Throughput Performance of Asymmetrically loaded FDDI Networks

IZHAK RUBIN and JAMES C.-H. WU

Electrical Engineering Department, 58-115 ENGR IV
University of California, Los Angeles, CA 90024-1594

ABSTRACT We present exact throughput and stability analysis of the FDDI network system under general traffic configurations and a single asynchronous priority level of service. We obtain results relating each station’s throughput to the system throughput. By exploiting these relations, the maximum throughput of a symmetrically loaded system and the throughput of a (nonsymmetrically) heavily or partially-heavily loaded system are derived. We develop a procedure for conducting throughput and stability analysis. We identify stable stations and compute the throughput levels attained by heavily loaded stations.

1 Introduction

The Fiber Distributed Data Interface (FDDI) [1] is a 100 Mbps local area network using optical fiber as the transmission medium in a ring configuration. Media Access Control (MAC) of the FDDI network is based on a Timed-Token Protocol (TTP) which supports a synchronous traffic class with guaranteed bandwidth and response time and an asynchronous traffic class with dynamic bandwidth sharing of the system.

Throughput analyses of the timed-token protocol have been studied under various system configurations. Dykeman and Bux [2] and Jayasumana [3] conduct approximate throughput analyses of FDDI under multiple priority levels of the asynchronous service and general traffic loading conditions. By characterizing the sets of recurrent states representing the token dwell times at the stations, Pang and Tobagi [4] present exact throughput analysis of the IEEE 802.4 timed-token bus [5] under heavy load. See [6] and the references therein for packet delay analysis.

In this paper, we carry out exact throughput and stability analysis of a heterogeneous FDDI network operating under a single asynchronous priority level of service and general traffic configurations.

2 System Description

We consider an FDDI network supporting K stations (denoted as station-0, station-1, . . ., station-(K − 1)), each of which has a buffer of infinite capacity.

A token circulates around the ring and visits station-0, station-1, . . ., station-(K − 1) cyclically. The average token walk time (ring latency) for a complete circulation of the token around the ring (excluding the packet transmission time) is equal to R [msec]. A target token rotation time (TTRT) is selected during ring initialization. A (common) single priority level of asynchronous service is provided for the stations. The asynchronous token holding time, denoted as T [msec], is set equal to TTRT (in accordance with the FDDI standard).

Tokens arrive at the station's buffer according to a general stochastic traffic process. The offered load at station-k is denoted by p_k, k = 0, 1, . . ., K − 1. The throughput at station-k is represented as p_k, k = 0, 1, . . ., K − 1. We denote the overall offered load and the overall throughput of the system by p^O and p respectively. Note that p^O = \sum_{k=0}^{K-1} p_k and p = \sum_{k=0}^{K-1} p_k < 1.

The timed token protocol is described as follows. A token rotation timer (TTRT) is used at each station to measure the time between successive token arrivals at that station. Let C denote this token interarrival time at the station. A token holding timer measured from the initial time of the j-th token arrival at station-k to the time the token leaves the station when the station has no packets queued or the timer has expired. Note that, to simplify the analysis, we assume transmission overruns, which occur when the token holding timer expires during a packet transmission, are negligible. Such occurrences have minor effects on the system's throughput (See also [4]).

3 Station throughput vs. system throughput

In this section, we derive an expression relating the throughput level at a station to the system’s throughput.

Let C'_k denote the (cycle) time elapsed between the j-th and (j+1)-st token arrivals at station-k, j \geq 0; k = 0, 1, . . ., K − 1. We use C_k^j to denote the j-th token dwell time at station-k. The average token dwell time (j-th token visit) provided during this time is T < TTRT. The token departs from the station when the station has no packets queued or the timer has expired. Note that, to simplify the analysis, we assume transmission overruns, which occur when the token holding timer expires during a packet transmission, are negligible. Such occurrences have minor effects on the system's throughput (See also [4]).

The average token dwell time at station-k (which is assumed to co-exist and is calculated as an ensemble or time-average), denoted as G_k, is thus bounded by

G_k \leq T - C_k. \quad (2)

where G_k denotes the average (defined as the above) token interarrival (cycle) time at station-k.

According to Eq. (2), we distinguish two types of traffic loading conditions at a station. Station-k is said to be "heavily loaded" if G_k = T - C_k; station-k is said to be "stable" if G_k < T - C_k. Thus, a heavily loaded station will (almost surely) transmit for the maximum allowable holding time, upon its receipt of a token. A "weaker" definition of a heavily loaded station requires the station's buffer to be never empty during each token visit to the station (as also assumed in [4]). Note that the latter property of a heavily loaded station implies the definition stated above.

It is well known that for a cyclic service system, we have

G_k = C = \frac{R}{1 - p}, \quad k = 0, 1, . . ., K - 1. \quad (3)

This work was supported by U.S. Navy NRD Contract No. N66001-92-C-7003, by NSF Grant NCR-8914690, and by Pacific Bell and University of California MICRO Grant No. 92-134.
The normalized throughput at station-\( k \) is given by
\[ \rho_k = \frac{G_k}{C_k}. \]  

(4)

Combining Eqs.(2) (3) (4), we derive the following inequality
\[ \rho_k \leq \frac{T}{R}(1 - \rho) - 1, \]  

(5)

where the equality holds if and only if station-\( k \) is heavily loaded.

Note that \( \rho_k \leq \rho^* \) if station-\( k \) is heavily loaded; \( \rho_k = \rho^* \) if station-\( k \) is stable.

By applying the inequality \( \rho_k \leq \rho \) to Eq.(5), we obtain the throughput level at station-\( k \) to be upper-bounded by
\[ \rho_k \leq \frac{T - R}{T + R}, \quad k = 0, 1, \ldots, K - 1. \]  

(6)

4 Maximum throughput for a \( K \)-station system

We calculate in this section the network’s maximum attainable throughput capacity.

By rewriting Eq.(5) for each \( k \in \{0, 1, \ldots, K - 1\} \) to yield Eq.(7), we can formulate the following Throughput Maximization problem (TM) in selecting the station throughput levels \( \{\rho_k, k = 0, 1, \ldots, K - 1\} \) to yield an overall maximum network throughput:

Maximize \( \rho = \sum_{k=0}^{K-1} \rho_k \),
subject to
\[ \rho_k(T + R) + \sum_{i \neq k} \rho_i T \leq T - R, \quad k = 0, 1, \ldots, K - 1. \]  

(7)

It is easy to verify that for the convex set of constraints in Eq.(7), the system throughput is maximized when all the stations are heavily loaded, which occurs when the equality holds for each constraint in Eq.(7). Therefore, we obtain the maximum throughput of a \( K \)-station system to be:
\[ \hat{\rho} = \frac{T - R}{T + R}. \]  

(8)

The maximum level is attained when all stations exhibit the same throughput level. In this case, \( \hat{\rho} \) is equally distributed to all the \( K \) stations to yield the station throughput:
\[ \rho_k = \frac{T - R}{K(T + R)}, \quad k = 0, 1, \ldots, K - 1. \]  

(9)

5 Throughput of a partially heavily loaded system

In this section, we derive the maximum attainable throughput level and the per-station throughput levels when the network contains a heterogeneous set of stations experiencing different levels of throughput.

Let \( H \) denote the set of heavily loaded stations and let \( L \) denote the set of stable stations. Note that \( H \cup L = \{0, 1, \ldots, K - 1\} \). Let \( K_H \) denote the number of heavily loaded stations.

We define a partially heavily loaded system to be a system in which there is at least a single heavily loaded station, i.e., \( K_H \geq 1 \) (\( H \neq \emptyset \)). The throughput of a partially heavily loaded system is denoted by \( \rho_{p.h.l.} \). In the following, we calculate the throughput capacity and the attainable per-station throughput levels for such a network.

We formulate the following throughput maximization problem for a partially heavily loaded system:

Maximize \( \rho_{p.h.l.} = \sum_{k=0}^{K-1} \rho_k \),
subject to
\[ \rho_k(T + R) + \sum_{i \neq k} \rho_i T \leq T - R, \quad k \in H. \]  

(10)

\[ \rho_k = \rho^*, \quad k \in L. \]  

(11)

Since there is a unique solution to Eqs.(10)-(11) (as presented in the following), the above problem is equivalent to

Compute \( \rho_{p.h.l.} = \sum_{k \in H} \rho_k + \sum_{k \in L} \rho_k \),
by solving
\[ \rho_k(T + R) + \sum_{i \neq k} \rho_i T = T - R, \quad k \in H. \]  

(12)

The unique solution to Eq.(12) is
\[ \rho_k = \frac{T - R - \sum_{i \neq k} \rho_i T}{K_h T + R}, \quad k \in H. \]  

(13)

We thus obtain the throughput of a partially heavily loaded system to be
\[ \rho_{p.h.l.} = \sum_{k \in H} \rho_k + \sum_{k \in L} \rho_k = \frac{T - R + \frac{H}{K_h} \sum_{k \in L} \rho_k}{T + \frac{H}{K_h}}, \quad K_h \geq 1. \]  

(14)

Note that in a partially heavily loaded system, all the heavily loaded stations have the same throughput (\( \rho_k \) in Eq.(13)) regardless of the offered load.

For a partially heavily loaded system, we derive a lower bound for \( \rho_k, k \in H \), by considering the following two cases:

1) \( \sum_{i \neq k} \rho^*_i = 0 \). From Eq.(14), we note that \( \rho_k, k \in H \), is lower-bounded by \( \rho_k \geq \frac{T - R}{T + R} \), where the lower bound is achieved when \( K_h = K \). 2) \( \sum_{i \neq k} \rho^*_i > 0 \). Since \( \rho^*_k = \rho_k < \frac{T}{T + R} (1 - \rho) - 1 = \rho_k^* \), \( i \in L, k \in H \) (see Eq.(5)), we have \( \sum_{i \neq k} \rho^*_i < (K - K_h) \rho_k^*, k \in H \). Substituting this expression into Eq.(14), we obtain \( \rho_k > \frac{T - R}{K_h T + R} \), \( k \in H \). Thus we have
\[ \frac{T - R}{K_h T + R} \leq \rho_k \leq \frac{T - R}{T + R}, \quad k \in H. \]  

(15)

where the upper bound for \( \rho_k \) is given by Eq.(6).

A lower bound for \( \rho_{p.h.l.} \) can be obtained by setting \( K_h = 1 \) and \( \sum_{k \in L} \rho_k = 0 \) in Eq.(14). An upper bound for \( \rho_{p.h.l.} \) is given by \( \hat{\rho} \) in Eq.(8). Thus we have
\[ \frac{T - R}{T + R} \leq \rho_{p.h.l.} \leq \frac{T - R}{T + R}. \]  

(16)

Note that the upper bound for \( \rho_{p.h.l.}, k \in H \), and the lower bound for \( \rho_{p.h.l.} \), are attained when only a single station is active and heavily loaded, while the lower bound for \( \rho_k, k \in H \), and the upper bound for \( \rho_{p.h.l.} \), are approached when all the stations are heavily loaded.

From Eq.(15), we observe that if a station’s offered load, \( \rho^*_k \), is less than \( \frac{T}{T + R} \), then the station is stable. From Eq.(16), we can make the following observation. If the overall offered load of a system, \( \rho^* \), is less than \( \frac{T}{T + R} \), then there exists no heavily loaded station and the system is stable.
6 Stability analysis

In this section, we conduct stability analysis to determine whether a station is stable or heavily loaded.

We first sort the stations into J groups, denoted as \( S_i, i = 0, 1, \ldots, J - 1 \), in accordance with the offered load levels. Each station in group \( S_i \) has a distinctive offered load level, \( \rho^{(i)}, i = 0, 1, \ldots, J - 1 \), where \( \rho^{(i)} > \rho^{(j)}, i < j \). The number of stations in \( S_i \) is denoted by \( |S_i|, i = 0, 1, \ldots, J - 1 \).

We introduce the following supplementary parameters of traffic loading:

\[
\rho^{(m)}_K = \rho^{(m)} \sum_{i=0}^{m} |S_i| + \sum_{j=m+1}^{J-1} \rho^{(j)}|S_i|, \quad m = 0, 1, \ldots, J - 1. \tag{17}
\]

Note that \( \rho^{(i)} > \rho^{(j)} \), \( i < j \).

From Eqs. (5), (10), the following relations hold for stable stations and heavily loaded stations respectively:

\[
\rho_k R + \rho T < T - R, \quad k \in L, \tag{18}
\]

\[
\rho_k R + \rho T = T - R, \quad k \in H. \tag{19}
\]

Let \( \hat{\rho} \) denote the (unique) throughput level at the heavily loaded stations (given by Eq. (13)).

According to the definitions of \( \{\rho^{(i)}\} \) and \( \{\rho^{(m)}_K\} \), we have

\[
\rho^{(0)} R + \rho^{(0)}_K T > \rho^{(1)} R + \rho^{(1)}_K T > \ldots > \rho^{(J-1)} R + \rho^{(J-1)}_K T. \tag{20}
\]

By exploiting the relation in Eq. (20), we can determine the stability conditions at the stations as follows:

- For \( \rho^{(0)} R + \rho^{(0)}_K T < T - R \):

  From Eq. (20), we obtain the following set of inequalities:

  \[
  \rho^{(i)} R + \rho^{(i)}_K T < T - R, \quad i = 0, 1, \ldots, K - 1.
  \]

  Since the offered load at every station satisfies the condition for a stable station (see Eq. (18)), every station in the system is stable.

  \( H = \emptyset \), \( L = \{0, 1, \ldots, K - 1\} \).

- For \( \rho^{(m)} R + \rho^{(m)}_K T < T - R \leq \rho^{(m-1)} R + \rho^{(m-1)}_K T \), \( m \in \{1, 2, \ldots, J - 1\} \):

  We can derive the relations,

  \[
  \rho^{(m)}_K < \rho < \rho^{(m-1)}_K.
  \]

  We can thus classify the set of heavily loaded stations and the set of stable stations, respectively, noting that \( \rho^{(0)} > \rho^{(1)} > \ldots > \rho^{(m-1)} > \rho \), and \( \rho^{(J-1)} < \rho^{(J-2)} < \ldots < \rho^{(m)} < \rho \), to be given by

  \[
  H = \bigcup_{i=m}^{J-1} S_i, \quad L = \bigcup_{i=m}^{J-1} S_i.
  \]

- For \( \rho^{(J-1)} R + \rho^{(J-1)}_K T \geq T - R \):

  From the definition of \( \{\rho^{(i)}\} \) and the constraint for a heavily loaded station in Eq. (19), we have

  \[
  \rho^{(0)} > \rho^{(1)} > \ldots > \rho^{(J-1)} \geq \hat{\rho}.
  \]

  Since the offered load at each station exceeds the maximum per-station throughput level \( \hat{\rho} \), we conclude that all the stations in the system are heavily loaded.

  \( H = \{0, 1, \ldots, K - 1\} \), \( L = \emptyset \).

By rewriting \( \rho^{(m)} R + \rho^{(m)}_K T \geq T - R \), we obtain \( \rho^{(m)} \geq \frac{T}{R}(1 - \rho^{(m)}_K) - 1 \), which is used in checking the stability condition at a station in the next section.

7 Procedure for throughput and stability analysis

In this section, we present an efficient procedure to calculate the throughput levels and analyze the stability conditions of the stations in a non-symmetric system. It is based on the analysis described in the previous sections.

We first conduct stability analysis. We determine whether a station is stable by starting from the group of stations with the highest level of offered load (\( S_0 \)) and then descending to the group of stations with the second highest level of offered load (\( S_1 \)), etc., until the stability condition given by Eq. (5) is satisfied. If the system is partially heavily loaded, the throughput analysis expressed in Eqs. (13) and (14) is then used to determine the appropriate throughput levels for the heavily loaded stations. At the completion of this procedure, we obtain the system throughput \( \rho \), the throughput level at each station \( \rho_k \), the number of heavily loaded stations \( K_h \) and the disjoint sets of stable stations \( L \) and heavily loaded stations \( H \).

**Procedure 1 (Throughput and Stability Analysis)**

0. (initialize)

- determine \( \rho^{(i)}, i = 0, 1, \ldots, J - 1 \), by ranking the offered load:
  \[
  \{\rho^{(m)}\} \quad m = 0, 1, \ldots, J - 1;
  \]

- (start with the maximum offered load group) \( m = 0 \).

1. (check for heavily loaded stations)

   while \( \left( \rho^{(m)} \geq \frac{T}{R}(1 - \sum_{k=0}^{K-1} \rho_k) - 1 \right) \) do

   - \( m = m + 1 \).

   - if \( (m = J) \) then goto 3.

   else \( \rho_k \leftarrow \rho^{(m)} \), \( k \in S_i, i = 0, 1, \ldots, m - 1 \).

   goto 4.

2. (a stable system)

   if \( (m = 0) \) then \( K_h = 0, H = \emptyset \), \( L = \{0, 1, \ldots, K - 1\} \).

   goto 4.

3. (a partially heavily loaded system)

   - (number) \( K_h = \sum_{i=0}^{J-1} |S_i| \).

   - (two disjoint groups of stations)

     \[
     H = \bigcup_{k=0}^{m-1} S_k, \quad L = \bigcup_{k=m}^{J-1} S_k.
     \]

   - (throughput) \( \rho_k = \frac{T - R - \sum_{k=0}^{J-1} \rho^{(m)}_k T}{K_h T + R}, \quad k \in H. \)

4. (system throughput) \( \rho = \sum_{k=0}^{K-1} \rho_k \).

8 Application Examples

We consider an FDDI network with 100 stations (\( K = 100 \)) operating under general traffic conditions.

1. Network performance under various parameters
In Fig.1, we show the stability regions of the network in terms of the network offered load \( (\rho^o) \) by varying the target token rotation time \( (T) \) from \( R \) (ring latency) to \( 20 \times R \) (which corresponds to a maximum achievable network throughput of 0.9495).

We consider the following three stability regions of the network.

1) For the region of \( \rho^o < \frac{P_T}{R + R/K_h} \), the system is stable (no heavily loaded station).

2) For the region of \( \frac{P_T}{R + R/K_h} \leq \rho^o < \frac{P_T}{R} \), there is at least one heavily loaded station and thus the system is unstable (i.e., it is partially heavily loaded).

3) For the region of \( \rho^o \geq \frac{P_T}{R} \), the stability condition of the system depends on the loading situations of the stations. If the stations are symmetrically loaded, the system is stable for \( \rho^o < \frac{P_T}{R} \). If there is only a single active station contributing to the overall system traffic (i.e., the extreme case of nonsymmetric loading), the system is unstable for \( \rho^o \geq \frac{P_T}{R} \).

For other traffic configurations, the stability conditions of the network can be determined by using Procedure 1.

In Figs.2 and 3, we exhibit the effects of the overall traffic loading at the stable stations. \( \rho_s = \sum_k E_k P_k \) on the throughput performance of the heavily loaded stations and the system by using Eqs.(13) and (14), respectively. The number of heavily loaded stations \( (K_A) \) is varied from \( K_A = 1 \) to \( K_A = 20 \), while the number of stable stations \( (K - K_A) \) is varied, respectively, from 99 to 80. The target token rotation time is selected to be equal to \( 20 \times R \).

From Eq.(13), we note that the available system bandwidth is distributed uniformly to all of the heavily loaded stations. In Fig.2, we observe, as expected, that the throughput at a heavily loaded station \( (\rho_s, k \in H) \) is decreased when we increase the number of heavily loaded stations. We notice that \( \rho_s, k \in H \), is lower-bounded by \( \frac{P_T}{R + R/K_h} = 0.009495 \) given by Eq.(9). As the overall loading at the stable stations is increased, the available system bandwidth (expressed as the numerator of the right-hand side in Eq.(13)) is decreased; consequently, the throughput at a heavily loaded station is decreased.

We illustrate the applications of Eq.(14) in Fig.3 by making the following observations. As we increase the loading level at the stable stations \( (\rho_s) \) or the number of heavily loaded stations \( (K_A) \), we decrease the variations of the traffic loading levels between the stations and thus the system throughput \( (\rho_p, h, k) \) is increased. We note that \( \rho_p, h, k \) is upper-bounded by \( \frac{P_T}{R} = 0.9495 \).

2. Applications of Procedure 1

We demonstrate the applications of Procedure 1 in the following two examples. The target token rotation time is set equal to \( T = 20 \times R \), which corresponds to a maximum system throughput of 0.9495.

First, we consider a nonsymmetric FDDI network with four groups of stations, denoted as \( S_0, S_1, S_2, S_3 \), which are distinguished by equal traffic loading levels of stations within each group. The number of stations in groups \( S_0, S_1, S_2, S_3 \), respectively, are selected to be equal to 10, 10, 30, and 50. We use Procedure 1 to determine the throughput at each group of stations under various offered loading conditions (see Table 1).

The maximum system throughput is observed to be 0.9487 under these traffic configurations. When the system is partially heavily loaded, we observed that the heavily loaded stations (in groups \( S_0 \) and \( S_1 \)) have the same throughput level, regardless of the offered load. The throughput levels of the stable stations are insensitive to variations in system throughput levels.

We next analyze the throughput performance of a symmetric system with 100 stations (see Table 2) by employing Procedure 1. When the overall offered load of the system is below the maximum system throughput, 0.9495, the stations are stable; otherwise, all the stations are heavily loaded.

### References

1. [International standard, ISO 9314-2, “FDDI token ring media access control (MAC),” 1989.]


### Tables

#### Table 1: Throughput performance of the FDDI network with four groups of stations.

<table>
<thead>
<tr>
<th>#stations</th>
<th>overall</th>
<th>( S_0 )</th>
<th>( S_1 )</th>
<th>( S_2 )</th>
<th>( S_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>offered load</td>
<td>throughput</td>
<td>0.9487</td>
<td>0.025935</td>
<td>0.025935</td>
<td>0.011</td>
</tr>
<tr>
<td>offered load</td>
<td>throughput</td>
<td>0.95</td>
<td>0.026</td>
<td>0.026</td>
<td>0.011</td>
</tr>
<tr>
<td>offered load</td>
<td>throughput</td>
<td>0.95</td>
<td>0.026</td>
<td>0.026</td>
<td>0.011</td>
</tr>
<tr>
<td>offered load</td>
<td>throughput</td>
<td>0.94866</td>
<td>0.026866</td>
<td>0.025</td>
<td>0.011</td>
</tr>
<tr>
<td>offered load</td>
<td>throughput</td>
<td>0.76</td>
<td>0.021</td>
<td>0.012</td>
<td>0.011</td>
</tr>
<tr>
<td>offered load</td>
<td>throughput</td>
<td>0.76</td>
<td>0.021</td>
<td>0.012</td>
<td>0.011</td>
</tr>
</tbody>
</table>

#### Table 2: Throughput levels of the system and the station in a symmetric FDDI network.

| offered load | throughput | 0.9495 | 0.009495 |
| offered load | throughput | 0.9495 | 0.009495 |
| offered load | throughput | 0.9495 | 0.009495 |
| offered load | throughput | 0.9495 | 0.009495 |
| offered load | throughput | 0.9495 | 0.009495 |
| offered load | throughput | 0.5 | 0.005 |
| offered load | throughput | 0.5 | 0.005 |
Figure 1: Stability regions of the FDDI network with 100 stations: network offered load ($\rho_0$) vs. target token rotation time ($T$).

Figure 2: Throughput at a heavily loaded station ($\rho_k$, $k \in H$) vs. the number of heavily loaded stations ($K_h$) under various loading conditions at the stable stations ($\rho_L$).

Figure 3: Throughput levels of the partially heavily loaded system ($\rho_{p,h,l}$) vs. the number of heavily loaded stations ($K_h$) under various loading conditions at the stable stations ($\rho_L$).