before appears, the limited

class being now the "highest". The utilization in class 6 falls down, while it can grow on in the others, all reaching a limit when the maximum throughput is sent on the cable and the number of packets transmitted by each class becomes constant from one rotation to the other. Let maxtime6 be the maximum time used by class 6; then, the limit utilizations are lr*(maxtime6/TRT) for this class, and lr*(TRT - token_time - maxtime6)/(3*TRT) for each of the others. These values are respectively 19.2% (20.2% if the maximum rotation time is given its true value of 9.5 ms) and 22.7% (or 22.3%). The values predicted by the simulation model are 20.2% and 22.3%.

IV.4.4. The critical values of the target rotation time

In the configuration that has been studied here, with the same arrival process in all the access classes and the same TRT everywhere, it is possible to give the values of this TRT for which the three different behaviors that have been observed exist.

The three areas are the following:

(1) TRT - token_time < maxtime6
(2) maxtime6 < TRT - token_time < 4*maxtime6
(3) TRT - token_time > 4*maxtime6

The first condition was met with the 1.5 ms target rotation time; in such a case, the lower classes are eventually prevented from sending any packet and the access class utilizations look like on figures 11 and 12.

Condition (2) was true when the TRT was 3 ms; in such a situation, the lower classes have a non-zero limit utilization, which is smaller than the one for class 6 (this is exactly what formula (2) asserts).

Condition (3) was true with a 10 ms TRT; with such values, the limit data utilization of the lower classes is greater than the limit for class #6. Of course, it is equal when inequality (3) is in fact an equality; but the limit utilization of class 6 can be made as small as desired by setting the target rotation time to a large enough value. But it cannot be made null.

Figures 28 to 30 show the kind of behavior the global data utilization, the mean rotation time and the mean waiting time may have in these three situations.

IV.5. Sensitivity to the number of stations

All the results described here come from the set of experiments number 4, depicted in section III. Below is a study of the influence of the number of stations on the same variables as above. All the other parameters of the network have the same values as in the first set of experiments (commented in section IV.1), except the target rotation times in the lower access classes, which have been made proportional to the number of stations, because the time used by both the token passing and the highest access class are clearly so, and it has been seen above that what is important for many variables is the ratio between these parameters.

IV.5.1. Utilization versus number of stations

The main result is that the utilizations for the whole network are independent of the number of stations, as a good approximation.

Indeed, the relative difference between the information utilization for 4 stations and for 100 stations is less than 5% at all loads up to 200% (since the first set of experiments showed that in this kind of configuration, nothing changes noticeably for loads between 200% and 500%, the offered load is limited to 200% here). For the total utilization, the relative difference is less than 1%.

The values for the maximum data utilizations are 87.10%, 87.15%, 87.16%, and 87.17%, for 4, 20, 40, and 100 stations respectively. For the maximum total utilization, the results are 97.42% for 4 stations, and 97.46% for the other situations.

So, the formula for the maximum utilization given in section IV.1.1 receives a further confirmation :

maxutil = lr*((maxrt - token_time)/maxrt)

is approximately true for all the numbers of stations, because the time to pass the token is in this model 37.2*N microseconds, where N is the number of stations, and maxrt is close to the target rotation time, which has been chosen proportional to N. So, maxutil is independent of N. At loads below this value, the utilization is simply the offered load, with no noticeable loss. The same remarks can be made about the formula for the total utilization given in section IV.3.1.1.

About the access class utilizations:

the study that was made in section IV.4.3 can be made here as well, and all the results should be independent of the number of stations. Figure 31 shows that this is not exactly true: it displays the maximum utilization in the highest access class and its mean in the three others as a function of the number of stations. The value for class #6 decreases (slightly) when there are more stations. Since the utilization for the

whole network remains constant, its value in the lower classes increases. Section IV.4.3 presented the value lr*(maxtime6/TRT) for the maximum utilization in class 6. In fact, maxrt (the maximum rotation time) should be written instead of TRT (the target rotation time). And the next section will show that the maximum rotation time is not exactly proportional to the number of stations. But the relative difference between the values for 4 and 100 stations is only 5%.

IV.5.2. Rotation time versus number of stations

Figure 32 shows the variations of the mean rotation time when the number of stations in the logical ring increases.

The main result is that the rotation time is proportional to the number of stations.

Figure 32 shows this is true for offered loads equal to 20%, 40%, 60%, and 80% of the capacity of the channel. The analytical prediction of section IV.1.2 for the rotation time fits the results of the simulation for all the numbers of stations. More precisely, the relative difference between both is less than 4% for loads up to 80% with 20 stations, less than 3% for loads up to 70% (and 6% for 80%) with 40 stations, less than 3% for loads up to 80% with 100 stations. In most cases, the values given by the analytical prediction are a little below those provided by the simulation.

The maximum values are 9.5 ms, 49.5 ms, 99.4 ms, and 249.5 ms respectively. The ratios between these values and the target rotation times are 0.95, 0.99, 0.99, and 1.00. So, the proportionality is not absolutely true at all loads, and this explains the small variations noticed in the maximum utilization of class 6 for different numbers of stations. Using the maximum rotation times instead of the target rotation times yields maximum utilizations in class #6 of 20.2%, 19.4%, 19.4%, 19.2%, while the values from the simulation are 20.2%, 19.4%, 19.3%, and 19.2%, respectively.

About the coefficient of variation of the rotation time:

figure 33 shows the coefficient of variation of the rotation time as a function of the offered load for the four numbers of stations of this set of experiments. This variable increases slowly for loads below the maximum utilization, then falls down suddenly, and then decreases very slowly. It is not surprising to see the largest values for offered loads around the maximum utilization, because this is a critical region for all the variables, as the previous sections alrady showed. The small values beyond this critical point reflect the big regularity of what happens in a rotation once the network is overloaded, although there can be larger differences from one station to another within the same rotation.

At all loads, the coefficient of variation of the rotation time is a decreasing function of the number of stations. Very approximately, it

looks to be proportional to the reciprocal of the square root of the number of stations. But, depending on the offered load, the fit is not very good. It has been possible to find this kind of approximation with a relative error below 5% for loads up to 80%, but not beyond (not enough precision in the measurement of the coefficient of variation, or not enough data points?).

In fact, the rotation time (minus token_time) is the sum of all the times spent by the right to transmit in all the access classes of all the stations. If these (random) variables were independent and identically distributed, their coefficient of variation would be proportional to 1/sqrt(N) . But the independence is not true because of the target rotation time mechanism: if a class uses much time, the next may not be allowed to send packets, for instance. This is why the fit is not very good.

The larger coefficient of variation for 4 stations may be related to the fact that the maximum rotation time is only 95% of its target value in this configuration: if there are noticeable differences from one rotation to another, the mean value can be smaller than the maximum, which is the target rotation time. See section IV.1.2 for the discussion. When there are more stations, at high loads, all the rotations have almost the same duration, which is very close to the TRT because the network is overloaded.

IV.5.3 Waiting time versus number of stations

Figures 34 and 35 display the behavior of the mean waiting time of the data packets in the queues of the IFM as a function of the number of stations, for various offered loads.

The waiting time of the data packets is approximately a linear function of the number of stations, at all offered loads.

Figure 36 shows the coefficient of variation of the waiting time as a function of the offered load for the four numbers of stations considered in this set of experiments.

The behavior is not the same as for the rotation time. The coefficient of variation increases slowly for loads up to 80%, then has a peak around a 90% load,and then decreases slowly. This is not exactly true in the situation with 4 stations, for which there is no peak but only a strong fall, which happens sooner. It can also be noticed that for 100 stations, the value is much bigger for a 200% load than for a 150% load, which is not true with the other numbers of stations. Section IV.1.3 showed that the coefficient of variation increases for loads above 200%, and eventually reaches 1 (with 4 stations). This is probably true for any number of stations, although the starting point of this increase can vary.

For loads up to 100%, the coefficient of variation of the waiting time is a decreasing function of the number of stations in the logical ring.

For higher loads, no clear behavior can be observed. It has not been possible to find a simple function giving a good approximation of this variable as a function of the number of stations. But here, the situation is more complex than for the rotation time: the waiting time of a data packet cannot be easily split in a sum of terms with identical and independent distributions, and so what was said for the rotation time no longer holds here.

IV.5.4. Queue length versus number of stations

The mean queue lengths do not differ much when the number of stations in the logical ring changes.

For most loads, it is a decreasing function of the number of stations: in the highest access class, the relative difference between the values with 4 stations and 100 stations varies from -5% to 2%, being mostly negative, but with a -26% peak when the offered load is 90% of the capacity of the channel (i.e. in the critical region). In the three lower classes, the values are almost the same from one to the other (Cf. section IV.1.4), and have the same behavior as for the upper access class as functions of the number of stations: the relative difference from 4 stations to 100 stations is mostly between -7% and 0%, with a 35% peak for a 90% offered load.

So, the queue length is a slightly decreasing function of the number of stations.

V) CONCLUSION

V.1. Summary

A simulation model of token bus networks has been designed to be able to predict their perfomances. It has been used to study the influence of the offered load, the packet length, the target rotation times, and the number of stations on the utilization of the network, the rotation time, the waiting time of data packets, and the queue lengths, in some configurations.

This work has shown the following features of token bus networks:

1. It is possible to predict the total and information utilizations as functions of the offered load.

2. It is possible to predict the rotation time as a function of the offered load when the load is less than the maximum data utilization.

3. The mean waiting time of data packets and the mean queue lengths are slowly increasing functions of the load when it remains below the maximum data utilization.

4. The maximum data utilization is an increasing function of the packet length.

5. The total utilization, the rotation time, and the waiting time are decreasing functions of the packet length for loads below the maximum data utilization (the variations remain small); but it is the contrary for their limit values.

6. The measured variables are approximately not sensitive to the packet length distribution.

7. The target rotation times are key parameters of the network: the maximum utilization in each access class can be predicted; when the target rotation tiime is the same in all the lower access classes of all the stations, three domains of its value can be identified, determining the maximum utilization in each class, which can be zero for the lower classes, or larger in the lower classes than in the upper.

8. At low loads, the measured variables are not sensitive to the target rotation times.

9. The total and data utilizations, and the queue lengths are approximately not sensitive to the number of stations.

10. The rotation time and the waiting time are proportional to the number of stations, at all offered loads.

V.2. Areas of future study

1. Test the model by comparisons with the measures of a laboratory on actual products, to validate it better.

2. Use the model in new configurations such as non homogeneous distributions of the packet arrivals and of the target rotation times, or arrival processes closer to what exists in the "real world".

3. Extend the number of configurations covered by the model, particularly by allowing stations to enter and leave the logical ring during the simulation run, to analyze the stability of the network (does it quickly reach its new state of equilibrium when the number of stations changes?), and evaluate the amount of time consumed by the protocol processes. Also, simulate connections between peer Logical Link Control sublayers, with response packets.

4. Simulate the connection of this model with higher level protocols or other kinds of local area networks.

5. Enhance the man/machine interface, especially the presentation of the results and the capabilities to perform statistical analysis.

This study provides a first understanding of the behavior of token bus networks, but much remains to do to really evaluate all the characteristics of this access method.

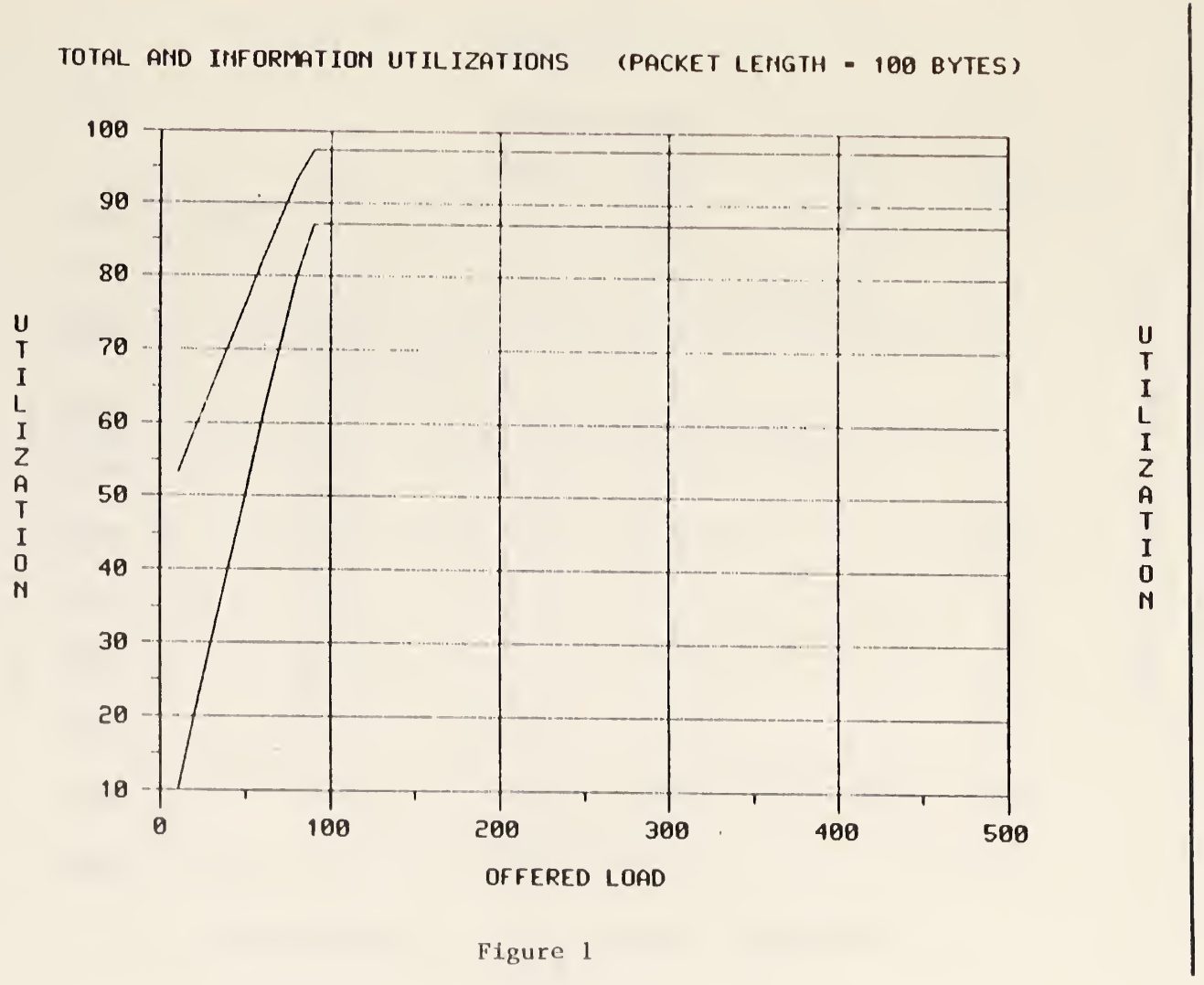

TOTAL AND INFORMATION UTILIZATIONS (PACKET LENGTH = 100 BYTES)
UTILIZATION
100
90
80
70
60
50
40
30
20
10
0
100
200
300
400
500
OFFERED LOAD
UTILIZATION

Figure 1

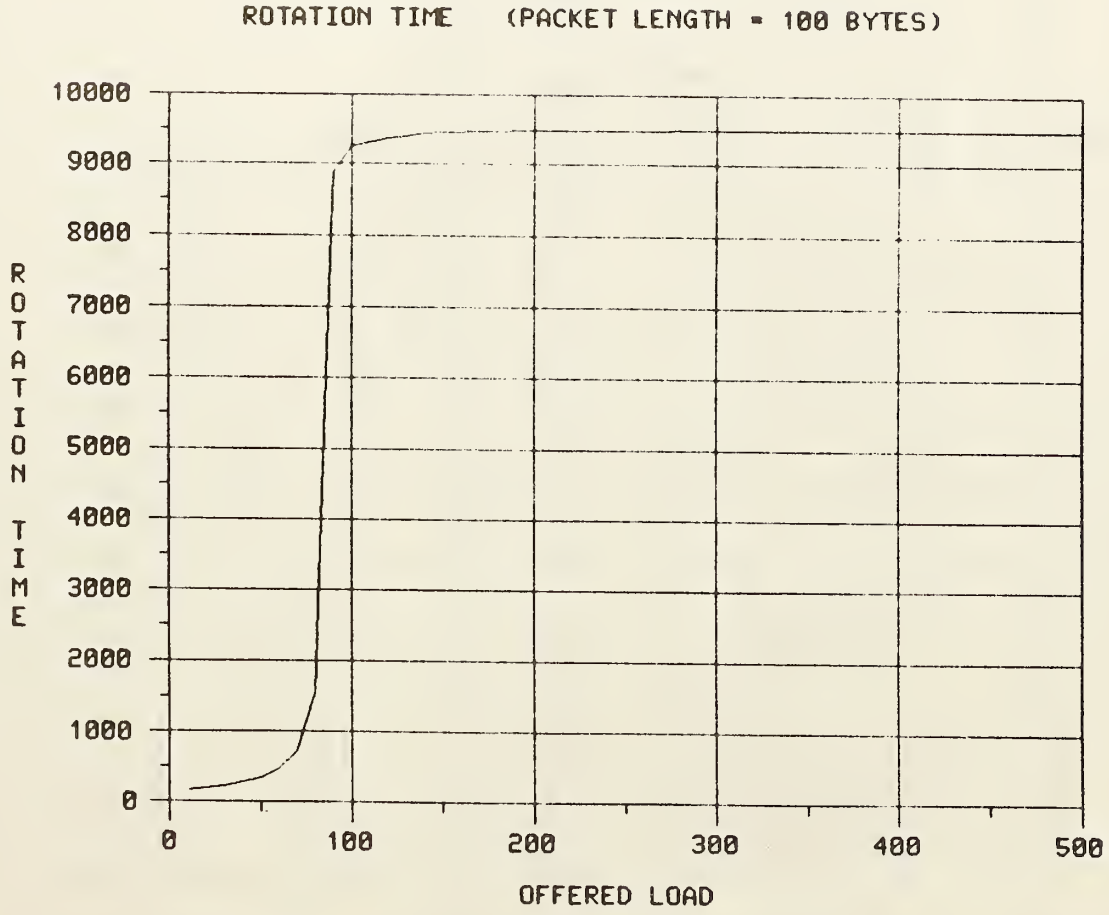


Figure 2

ROTATION TIME

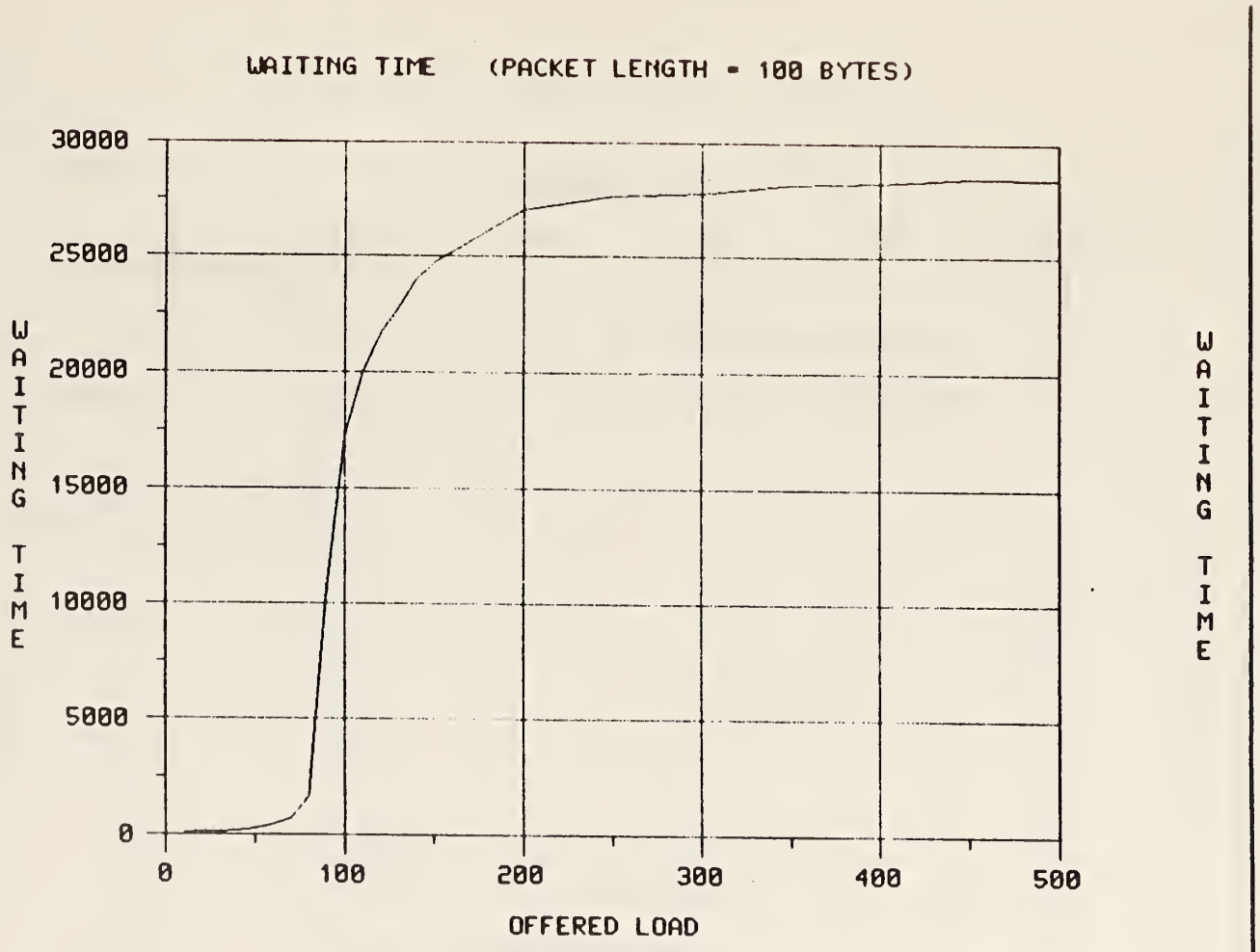

WAITING TIME (PACKET LENGTH = 100 BYTES)
30000
25000
20000
15000
10000
5000
0
WAITING TIME
0
100
200
300
400
500
OFFERED LOAD
WAITING TIME

Figure 3

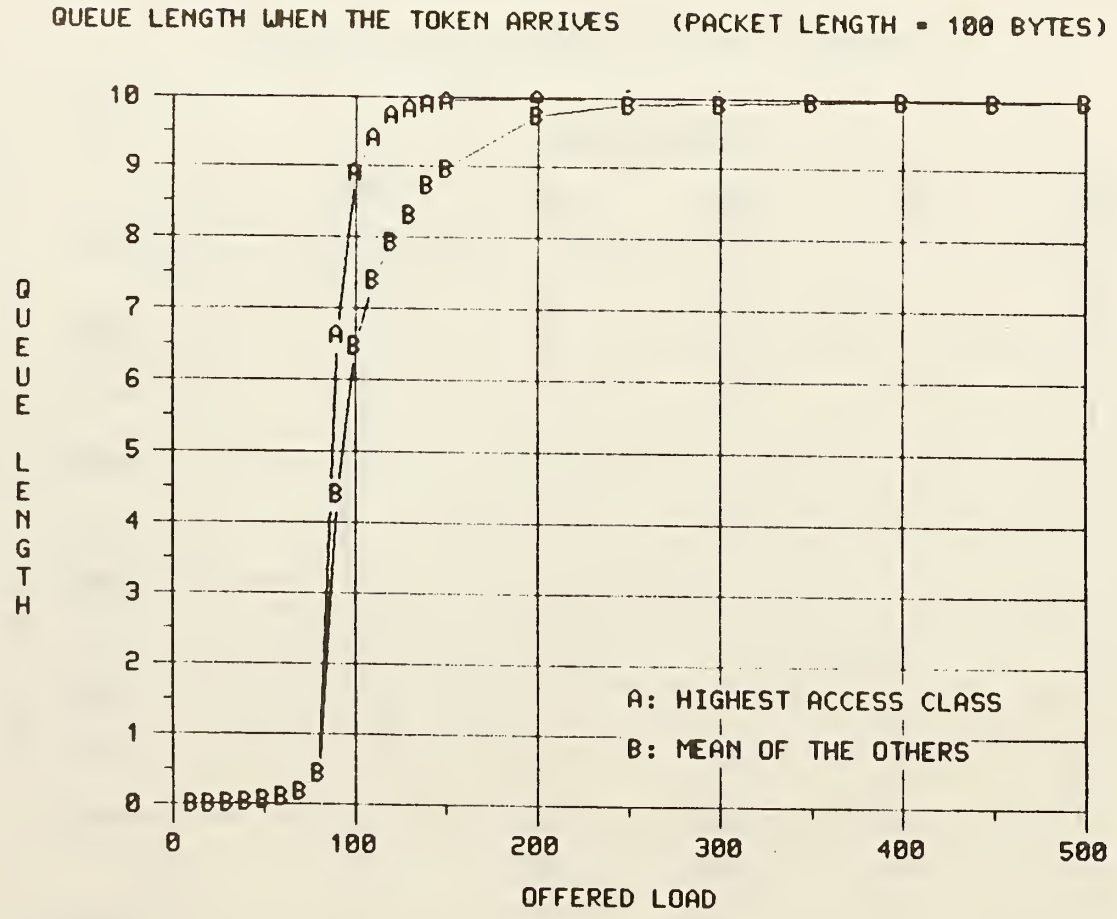

QUEUE LENGTH WHEN THE TOKEN ARRIVES (PACKET LENGTH = 100 BYTES)
QUEUE LENGTH
10
9
8
7
6
5
4
3
2
1
0
0
100
200
300
400
500
OFFERED LOAD
A: HIGHEST ACCESS CLASS
B: MEAN OF THE OTHERS

Figure 4

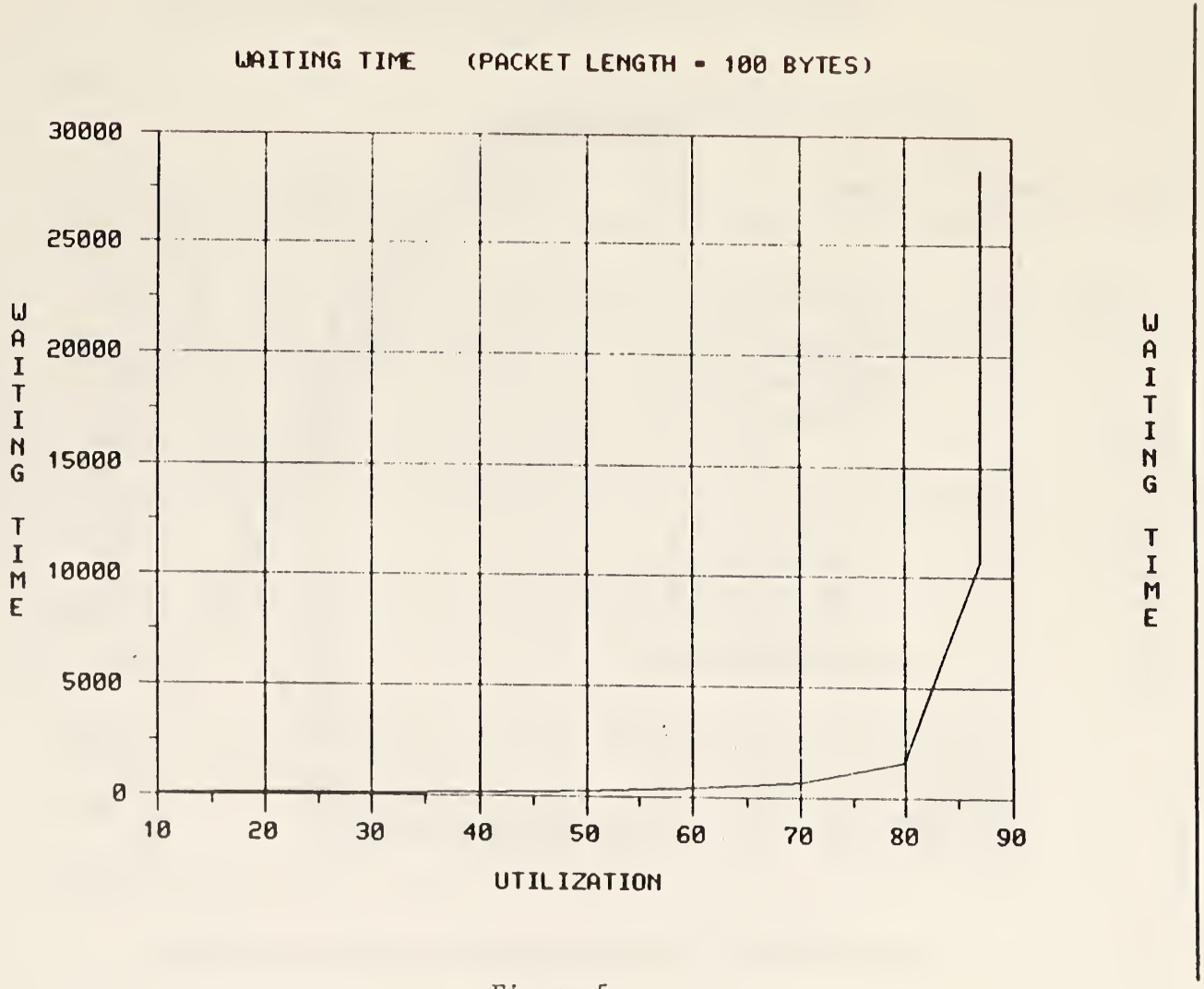

WAITING TIME (PACKET LENGTH = 100 BYTES)
30000
25000
20000
15000
10000
5000
0
WAITING TIME
10
20
30
40
50
60
70
80
90
UTILIZATION
WAITING TIME

Figure 5

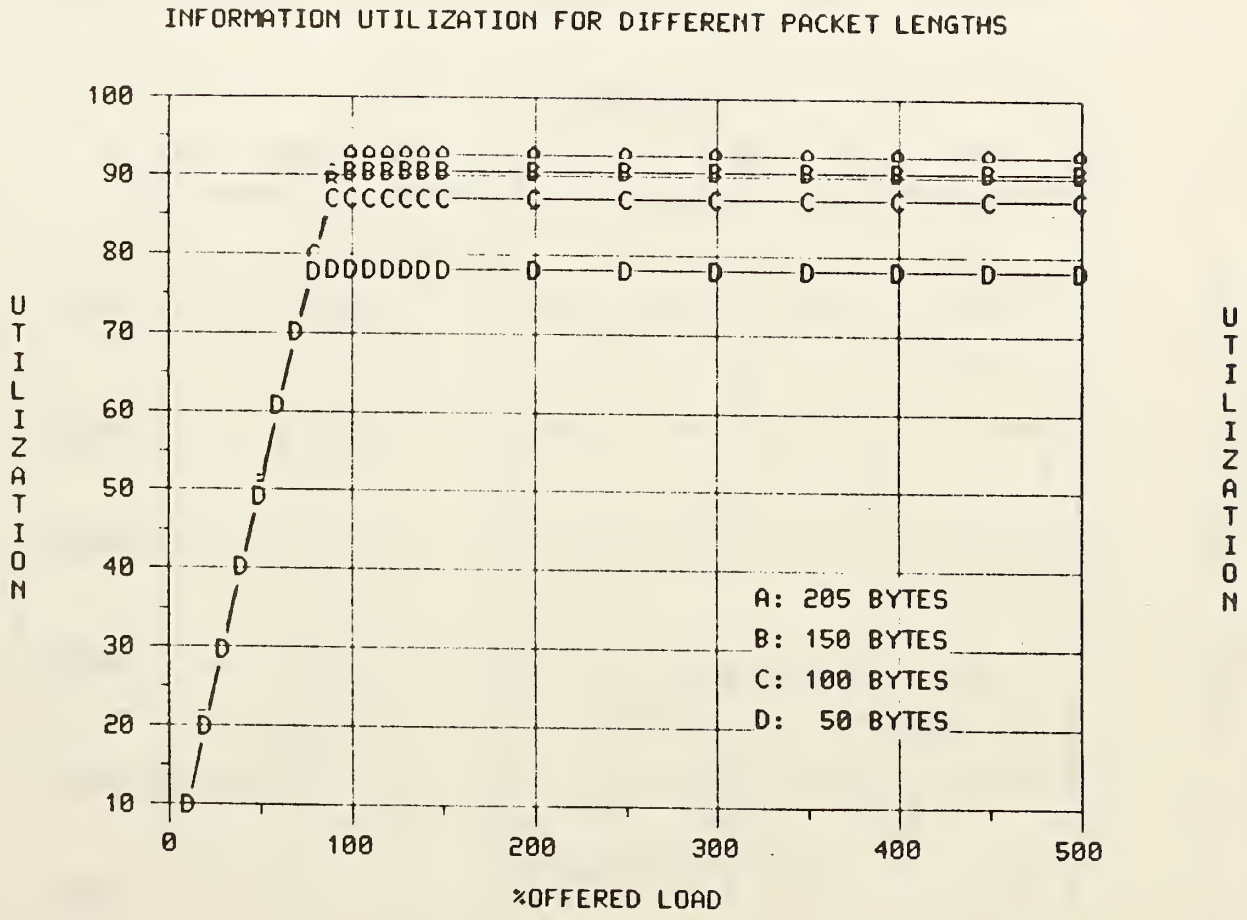

INFORMATION UTILIZATION FOR DIFFERENT PACKET LENGTHS
UTILIZATION
100
90
80
70
60
50
40
30
20
10
0
100
200
300
400
500
%OFFERED LOAD
A: 205 BYTES
B: 150 BYTES
C: 100 BYTES
D: 50 BYTES

Figure 6

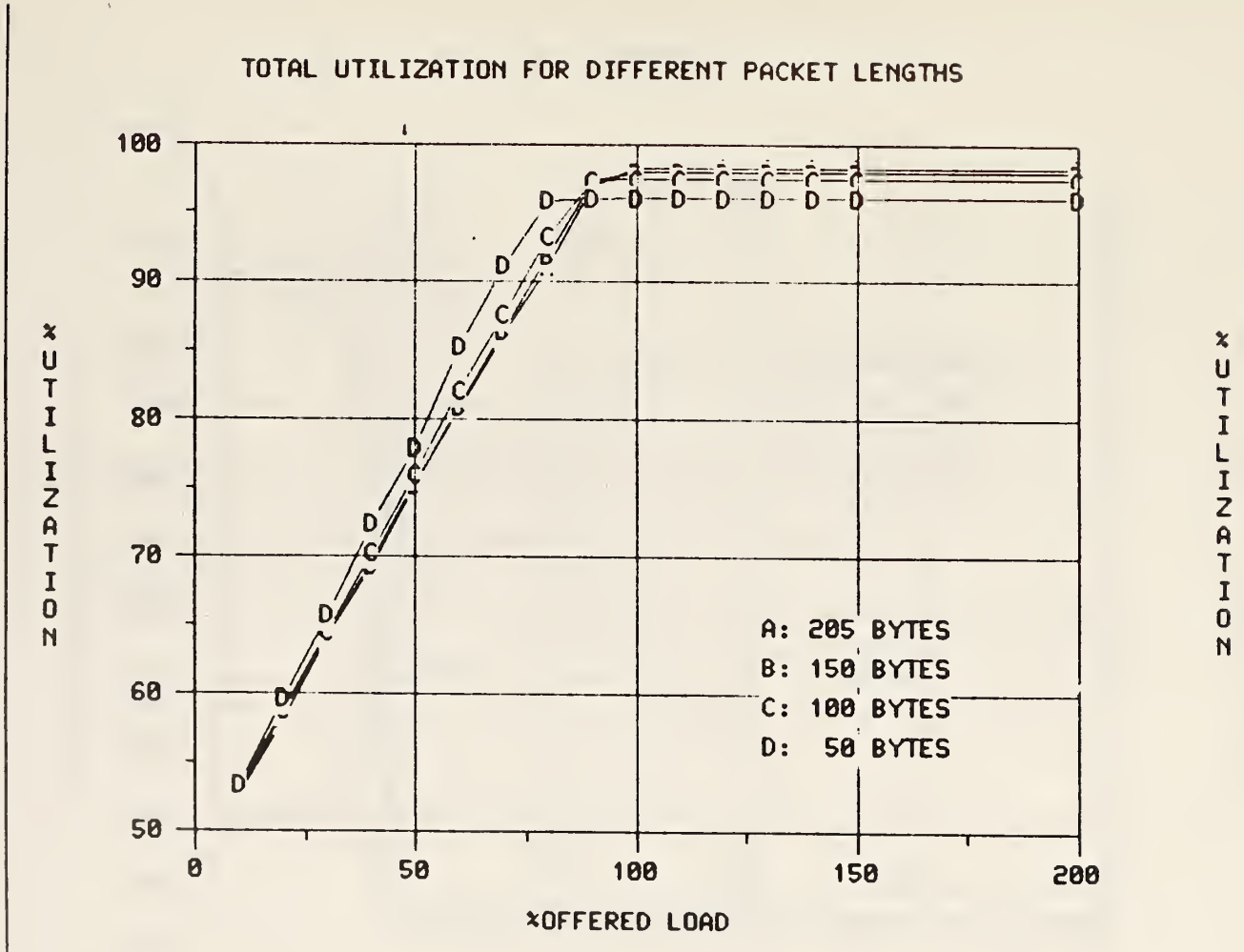

TOTAL UTILIZATION FOR DIFFERENT PACKET LENGTHS
%UTILIZATION
100
90
80
70
60
50
0
50
100
150
200
%OFFERED LOAD
A: 205 BYTES
B: 150 BYTES
C: 100 BYTES
D: 50 BYTES
%UTILIZATION

Figure 7

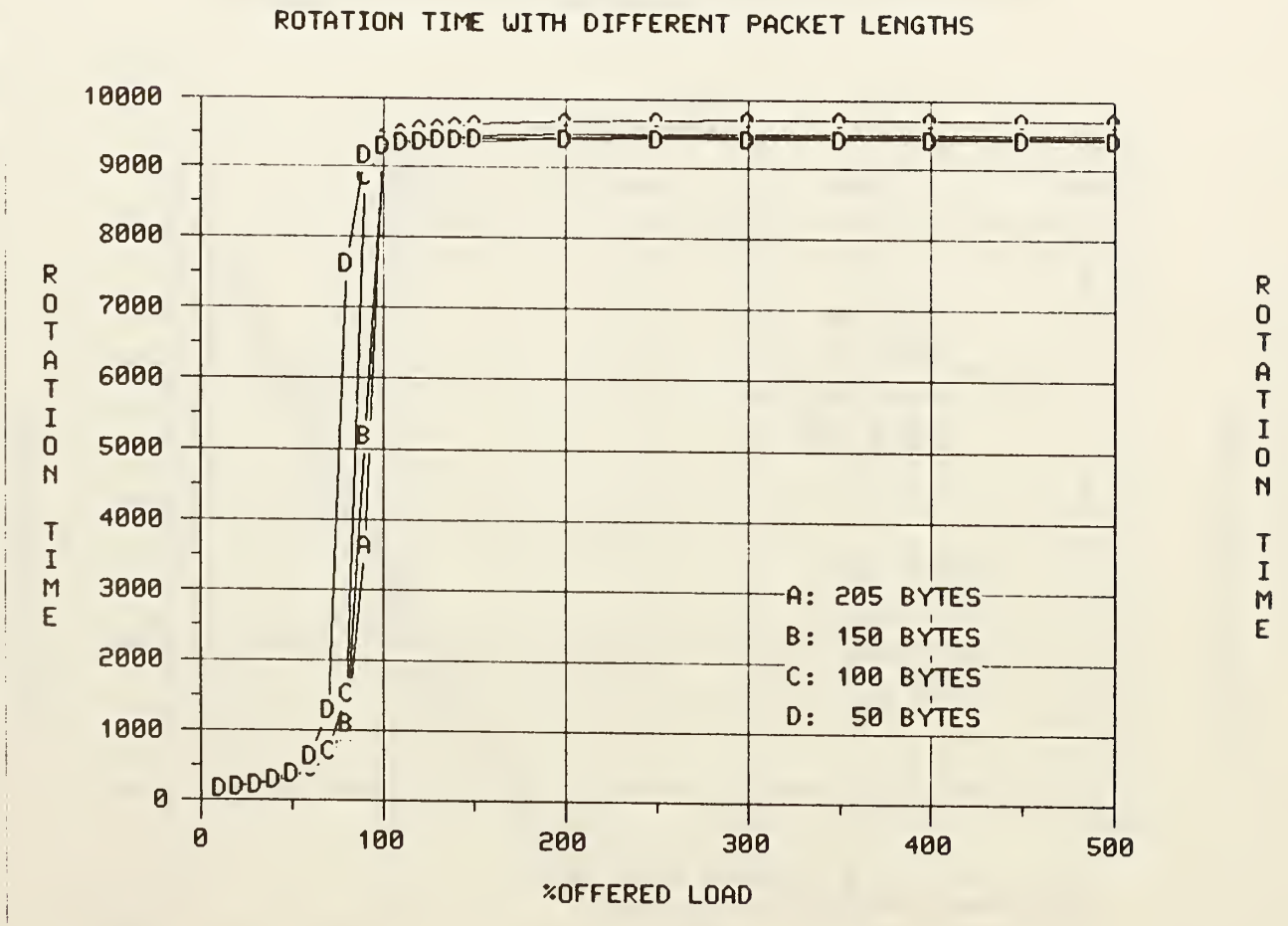

ROTATION TIME WITH DIFFERENT PACKET LENGTHS
ROTATION TIME
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0
100
200
300
400
500
%OFFERED LOAD
A: 205 BYTES
B: 150 BYTES
C: 100 BYTES
D: 50 BYTES

Figure 8

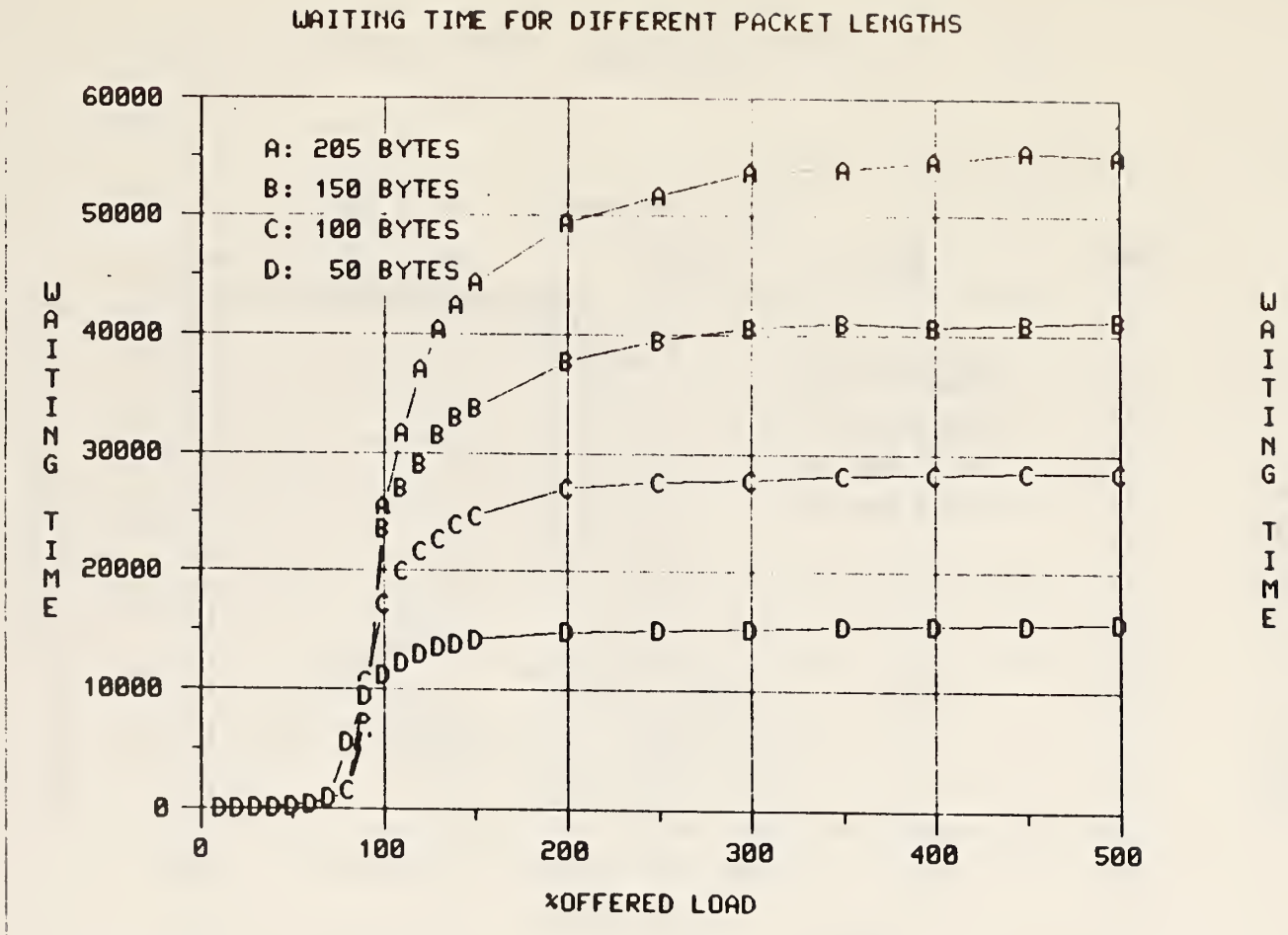

WAITING TIME FOR DIFFERENT PACKET LENGTHS
A: 205 BYTES
B: 150 BYTES
C: 100 BYTES
D: 50 BYTES
WAITING TIME
60000
50000
40000
30000
20000
10000
0
0
100
200
300
400
500
%OFFERED LOAD
WAITING TIME

Figure 9

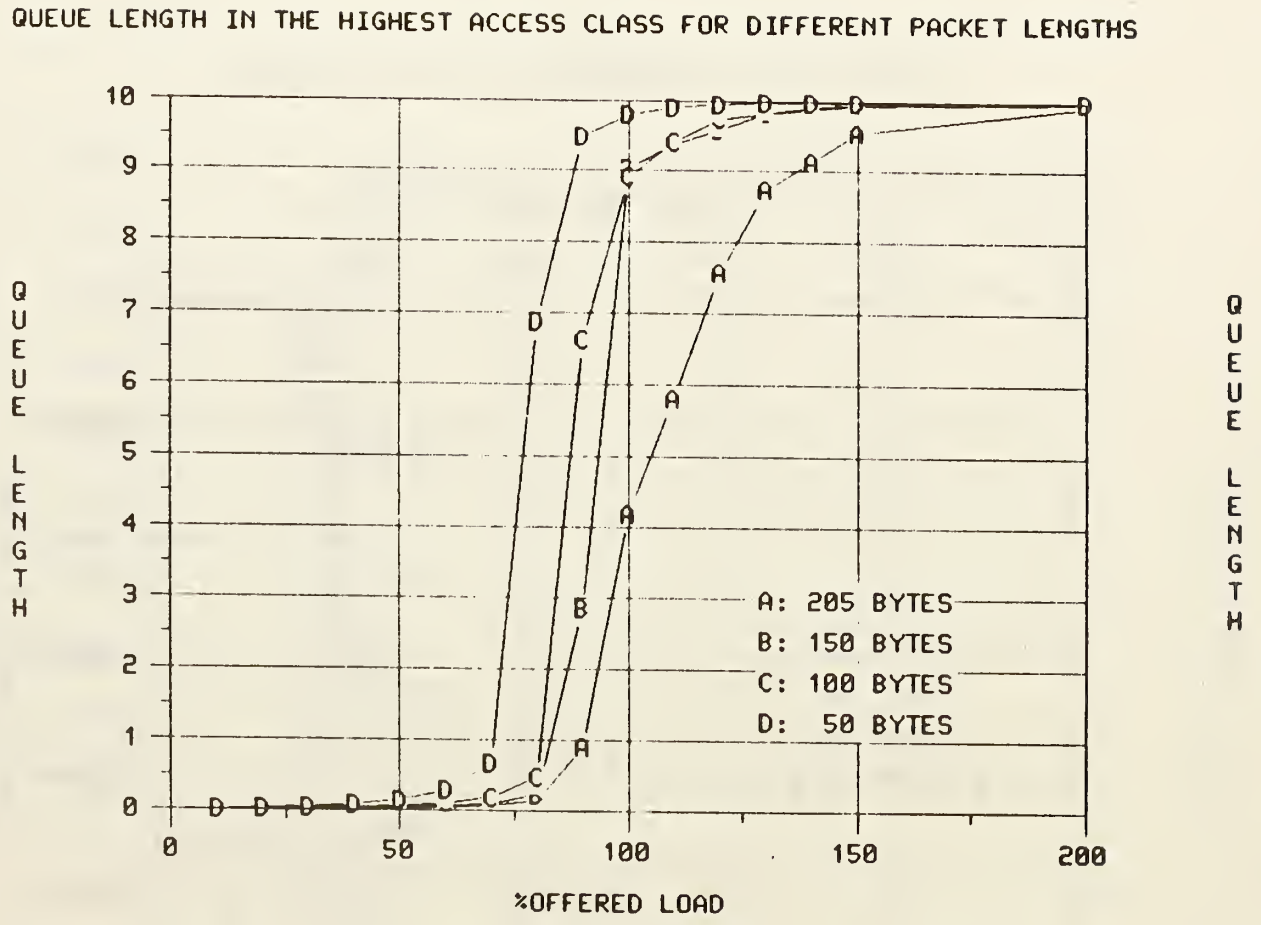

QUEUE LENGTH IN THE HIGHEST ACCESS CLASS FOR DIFFERENT PACKET LENGTHS
QUEUE LENGTH
QUEUE LENGTH
10
9
8
7
6
5
4
3
2
1
0
0
50
100
150
200
%OFFERED LOAD
A: 205 BYTES
B: 150 BYTES
C: 100 BYTES
D: 50 BYTES

Figure 10

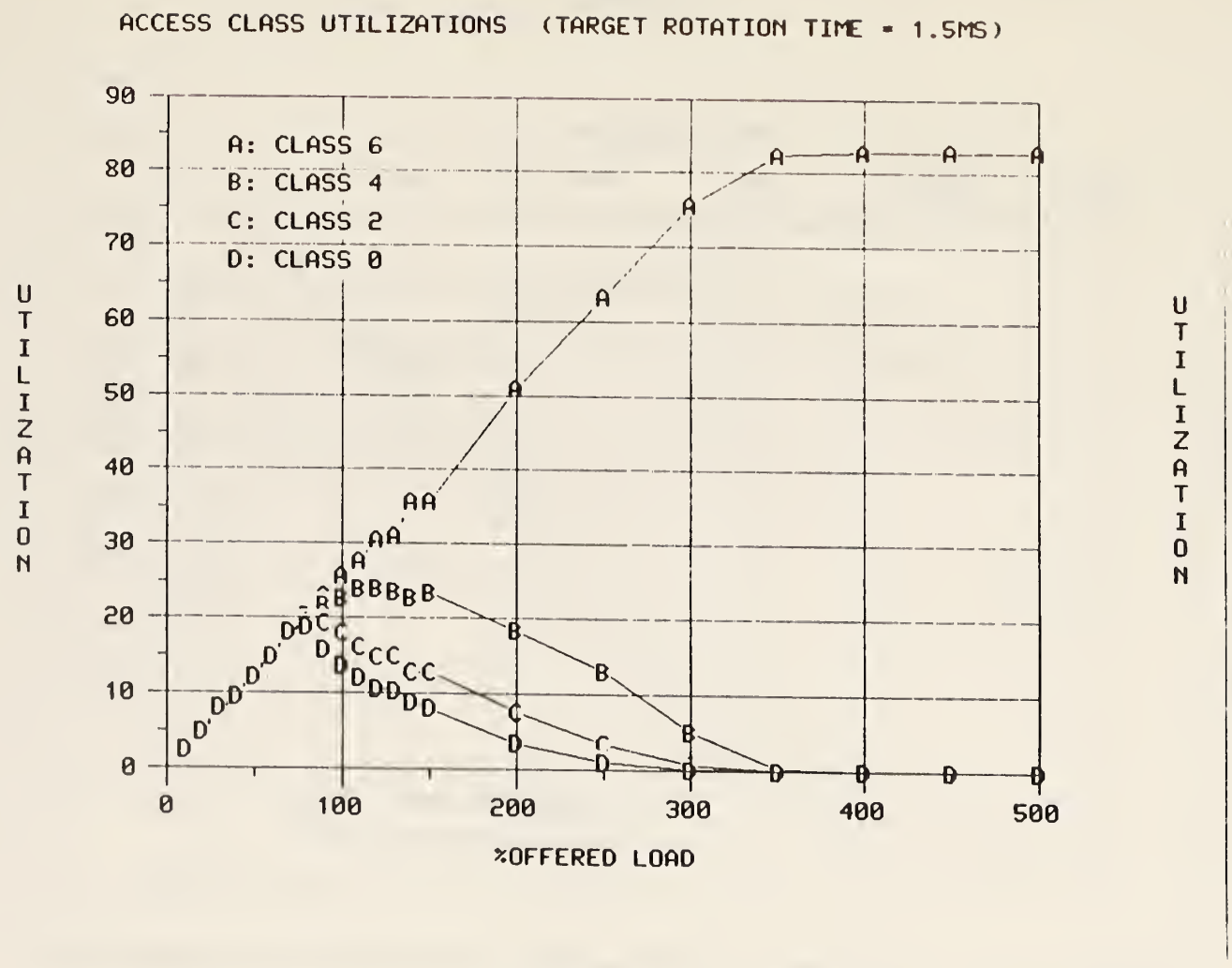

ACCESS CLASS UTILIZATIONS (TARGET ROTATION TIME = 1.5MS)
A: CLASS 6
B: CLASS 4
C: CLASS 2
D: CLASS 0
UTILIZATION
UTILIZATION
90
80
70
60
50
40
30
20
10
0
0
100
200
300
400
500
%OFFERED LOAD

Figure 11

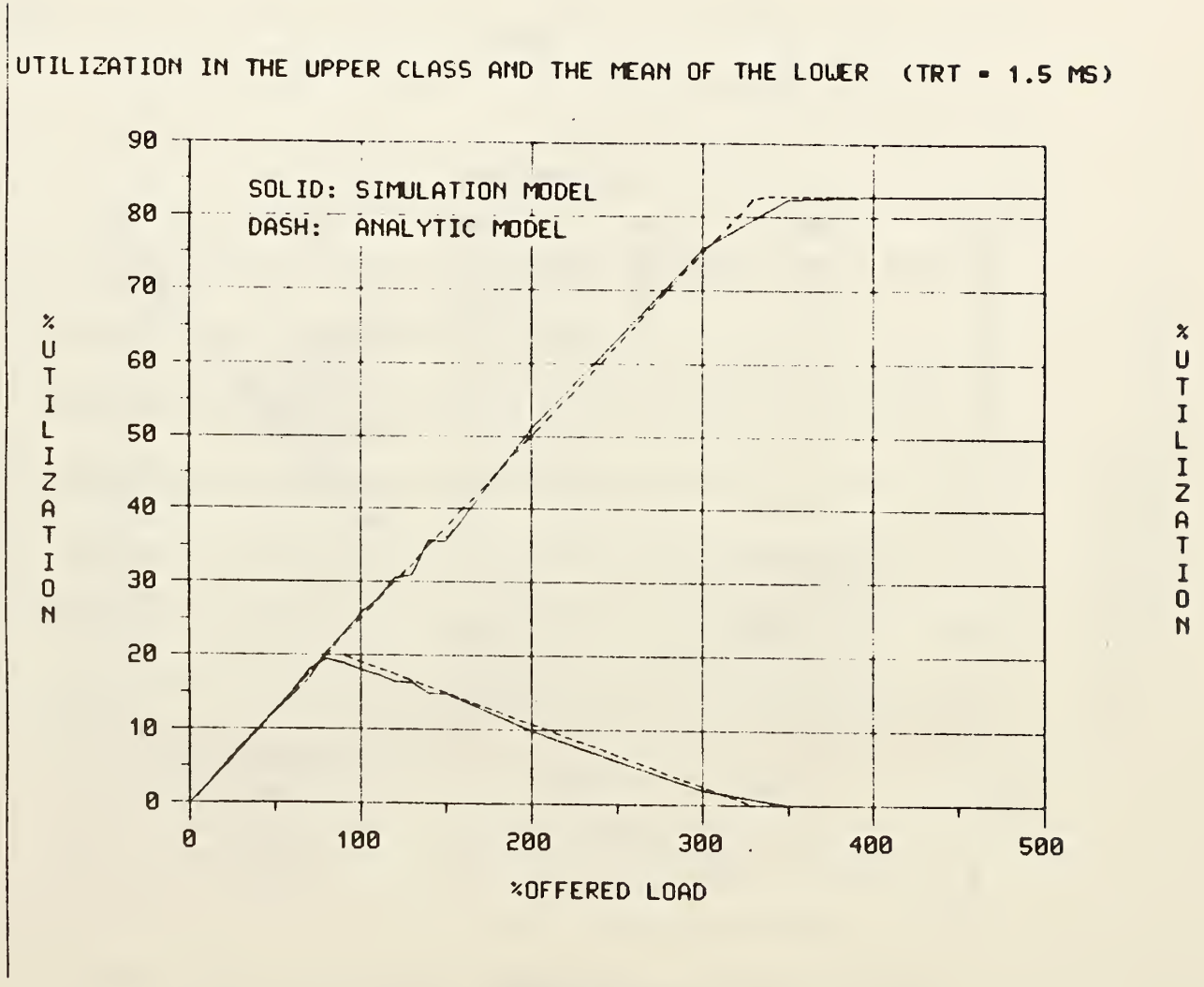

UTILIZATION IN THE UPPER CLASS AND THE MEAN OF THE LOWER (TRT = 1.5 MS)
SOLID: SIMULATION MODEL
DASH: ANALYTIC MODEL
%UTILIZATION
%UTILIZATION
90
80
70
60
50
40
30
20
10
0
0
100
200
300
400
500
%OFFERED LOAD

Figure 12

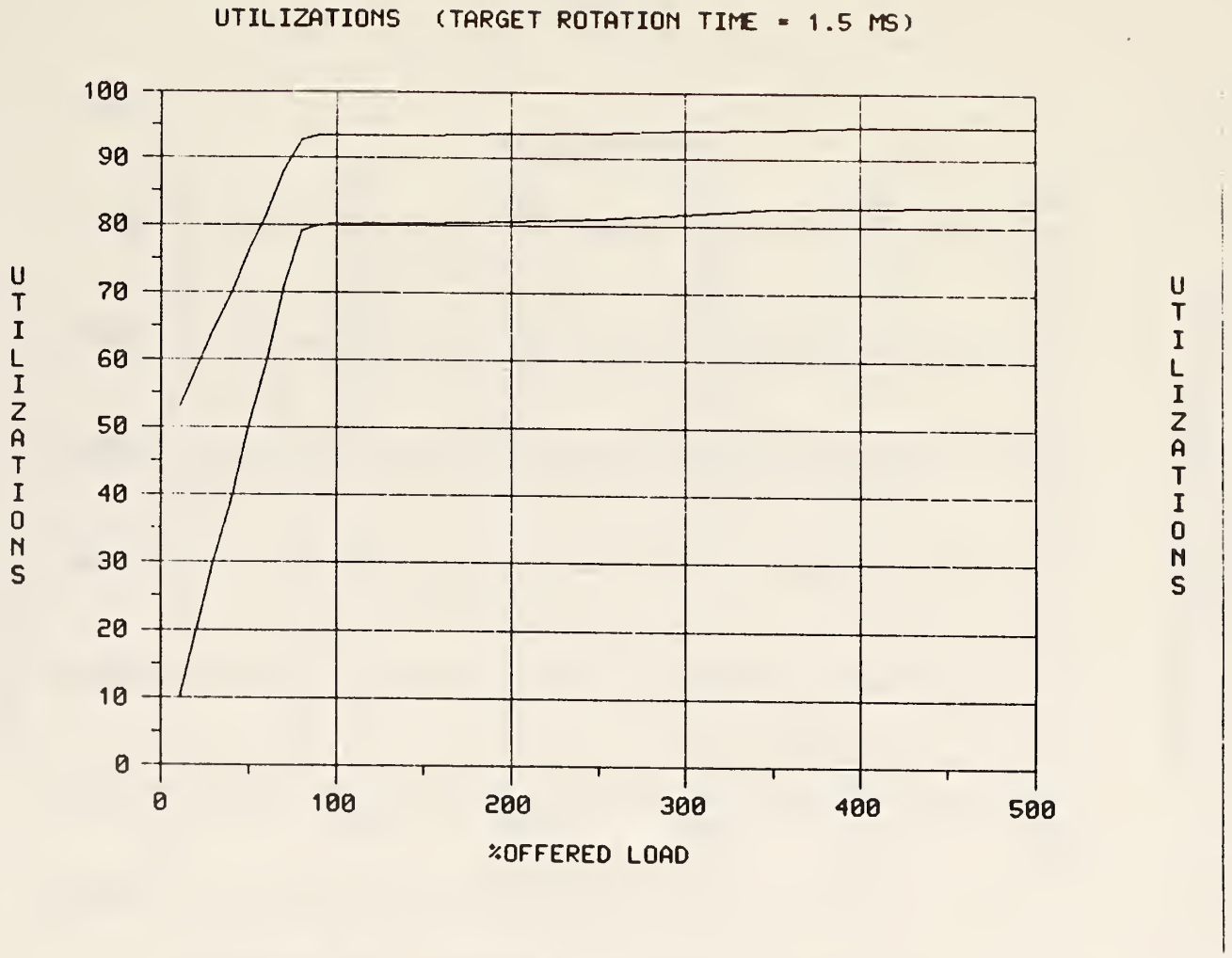

UTILIZATIONS (TARGET ROTATION TIME = 1.5 MS)
UTILIZATIONS
100
90
80
70
60
50
40
30
20
10
0
0
100
200
300
400
500
%OFFERED LOAD
UTILIZATIONS

Figure 13

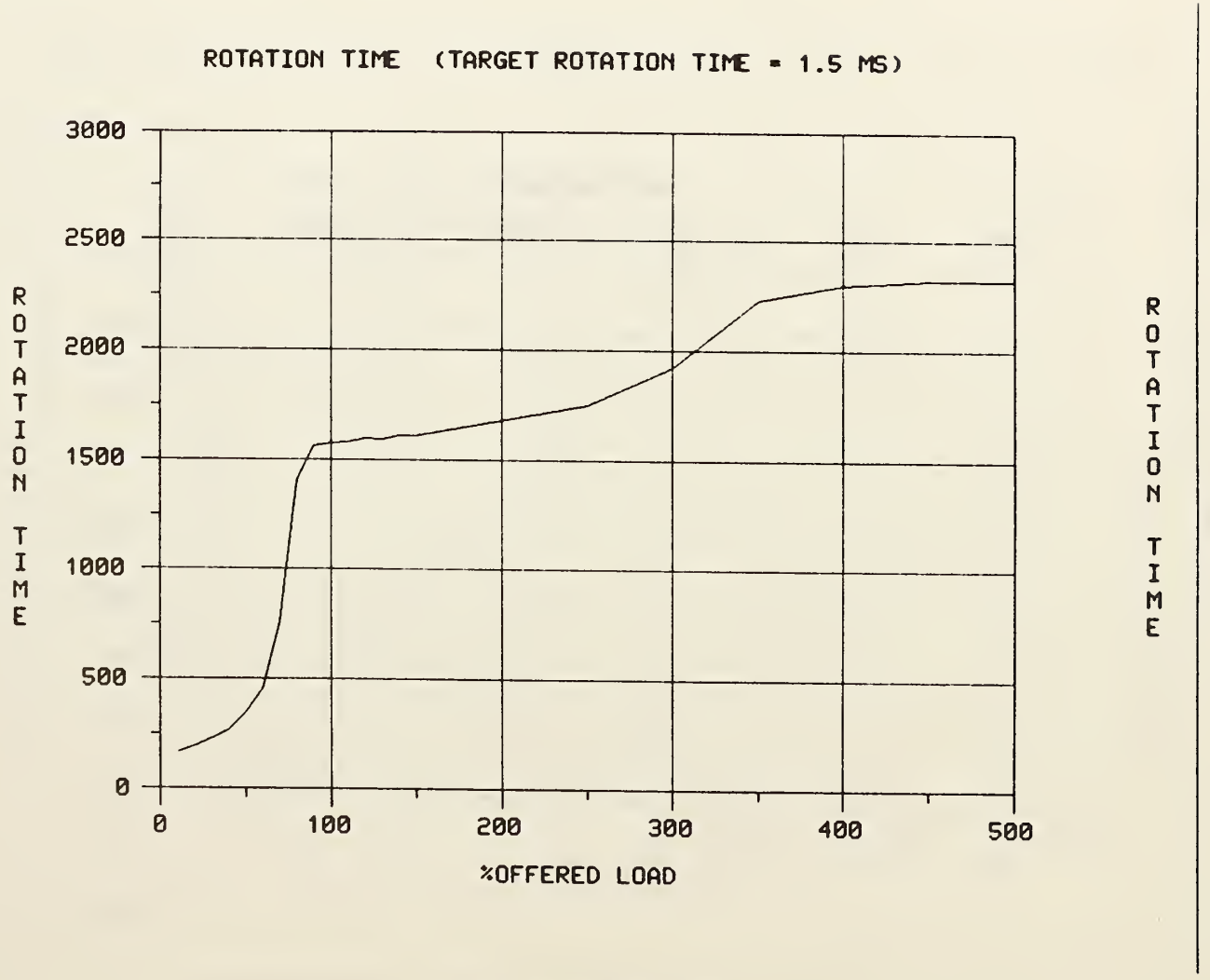

ROTATION TIME (TARGET ROTATION TIME = 1.5 MS)
ROTATION TIME
3000
2500
2000
1500
1000
500
0
0
100
200
300
400
500
%OFFERED LOAD
ROTATION TIME

Figure 14

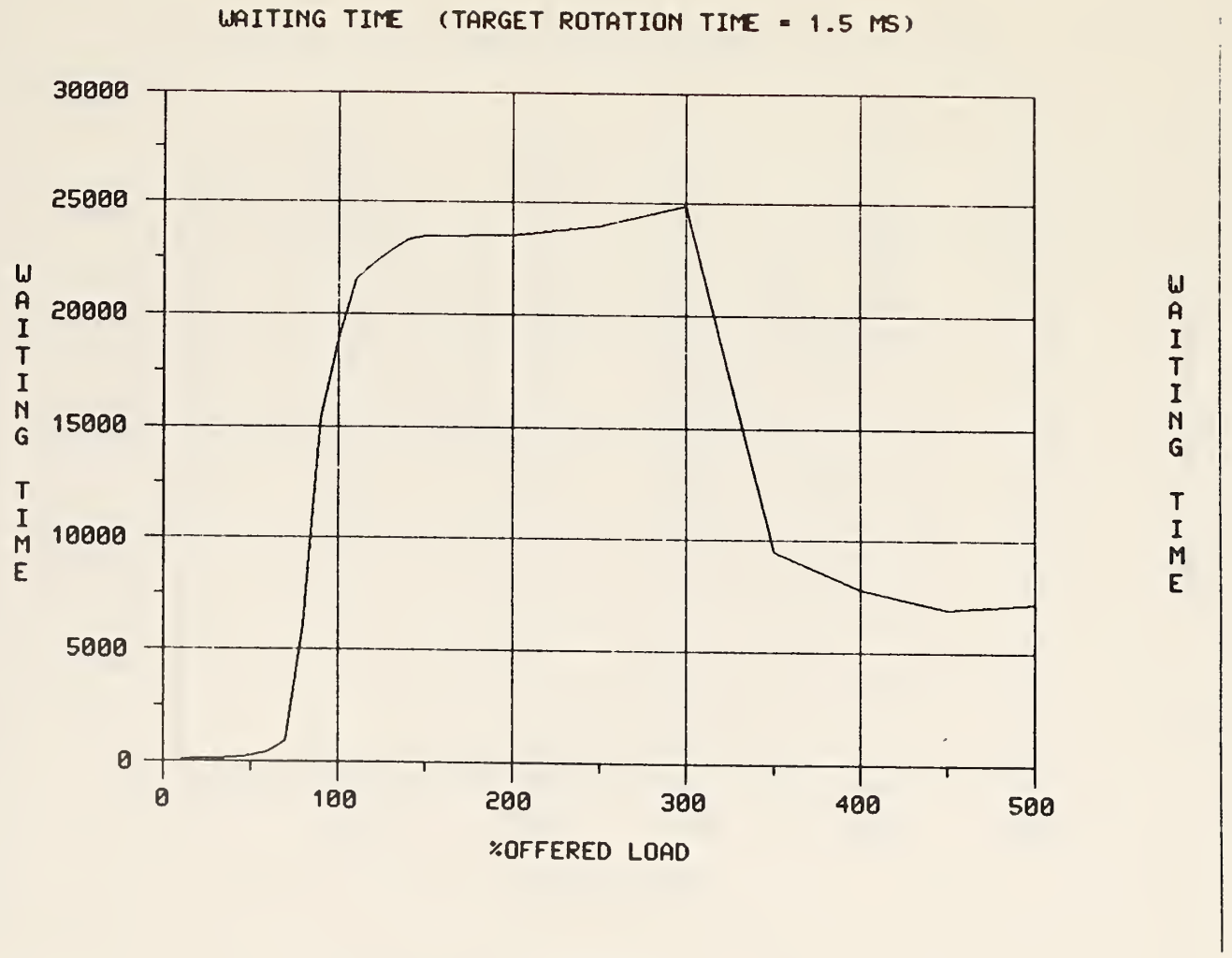

WAITING TIME (TARGET ROTATION TIME = 1.5 MS)
30000
25000
20000
15000
10000
5000
0
WAITING TIME
0
100
200
300
400
500
%OFFERED LOAD
WAITING TIME

Figure 15

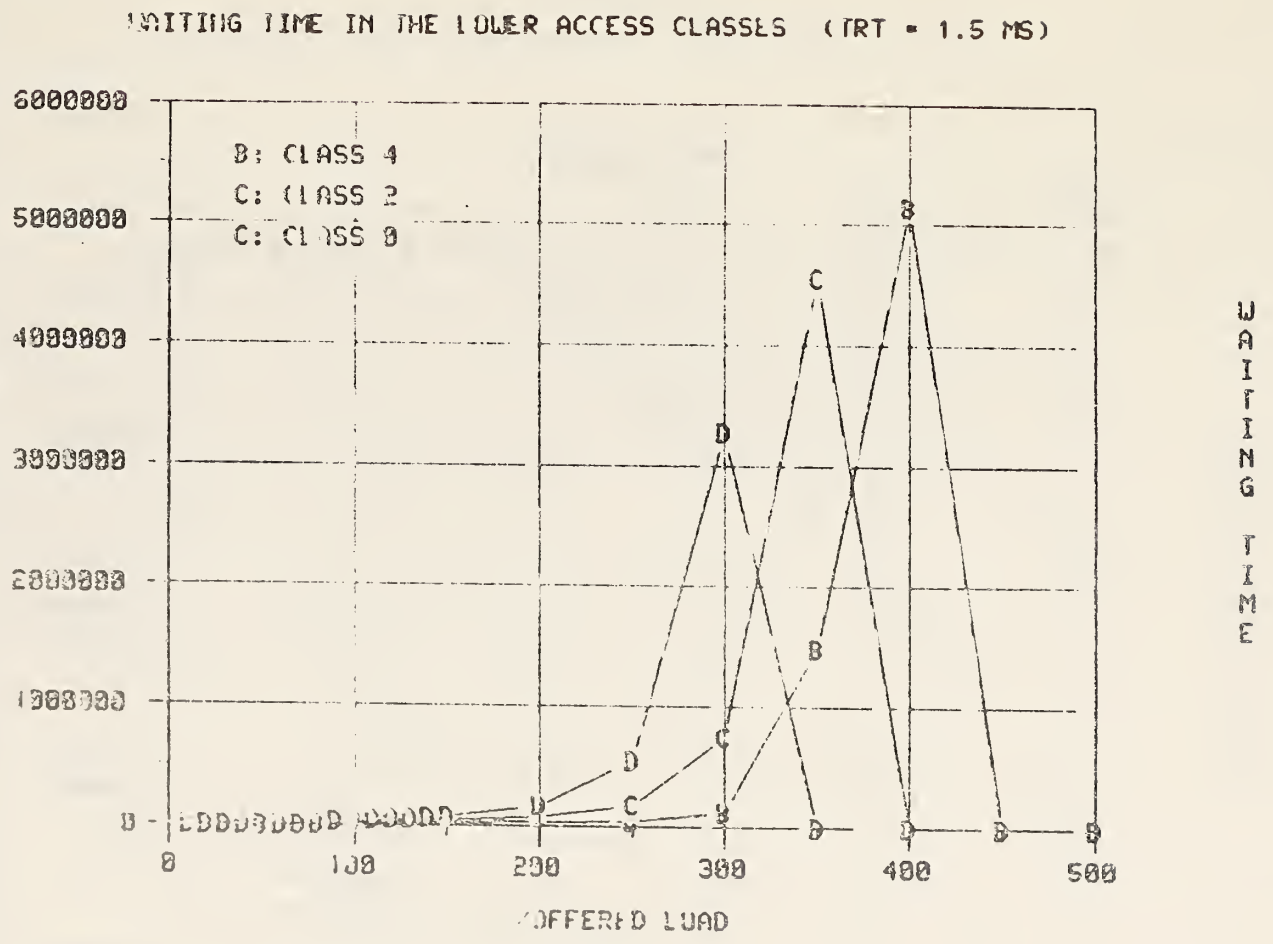

WAITING TIME IN THE LOWER ACCESS CLASSES (TRT = 1.5 MS)
6000000
5000000
4000000
3000000
2000000
1000000
0
B: CLASS 4
C: CLASS 2
D: CLASS 0
0
100
200
300
400
500
OFFERED LOAD
WAITING TIME

Figure 15

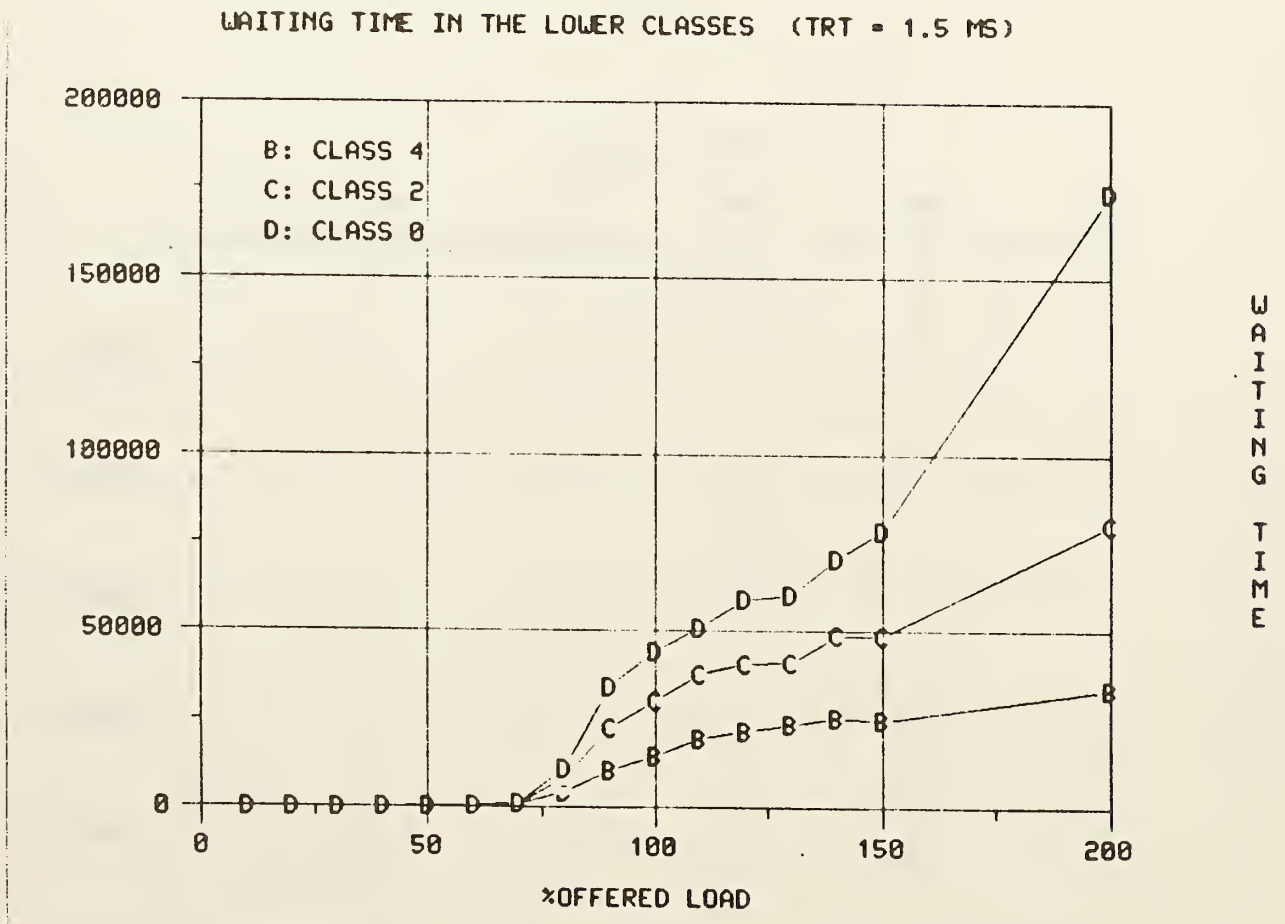

WAITING TIME IN THE LOWER CLASSES (TRT = 1.5 MS)
B: CLASS 4
C: CLASS 2
D: CLASS 0
200000
150000
100000
50000
0
0
50
100
150
200
%OFFERED LOAD
WAITING TIME

Figure 16

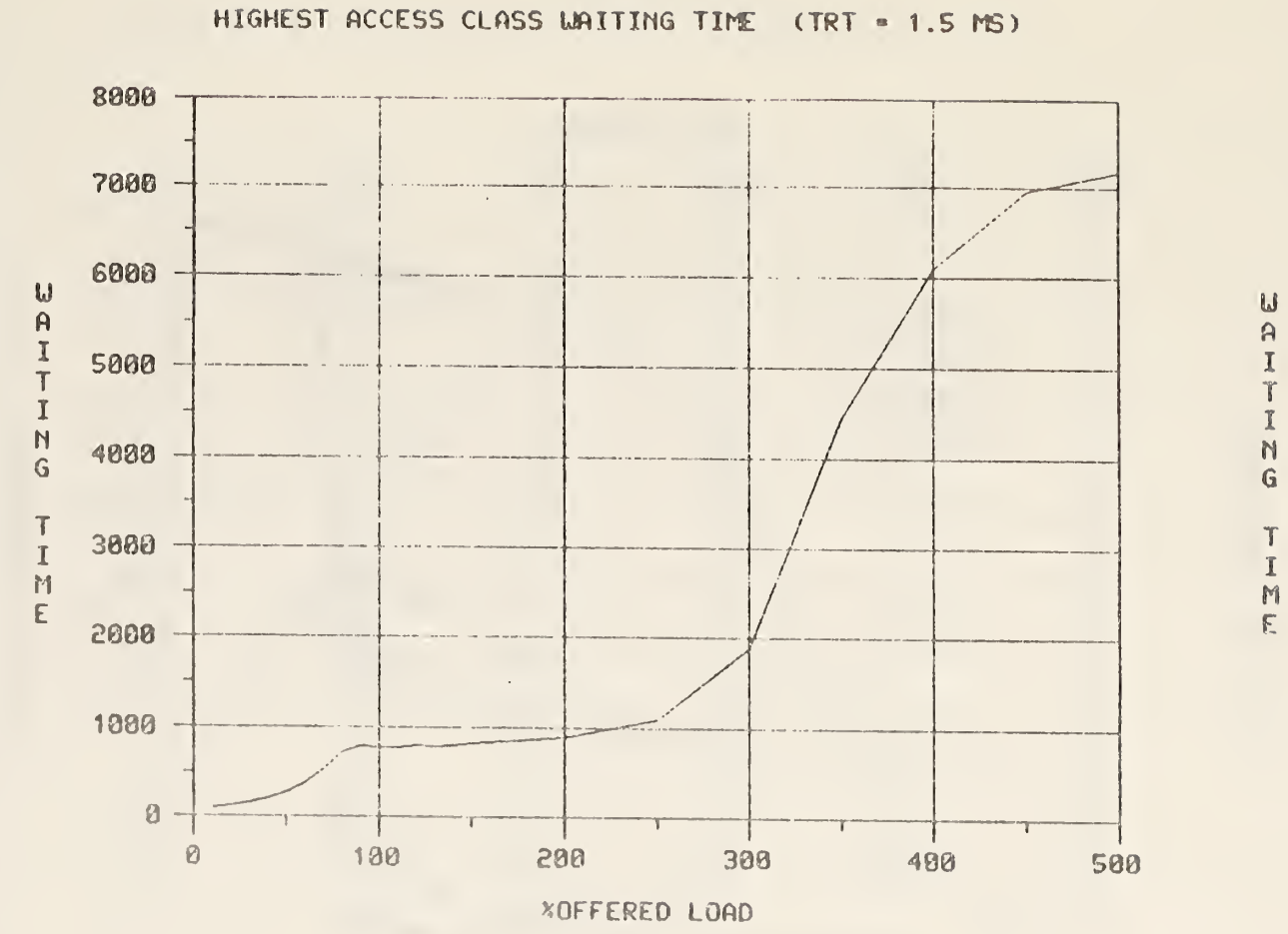

HIGHEST ACCESS CLASS WAITING TIME (TRT = 1.5 MS)
WAITING TIME
8000
7000
6000
5000
4000
3000
2000
1000
0
0
100
200
300
400
500
%OFFERED LOAD
WAITING TIME

Figure 17

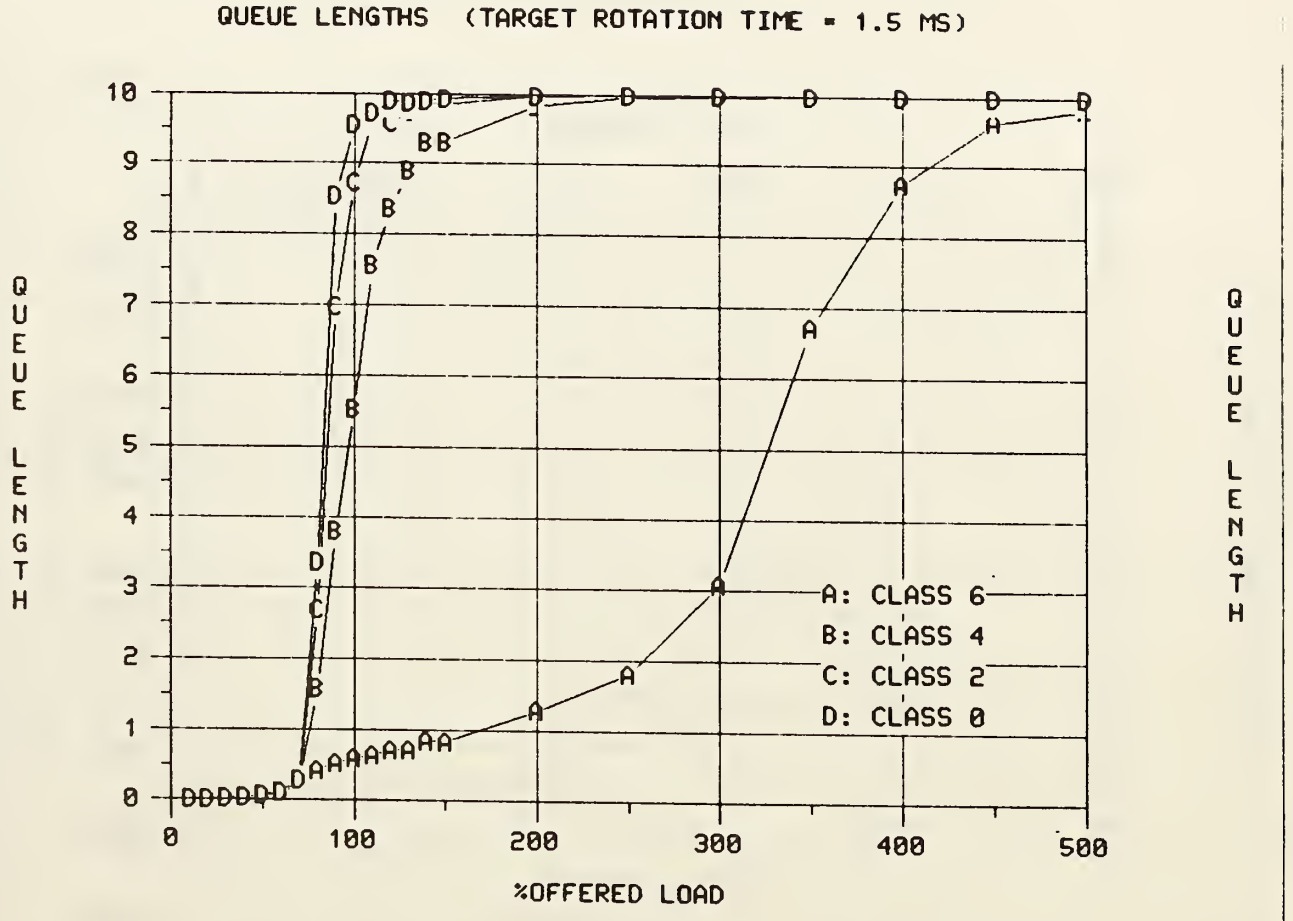

QUEUE LENGTHS (TARGET ROTATION TIME = 1.5 MS)
QUEUE LENGTH
QUEUE LENGTH
%OFFERED LOAD
0
100
200
300
400
500
A: CLASS 6
B: CLASS 4
C: CLASS 2
D: CLASS 0

Figure 18

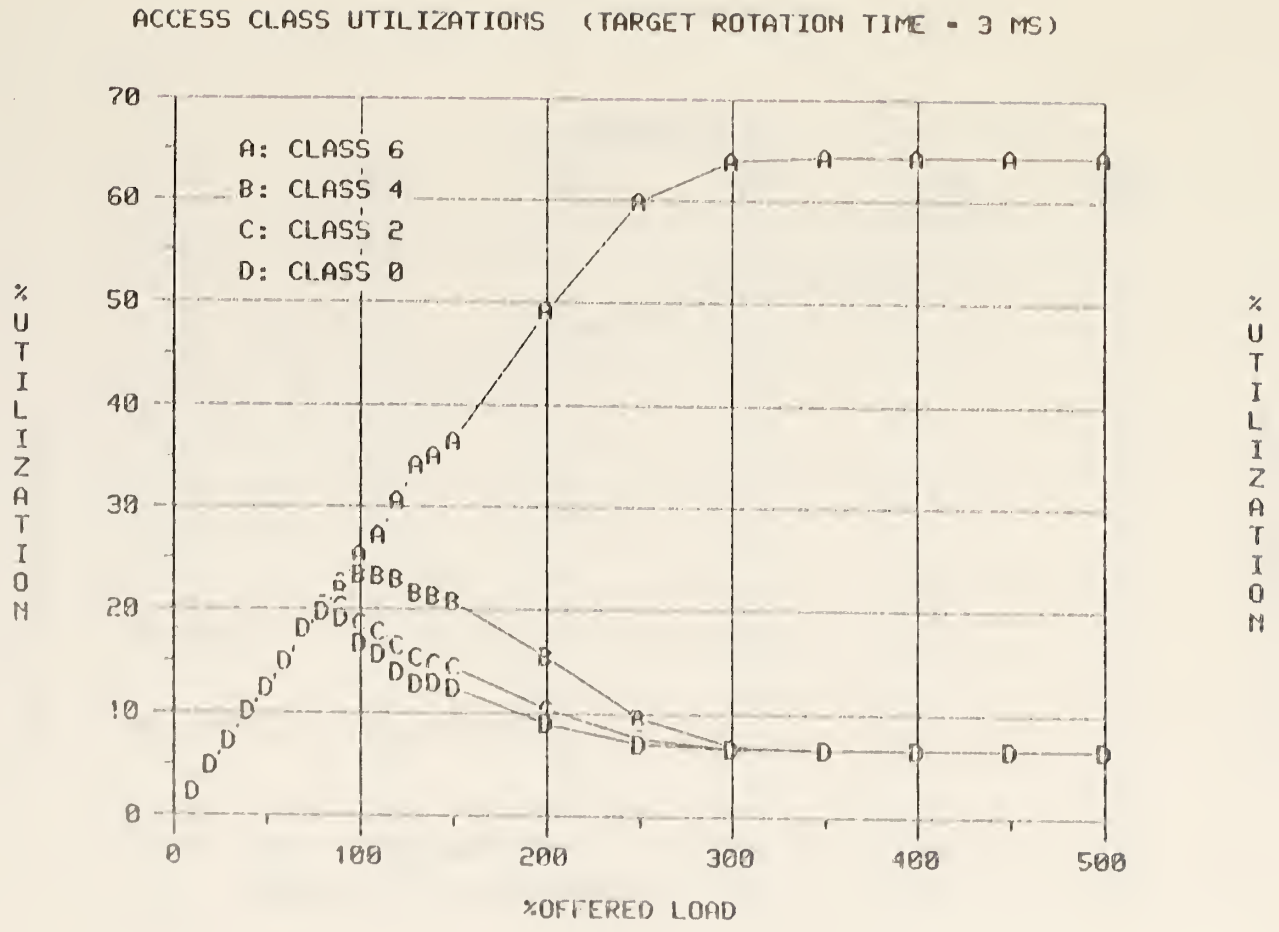

ACCESS CLASS UTILIZATIONS (TARGET ROTATION TIME = 3 MS)
A: CLASS 6
B: CLASS 4
C: CLASS 2
D: CLASS 0
% UTILIZATION
% UTILIZATION
70
60
50
40
30
20
10
0
0
100
200
300
400
500
%OFFERED LOAD

Figure 19

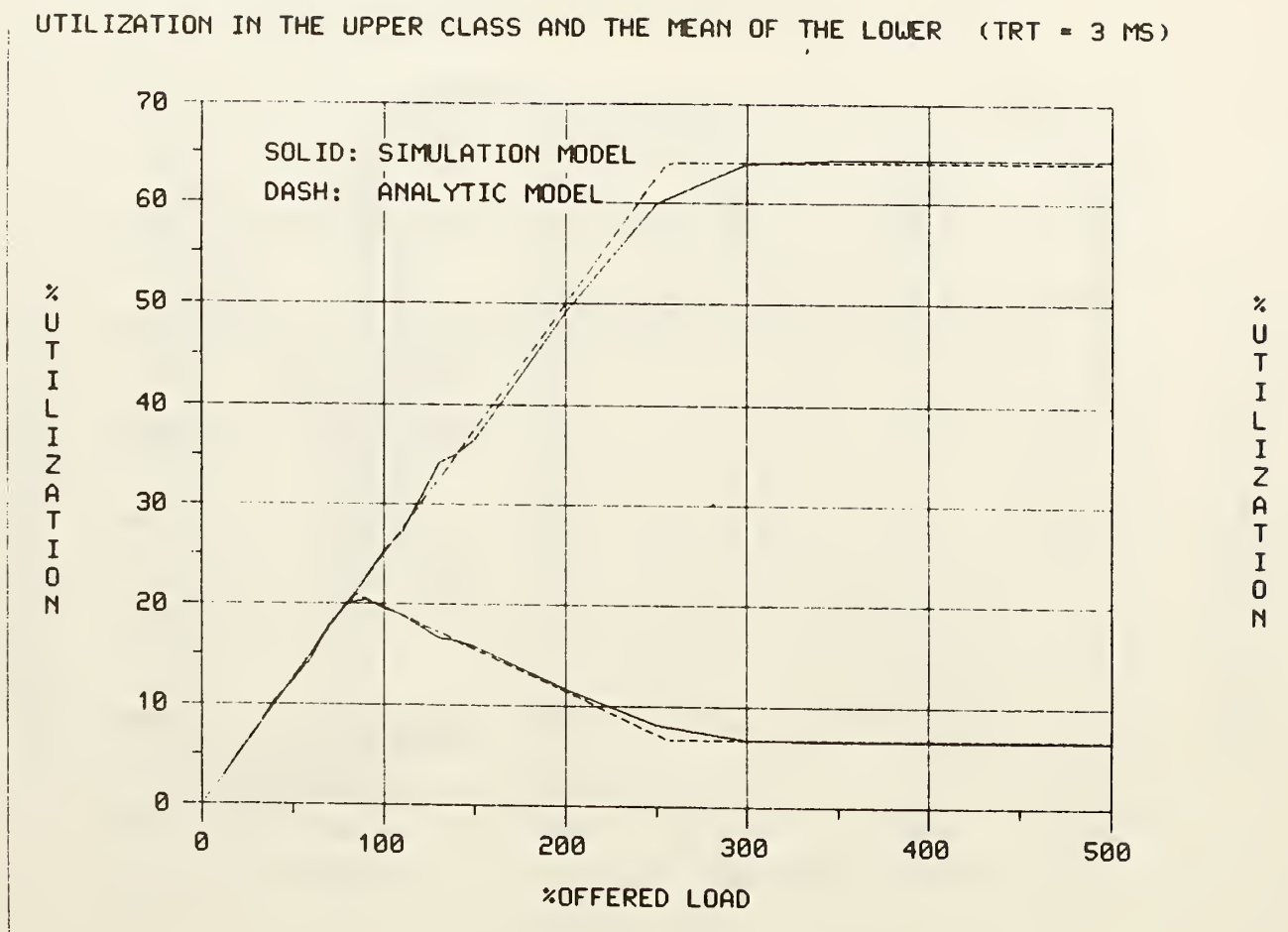

UTILIZATION IN THE UPPER CLASS AND THE MEAN OF THE LOWER (TRT = 3 MS)
SOLID: SIMULATION MODEL
DASH: ANALYTIC MODEL
%UTILIZATION
%UTILIZATION
%OFFERED LOAD
0
10
20
30
40
50
60
70
0
100
200
300
400
500

Figure 20

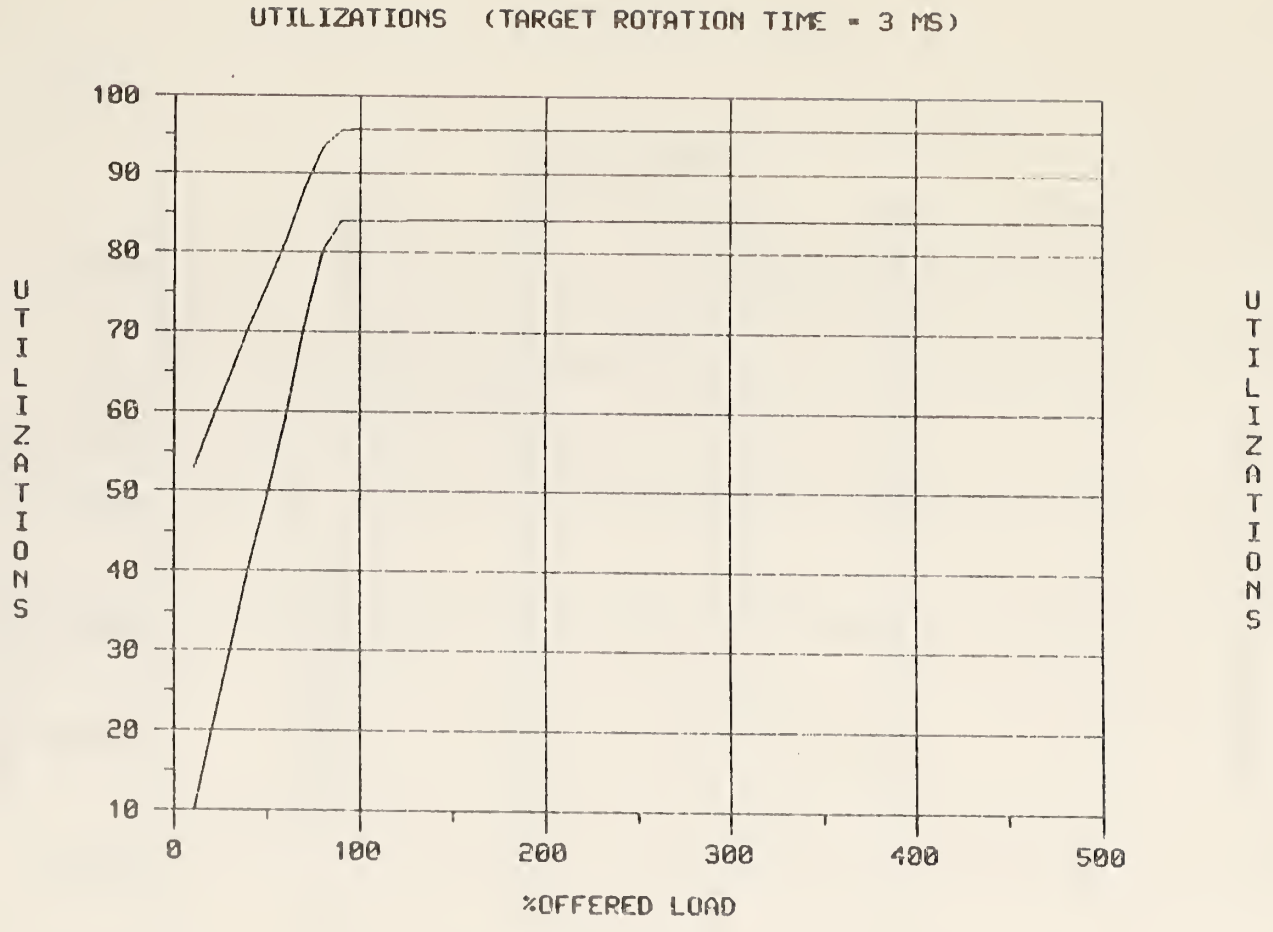

UTILIZATIONS (TARGET ROTATION TIME = 3 MS)
UTILIZATIONS
100
90
80
70
60
50
40
30
20
10
0
100
200
300
400
500
%OFFERED LOAD
UTILIZATIONS

Figure 21

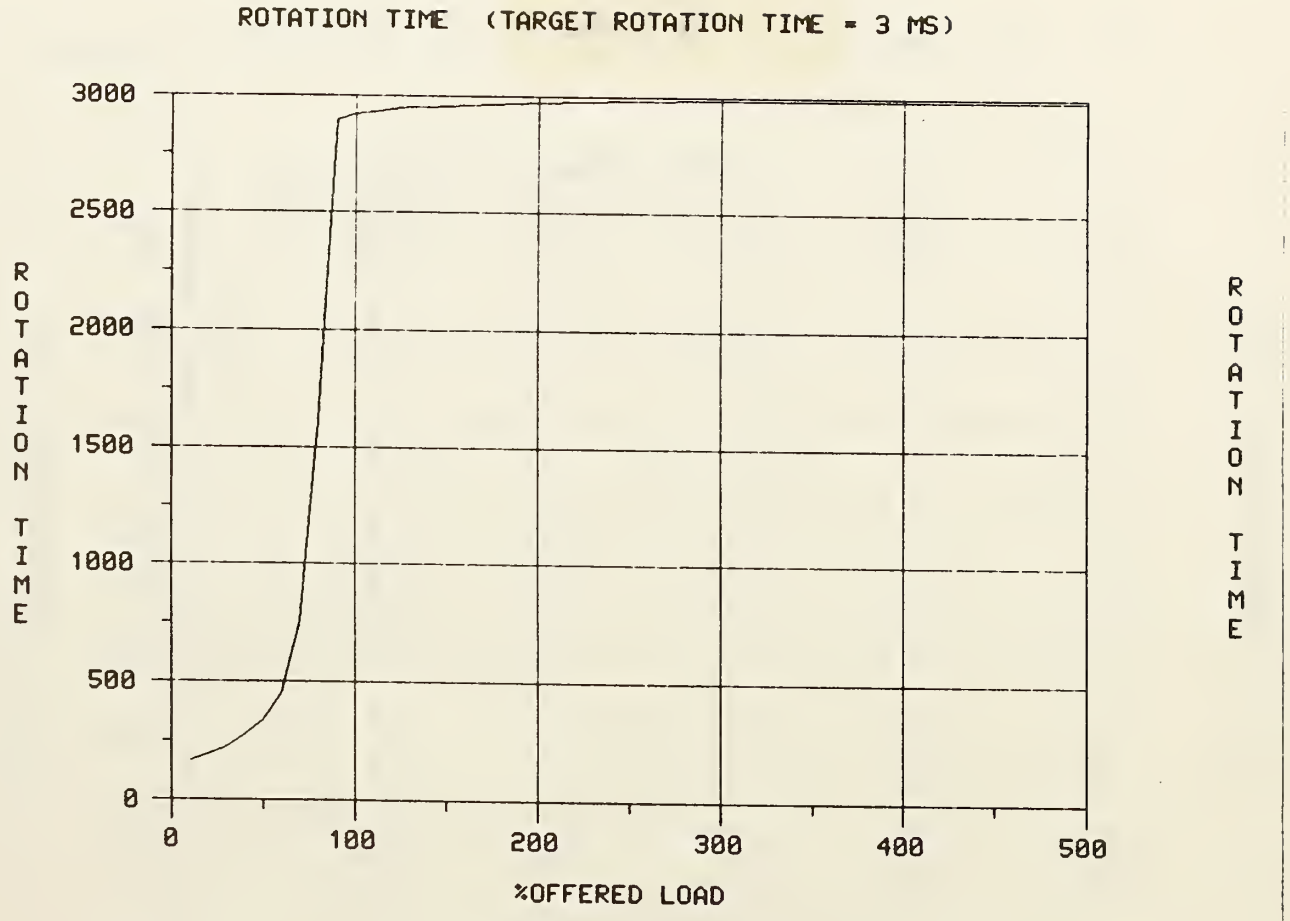

ROTATION TIME (TARGET ROTATION TIME = 3 MS)
ROTATION TIME
3000
2500
2000
1500
1000
500
0
0
100
200
300
400
500
%OFFERED LOAD
ROTATION TIME

Figure 22

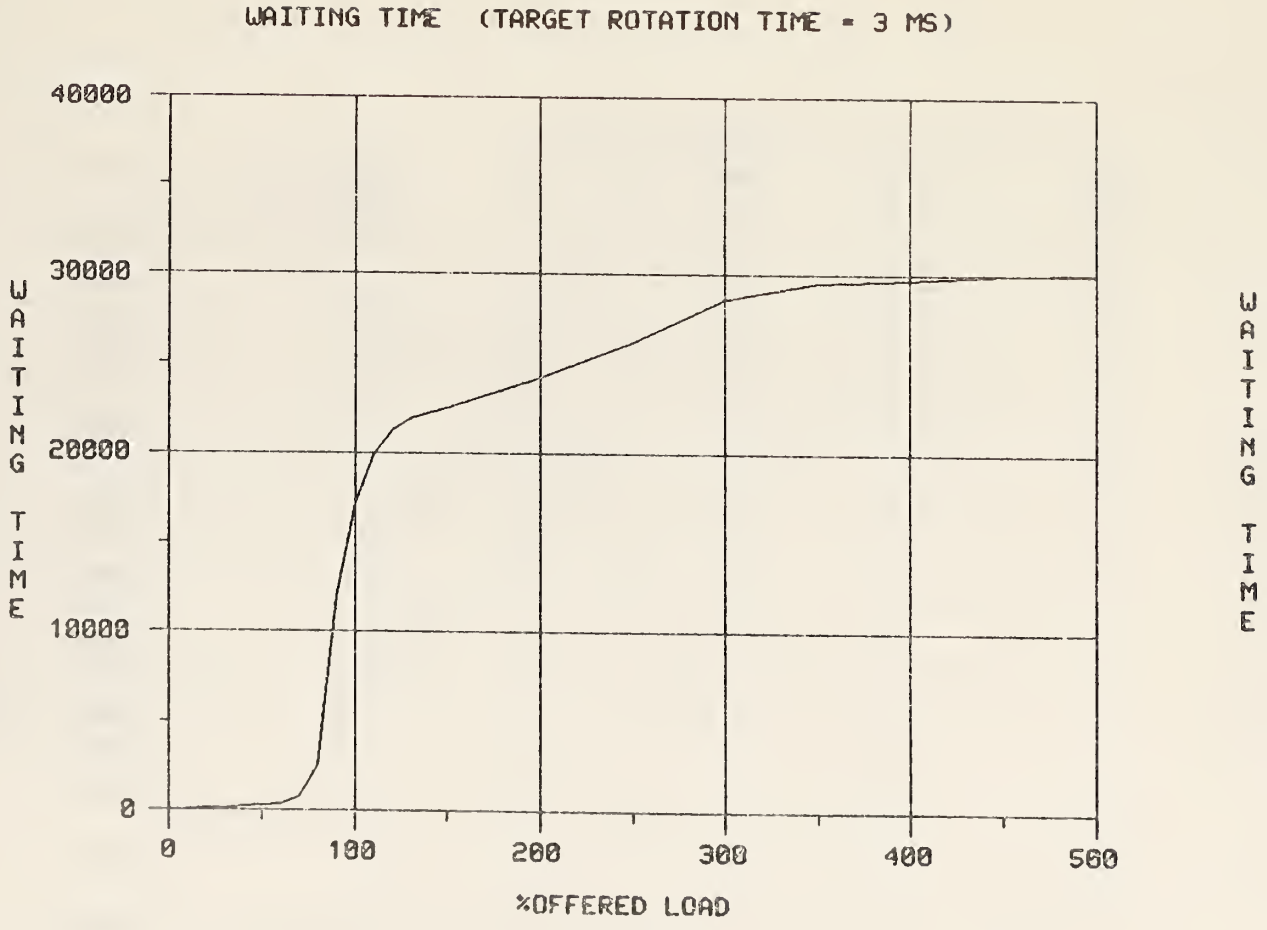

WAITING TIME (TARGET ROTATION TIME = 3 MS)
40000
30000
20000
10000
0
WAITING TIME
0
100
200
300
400
500
%OFFERED LOAD
WAITING TIME

Figure 23

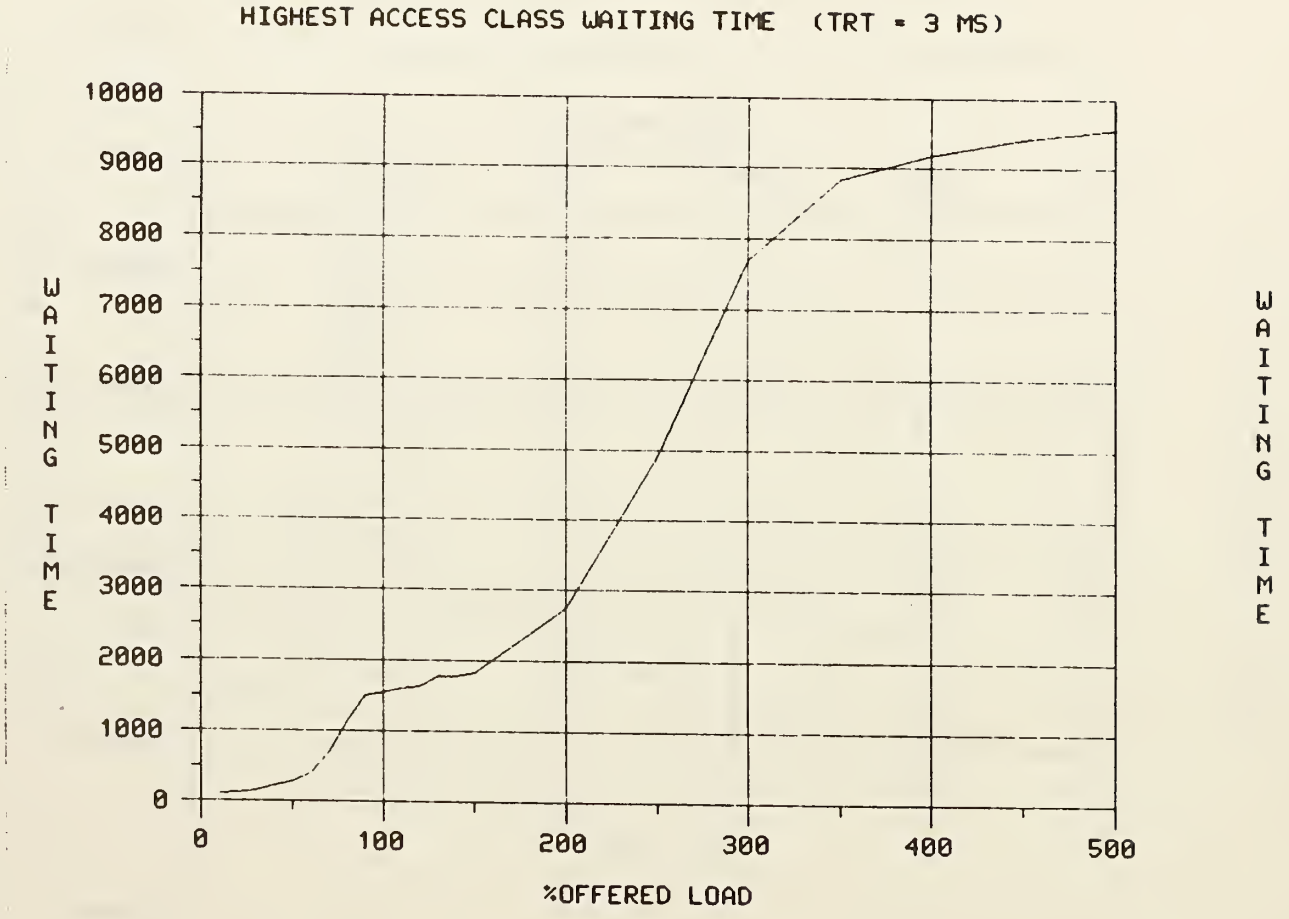

HIGHEST ACCESS CLASS WAITING TIME (TRT = 3 MS)
WAITING TIME
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0
100
200
300
400
500
%OFFERED LOAD
WAITING TIME

Figure 24

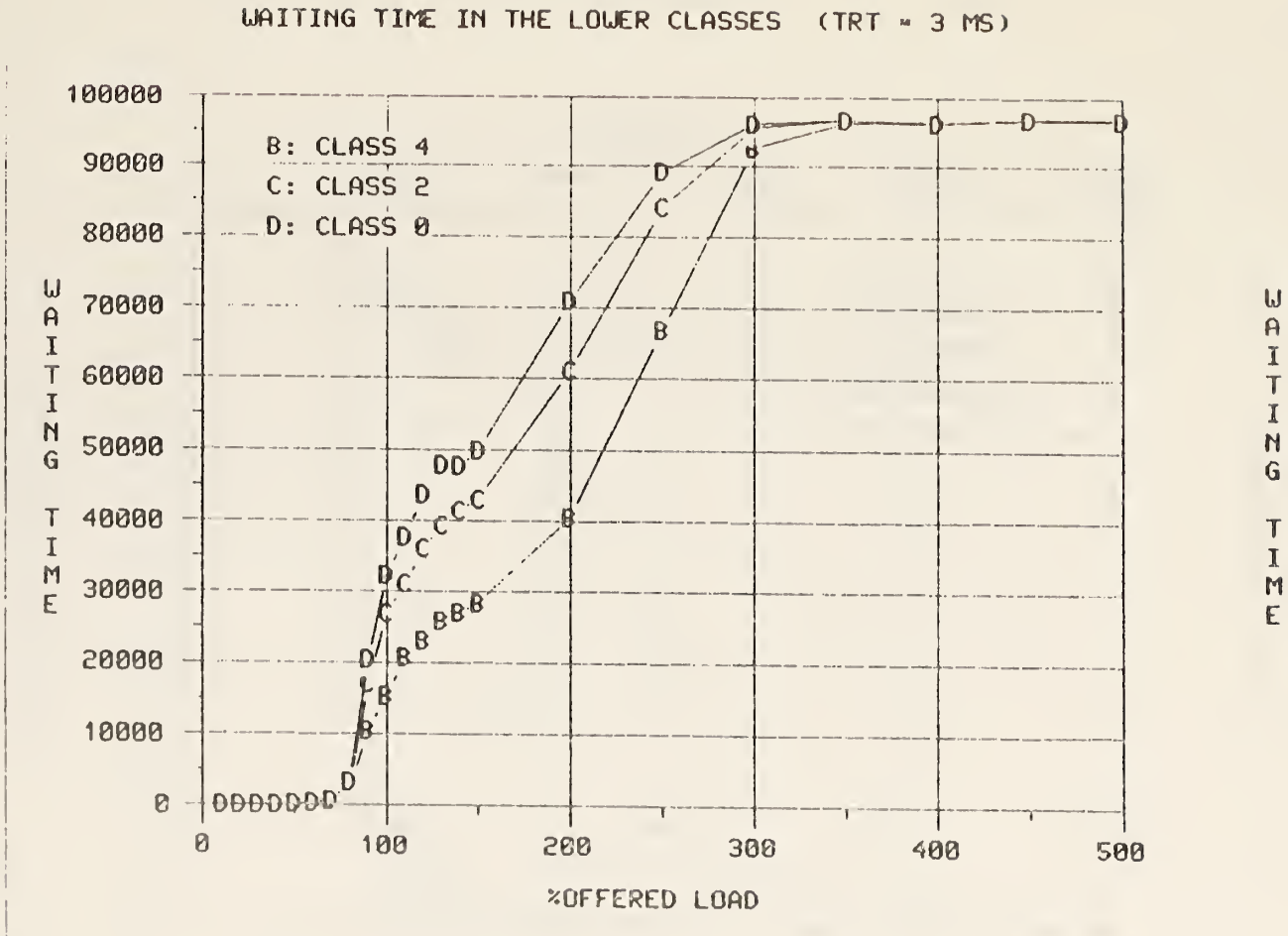

WAITING TIME IN THE LOWER CLASSES (TRT = 3 MS)
B: CLASS 4
C: CLASS 2
D: CLASS 0
WAITING TIME
100000
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
0
100
200
300
400
500
%OFFERED LOAD

Figure 25

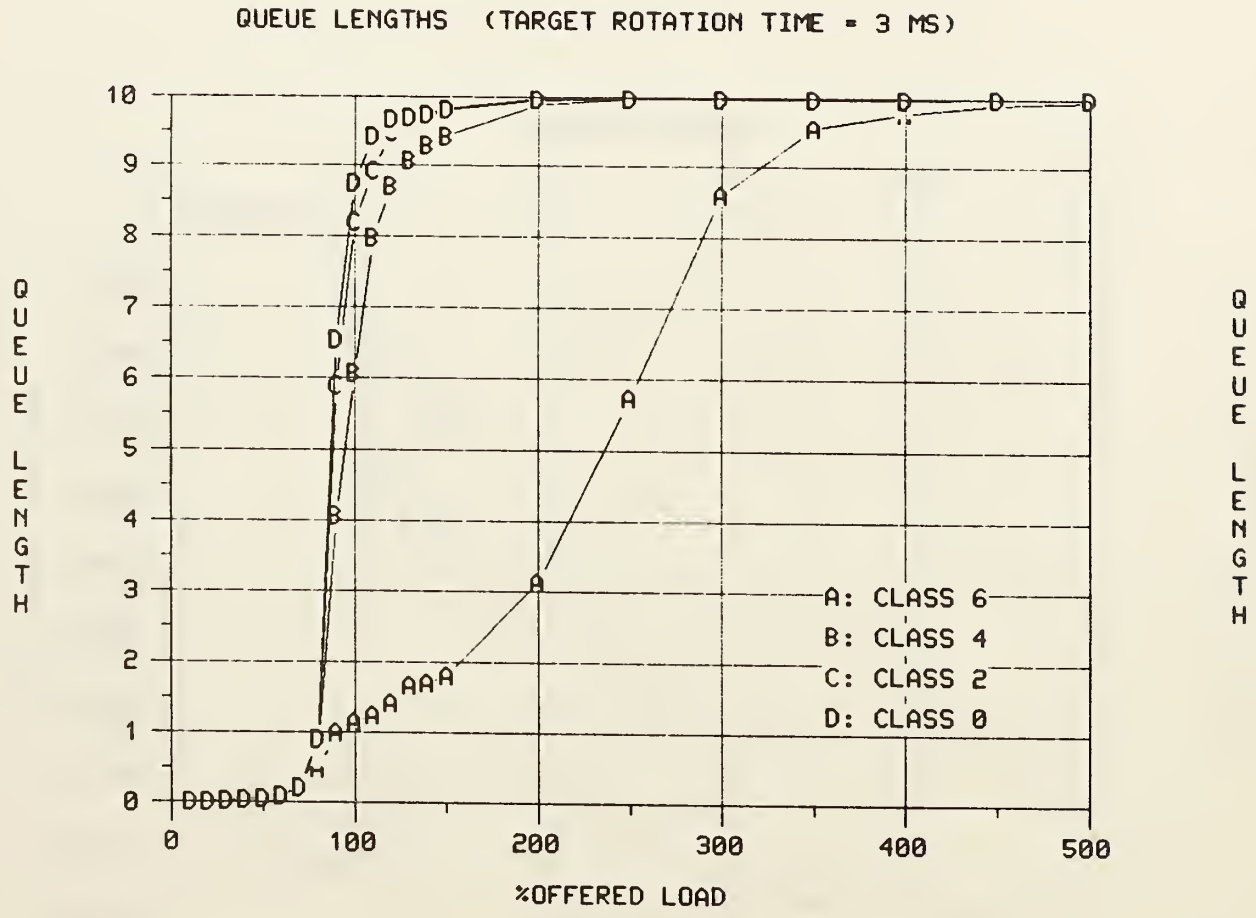

QUEUE LENGTHS (TARGET ROTATION TIME = 3 MS)
QUEUE LENGTH
QUEUE LENGTH
%OFFERED LOAD
0
100
200
300
400
500
0
1
2
3
4
5
6
7
8
9
10
A: CLASS 6
B: CLASS 4
C: CLASS 2
D: CLASS 0

Figure 26

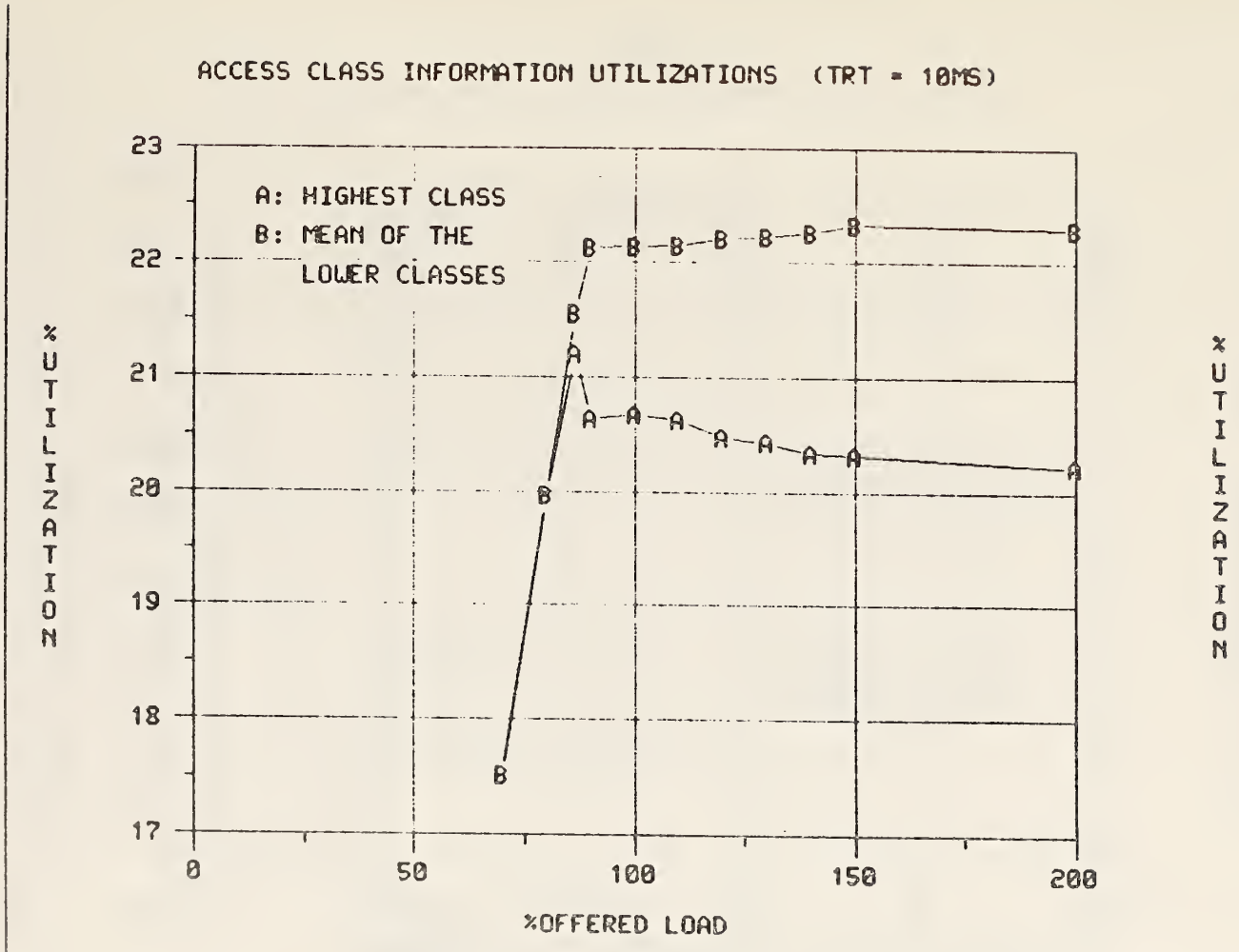

ACCESS CLASS INFORMATION UTILIZATIONS (TRT = 10MS)
A: HIGHEST CLASS
B: MEAN OF THE LOWER CLASSES
%UTILIZATION
%UTILIZATION
%OFFERED LOAD
23
22
21
20
19
18
17
0
50
100
150
200

Figure 27

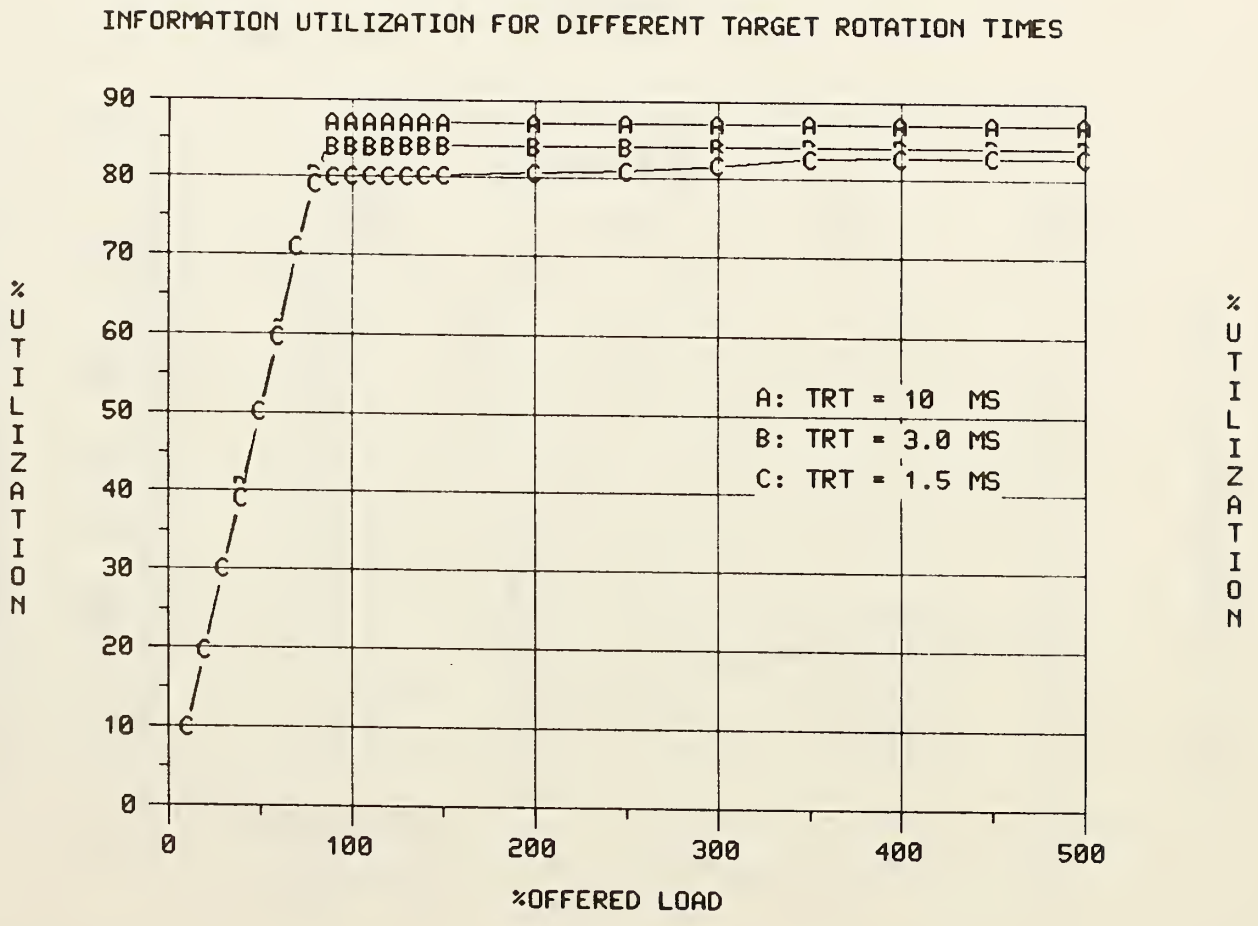

INFORMATION UTILIZATION FOR DIFFERENT TARGET ROTATION TIMES
%UTILIZATION
%UTILIZATION
90
80
70
60
50
40
30
20
10
0
0
100
200
300
400
500
%OFFERED LOAD
A: TRT = 10 MS
B: TRT = 3.0 MS
C: TRT = 1.5 MS

Figure 28

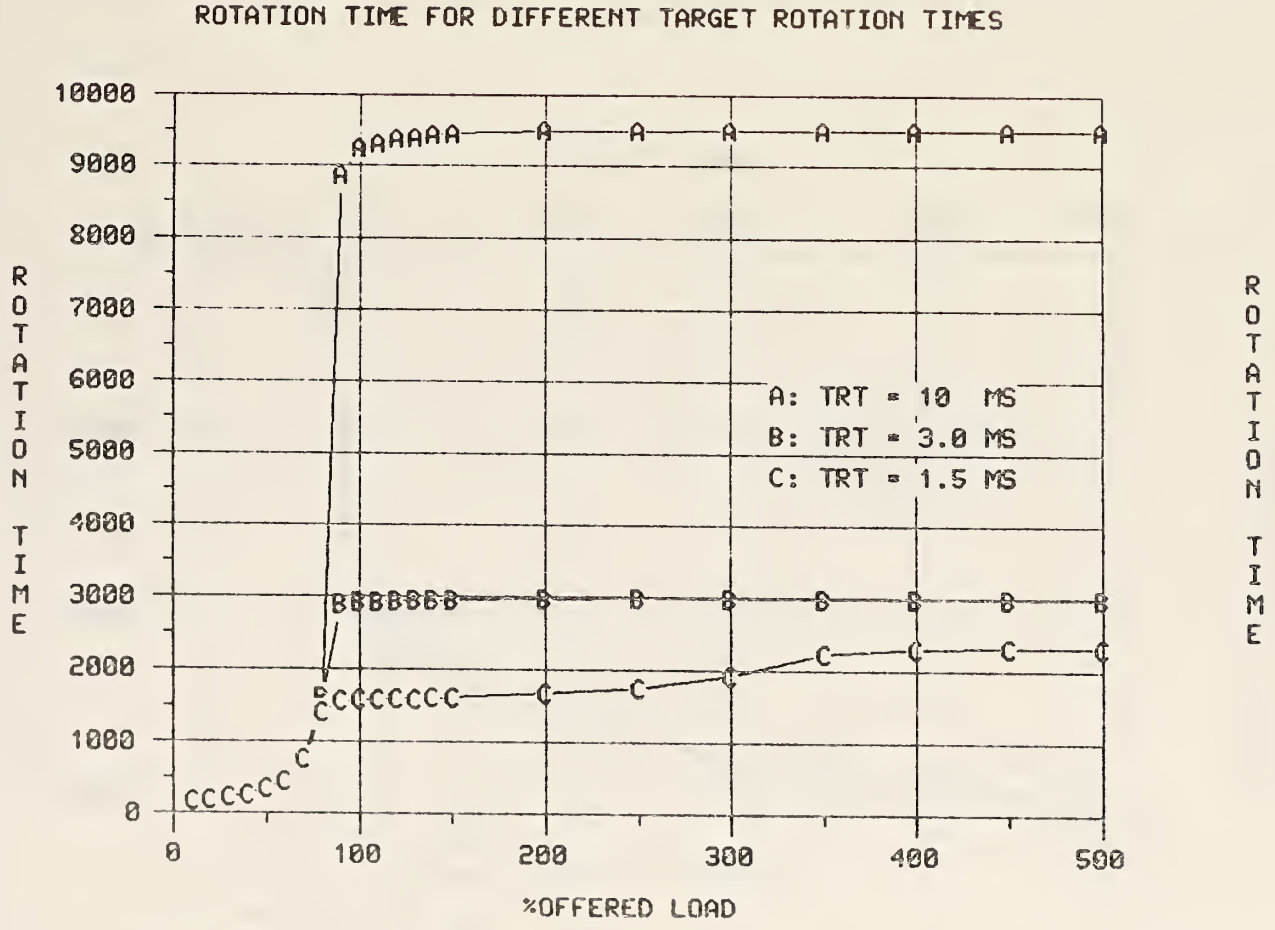

ROTATION TIME FOR DIFFERENT TARGET ROTATION TIMES
ROTATION TIME
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
A: TRT = 10 MS
B: TRT = 3.0 MS
C: TRT = 1.5 MS
0
100
200
300
400
500
%OFFERED LOAD
ROTATION TIME

Figure 29

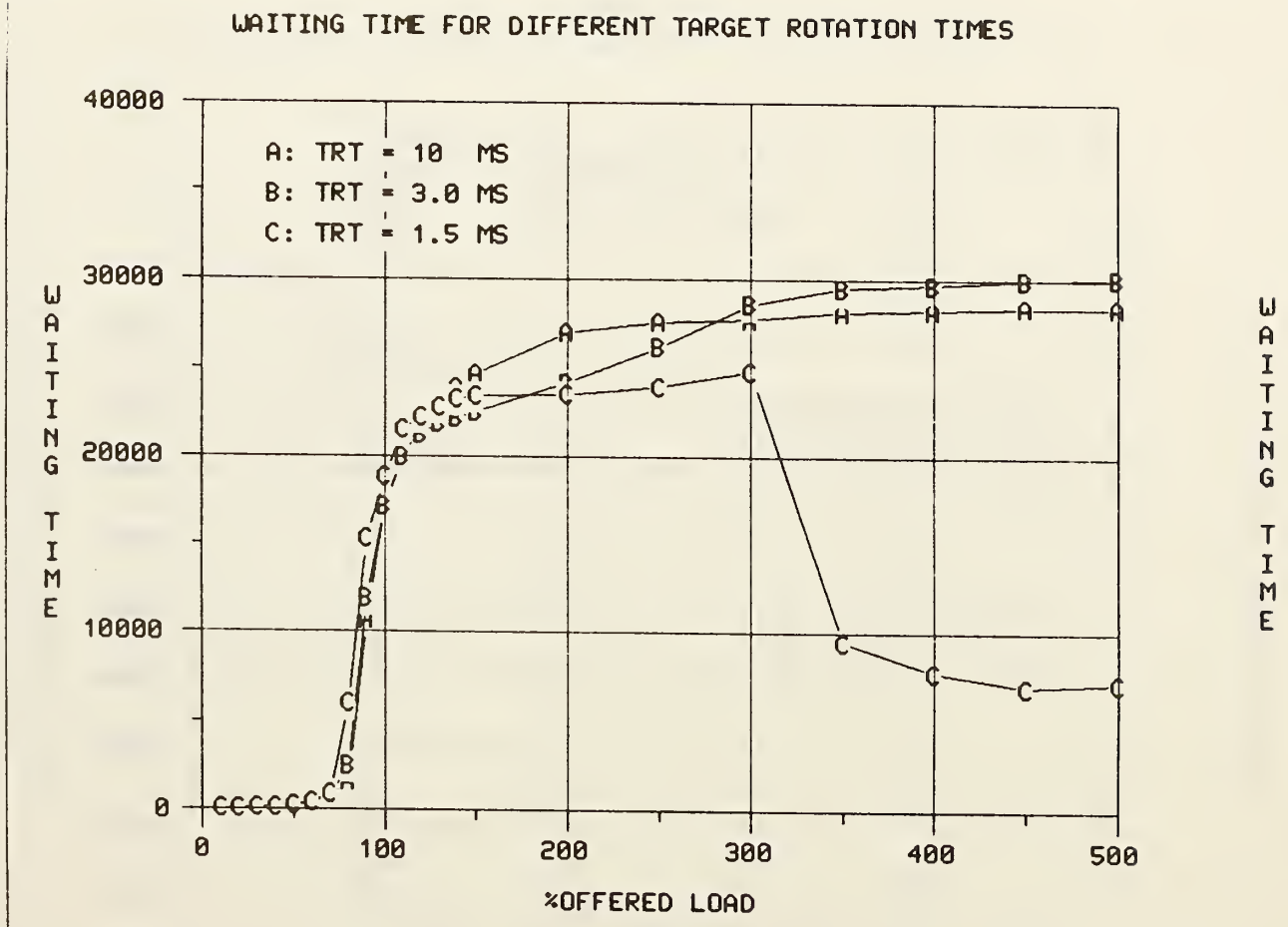

WAITING TIME FOR DIFFERENT TARGET ROTATION TIMES
A: TRT = 10 MS
B: TRT = 3.0 MS
C: TRT = 1.5 MS
WAITING TIME
40000
30000
20000
10000
0
0
100
200
300
400
500
%OFFERED LOAD
WAITING TIME

Figure 30

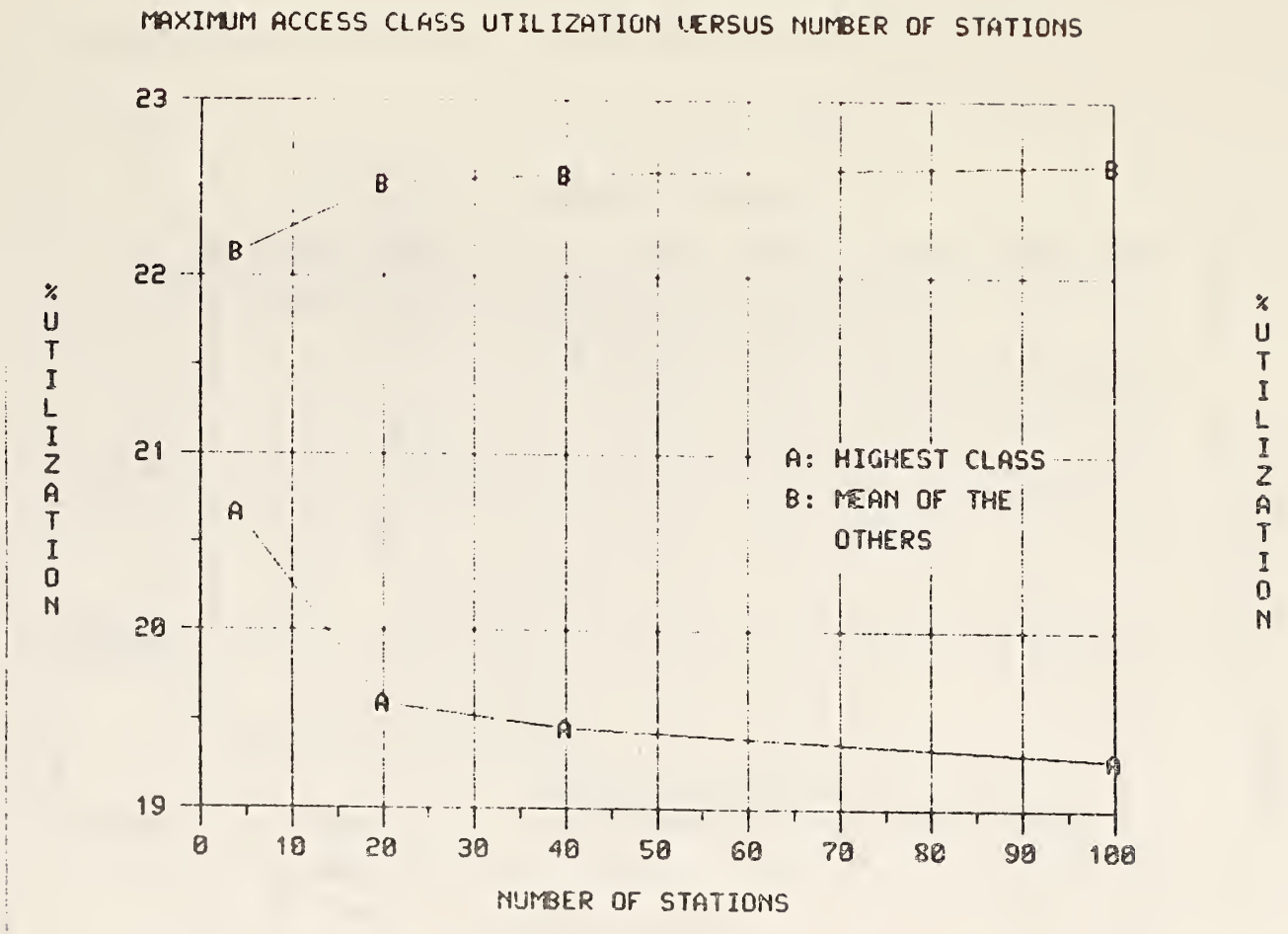

MAXIMUM ACCESS CLASS UTILIZATION VERSUS NUMBER OF STATIONS
% UTILIZATION
% UTILIZATION
23
22
21
20
19
0
10
20
30
40
50
60
70
80
90
100
NUMBER OF STATIONS
A: HIGHEST CLASS
B: MEAN OF THE OTHERS
A
B

Figure 31

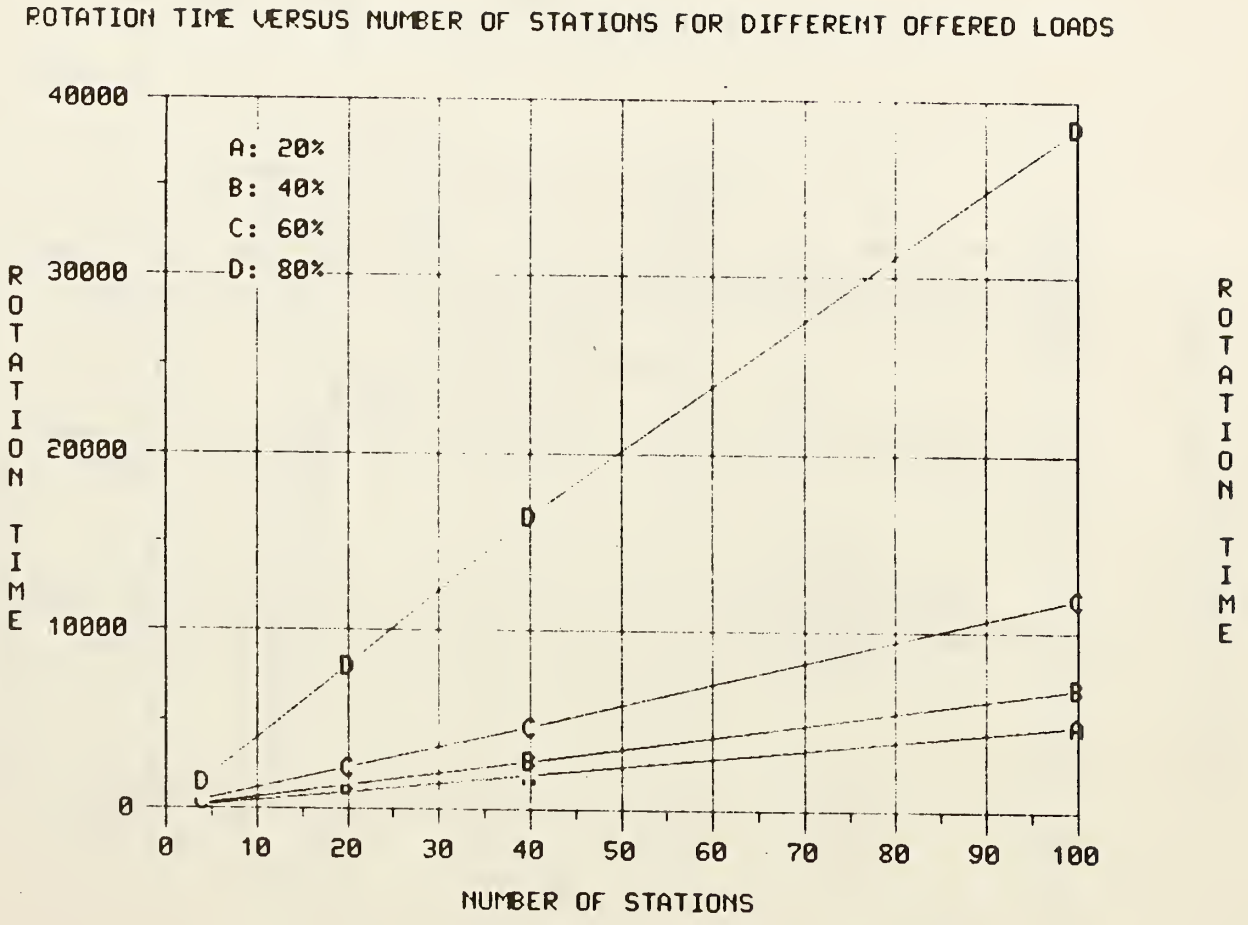

ROTATION TIME VERSUS NUMBER OF STATIONS FOR DIFFERENT OFFERED LOADS
A: 20%
B: 40%
C: 60%
D: 80%
ROTATION TIME
40000
30000
20000
10000
0
0
10
20
30
40
50
60
70
80
90
100
NUMBER OF STATIONS
ROTATION TIME

Figure 32

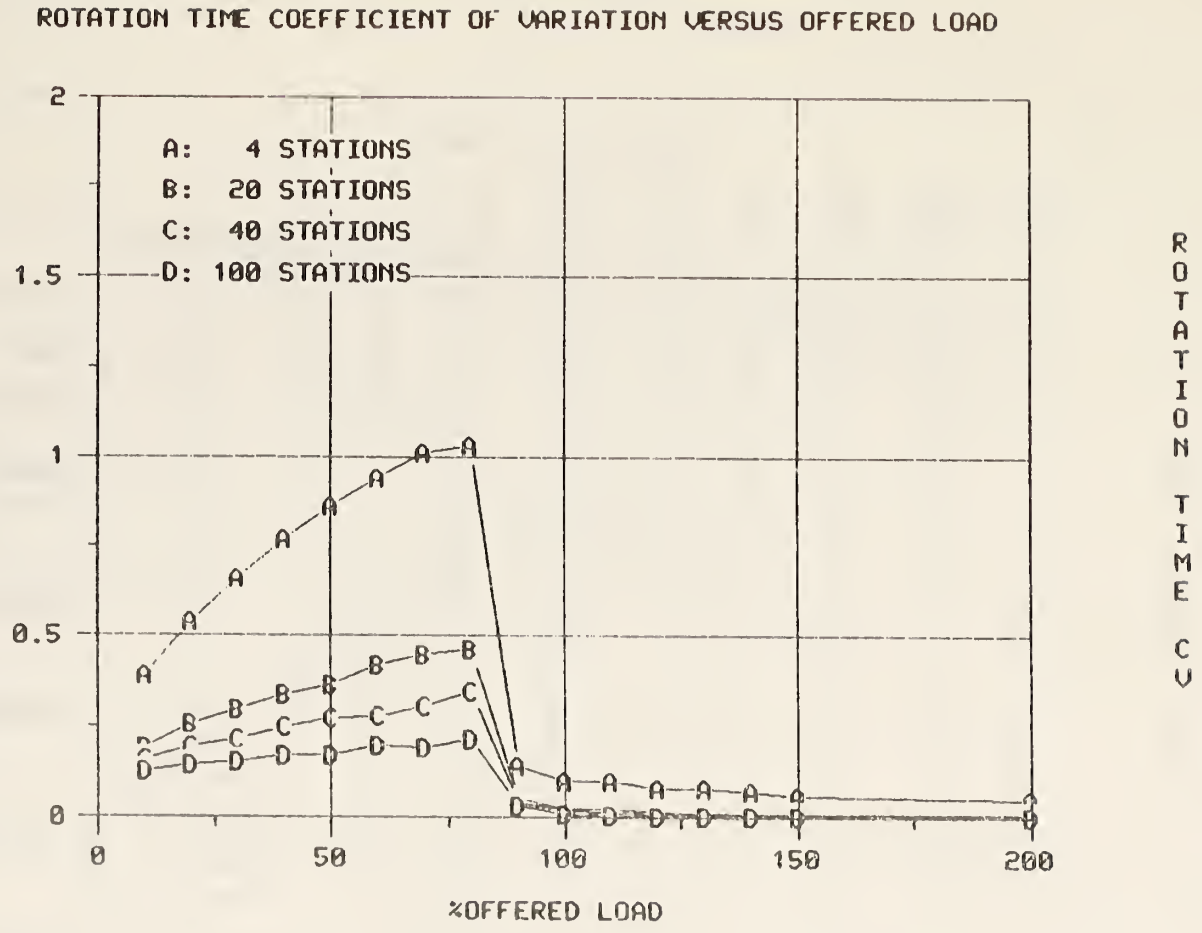

ROTATION TIME COEFFICIENT OF VARIATION VERSUS OFFERED LOAD
A: 4 STATIONS
B: 20 STATIONS
C: 40 STATIONS
D: 100 STATIONS
2
1.5
1
0.5
0
0
50
100
150
200
%OFFERED LOAD
ROTATION TIME CV

Figure 33

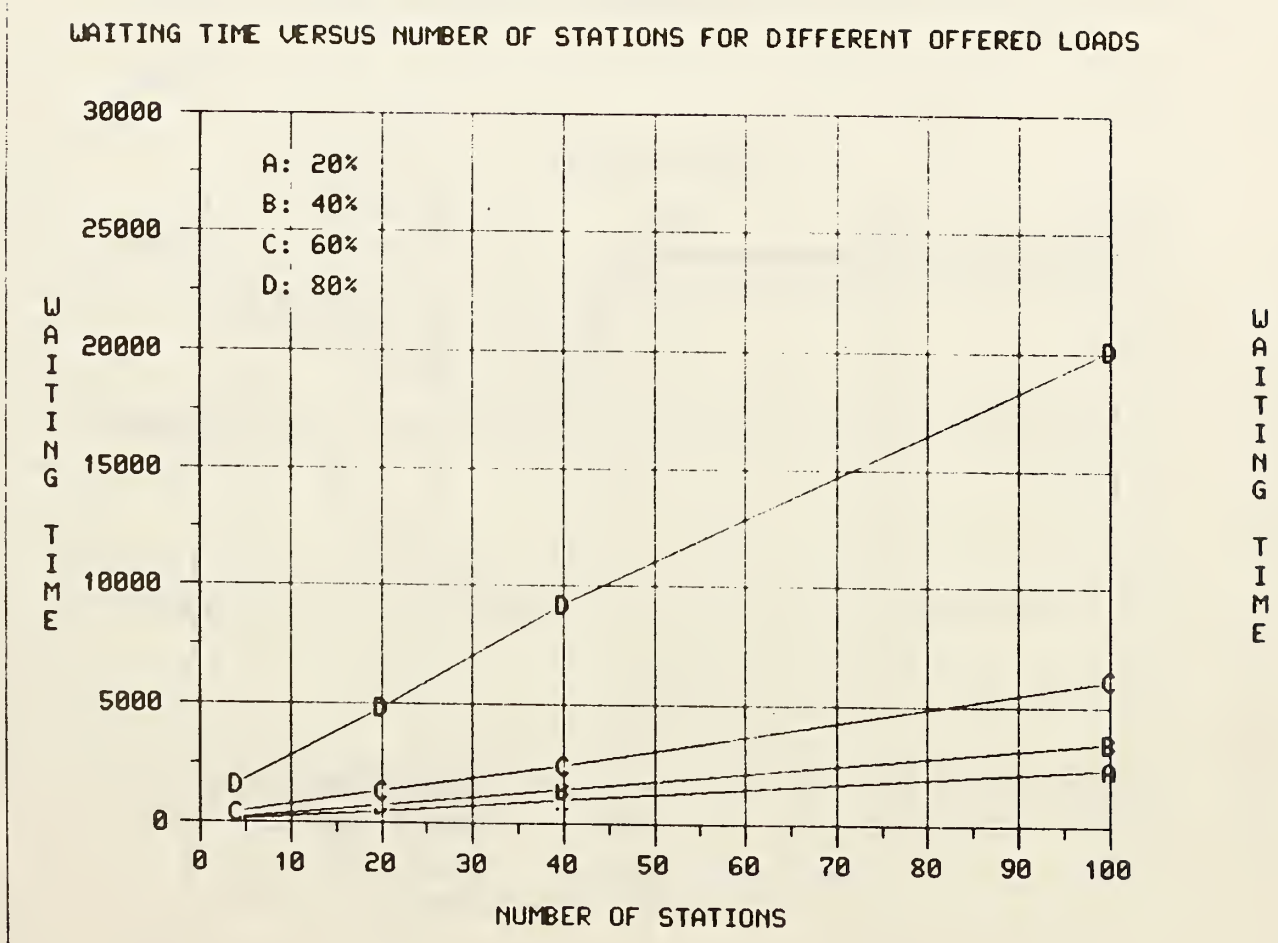

WAITING TIME VERSUS NUMBER OF STATIONS FOR DIFFERENT OFFERED LOADS
A: 20%
B: 40%
C: 60%
D: 80%
WAITING TIME
30000
25000
20000
15000
10000
5000
0
0
10
20
30
40
50
60
70
80
90
100
NUMBER OF STATIONS
WAITING TIME

Figure 34

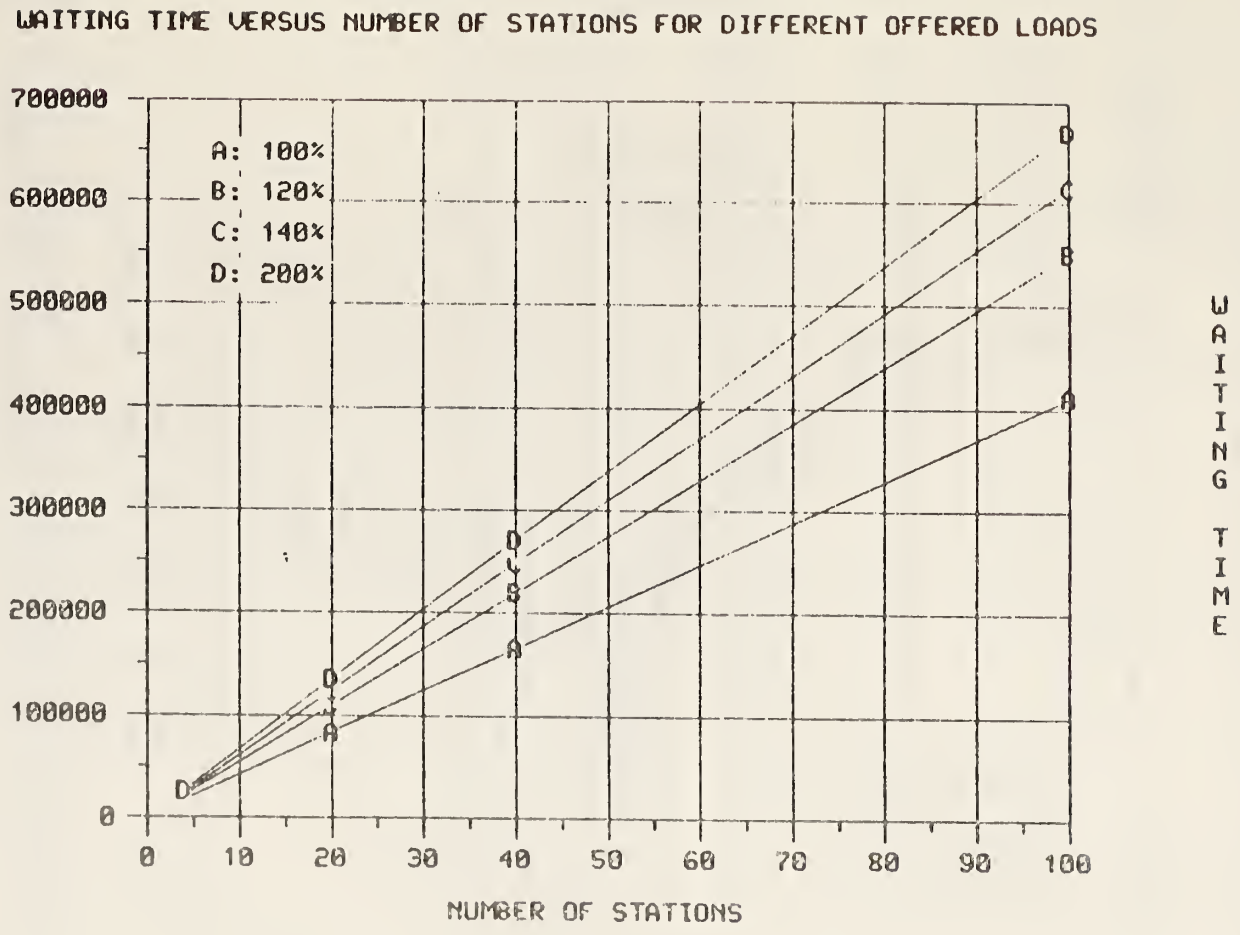

WAITING TIME VERSUS NUMBER OF STATIONS FOR DIFFERENT OFFERED LOADS
A: 100%
B: 120%
C: 140%
D: 200%
700000
600000
500000
400000
300000
200000
100000
0
0
10
20
30
40
50
60
70
80
90
100
NUMBER OF STATIONS
WAITING TIME

Figure 35

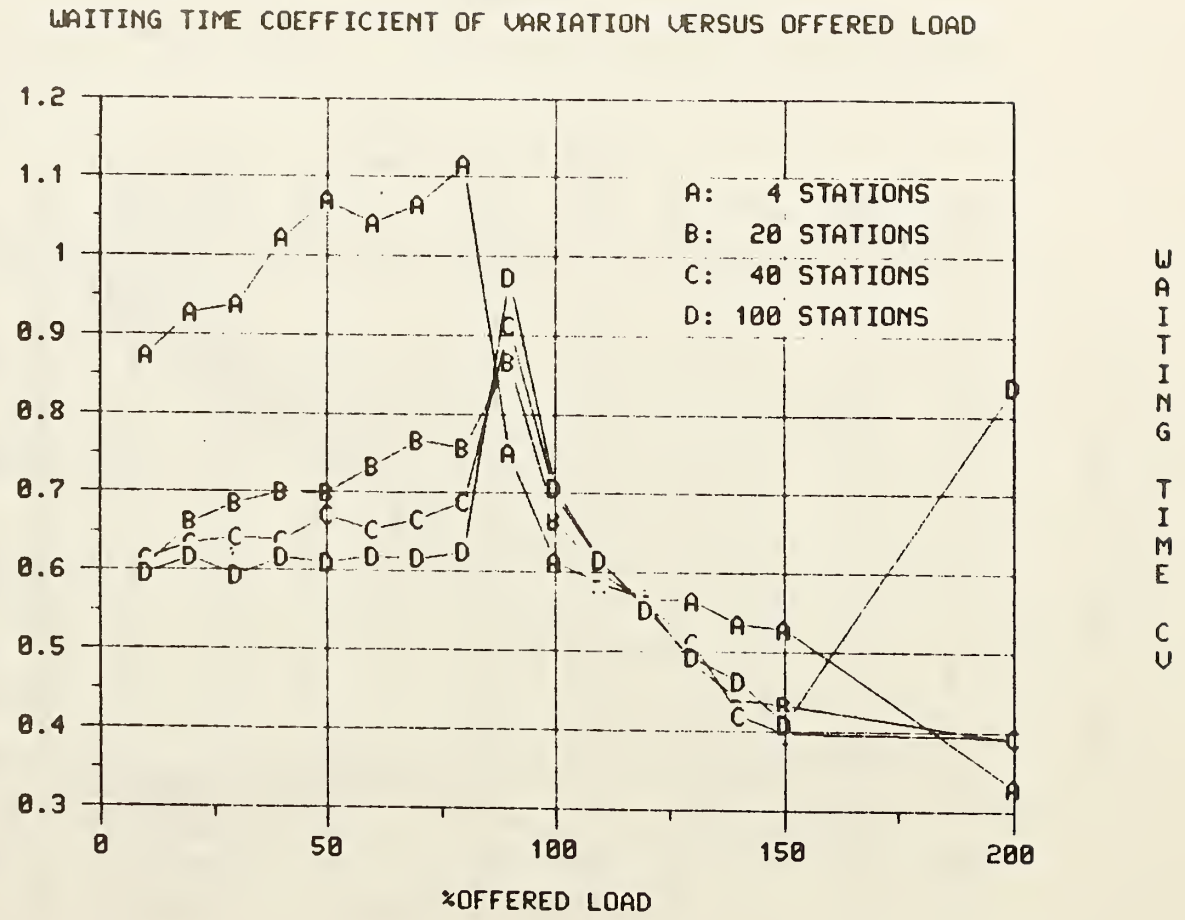

WAITING TIME COEFFICIENT OF VARIATION VERSUS OFFERED LOAD
A: 4 STATIONS
B: 20 STATIONS
C: 40 STATIONS
D: 100 STATIONS
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
50
100
150
200
%OFFERED LOAD
WAITING TIME CV

Figure 36

NBS-114A (REV. 2-8C)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET (See instructions)	1. PUBLICATION OR N REPORT NO. NBSIR 84-2966	2. Performing Organ. Report No. NBS	3. Publication Date November 1984

4. TITLE AND SUBTITLE

An IEEE 802.4 Token Bus Network Simulation

5. AUTHOR(S)

Jean-Luc Archambault

6. PERFORMING ORGANIZATION (If joint or other than NBS, see instructions)

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

7. Contract/Grant No.

8. Type of Report & Period Covered

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (Street, City, State, ZIP)

National Bureau of Standards
Bldg. 225, Rm. B218
Gaithersburg, MD 20899

10. SUPPLEMENTARY NOTES

☐ Document describes a computer program; SF-185, FIPS Software Summary, is attached.

11. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here)

A discrete event simulation of token bus networks has been designed and implemented at the National Bureau of Standards, and used to analyze the performance of local area networks compliant with the IEEE 802.4 specifications. The model measures the utilization of the network, the rotation time, the waiting time of the data packets, and the queue lengths in the Medium Access Control sublayer. This paper presents the model, and studies the sensitivity of these variables to the offered load, the packet length, the target rotation times, and the number of stations.

12. KEY WORDS (Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons)

Local Area Networks; Modelling; Performance Analysis; Simulation; Standards; Token Bus.

13. AVAILABILITY

☐ Unlimited

☐ For Official Distribution. Do Not Release to NTIS

☐ Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.

☒ Order From National Technical Information Service (NTIS), Springfield, VA. 22161

14. NO. OF PRINTED PAGES 73

15. Price $10.00