Fiber Distributed Data Interface: Throughput Evaluation with Multiple Classes of Traffic

Priya N. Werahera, Member, IEEE and Anura P. Jayasumana, Senior Member, IEEE

Abstract—The Fiber Distributed Data Interface (FDDI) standard supports synchronous and asynchronous data transmissions. It allows each station to have multiple classes of asynchronous data, and meets the requirements of different classes by means of a timer-based priority scheme. The performance of a FDDI network carrying multiple classes of traffic is considered. An analytical model is presented to evaluate the throughput of synchronous traffic and asynchronous traffic. The model can be used to evaluate the throughput of individual priority classes and the mean token-cycle time of the network, when the network offered load varies from very low values to high values. The governing equations for the throughput characteristics and mean token-cycle time are strictly functions of network parameters. Simulations are used to verify the analytical model. Simulation results are also used to examine the variation of the mean access delay characteristics of various classes and the token-cycle time distribution of the priority scheme. Limitations of the model are outlined.

Keywords—FDDI, token-ring, local-area networks, priority classes, integrated services, voice/data integration.

I. INTRODUCTION

The Fiber Distributed Data Interface (FDDI) provides high bandwidth general purpose interconnections using optical fiber as the physical medium in a token-ring configuration [1,2]. It supports two major classes of traffic; synchronous and asynchronous. Synchronous class provides a preallocated bandwidth and a guaranteed response time, and may be used for time-critical messages such as voice and video signals. The remaining bandwidth is dynamically allocated to asynchronous classes with non-time critical messages such as data by means of a timer-based priority mechanism. This priority mechanism is very similar to that of the IEEE 802.4 token-passing scheme. Analytical and simulation models have been used to evaluate the performance of the 802.4 priority scheme for homogeneous token-passing bus networks supporting multiple classes of traffic [2-5]. Performance of 802.4 networks for time critical messages has been investigated in [6-8]. A heterogeneous token-passing bus network under heavy traffic is analyzed in [9]. The model presented in this paper predicts the throughput of FDDI priority scheme under both normal and heavy load conditions.

The basic idea of timed-token protocol was presented by Grow [10] and then studied by Ulm [11] considering only two priority classes. Simulation results are presented in [12] for a homogeneous FDDI network supporting packetized voice. A performance comparison of FDDI network and the token-ring network with a reservation scheme is given in [13]. Simulation models are utilized in [14] to evaluate the performance of several FDDI networks under normal and faulty operating conditions. A simulation model has been used in [15] to analyze and verify synchronous and asynchronous recovery times of FDDI networks.

An analytical model is proposed in [16] to predict the throughput of a heterogeneous network. However, it allows only one priority class per station and does not consider synchronous traffic. It is important to consider synchronous traffic to fully understand the behavior of FDDI, since synchronous traffic can affect the throughput and the access delay of asynchronous traffic. The model presented in this paper allows multiple classes of priorities at a given station. It does not, however, explicitly cover the case considered in [16]. The analytical model proposed in [17] can be used to calculate the throughput of individual priority classes of a homogeneous FDDI network with the network load at a station asymmetrically distributed among individual priority classes. For the model derived in this paper, the FDDI network need not be homogeneous. It allows multiple priority classes and asymmetrical distribution of load among classes.

A method is presented in [18] to properly select timed-token protocol parameters, considering the tradeoff between throughput of asynchronous messages and access delay of synchronous messages. However, only one asynchronous class is considered. The model presented in this paper may be used for parameter selection of an FDDI network supporting real-time traffic and synchronous traffic with multiple classes of priorities. A mean-value analysis for throughput and delay is presented in [19].

In this paper, we evaluate the performance of a class of heterogeneous FDDI token-ring networks, in the presence of synchronous traffic and asynchronous traffic. It extends the models presented in [16] and [17] as those models cannot predict the performance of the larger class of networks considered in this paper. Voice signals are used to represent synchronous traffic, but any other type of synchronous traffic may be used instead. The model predicts the throughput of FDDI priority scheme at low, moderate and heavy network load conditions. The model assumes that the load distribution among individual priority classes at a station is known. All equations related to the throughput charac-
teristics of the network are expressed strictly in terms of network parameters.

The FDDI protocol is described in Section II. Performance of FDDI protocol is evaluated in Section III. FDDI protocol parameters are calculated to meet specific traffic requirements in Section IV. The analytical results are verified using simulations in Section V. Mean access delay for different classes and the token-cycle time distribution for the priority scheme obtained from simulations are presented. The limitations of this work are outlined in Section VI.

II. FDDI MAC PROTOCOL
FDDI is a token-ring network [1] in which any station may capture the token by removing it from the ring. After the removal of the token, the station may begin to transmit information frames. When the information transmission is completed, the station immediately issues a new token.

A set of timers are used to limit the length of time a station may transmit messages before passing the token to the next station and the duration of information transmission for each class within a station. Each station maintains two hardware timers, Target-Token-Rotation-Time (TRT) and Token-Holding-Time (THT). TRT at station j is used to measure the time a token takes to circulate around the ring starting from station j. During ring initialization, all the stations negotiate a value for the parameter called the Target-Token-Rotation-Time (TRT). At the end of the negotiation period, the smallest TRT requested becomes the operative TRT for the network and the T_Opr in each station is set to this value. If a station captures the token before its TRT reaches the value of TRT, it is an "early" token. If it captures the token after TRT has exceeded the value of TRT, then it is a "late" token. When station j captures an "early" token, THT is reset and restarted immediately. Before resetting TRT, its current value is assigned to THT. THT becomes active only during asynchronous message transmission at station j. When the token is passed to the next station, and it becomes inactive while TRT continues to run until the token arrives at station j again. When station j captures a "late" token, TRT is not reset, but is allowed to continue in order to accumulate the "lateness" of the network.

An "early" token may be used to transmit both synchronous and asynchronous traffic while a "late" token may only be used for synchronous traffic. We use the term High-Priority-Token-Time (T_h) to denote the duration for which a station is allowed to initiate the transmission of the synchronous messages. If the token is "late", the station will issue a new token immediately after transmitting synchronous traffic. For an "early" token, the station is allowed to transmit asynchronous traffic provided the current value of THT is less than the value of TRT. The difference between the current value of THT and TRT determines the asynchronous bandwidth available to this station. However, a station cannot hold the token longer than TRT to initiate any message transmission. It has been stated that the protocol guarantees an average response time for synchronous traffic not greater than TRT, and a maximum response time not greater than twice TRT [20,21].

FDDI priority scheme also supports multiple classes of asynchronous traffic. Each priority class at a station has a threshold value T_Pri(i) (i=1,...,n). Class 1 is assumed to have the highest priority, and class n the lowest priority among asynchronous classes of traffic. Transmission of messages of an asynchronous class begins with class 1 and continues with the lower priority classes sequentially. The asynchronous traffic of class i may only be transmitted if the current value of THT is less than the class threshold value T_Pri(i). Since, the difference between the current value of THT and TRT reflects the asynchronous bandwidth, the maximum value that can be assigned to T_Pri(i) of class i is restricted to TRT. If there are no messages in class i, or if the THT has exceeded the T_Pri(i), then the next lower class is served. A new token is issued when there are no more messages, or after the lowest priority class has been served according to the above scheme.

III. PERFORMANCE EVALUATION OF FDDI
In this section, we derive an analytical model for predicting the throughput of a class of heterogeneous FDDI networks. First we state assumptions and derive an analytical model for a network with synchronous traffic and three asynchronous classes of priorities. Then, we extend the model for an arbitrary number of classes.

A. Assumptions
The analytical model allows three types of stations to be connected to the ring: those that support synchronous traffic or "voice only" stations, those that support asynchronous traffic or "data only" stations, and those that support both synchronous traffic and asynchronous traffic or "voice and data" stations. N_s is the total number of stations that support asynchronous traffic and N_v is the total number of stations that support synchronous traffic. Traffic of a given class is assumed to be equally distributed among the stations that support that class of traffic. The asynchronous traffic load at a station is distributed among three classes of priorities. In the analysis presented, class 0 corresponds to synchronous traffic and classes 1, 2 and 3 correspond to asynchronous traffic. The asynchronous traffic may be asymmetrically distributed among these three classes of priorities, i.e., the asynchronous traffic at a station (when present) is divided among class 1, class 2 and class 3 in a ratio of l_1 : l_2 : l_3 with l_1 + l_2 + l_3 = 1.

The analysis assumes error free operation of the network. A station has to wait until it completely removes the token from the ring before transmitting messages. This so-called token-capture time is very small and is neglected in the analytical model, but it is taken into account in the simulation model. For asynchronous class i, it is assumed that the message length is exponentially distributed with a mean length of T_i and that the message arrivals follow a Poisson process. Voice conversations are used to model
synchronous traffic. All time values are expressed in bit times, the time to transmit a single bit. The terminology and the network parameters used in the model are summarized in Table 1.

The relative magnitudes of the class thresholds affect the performance of a timer-based priority scheme [2-6]. Assume that the class thresholds are such that,

\[ TTRT > T.Pri(1) > T.Pri(2) > T.Pri(3). \]  

This condition ensures that class 1 messages receive a higher bandwidth allocation than classes 2 and 3. Further, it is assumed that the individual priority threshold values differ at least by several packet lengths from each other. The FDDI standard, however, does not impose any restriction on the relative magnitudes of the priority thresholds. If the condition \( TTRT > T.Pri(3) > T.Pri(1) > T.Pri(2) \) were satisfied for example, then class 3 messages would receive a higher bandwidth allocation than classes 1 and 2. The model can be extended to cover these other cases.

### Table 1: Network Parameters and Terminology

| \( N \) | total number of stations |
| \( N_d \) | number of stations transmitting asynchronous traffic |
| \( N_v \) | number of stations transmitting synchronous traffic |
| \( TTRT \) | target token rotation time (bit times) |
| \( TRT \) | token rotation timer (bit times) |
| \( THT \) | token holding timer (bit times) |
| \( T.Pri(i) \) | priority threshold value for asynchronous class \( i \) (bit times) |
| \( T_s \) | high priority token time (bit times) |
| \( C \) | mean token cycle time (bit times) |
| \( C_0 \) | mean token circulation time (bit times) |
| \( P_v \) | packetization time of voice messages (bit times) |
| \( L_v \) | packet length of voice messages (bits) |
| \( C_v \) | number of voice traffic channels |
| \( T_i \) | mean message length of class \( i \) traffic (bits) |
| \( T_{oh} \) | number of overhead bits added to any packet (bits) |
| \( \alpha_i \) | utilization factor of class \( i \) |
| \( t_i \) | fraction of asynchronous load in class \( i \) |
| \( S_i \) | total throughput of class \( i \) |
| \( U \) | total network utilization |
| \( U_i^{(x)} \) | utilization of class \( i \) at point \( x \) |
| \( G \) | total network offered load |
| \( G_i \) | total offered load of class \( i \) |
| \( G^{(x)} \) | total offered load at point \( x \) |
| \( P_i^{(x)} \) | scaled utilization factor of class \( i \) at point \( x \) |
| \( E_i^{(x)} \) | effective threshold value of class \( i \) at point \( x \) |

### B. Model Description

Throughput of a network \( S \) is defined as the total number of data bits received at the destination per second, expressed as a fraction of the total network bandwidth. Utilization \( U \) refers to the fraction of time the network spends carrying data packets including header information. In an error free network, \( (1 - U) \) is the fraction of time spent on token transmission. The offered load \( G \) is defined as the total number of data bits generated by all active stations per unit time (expressed as a fraction of the bandwidth).

The throughput of class \( i \), \( S_i = \sum_j s_{ij} \) where \( s_{ij} \) is the throughput of class \( i \) at station \( j \) and \( N \) is the number of stations. \( U_i \) and \( G_i \) can be defined in a similar manner. As a data frame includes header information, checksum, etc., in addition to the data field, the throughput is always less than the utilization. The relation between throughput and utilization is given by

\[ S_i = \alpha_i \ U_i \quad \text{for } i = 0, 1, 2, 3 \]  

where

\[ \alpha_i = \frac{T_i}{T_i + T_{oh}}. \]

Token-cycle time \( C \) is defined as the time elapsed from the instant a station receives the token till the next instant the same station receives the token. This definition of token-cycle time used in the following analysis is different from \( TRT \). \( C \) is the mean token cycle time averaged over all the stations. \( C_0 \) is the no load token-cycle time defined as the sum of optical fiber path delay and station latencies. The network utilization \( U \) is given by [2]

\[ U = \frac{(\bar{C} - C_0)}{C}. \]

In this paper, we select the network parameters such that all synchronous traffic generated is transmitted with bounded delay at any load condition. Hence, \( S_0 = G_0 \). In our analysis, we assume a constant synchronous load, i.e., that \( G_0, S_0 \) and \( U_0 \) remain constant.

Now consider a network with synchronous traffic and asynchronous traffic. Under very low load conditions, all messages generated in all classes will be transmitted. Hence \( S_i = G_i \) for \( i = 0, 1, 2, 3 \). As asynchronous traffic intensity is gradually increased, the network will transmit messages of all four classes until the total network load exceeds a certain threshold value. Then the network will no longer be able to transmit all the traffic. Since \( T.Pri(1) > T.Pri(2) > T.Pri(3) \), the network will now transmit all of class 0, class 1 and class 2 messages, but will not transmit some of class 3 messages. The throughput for class 3 reaches a peak when all of class 3 messages in the network get barely transmitted over the transmission medium. Further increase in asynchronous traffic load will result in a decrease in class 3 throughput to accommodate traffic of classes 1 and 2. Out of the three types of stations considered, the class 3 traffic at the "voice and data" and "data only" stations reach this limit first with equal probability.
This is illustrated in Fig. 1 for “voice and data” and “data only” stations.

A station is allowed to complete transmission of a message even if the corresponding timer expires during the transmission. Since the message length of asynchronous traffic is exponentially distributed, the portion of the message remaining for transmission after \(THT\) exceeds the respective priority threshold value is also exponentially distributed with the same mean message length. Hence, as shown in Fig. 1, \(THT\) is equal to \(T\_Pri(3) + T_3\) after transmitting message of classes 1, 2 and 3. When all class 3 messages generated at “voice and data” or “data only” stations barely get transmitted (an “early” token arrival), \(\bar{C}\) is such that

\[
T\_Pri(3) + T_3 = \bar{C} + (X_a(1) + Y_a(1)) + (X_a(2) + Y_a(2)) + (X_a(3) + Y_a(3)), \tag{5}
\]

where \(X_a(i)\) is the mean number of data bits transmitted by the asynchronous class \(i\) of a station during a cycle and \(Y_a(i)\) is the number of overhead bits required to transmit \(X_a(i)\) data bits for \(i = 1, 2, 3\). Note that the message transmission duration of the synchronous class \(T_s\) at a “voice and data” station is not included in Equation (5) since \(THT\) is disabled during this period [1]. Therefore, Equation (5) is valid for both “voice and data” and “data only” stations. Since the offered load in classes 1, 2 and 3 are distributed in a ratio of \(X_a(1) : X_a(2) : X_a(3) = l_1 : l_2 : l_3\), from Equation (3)

\[
\alpha_1(X_a(1) + Y_a(1)) : \alpha_2(X_a(2) + Y_a(2)) : \alpha_3(X_a(3) + Y_a(3)) = l_1 : l_2 : l_3. \tag{6}
\]

Using Equations (5) and (6)

\[
T\_Pri(3) + T_3 = \bar{C} + \left(\frac{l_1}{\alpha_1} + \frac{l_2}{\alpha_2} + \frac{l_3}{\alpha_3}\right) \left(X_a(3) + Y_a(3)\right), \tag{7}
\]

The utilization of an asynchronous class \(i\) is given by

\[
U_i = \frac{(X_a(i) + Y_a(i) N_d)}{\bar{C}} \quad \text{for } i = 1, 2, 3. \tag{8}
\]

Using Equation (7) and Equation (8) with \(i = 3\), \(\bar{C}\) can be expressed as

\[
\bar{C} = \frac{E_3^{(a)}}{1 + P_3^{(a)} U_3}, \tag{9}
\]

where

\[
E_3^{(a)} = \left[ T\_Pri(3) + T_3 \right], \tag{10}
\]

\[
P_3^{(a)} = \frac{\alpha_3}{l_3} \left\{ \frac{1}{N_d} \sum_{i=1}^{3} \frac{l_i}{\alpha_i} \right\}. \tag{11}
\]

The total network utilization \(U\) is

\[
U = U_0 + U_1 + U_2 + U_3, \tag{12}
\]

where \(U\) is given by Equation (4). From Equations (4), (6) and (8),
The limiting case for which class 3 throughput becomes negligible at "voice and data" and "data only" stations ($T_s = 0$ for "data only" stations).

$$\bar{C} - C_0 = \frac{(l_1/\alpha_1) + (l_2/\alpha_2) + (l_3/\alpha_3)}{(l_3/\alpha_3)} U_3 + U_0 .$$  

Using Equation (9) for $\bar{C}$, we can solve Equation (13) for $U_3$ as

$$U_3^{(a)} = \frac{(E_3^{(a)}(1 - U_0) - C_0)}{(E_3^{(a)} N_d + C_0) P_3^{(a)}} .$$  

Equations (6), (8) and (14)-(16) respectively as

$$U_1^{(a)} = \frac{(E_3^{(a)}(1 - U_0) - C_0)}{(E_3^{(a)} N_d + C_0) P_1^{(a)}} ,$$

$$U_2^{(a)} = \frac{(E_3^{(a)}(1 - U_0) - C_0)}{(E_3^{(a)} N_d + C_0) P_2^{(a)}} ,$$

where

$$P_1^{(a)} = \frac{\alpha_1}{l_1} \left\{ \frac{1}{N_d} \sum_{i=1}^{3} \frac{l_i}{\alpha_i} \right\} ,$$

$$P_2^{(a)} = \frac{\alpha_2}{l_2} \left\{ \frac{1}{N_d} \sum_{i=1}^{3} \frac{l_i}{\alpha_i} \right\} .$$

Note that, $U_i^{(a)}$ denotes the utilization of class $i$ for $i = 1, 2, 3$ when all the arriving packets get transmitted. Hence, $G_i = S_i$ for $i = 1, 2, 3$. Therefore, from Equations (2), (3) and (14)-(16) the maximum total network load $G^{(a)}$,

$$G^{(a)} = \frac{(E_3^{(a)}(1 - U_0) - C_0)}{(E_3^{(a)} N_d + C_0)} \left\{ \sum_{i=1}^{3} \frac{1}{\alpha_i} \right\} + \alpha_0 U_0 .$$

Hence, $S_i = G_i$ for $i = 1, 2, 3$ in the range $0 \leq G \leq G^{(a)}$, where $G^{(a)}$ is given by Equation (19).

When the asynchronous traffic intensity is further increased, the throughput of class 3 will decline in order to accommodate the increasing loads of classes 1 and 2. When the network load exceeds a certain value, class 3 throughput becomes negligible. When either one of "voice and data" and "data only" stations cannot transmit any class 3 traffic, no class 3 traffic ever gets transmitted in the network. The condition that $T_{Pri(3)}$ is at least several messages lengths less than $T_{Pri(2)}$ and $T_{Pri(1)}$ ensures that the throughput of class 3 is zero. As shown in Fig. 2, this situation arises when $THT$ is at least equal to $T_{Pri(3)}$ after transmitting class 1 and class 2 traffic. For this case, $\bar{C}$ is given by,

$$T_{Pri(3)} = \bar{C} + (X_s(1) + Y_s(1)) + (X_s(2) + Y_s(2)) ,$$

where $X_s(i)$ is the mean number of data bits transmitted by the asynchronous class $i$ during a cycle and $Y_s(i)$ is the number of overhead bits. Therefore,

$$\alpha_1 (X_s(1) + Y_s(1)) : \alpha_2 (X_s(2) + Y_s(2)) = l_1 : l_2 .$$
and the utilization of class 1 and class 2 for this case is given by,

$$U_i = \frac{(X_b(i) + Y_b(i)) N_d}{C} \quad \text{for} \quad i = 1, 2. \quad (22)$$

Solving Equations (20), (21) and (22), we get following expressions for $C$:

$$C = \frac{E_3^{(b)}}{\left(1 + \frac{1}{N_d} \sum_{i=1}^{2} \frac{l_i}{\alpha_i} \right) U_2}, \quad (23)$$

where

$$E_3^{(b)} = T_{-\text{Pri}}(3), \quad (24)$$

$$p_2^{(b)} = \frac{\alpha_2}{l_2} \left\{ \frac{1}{N_d} \sum_{i=1}^{2} \frac{l_i}{\alpha_i} \right\}, \quad (25)$$

The total network utilization $U$ can be written as

$$U = U_0 + U_1 + U_2. \quad (26)$$

$U_1$ and $U_2$ for this case can be evaluated using Equations (4), (22), (23) and (26) as follows:

$$U_1^{(b)} = \frac{(E_3^{(b)} (1 - U_0) - C_0)}{(E_3^{(b)} N_d + C_0) p_1^{(b)}}, \quad (27)$$

$$U_2^{(b)} = \frac{(E_3^{(b)} (1 - U_0) - C_0)}{(E_3^{(b)} N_d + C_0) p_2^{(b)}}, \quad (28)$$

where

$$p_1^{(b)} = \frac{\alpha_1}{l_1} \left\{ \frac{1}{N_d} \sum_{i=1}^{2} \frac{l_i}{\alpha_i} \right\}. \quad (29)$$

The throughput of individual classes can be found from Equation (2). Since all class 1 and 2 messages generated get transmitted, the offered load of class 1 and 2 is equal to respective class throughput: $G_i = S_i$ for $i = 1, 2$ and $S_3 = 0$. Therefore, the total network load $G^{(b)}$ can be written as

$$G^{(b)} = \frac{(E_3^{(b)} (1 - U_0) - C_0)}{(E_3^{(b)} N_d + C_0) \left\{ \frac{N_d}{\sum_{i=1}^{2} \frac{l_i}{\alpha_i}} \right\} + \alpha_0 U_0}. \quad (30)$$

Hence, $S_i = G_i$ for $i = 1, 2$ in the range $G^{(b)} \leq G \leq G^{(b)}$. $S_3$ in this range is found using linear interpolation of values at $G^{(a)}$ and $G^{(b)}$.

A similar analysis gives the following results for the case when the throughput of class 2 reaches its maximum.

$$\tilde{C} = \frac{E_2^{(c)}}{(1 + p_2^{(c)} U_2)}, \quad (31)$$

and the utilization of class 1 and class 2 for this case is given by,

$$U_1^{(c)} = \frac{(E_2^{(c)} (1 - U_0) - C_0)}{(E_2^{(c)} N_d + C_0) p_1^{(c)}}, \quad (32)$$

$$U_2^{(c)} = \frac{(E_2^{(c)} (1 - U_0) - C_0)}{(E_2^{(c)} N_d + C_0) p_2^{(c)}}, \quad (33)$$

where

$$E_2^{(c)} = \left[ T_{-\text{Pri}}(2) + T_2 \right], \quad (34)$$

$$p_1^{(c)} = \frac{\alpha_1}{l_1} \left\{ \frac{1}{N_d} \sum_{i=1}^{2} \frac{l_i}{\alpha_i} \right\}, \quad (35)$$

$$p_2^{(c)} = \frac{\alpha_2}{l_2} \left\{ \frac{1}{N_d} \sum_{i=1}^{2} \frac{l_i}{\alpha_i} \right\}. \quad (36)$$

Then by a similar argument, the throughputs are given by

$$G_i = S_i \quad \text{for} \quad i = 1, 2 \quad \text{and} \quad S_3 = 0. \quad \text{Therefore, the total network load} \quad G^{(c)} \quad \text{for this case can be written as}$$

$$G^{(c)} = \frac{(E_2^{(c)} (1 - U_0) - C_0)}{(E_2^{(c)} N_d + C_0) \left\{ \frac{N_d}{\sum_{i=1}^{2} \frac{l_i}{\alpha_i}} \right\} + \alpha_0 U_0}. \quad (37)$$

Hence, $S_i = G_i$ for $i = 1, 2$ and $S_3 = 0$ in the range $G^{(c)} \leq G \leq G^{(c)}$.

The condition that $T_{-\text{Pri}}(2)$ is at least several message lengths less than $T_{-\text{Pri}}(1)$ ensures that the throughput of class 2 is zero when the asynchronous traffic load is further increased. Then $\tilde{C}$ and the utilization of class 1 can shown to be

$$\tilde{C} = \frac{E_2^{(d)}}{(1 + p_2^{(d)} U_1)}, \quad (38)$$

$$U_1^{(d)} = \frac{(E_2^{(d)} (1 - U_0) - C_0)}{(E_2^{(d)} N_d + C_0) p_1^{(d)}}, \quad (39)$$

where

$$E_2^{(d)} = T_{-\text{Pri}}(2), \quad (40)$$

$$p_1^{(d)} = \frac{\alpha_1}{l_1} \left\{ \frac{1}{N_d} \right\}. \quad (41)$$

Then, $G_i = S_i$ and $S_2 = S_3 = 0$. Therefore, the total network offered load $G^{(d)}$ can be written as

$$G^{(d)} = \frac{(E_2^{(d)} (1 - U_0) - C_0)}{(E_2^{(d)} N_d + C_0) \left\{ \frac{N_d}{\sum_{i=1}^{2} \frac{l_i}{\alpha_i}} \right\} + \alpha_0 U_0}. \quad (42)$$

Hence, $S_1 = G_1$ and $S_3 = 0$ in the range $G^{(c)} \leq G \leq G^{(d)}$. $S_2$ in this range is found using linear interpolation of values at $G^{(c)}$ and $G^{(d)}$.

A similar analysis can be carried out for the case where the throughput of class 1 reaches maximum as a result of increased load. $\tilde{C}$ and the utilization of class 1 for this case is

$$\tilde{C} = \frac{E_1^{(e)}}{(1 + p_1^{(e)} U_1)}, \quad (43)$$

$$U_1^{(e)} = \frac{(E_1^{(e)} (1 - U_0) - C_0)}{(E_1^{(e)} N_d + C_0) p_1^{(e)}}, \quad (44)$$

where
Table 2: Offered Load Versus Throughput Characteristics

<table>
<thead>
<tr>
<th>Region</th>
<th>( S_1 )</th>
<th>( S_2 )</th>
<th>( S_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: ( 0 \leq G \leq G^{(a)} )</td>
<td>( G_1 )</td>
<td>( G_2 )</td>
<td>( \left{ \frac{\alpha_s U^{(a)}}{G^{(a)} - G_3} \right} (G^{(b)} - G_3) )</td>
</tr>
<tr>
<td>2: ( G^{(a)} \leq G \leq G^{(b)} )</td>
<td>( G_1 )</td>
<td>( G_2 )</td>
<td>0</td>
</tr>
<tr>
<td>3: ( G^{(b)} \leq G \leq G^{(c)} )</td>
<td>( G_1 )</td>
<td>( \left{ \frac{\alpha_s U^{(c)}}{G^{(c)} - G_2} \right} (G^{(c)} - G_2) )</td>
<td>0</td>
</tr>
<tr>
<td>4: ( G^{(c)} \leq G \leq G^{(d)} )</td>
<td>( G_1 )</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5: ( G^{(d)} \leq G \leq G^{(e)} )</td>
<td>( \alpha_1 U^{(e)} )</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Hence, the utilization of class \( i \) for \( 1 \leq i \leq r \) is

\[
U_i^{(l)} = \frac{\left( E_i^{(l)} \right) \left( 1 - U_0 \right) - C_0}{\left( E_i^{(l)} \right) \cdot N_d + C_0 \cdot p_i^{(l)}},
\]

where

\[
p_i^{(l)} = \frac{\alpha_i}{l_i} \left\{ \frac{1}{N_d} \sum_{i=1}^{r} \frac{l_i}{\alpha_i} \right\}.
\]

When the offered load is further incremented, the throughput of class \( r \) becomes negligible at an offered load \( G^{(m)} \) given by

\[
G^{(m)} = \frac{\left( E_r^{(m)} \right) \left( 1 - U_0 \right) - C_0}{\left( E_r^{(m)} \right) \cdot N_d + C_0} \left\{ \frac{N_d}{\sum_{i=1}^{r} \frac{1}{\alpha_i}} \right\} + \alpha_0 U_0,
\]

where

\[
E_r^{(m)} = T_{Pri}(r).
\]

The corresponding value for class \( r+1 \), \( G^{(k)} \) can be derived similarly.

The throughput characteristics of classes \( r+1, r \) and \( r-1 \) is shown in Fig. 3. The throughput versus offered load characteristics for this region can be expressed using the following equations.

C. Generalization of Model

The above results can easily be extended to cover FDDI networks with \( n \) classes of priorities. Assume that the priority threshold values are such that

\[ TTRT > T_{Pri}(1) > T_{Pri}(2) > \ldots \]

\[ \quad > T_{Pri}(n) \] (48)

Further, the asynchronous traffic at a station is assumed to be divided among \( n \) classes to a ratio of \( l_1 : l_2 : \ldots : l_n \) and \( l_1 + l_2 + \ldots + l_n = 1 \). Consider the scenario at "voice and data" and "data only" stations when all of the class 1 (\( 1 < r < n \)) messages generated barely get transmitted. The total offered load for this case is given by

\[
G_i^{(l)} = \frac{\left( E_i^{(l)} \right) \left( 1 - U_0 \right) - C_0}{\left( E_i^{(l)} \right) \cdot N_d + C_0} \left\{ \frac{N_d}{\sum_{i=1}^{r} \frac{1}{\alpha_i}} \right\} + \alpha_0 U_0.
\]

where

\[
E_i^{(l)} = [T_{Pri}(r) + T_r].
\]
IEEE TRANSACTIONS ON COMMUNICATIONS, VOL. 42, NO. 2/3/4, FEBRUARY/MARCH/APRIL 1994

Fig. 3 Throughput of class i vs. total offered load characteristics (i = r + 1, r, r - 1) in the regions R and R'.

The overall throughput characteristics can be obtained by Equations (55) and (56) for r = 1, ..., n. Note that in Equation (53), summation \( \sum_{r=1}^{n} \frac{1}{T_r} \) is zero for \( r = 1 \). Hence, the value of \( G^{(m)} \) for \( r = 1 \) becomes infinity. This is indeed valid as an increase in load after class 1 throughput has reached its maximum value causes throughput to stay constant at the value given by Equation (51).

IV. PARAMETER SELECTION

The bandwidth requirement for a single voice conversation is a very small fraction of the bandwidth of FDDI. Therefore, in order to provide a substantial voice load to the network, more than one voice channel per station is allowed in the model. This way, when the station captures the token, it is allowed to transmit one voice packet from each channel. Let the number of voice channels per station be \( C_v \). Assume that voice packets of fixed length \( L_v \) are generated at fixed time intervals of \( P_v \) in all voice channels per station after sampling, quantization and possible text compression. The total voice load of the network \( G_0 \) is given by

\[
G_0 = \left\{ \frac{C_v L_v}{P_v} \right\} N_v \quad (57)
\]

The high-priority-token-time \( T_s \) is thus given by

\[
T_s = C_v \left( \frac{L_v}{P_v} + T_{oh} \right) \quad (58)
\]

The mean of time required to transmit voice packets during a cycle, \( T_{sync} \), and the mean of time available for asynchronous traffic during a cycle, \( T_{async} \), are given by

\[
T_{sync} = N_v T_s \quad (59)
\]

\[
T_{async} = T_{TRT} - T_{sync} - C_0 \quad (60)
\]

Table 3: Network Parameter Values used for Performance Analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_v )</td>
<td>5.0 \times 10^9 (bit times)</td>
</tr>
<tr>
<td>( L_v )</td>
<td>3200 (bits)</td>
</tr>
<tr>
<td>( T_i )</td>
<td>2048 (bits) for i=1,2,3</td>
</tr>
<tr>
<td>( T_{oh} )</td>
<td>160 (bit times)</td>
</tr>
<tr>
<td>( T_{TRT} )</td>
<td>2.5 \times 10^6 (bit times)</td>
</tr>
<tr>
<td>( T_{Pri}(1) )</td>
<td>2.3 \times 10^6 (bit times)</td>
</tr>
<tr>
<td>( T_{Pri}(2) )</td>
<td>1.6 \times 10^6 (bit times)</td>
</tr>
<tr>
<td>( T_{Pri}(3) )</td>
<td>1.0 \times 10^6 (bit times)</td>
</tr>
<tr>
<td>( T_i )</td>
<td>3360 \times C_v (bit times)</td>
</tr>
</tbody>
</table>

A mean access delay of \( T_{TRT} \) for synchronous traffic at a station [20]. Therefore, if \( T_{TRT} \) is \( \frac{1}{2} P_v \), then synchronous traffic will have the above bounded delays.

The values for \( T_{Pri}(i) \) for \( i = 1,2,3 \) depend on the particular application, but cannot exceed the value of \( T_{TRT} \). The largest value assigned for any \( T_{Pri}(i) \) has a significant effect on the ‘asynchronous over-run’ problem [20]. The parameters have been selected such that the synchronous traffic transmission is strictly limited to the duration of \( T_s \) after a station captures the token. Once a station begins to transmit a message from an asynchronous priority class \( i \), it is allowed to complete the transmission of this message even when the \( THT \) exceeds the respective class threshold \( T_{Pri}(i) \). Although we do not provide a parameter tuning procedure for \( T_{Pri}(i) \) as given in [12,18], our analytical model can be used to find a set of suitable values for \( T_{Pri}(i) \) to match the desired characteristics for the given application.

V. RESULTS

In this section, we present a comparison of the results obtained from the analytical model described in Section III with the results obtained via a computer simulation model. The offered load \( G \) and throughput \( S \) are normalized with respect to the network bandwidth.

A. Parameter Values Used for the Model Verification

The timer values and parameters for both models are calculated according to the description given in Section IV. The asynchronous load at a station is arbitrarily assumed to be distributed among the classes 1, 2 and 3 according to the ratio 3:2:1. The data messages follow Poisson arrivals, and the length is assumed to be exponentially distributed with a mean of 2048 bits. Length of voice packets is 3200 bits, and are generated at 50 ms time intervals in each channel. All stations supporting voice traffic are assumed to have identical statistics for voice message generation. 160 overhead bits are added to each packet before transmission. The token-capture time is neglected in the analytical model, but is taken into account in the simulation model. A summary of all the values used are given in Table 3.

From Equation (57), it can be shown that for \( N_v = 79 \) voice stations, 2 voice channels per station are required to generate a \( G_0 = 10\% \) voice load for the network. When
it is necessary to raise the voice load, the number of voice channels allowed at a station is increased. A total of 200 stations are considered, with 50 "voice only" stations, 29 "voice and data" stations, and the rest being "data only" stations. The stations are assumed to be evenly spaced in a 20 km fiber path with a propagation delay of 5.09 μs/km and a station latency of 0.6 μs/station. Hence, \( C_0 = 0.22 \text{ms} \).

B. Performance Results of the Model

The offered load versus throughput characteristics for a 10% voice load is illustrated in Fig. 4. The network is able to transmit all class 0 messages even when the network is overloaded. Fig. 5 and Fig. 6 illustrate the offered load versus throughput characteristics for voice loads of 30% and 50% respectively. Note that when the voice load is raised from 10% to 50%, there is a considerable reduction in the throughput of all asynchronous classes since the bandwidth allocation for asynchronous traffic is reduced. The simulation results for all the cases confirm the results provided by the analytical model. At high loads, only classes 0 and 1 sustain a non-zero throughput due to the restriction imposed by Equation (1).

![Fig. 4 Throughput of class i vs. total offered load (i= 0, 1, 2, 3) for 10% voice load.](image1)

![Fig. 5 Throughput of class i vs. total offered load (i= 0, 1, 2, 3) for 30% voice load.](image2)

For a 30% voice load, the mean access delay (or the access delay) of various classes is shown in Fig. 7. The delay evaluated in this case is the queuing delay, defined as the time interval between the instant a message is assigned to a queue to the time its transmission is started. The mean access delay is obtained for class \( i \) by averaging the sum of all queuing delay samples over the total number of messages transmitted in class \( i \) during the time interval considered. Note that the access delay of class 0 increases from 5 ms to 18 ms when the total network load is increased from 30% to 200%. The access delay of class 0 at higher loads is 12 and 19 ms for 10% and 50% voice loads respectively. These results show the capability of the FDDI network to deliver its synchronous messages with a predetermined access delay. Due to the assumption of infinite buffer capacity, the access delay of asynchronous classes 1, 2 and 3 become infinitely large once the offered load exceeds the maximum throughput of the respective class.

![Fig. 6 Throughput of class i vs. total offered load (i= 0, 1, 2, 3) for 50% voice load.](image3)

Our results show that FDDI network can support 158 simultaneous voice conversations when the voice load is a mere 10%. This can be raised up to 790 simultaneous conversations when the voice load is 50%. The throughput of asynchronous classes is drastically reduced with the increasing voice load. From mean access delay characteristics, it can also be concluded that no voice packets are lost or discarded due to excessive delays.

Token-cycle time plays an important role in the priority scheme. The normalized token-cycle time distribution for FDDI network with 30% voice load is given in Fig. 8. The token-cycle time distribution is obtained for the entire network by averaging all the samples over the total number of
samples collected at every station.

Fig. 7 Mean access delay of class i vs. total offered load (i= 0, 1, 2, 3) for 30% voice load.

Fig. 8 Normalized Token-Cycle Time Distribution for 30% Voice Load.

The vertical axis is the normalized frequency of occurrence of the token-cycle time in milliseconds. When the network traffic load is low, most of the time the token rotates freely and hence the token-cycle time is equal to $C_0$ more frequently than any other value. The peaks A and B in Fig. 8 illustrate this scenario. When the throughput of class 3 is maximum, token-cycle time records the value of $T_{Pri}(3)$ more frequently than any other value. The token must rotate fast enough to avoid any backlogging of class 3 messages. The peak C indicated on the curve corresponding to 92% offered load, for example, illustrates this scenario. Similar argument can be presented for classes 1 and 2. The peaks D and E on 105% and 160% load curves illustrate the token-cycle time values when the throughputs of classes 2 and 1 have reached their respective peaks. These peaks D and E occur when the token-cycle time is in the vicinity of $T_{Pri}(2)$ and $T_{Pri}(1)$ respectively. Note that the token-cycle time for this particular network has recorded values exceeding $TT_RT$ (=25ms), i.e., the token has arrived “late” on these instances. The token-cycle time distribution shown in Fig. 8 characterizes the behavior of timer-based priority schemes similar to FDDI.

The model results are verified using the results obtained from an event-driven simulator written in FORTRAN. The main program utilizes several subroutine calls for message generation, message transmission, token propagation, statistical data collection and to update the clocks and the event list.

The initial statistical data of the network is discarded and the data used for model verification are collected after the network has reached steady state. For each parameter/variable of interest, four sample values are calculated in four consecutive equal time intervals. The sample mean is used as the estimate of the parameter. The popular 95% confidence interval is calculated using the $t$-distribution and is given by

$$
(\bar{X} - t(n-1)\sigma_\bar{X}/\sqrt{n}, \bar{X} + t(n-1)\sigma_\bar{X}/\sqrt{n})
$$

where $n$ is the number of samples and $\bar{X}$ and $\sigma_\bar{X}$ are the
mean and the standard deviation of the samples respectively. For $n = 4$, from t-distribution tables, $t(3) = 3.182$. In Fig. 7, the class 2 messages at 92% total load have the worst case confidence interval at $(13.59, 16.38)$ ms with a sample mean of 14.98 ms. Hence, the variation of sample mean for mean access delay is less than or equal to ±1.4 ms. More details about the simulator can be found in [23].

VI. CONCLUSIONS

Performance of a class of heterogeneous FDDI token-passing ring networks has been evaluated in the presence of both synchronous and asynchronous traffic. The results show that the model presented can accurately predict the throughput characteristics at low, moderate and high network load conditions. A scheme has been proposed to calculate the network parameters to meet the desired performance requirements. Since the computer simulation models consume a large amount of CPU time, the analytical model provides an efficient way to calculate the throughputs of all the classes with high accuracy. The equations governing the throughput characteristics are strictly functions of network parameters such as $T_{Pri}(i)$s, $C_0$, $\alpha_i$, ratio of load distribution at a station, etc. The use of this model as an inverse model is not straightforward, i.e., the model cannot be used to calculate the class threshold values $T_{Pri}(i)$ to obtain the desired throughputs. However, such applications of this model have been considered in [23].

The analytical model has been derived under the assumption that the network is heterogeneous. In order to use the model to evaluate the throughput of individual classes of priorities, the respective class threshold values ($T_{Pri}(i)$s) should differ at least by several message lengths. Otherwise, the model can be used to evaluate the combined throughput of two or more classes that have equal or approximately equal $T_{Pri}(i)$s. Further, this model is valid only if the condition $TTRT > T_{Pri}(1) > T_{Pri}(2) > T_{Pri}(3)$ is satisfied. However, this model can easily be extended to cover possibilities such as $TTRT > T_{Pri}(2) > T_{Pri}(1) > T_{Pri}(3)$ [23]. The results presented here can also be extended to predict the performance of IEEE 802.4 token-bus priority scheme under similar conditions. This model cannot be used to evaluate the throughput characteristics of FDDI networks when the number of priority classes per station is different among stations. The analytical model has been derived under the assumption that the behavior of class $i$ at all stations are identical. However, this assumption is not true when different stations have different number of priority classes in an FDDI network.

ACKNOWLEDGEMENT

The authors wish to thank James Faber, a Professor in Electrical Engineering at Colorado State University for his comments to this manuscript. The authors also thank the reviewers for providing many useful comments and suggestions.

REFERENCES

[22] M. J. Johnson, "Fairness of Channel Access For Non-Time-


Priya N. Werahera (S'85,M'94) received the B.Sc. degree in Electrical Engineering with honors, from the University of Moratuwa, Moratuwa, Sri Lanka in March 1982, and the M.S. degree in Electrical and Computer Engineering from the Oregon State University, Corvallis, Oregon in August 1986. He joined the Department of Electrical Engineering, University of Moratuwa as an Assistant Lecturer in March 1982. In September 1983, he joined the Telecommunication Department of Saudi Arabia as a Telecommunication Engineer. He is currently working as a Research Scientist for the University of Colorado Health Sciences Center. He has developed a three dimensional reconstruction algorithm to reconstruct prostate glands using cross sectional information. The results of his work will be used to study the biomorphometry of human prostatic cancer.

Priya Werahera is a candidate for the Ph.D. degree Electrical Engineering at the Colorado State University, Fort Collins, Colorado. He received Teaching Fellowship Award twice from the Electrical Engineering Department, Colorado State University in 1987 and 1989. His research interests include stochastic processes, queueing theory, communication networks, neural networks, three dimensional modeling and visualization of human organs, computer architecture, microprocessor based system development CMOS VLSI designs and design automation. He is a member of Eta Kappa Nu.

Dr. Anura Jayasumana (S'83,M'85,SM'90) is an Associate Professor in the Electrical Engineering Department and the Computer Science Department at the Colorado State University. He received his Ph.D. degree in electrical engineering and the M. S. degree in electrical engineering from the Michigan State University in 1984 and 1982. He received his B.Sc. degree in electronics and telecommunication engineering, with first class honors, from the University of Sri Lanka, Moratuwa, in 1978. Anura worked as an electrical engineer at the National Engineering Research and Development Center of Sri Lanka and as an assistant lecturer in the electronics and telecommunication engineering department at the University of Sri Lanka, from 1979 to 1980. His research interests include high-speed network protocols, performance evaluation, and VLSI design and testing. He has authored/coauthored over eighty publications in these areas and a book 'FDDI and FDDI-II: Architecture, Protocols and Performance,' Published by Artech House. He has served as a consultant to NCR Microelectronic Division since 1988.

Dr. Jayasumana is a member of Phi Kappa Phi, the IEEE Computer Society, and the IEEE Communications Society. He is an associate editor of the IEEE Network. He was the winner of the award for the best student in electrical engineering at the University of Sri Lanka, Moratuwa, in 1978, and the College of Engineering Outstanding Academic Achievement Award, Michigan State University, in 1982 and 1983. In January 1990, he received the Outstanding Faculty of the Year Award from the Mountain States Council of the American Electronics Association.