AVERAGE WAITING TIME IN FDDI NETWORKS
WITH LOCAL- PRIORITIES

Gokhan Gercek
Lockheed Engineering and Sciences Company

Abstract: A method is introduced to compute the average queuing delay experienced by different priority group messages in an FDDI node. It is assumed that no FDDI MAC layer priorities are used. Instead, a priority structure is introduced to the messages at a higher protocol layer (e.g. network layer) locally. Such a method was planned to be used in Space Station Freedom FDDI network. Conservation of the average waiting time is used as the key concept in computing average queuing delays. It is shown that local priority assignments are feasable specially when the traffic distribution is asymmetric in the FDDI network.

I. INTRODUCTION

Analysis of priority based queuing systems have attracted first interests from mathematicians and operation research analysts. With the explosive growth of computing systems and computer networks it has found new application grounds and thus has become a research area in these fields as well. Priority queuing disciplines are more complex in nature than their non-priority counterparts due to multidimensional state variables involved in the description of the system.

Among the queuing disciplines that impose a static priority structure (exogenous), head-of-the-line (HOL) discipline is perhaps the most common and most intuitive one. HOL discipline was first studied by Cobham [1] who derived an expression for the average waiting times associated with different priority groups in the queue. Later, Kesten [2] derived a formula for calculating the Laplace transform of the waiting time in an HOL queue, which is usually very difficult to invert. Davis [3] introduced a method to evaluate the waiting time distributions associated with a two level priority queue. A way of calculating the probability density functions for a multilevel HOL queue under identical service time assumption is given in [4]. An important study concerning the moments of the waiting times in an HOL queue is performed by Miller [5].

In computer network performance analysis, the derivation of an average waiting time expression associated with a computer network is an elaborated task. Under priority access, the derivation becomes even more complex and in most cases intractable. As will be shown in the subsequent section, it is possible to find the waiting time for an FDDI network with local priorities since an analysis for the no-priority counterpart is readily available.

The underlying assumptions in the average message time derivation are:

1. Exhaustive service discipline: The FDDI station which is using the communication medium transmits all its messages starting with the highest priority ones. It should be noted that in practice, the token holding time used by MAC layer does not permit exhaustive transmission of messages. Therefore, exhaustive service assumption is an approximation. In implementation,

however, this issue can be circumvented by setting the token holding time as large as possible.

2. Local Priority Assignment: Message priority assignments have only local meanings. That is other FDDI stations are not aware of the message priorities in the buffers of a particular station.

Organization of the paper:

The subsequent sections are organized in the following way. In section II a method is introduced to transform the FDDI network into a classical priority queuing system. Then the solution for the resulting priority queuing system is given. In section III the method is applied to FDDI networks with symmetric and asymmetric traffic patterns. Section IV is the conclusion.

II. WAITING TIME IN FDDI NETWORKS WITH LOCAL PRIORITIES

A. Modeling of FDDI networks with Local Priorities.

As will be explained in the following paragraphs, the FDDI network considered in this paper can be analyzed in terms of standard priority queues. The results from HOL priority systems will constitute the framework of the subsequent discussions.

Average Waiting Time in HOL queues:
For an N level HOL queue, with level N as the highest and level 1 as the lowest priority level (Fig. 1), the average waiting time, W_p, associated with each level can be found according to the formula [1]:

$$W_p = \frac{W_R}{(1 - \sigma_p)(1 - \sigma_{p+1})} \quad , \quad p = 1, 2, \ldots \ldots N$$

(1)

where

$$\sigma_p = \sum_{k=p}^{N} \rho_k \quad and \quad \sigma_{N+1} = 0$$

W_R is the remaining service time of the entity found it the service and ρ_k is the traffic intensity into level k. Using renewal theory arguments [6], and assuming Poisson arrivals, W_R can be calculated as:

$$W_R = \sum_{j=1}^{N} \rho_j \frac{b_j^{(2)}}{b_j}$$

(2)

where $b_j^{(2)}$ and b_j are the second and first moments of the message length distributions respectively. As can be noticed from Eq. 1, the waiting time of a level k entity is not affected by the entities in levels k-1, k-2,...,1, except for the contribution of these levels to W_R. Furthermore assuming Poisson arrivals to all levels, the average waiting times satisfy the conservation law of the waiting times [7].
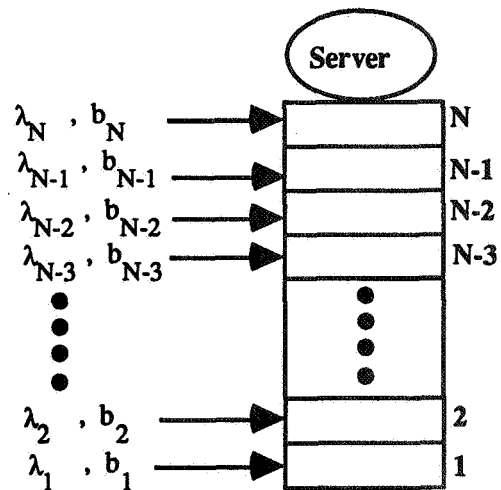

Figure 1. Head Of the Line (HOL) queue.

B. Vacation Time of the Server

Here we consider an FDDI network with M stations sharing the common fiber optic communication medium (the server). When a station "i" gains the access right to the transmission medium, it transmits all of its messages exhaustively and then passes the token to the next station according to FDDI protocol. At this point, from the viewpoint of the station "i", the server is considered to be in vacation. The vacation time of the server is a random variable whose distribution is generally unknown. The station "i" can reaccess to the medium at the completion of the vacation time (Fig. 2).
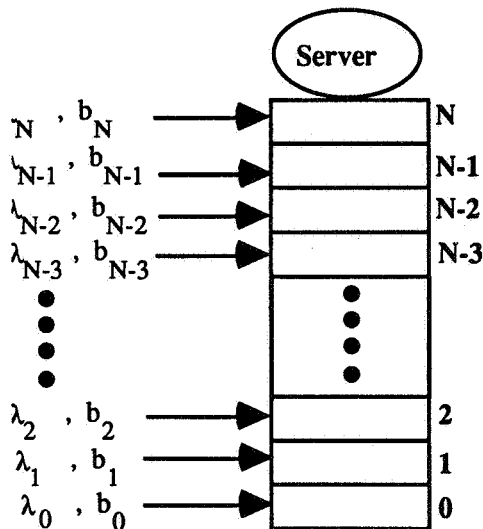

Figure 2. Representation of an M station local priority LAN in terms of an HOL queue.

C. Incorporating Vacation Time into Priority Queue Model

From the viewpoint of a station i, the vacation time of the server can be treated as just another cause (or a hypothetical message) which keeps the server busy. Moreover this message is processed always after the completion of the services of the other messages. Therefore if the service time of the

hypothetical message representing the vacation time is a random variable b_0, then the vacation time can be incorporated into the priority queue structure associated with the station of interest as an additional priority level (level 0). It should be noted that, in order to properly mimic vacation time, there must always be a message available to service in this priority level. When the server completes the service of all messages in N levels, if it can not find a message in the level 0, then the modeling will not be valid.

Fig. 2 shows modeling of a local area network in terms of an N+1 level HOL queue. The requirement that there must be a pool of entities in the priority level 0 can be satisfied by adjusting the Poisson arrival rate into this level so that the overall traffic rate approaches unity. The waiting times associated with the new N+1 level priority queue are given by:

$$W_p = \frac{\overline{W_R}}{(1 - \sigma_p)(1 - \sigma_{p+1})} \quad , \quad p = 0,1,2,\ldots\ldots N$$

(3)

where

$$\overline{W_R} = \rho_0 \frac{b_0^{(2)}}{b_0} + \sum_{j=1}^{N} \rho_j \frac{b_j^{(2)}}{b_j}$$

(4)

It should be noted that the terms $b_j^{(2)}$ and b_0, in Eq-4 are not known since the distribution of the vacation time is not available. Nevertheless W_R can be determined by using the results for a queue with N=1 (i.e. non-priority) which has a traffic intensity same as that of the N level queue. Now, let's assume an FDDI network which has the same access protocol as the assumed N priority level FDDI network. Also let us assume that the overall traffic intensity at a station 'k' are same for all stations in both cases. Remembering that the vacation time can be modeled as an additional level, we can write the following relation using the conservation law of the waiting time.

$$\rho_0 W_0 + \sum_{j=1}^{N} \rho_j W_j = \hat{\rho}_0 \hat{W}_0 + \hat{\rho}_1 \hat{W}_1 \qquad (5)$$

The variables marked with caret belong to N=1 case. Obviously W_0 is same as W_0, and ρ_0 and ρ_0 are equal. Therefore;

$$\sum_{j=1}^{N} \rho_j W_j = \hat{\rho}_1 \hat{W}_1 \qquad (6)$$

The term W_1 in Eq. 6 is the waiting time of the messages in a non-priority FDDI network and it should satisfy Eq. 3 for N=1. Assuming that ρ_1 and σ_1 are the same, we have;

$$\sum_{j=1}^{N} \lambda_j b_j = \hat{\lambda}_1 \hat{b}_1 \qquad (7)$$

and if further assume that W_R's are the same in both networks then:

$$\rho_0 \frac{b_0^{(2)}}{b_0} + \sum_{j=1}^{N} \rho_j \frac{b_j^{(2)}}{b_j} = \hat{\rho}_0 \frac{\hat{b}_0^{(2)}}{\hat{b}_0} + \hat{\rho}_1 \frac{\hat{b}_1^{(2)}}{\hat{b}_1} \qquad (8)$$

but since the priority level-0 arrival and service length distributions are the same;

$$\sum_{j=1}^{N} \rho_j \frac{b_j^{(2)}}{b_j} = + \hat{\rho}_1 \frac{\hat{b}_1^{(2)}}{\hat{b}_1} \qquad (9a)$$

or

$$\sum_{j=1}^{N} \lambda_j \frac{b_j^{(2)}}{2} = \hat{\lambda}_1 \frac{\hat{b}_1^{(2)}}{2} \qquad (9b)$$

If $(\rho_1 + \rho_0)$ is allowed to approach to unity, then the queue lengths in level-0 become instable and thus in these levels constant presence of entities will be assured.

Now we are in a position to summarize the method for finding waiting times in an FDDI station with N priority levels. It is assumed that the arrival and message length distributions are known for all M stations in the network and average waiting time expression is available for the non-priority version of the same network. The following three steps outline the approach.

1. Treating the station as a non-priority station and using the average waiting time is calculated.

$$\hat{b}_1 = \sum (\lambda_j \, b_j) / \hat{\lambda}_1$$

$$b_1^{(2)} = \sum (\lambda_j \, b_j^{(2)}) / \hat{\lambda}_1$$

2. Using Eq. 3 for N=1, W_R is calculated

$$\hat{W}_R = W_1 (1 - \hat{\rho}_1)$$

3. W_R substituted in Eq. 3 to determine W_p's for p=1,2,....N.

$$W_p = \frac{W_R}{(1 - \sigma_p)(1 - \sigma_{p+1})} \quad , \qquad p = 1,2,.......N \qquad (10)$$

III. COMPUTATION OF LOCAL PRIORITY FDDI NETWORK AVERAGE WAITING TIMES

In this section we apply the algorithm developed above to two types of traffic patterns: an FDDI network with a symmetric traffic pattern and an FDDI network with an asymmetric pattern.

A. FDDI network with symmetric traffic pattern:

The network is assumed to have a symmetric overall traffic pattern for all M stations. Stations may have different number of priority levels with the provision that the overall traffic intensity to all N levels is constant and same for all the stations. The average waiting time for the non-priority version of the polling protocols is given by [8]:

$$\hat{W}_1 = \frac{1}{2} \left(\frac{M\hat{\lambda}_1 \hat{b}_1^{(2)}}{(1-M\hat{\rho}_1)} + rC_r^2 + \frac{(1-\hat{\rho}_1)Mr}{(1-M\hat{\rho}_1)} \right)$$

(11)

where M is the number of stations, $\hat{\lambda}_1$ is the arrival rate into a station, $\hat{\rho}_1$ is the traffic intensity into a station, $\bar{r}$ is the average walking time of the token (time to transfer the access right from one station to the next one) and C is the coefficient of variation of the walking time.
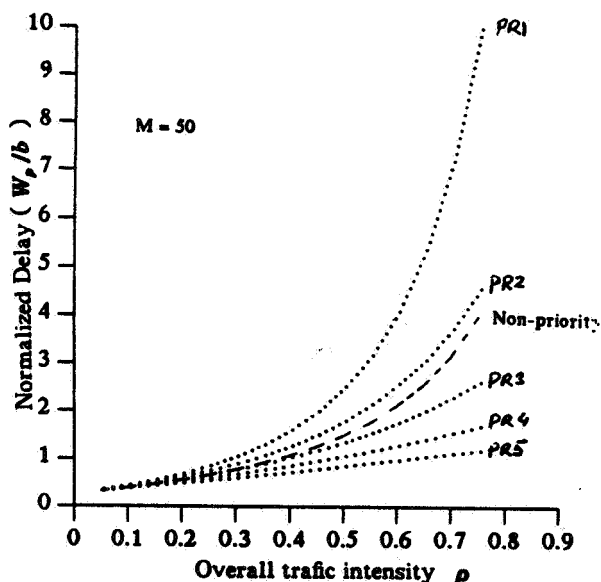

Figure 3. Local Priority FDDI protocol with symmetric arrivals ($\bar{r}$=0.01).

Substituting $\hat{W}_1$ from Eq. 11 in Eq. 10 yields the average waiting times associated with a FDDI protocol which

has N distinct priority levels. In Fig. 3 and Fig. 4 the results for a 5 priority level FDDI protocol are shown. The message lengths are assumed to be exponentially distributed and arrivals are assumed to be Poisson distributed for all levels with identical parameters.

The figures also give the average waiting times associated with a non-priority FDDI network. In both cases, r is assumed to be constant. As can be observed it is possible to achieve significant improvements in the average waiting time of the high priority entities at the expense of the low priority ones.

B. FDDI network Asymmetric Traffic Patterns:

In this case the FDDI network is assumed to have a heavily unbalanced arrival pattern. While a particular station generates all the traffic in the network, other stations idle and pass the token to the subsequent station. Other assumptions are same as the symmetric case. The average delay expression for this case is given by [8]:

$$\hat{W}_1 = \frac{1}{2} \left(\frac{\hat{\lambda}_1 \hat{b}_1^{(2)}}{(1-M\hat{\rho}_1)} + \bar{r}C_r^2 + M\bar{r} \right)$$

(12)

The definitions for $\bar{r}$, C and M are same as before. $\hat{\lambda}_1$ is the arrival rate into the station, $\hat{\rho}_1$ is the traffic intensity into the station, $\hat{b}_1^{(2)}$ is the second moment of the message length distribution. Proceeding in similar way as done for symmetric arrival pattern, the average waiting time associated with different priority levels can be determined. Fig. 5 shows the results for an asymmetric FDDI protocol under similar assumptions as the previous case.

In Fig. 4 an interesting (and counter intuitive) trend is observed for high priority messages (priority 5 and 4).

647

The delays experienced for those classes of users decrease with the traffic intensity. A qualitative explanation for this situation can be given as follows. At low traffic intensities, waiting time of a message is due to token circulation time. When ρ approaches to 0 all the entities experience an average of $M\bar{r}/2$ delay. At high traffic intensities token spends less time circulating freely and more time serving the station since with a high probability there will be some entities waiting in the buffer. And high priority entities enjoy the increased availability of the token by not waiting for the token to arrive. This effect becomes less visible as Mr decreases.
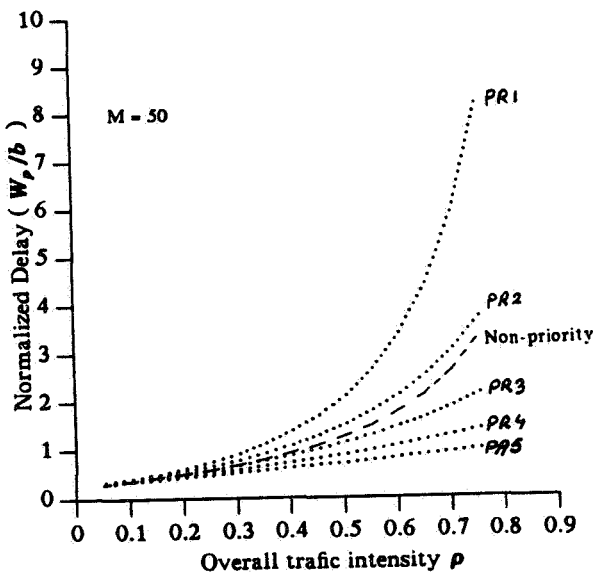

Figure 4. Local priority FDDI protocol with assymetric arrivals ($\bar{r}$=0.01)

IV. CONCLUSIONS

Average waiting time analysis of FDDI networks under local priority assignments become possible by modeling the vacation time of the server as an additional priority level. Then the problem can be treated as a standard HOL priority queuing problem. Lack of information about the distribution of the vacation time of the server can be overcome since analytical results due to

the non-priority version of the same FDDI network is readily available. Two important assumptions which affect the validity of the results are the locality of the priority assignments and the exhaustiveness of the service discipline.

REFERENCES

[1]   A. Cobham, "Priority Assignment in Waiting Line Problems," *Operations Research-ch*, 2, pp. 70-76, 1954.

[2]   H. Kesten and J. T. Runnenberg, "Priority in Waiting-Line Problems I and II," *Nederlandse Akademie van Wetenschappen*, Amsterdam, Proceedings, Series A, 60, pp. 312-324, 325-336, 1957.

[3]   R. Davis, "Waiting time distribution of a multiserver queuing system," *Operations Research*, vol. 14, pp. 133-136, 1966.

[41   G. Gercek and G. Davis, "Waiting time pdf's in multilevel static priority queues," *IEEE Phoenix Conference on Computers and Communications*, 1988 .

[5]   R. G. Miller, "Priority queues," Ann-Math.Stat., vol. 31, pp. 86-103.

[6]   D. R. Cox,*Renewal Theory*. London: Methuen, 1962.

[7]   L. Kleinrock, "A conservation law for a wide class of queuing disciplines *Naval Research Logistics Quarterly*, vol. 12, pp. 181-192, 1965

[8]   Rubin and L. F. M. De Moraes, "Message delay analysis for polling and token multiple-access schemes for local communication networks," IEEE *Journal on Selected Areas in Communicatons*, vol. SAC-1, No. 5, November, 1983.