**Bhupinder Kumar**
**Roll No. 04010210**

# A Queueing Model for Optimal Segmentation of a Token-Ring Local Area Network

(Author : Shaler Stidham,Jr.)

## Introduction

In a token-ring LAN the nodes (labeled i = l, . . . , n) are arranged logically in a ring with each node transmitting to the next node in the ring. There is a single token and a node must be in possession of it in order to transmit a packet. Each intermediate node receives the packet and retransmits it, with a time delay'd'. When the node completes transmitting the packet it appends the token to the end of the packet, thus indicating to the next (downstream) node that it may begin transmitting. If it does not have a packet to transmit, it passes on the token. The intended recipient of the packet both reads the packet into the node and relays it around the ring. When the entire packet has returned to the transmitting node, it starts to relay what is coming in. If some other node had a packet to send, then that packet is what is relayed; otherwise, the token is relayed. Travel of data around the ring is unidirectional as indicated by the arrows.

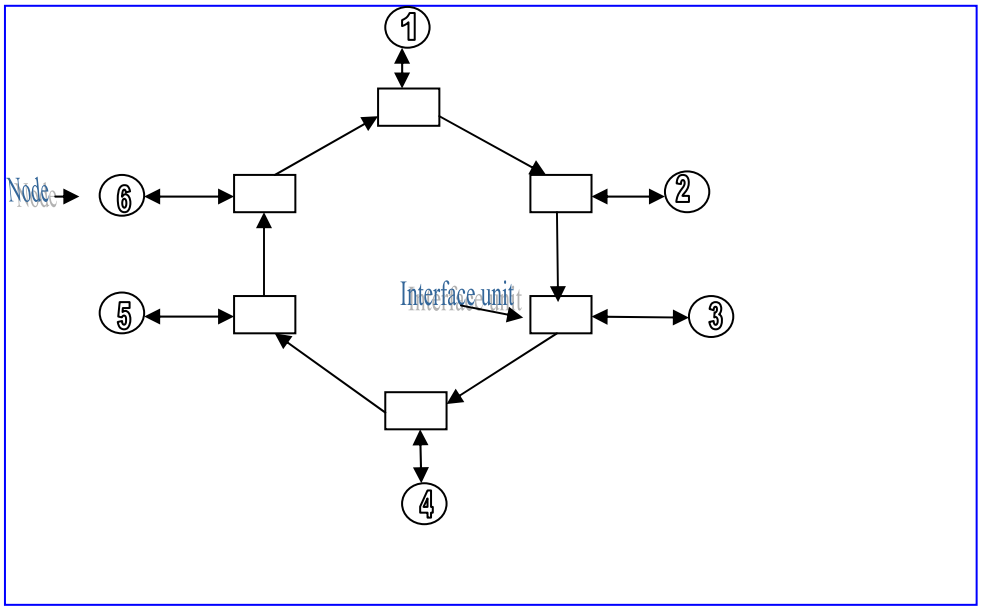


Figure: Typical token-ring network

## Husselbaugh's contribution

Husselbaugh proposed the use of an M/GI/l queueing model to help determine whether a token-ring local area network (LAN) should be broken up into two or more subnetworks,

connected by a bridge. As more nodes are added to a token ring and the nodes compete for the single token, the system may experience intolerable congestion or even become unstable (if the overall packet-generation rate exceeds the LAN‘s transmission capacity). By segmenting the network into separate domains, each configured as a token ring, one can reduce the waiting time for a token but at the expense of installing a bridge, packet switch, or router to connect the segments. He also examined the effects on congestion of various parameters, such as the number of stations in a segment, the transmission rate, packet size, and the packet demand rate. A novel feature of his model (compared to classical queueing design models) is that increasing the number of stations per segment increases not only the overall arrival rate (of packets) but also the service time of each packet, because of the time delay required for each station to receive and retransmit the packet. The theme in this that “using bandwidth techniques to predict network performance is fraught with inadequacies. One particular problem with bandwidth is its failure to take competition [for the token] into account.” In the language of queueing theory: a system in with adequate capacity (arrival rate less than service rate) may still have unacceptably high waiting times.

**Contribution of the present paper**

1. The author explicitly considers the trade-off between reduced congestion and the cost of inserting a bridge or other routing device.
2. He develope what we believe is a more accurate model for the traffic in each LAN segment as a function of the number of nodes in the segment (or, equivalently,as a function of the number of segments ).
3. He compare the relatively simple, single-class M/GI/1 model of to the more complicated, multi-class polling models that have been proposed in the literature on token-ring LANS.

He argues that the single-class model gives reasonably accurate estimates of performance measures such as average waiting time for a symmetric LAN with negligible switchover (walking) times between classes. Moreover, with this simple model it is possible to solve efficiently the design problem of selecting the optimal number of nodes to assign to each LAN segment for a wide range of parameter The author models a token-ring LAN or LAN segment as an M/GI/l queue. The customers are the packets which are being generated for transmission by the n nodes in the ring. Assuming symmetric demand, he assumes each node generates the same average number of packets per second(say, $\lambda_u$ ,) and also that these packet-generation processes are independent and Poisson. Then the traffic volume for the LAN segment m a whole is a Poisson process with arrival rate

$$\lambda = n\lambda_u \qquad (1)$$

The token is the server. Since the transmission protocol requires each packet to complete a round trip of the ring, the service time may be taken to be the time to transmit the packet all the way around the ring. Thus, assuming that the packet-size distribution is also the same for all nodes, the mean and variance of the service time are given, respectively, by

$$\frac{1}{\mu} = \frac{p}{r} + nd, \sigma^2 = \frac{v}{r^2} \qquad (2)$$

where p and v are the mean and variance, respectively, of the number of bytes per packet, and r is the transmission rate (in bytes/second). (Note the dependence of $1/\mu$ on n, which is due to the requirement that each of the n nodes must receive and retransmit the packet.)The average waiting time that a packet must wait for transmission is given by the Pollaczek-Khintchine formula:

$$W_q = \frac{\lambda(\sigma^2 + \frac{1}{\mu^2})}{2(1-\frac{\lambda}{\mu})} \qquad (3)$$

Because of our symmetry assumptions, this average waiting time is the same for all nodes.

Husselbaugh uses this formula (with minor variations) to measure congestion as a function of the number of inserted stations 'n' in the LAN. Curiously, in his analysis he holds $\lambda$ fixed, while allowing the mean service time to increase with 'n' according to the formula (2). In light of the formula (1) defining $\lambda$, this makes sense only if one assumes that the transmission rate of each node decreases in inverse proportion to the number of nodes, n. He concludes from this analysis that congestion ''is greatly influenced by the number of transmissions per second i.e. $\mu$, but hardly at all by the volume of data to be moved i.e. $\lambda$.'' Since he is artificially holding $\lambda$ constant, this is hardly a surprising conclusion, but it has little bearing on the behavior of a real LAN, in which increasing the number of nodes will have the double-whammy effect of increasing both $\lambda$ and $1/\mu$!.The implicit constraint used in to determine the maximum acceptable value of n is that the waiting time in the queue should not exceed the service time ($W_q \le 1/\mu$).The author proposes an alternative approach below, while at the same time taking explicit account of the dependence of both $\lambda$ and $1/\mu$ on 'n'.

## An Optimal Design Model for LAN Segmentation

Consider a network consisting of a fixed number of nodes N. We wish to consider segmenting the network into k separate token-ring LAN's, each consisting of n = N/L nodes, connected by a bridge. Let b(k) denote the amortized cost per unit time of the bridge, which he assumes to be a non-decreasing function of the number of segments, k. We incur a cost h per unit time spent by each packet while waiting for transmission and being transmitted. Let L denote the average number of packets in each segment, either waiting for transmission or being transmitted. Our objective is to choose a value of k to minimize the total cost per unit time:

$$c(k) = b(L) + khL .$$

Now we need to develop a formula for L. It follows from Little's Law that

$$L = \frac{\lambda}{\mu} + \lambda W_q$$

Since each of these variables is a non-decreasing function of n = N/k, it follows that L is a non-increasing function of k.This optimization problem can be solved in many cases by complete enumeration of the possible values of k. With some simplifying assumptions, we can solve the problem analytically, obtaining a closed-form solution that reveals

interesting properties of the optimal solution as a function of the problem parameters. To this end, let us assume that:
(AI) b(k)=b*k
(A2) The packet-size distribution is exponential.
(A3) The time delaty at each node is negligible; i.e.,d = 0.

Under these assumptions we have

$$c(k) = k\left[ b + h\left( \frac{\lambda}{\mu - \lambda} \right) \right] = k\left[ b + \frac{hN\lambda_u k}{k\mu - \lambda_u} \right]$$

If we treat k as a continuous variable then, using the fact that c(.) is a convex function, we can find the minimizing value of k by differentiating c(k) and setting the derivative equal to zero. This leads to the following formula for the optimal value, k':

$$k' = \frac{\lambda_u}{\mu}\left( 1 + \sqrt{\frac{Nh}{b}} \right)$$

The actual minimum value of k will be one of the feasible integer values on either side of k'. (Recall that in order for a value of k to be feasible, it must not only be an integer but must be a divisor of N, so that the number of nodes in each segment, n = N/k, is also an integer.)Note that k' is proportional to $\lambda_u$, the demand rate per node, and to $1/\mu$, the average packet transmission time. The dependence on N, h, and b is only through the ratio Nh/b, and, since this ratio enters into the formula through its square root, k' is relatively insensitive to these parameters.

Finally we note that, by varying the cost parameters, b and h, in this optimal design model, one can solve the constrained optimization problem, in which quality of service, as measured by the expected waiting time for a token, is guaranteed to be no larger than a given value.