#137

MEAN MESSAGE DELAY IN MULTI-HOP PACKET RADIO* NETWORKS USING SPATIAL-TDMA

Randolph Nelson and Leonard Kleinrock

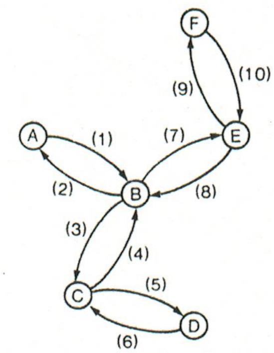
Computer Science Department
University of California
Los Angeles, California 90024

Abstract — **Spatial-TDMA is a collision-free access scheme for multi-hop packet radio networks. In this system, nodes are assigned specific intervals of time, from a cyclic time frame, during which they can transmit. This assignment creates a queueing system in each node consisting of a sequence of arrival, service and idle epochs. In this paper we develop an approximate formulation for the mean system time for messages in such networks. This approximation is compared to simulation results and shows a close fit.**

1. DESCRIPTION OF THE PROTOCOL

We assume the nodes in a multi-hop packet radio network are stationary (i.e. not mobile). Knowing their locations and the capture parameter of the system [1], allows one to generate a *compatibility matrix*. In such a matrix, a 1 in the (i, j) position indicates that arc i and arc j of the network can simultaneously transmit (specifically, their corresponding nodes being able to transmit in these directions) without causing a collision. Such an arc is said to be *enabled*. Using this compatibility matrix, one can generate a set of *cliques* containing arcs having the property that all arcs in the same clique can be simultaneously enabled without causing a collision. If we let C_i denote the i^{th} clique, we can form a *clique cover* $C = \{C_1, C_2, \cdots, C_k\}$ with the property that every arc is contained in at least one member of C. For each clique in the clique cover one can assign an interval of time, t_i, from a given time cycle (i.e. a frame whose structure repeats), during which arcs in that clique are enabled. Since any particular arc can be contained in more than one clique, the times during which an arc is enabled depend upon the times assigned to the cliques of which it is a member.

In Figure 1 we exemplify this construction. In this figure, the six nodes (A through E) have arcs which are labeled 1 through 10. When node B enables arc 7, nodes A, C, and E hear the transmission and B's message is addressed to node E. During this transmission, if we assume that receivers cannot capture transmissions, one of arcs 5 or 6 can also be enabled without causing collisions. In the compatibility matrix shown in the figure, then, the 7'th row contains a 1 in the 5'th, 6'th, and 7'th positions. Each row of this compatibility matrix is generated in a like manner. Using this matrix, we have generated all possible maximal cliques $C_1, C_2, \cdots, C_{10}$, and from these have selected a particular clique cover consisting of $C = \{C_1, C_2, \cdots, C_6\}$.



COMPATABILITY MATRIX

1	0	0	0	0	1	0	0	0	1
0	1	0	0	1	1	0	0	1	1
0	0	1	0	0	0	0	0	1	1
0	0	0	1	0	0	0	0	0	1
1	1	0	0	1	0	1	0	1	1
1	0	0	0	0	1	0	1	1	1
0	0	0	0	1	1	1	0	0	0
0	0	0	0	0	1	0	1	0	0
1	1	1	1	1	1	0	0	1	0
1	0	0	1	1	1	0	0	0	1

$C_1 = \{1, 6, 10\}$ $C_2 = \{2, 5, 9\}$ $C_3 = \{3, 9\}$ $C_4 = \{4, 10\}$ $C_5 = \{5, 7\}$

$C_6 = \{6, 8\}$ $C_7 = \{6, 9\}$ $C_8 = \{2, 5\}$ $C_9 = \{5, 9\}$ $C_{10} = \{5, 10\}$

$C = \{C_1, C_2, C_3, C_4, C_5, C_6\}$

FIGURE 1.

If we denote the cycle time of the frame as T time units, we can allocate time durations t_i, where $T = \sum_{i=1}^{6} t_i$ to each of the cliques during which their arcs are enabled. Each frame then consists of a set of durations during which cliques are enabled, and the sequence of the frame cycles every T time units. This describes the spatial-TDMA protocol defined in [2] and in this reference an optimal capacity assignment problem (an assignment of t_i's which minimizes the average delay of messages in the network) for these networks was solved.

2. QUEUEING APPROXIMATION

In this section we will describe an algorithm for approximating the average system time (queueing plus transmission) of messages passing through nodes using the spatial-TDMA protocol previously described. The queueing system at a given node created by this protocol consists of non-overlapping input, service, and idle periods which are enabled during specific periods of the frame. For example in Figure 2 we have shown a queue (contained in a single node of the network) with its corresponding frame, Figure 3.

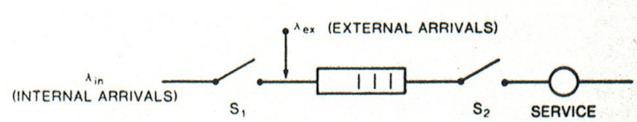


FIGURE 2 Model of Queue for Nodes in the Network

*This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract MDA 903-77-C-0272.

0536-1486/82/0000-0016 $00.75 © 1982 IEEE

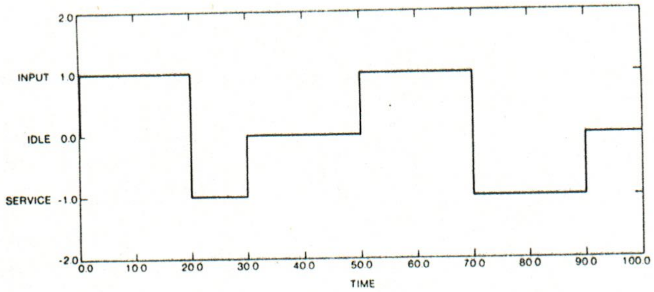


FIGURE 3 A Time Frame

The queue contains two switches s_1 and s_2 which are used to control the internal input and service processes. At most one of these switches can be closed during any part of the frame since we do not allow simultaneous reception and transmission by radios in the network. Over the entire frame, *external messages* (from the attached host) can arrive. As shown in Figure 2 these arrivals immediately enter the tail of the queue at a rate of λ_{ex}. This arrival process is assumed to have Poisson statistics and thus the probability that k external arrivals occur in m time units (denoted by $P[k\ ext|\ m]$) is given by:

$$P[k\ ext|\ m] = \frac{(m\lambda_{ex})^k e^{-(m\lambda_{ex})}}{k!}$$

During the first 20 time units of the frame in Figure 3, we see an *internal arrival interval*, during which s_1 is closed (s_2 will be open) to allow arrivals to enter from the network. Arrivals are assumed to occur in *message units* (assumed to be packets, bits, etc...) per unit time, where we normalize transmission rates so that each time unit accommodates at most one message unit. The number of internal arrivals to the queue during the first interval shown in the figure then will contain no more than 20 message units since it is 20 time units long. The statistics for the internal input process of nodes in a spatial-TDMA network is a complex aggregation of the outputs of all the previous queues in a packet's route. To generate this internal traffic in our simulation, we assumed the traffic statistics were given by a truncated Poisson distribution. In particular, if the total time from the frame allowed for internal arrivals is m time units, then the probability that k message units arrive during the frame (denoted by $P[k\,\text{int}|\ m]$) is given by:

$$P[k\ int|\ m] = \frac{(m\lambda_{in})^k e^{-(m\lambda_{in})}/k!}{\sum_{i=0}^{m} (m\lambda_{in})^i e^{-(m\lambda_{in})}/i!} \qquad k = 0, 1, \cdots, m$$

These k arrivals are then assumed to be proportionately distributed over all the internal arrival intervals of the frame. In the frame shown in figure 3, since $m = 40$, the number of arrivals to the first internal arrival interval will be given by $(20/40)\ k = k/2$. During the next phase of the frame, we see a *service interval* of 10 units in which s_2 is closed (s_1 is open) and message units are served at the rate of one per time unit. During this time a maximum of 10 message units can be serviced. The next phase we show is an *idle interval* during which both s_1 and s_2 are closed and no internal arrivals or services are allowed. Switches s_1 and s_2 are then turned on and off according to the time patterns depicted in Figure 3 and continue to cycle every 100 time units.

An exact analysis for the average system time is mathematically intractable and thus we seek an approximate formula. Fortunately a fluid approximation [3] to this system gives very good results and we will illustrate this method using Figure 4. In the fluid approximation, waiting times are calculated by assuming the actual backlog of packets in the queue is approximated by the average backlog.

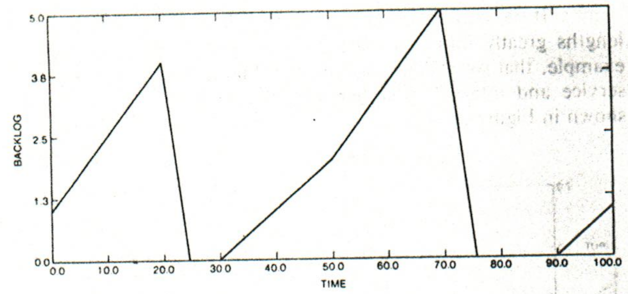


FIGURE 4

In figure 4 we have plotted the growth of the average backlog of message units in the system during the course of the frame shown in figure 3. For reasons to be explained later, we have started this growth pattern at the beginning of the last idle interval. During this interval, since internal arrivals are prohibited, only external arrivals can add to the backlog. If we assume the rate at which the queue increases during an idle period to be M_i, we see that the expected backlog grows linearly with this slope during this interval. As the frame progresses to the first interval, the growth rate for the backlog is given by M_a, which is the sum of M_i and the additional rate offered by the internal arrivals. A service period follows in the next interval and the backlog drops by a rate of $M_s < 0$ message units per time unit. Since external messages can arrive during any type of interval, M_s is equal to M_i less the service rate of 1 message unit per unit time. As seen in the figure, the backlog of message units in the queue drops to zero before the service interval is finished. Internal arrivals during this time are assumed to be immediately serviced and thus do not contribute to the backlog. This process continues in this manner until the last idle period, at which point it starts from a zero backlog once again. We will call a point on the frame a *zero-point* if, starting with an empty queue at this point, yields a backlog of zero after exactly one frame.

We now calculate the rates described in the previous section. First we make the following definitions:

Let T = Length of the time frame
T_i = Total time of idle intervals per frame
T_a = Total time of internal arrival intervals per frame
T_s = Total time of service intervals per frame

From our previous discussion we can write:

$$M_i = \lambda_{ex}/T$$
$$M_a = \lambda_{in}/T_a + M_i$$
$$M_s = M_i - 1$$

The area under the backlog curve represents the number of message-time units accumulated by messages arriving during the frame. Dividing this by the average number of message units that arrive during the frame, $\lambda_{in}T_a + \lambda_{ex}T$, gives, by Little's result [4], the average time spent in the queueing system. Because we began the calculation at a zero-point, message delays for all arrivals to the frame are counted. It is clear we can always find a zero-point on any frame satisfying $T_s \geqslant \lambda_{in}T_a + \lambda_{ex}T$.

This then describes the algorithm for the approximation which we summarize as:

1. Find a zero-point.

2. Calculate the area under the backlog curve.

3. Divide by $\lambda_{in}T_a + \lambda_{ex}T$ to arrive at the average system time.

It is clear that the ordering of the intervals and their lengths greatly influences the average system time. Suppose, for example, that we change the frame in Figure 3 by coalescing all the service and intervals together and placing them on the frame as shown in Figure 5.

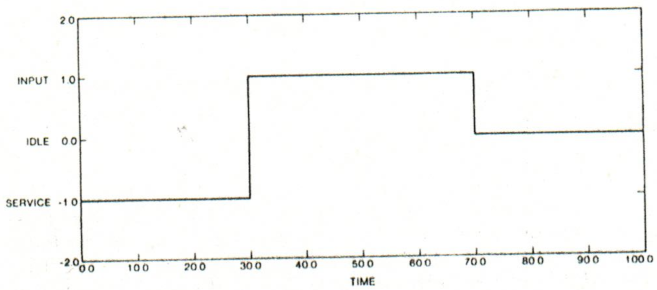


FIGURE 5

Although it has exactly the same interval lengths, this translation increases the average system time (in fact it is a worst case example). In this system, any messages that arrive during the internal arrival interval must wait at least 30 time units before being serviced. If the service and internal arrival intervals are interchanged, it is clear that the average system time would decrease by at least this much. In fact, the system that has the minimum average system time is one that spreads arrival and service intervals in infinitesimal units that alternate with each other. In this way, an arrival is immediately serviced in the following interval after accruing little waiting time. In a practical implementation of spatial-TDMA, however, there are limits to how small one can make interval sizes since radios have a finite switching time between transmission and reception, and choosing a frame pattern that minimizes the system time for one particular queue in the network does not in general decrease the total average message delay for messages in the entire network. In fact, finding the alternation that does achieve the minimum delay for all message in the network is a very difficult problem.

3. DISCUSSION OF RESULTS

In this section we compare simulation results with those found using the fluid approximation described in the previous section. Numerous frames were randomly generated that had the same input parameters (T, T_s, T_i, T_a, λ_{ex}, and λ_{in}) and results of the simulation checked against those of the approximation.

Three such frames are shown in Figure 6 where +1 steps correspond to internal arrival intervals, 0 to idle intervals, and −1 to service intervals. In all three frames $T=10000$, $T_i=6000$, $T_a=2000$, $T_s=2000$, the average service and internal arrival intervals have length 200, and the average external arrival rate is 200 message units over the frame time T. The mean service time as a function of ρ for these frames is shown in Figure 7.

There are several interesting features of these curves. We first see the close match between the simulation and approximation thus assuring us that the approximation is accurate. The variation between the mean system time for the three frames is very large which shows the dependency upon the ordering and size of the intervals of the frame. For example, for $\rho=.7$, set 1 has a mean system delay of about 650 time units whereas set 3 has a value of 2800, more than four times as much. The extreme delays of the third frame arise from the long periods (9000, 10000) and (0, 2000), during which there are no service intervals. All packets arriving during these periods create a backlog that cannot be depleted until much later in the frame. On the other hand, the fortuitous placement of intervals in frame 1 consists of groups of arrival intervals followed by service intervals that allow an accumulated backlog to be serviced quickly.

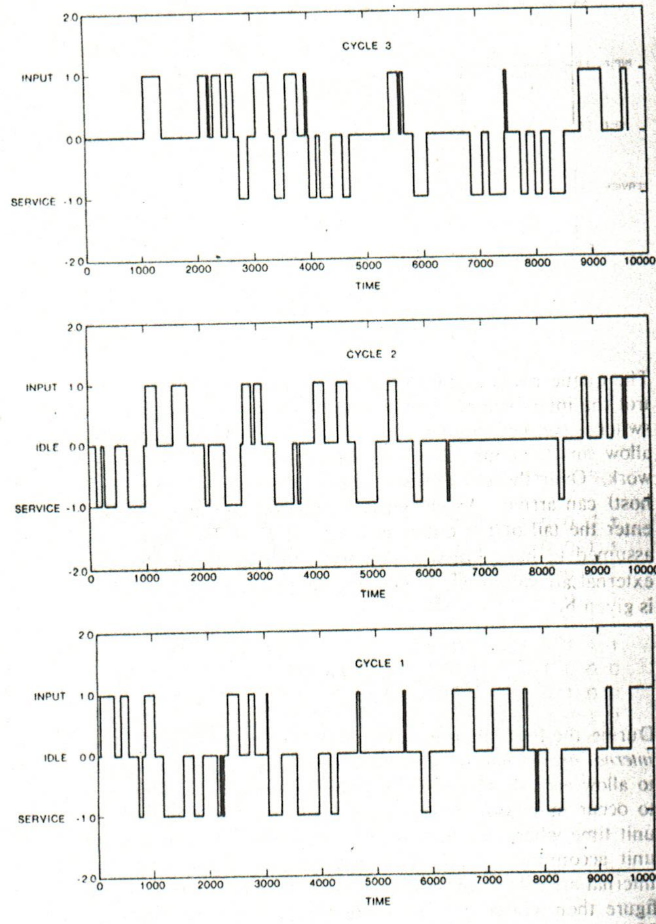


FIGURE 6

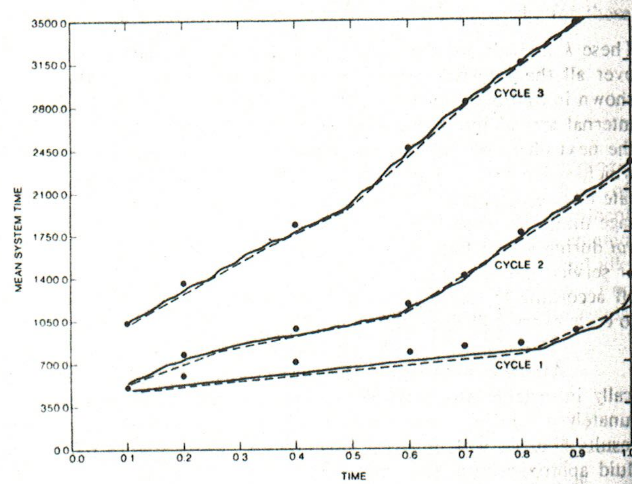


FIGURE 7

Even though there is a large variance in the curves, there is a striking similarity in the shapes of the curves. The curves are very well approximated by a piece-wise linear function (shown as a dashed line in the figure). We note that they do not approach infinity at $\rho=1$. Since other queueing systems have explosive growth as $\rho \rightarrow 1$, arising from the omni-present $1/(1-\rho)$ term in most queueing equations, this behavior is very unusual. The explosive growth for most queueing systems arises from the fact that there is variation in the arrival pattern of messages to the queue, and as $\rho \rightarrow 1$ the probability that a sequence of arrivals saturates the queue approaches 1. In spatial-TDMA however, the number of arrivals for each internal arrival interval is constrained to be no greater than the normalized interval length, and thus the variance of arrival statistics is also constrained. Besides reducing queueing delays, this restriction on the number of messages that can enter the system during an interval also acts as a natural flow control mechanism for messages in such a network.

The slope changes in the piece-wise linear approximation occur when the arrival rate is so large that a sufficient number of arrivals to an internal arrival interval cannot be serviced in the next set of service epochs. For example, the change about the point $\rho = .6$ for frame 2 arises from the fact that for $\rho > .6$, some arrivals over the interval (6000,7800) must wait until the next service set of intervals (1200, 2300) to be processed. For lesser values of ρ, $\rho < .6$, most arrivals to (6000, 7800) are serviced in the interval (7800, 9100) and thus suffer less delay. Naturally as ρ increases, the proportion of messages that must wait until (1200, 2300) to be serviced grows and so does the mean system time. Each of the breaks in the piece-wise linear approximation can be explained in this manner.

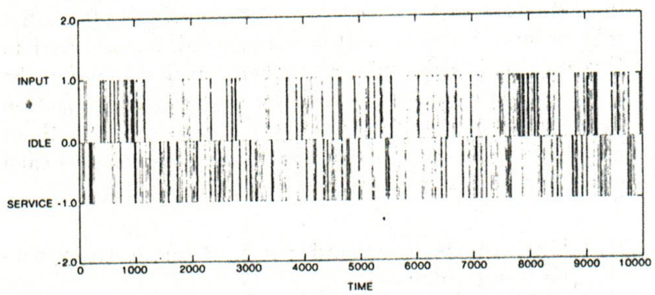


FIGURE 8

In Figure 8 we have shown a frame for the same input parameters but where the average size of the service and internal arrival intervals is equal to 20 time units instead of 200 as in those of Figure 6. The corresponding mean system time curve is shown in Figure 9. We see a marked decrease in the mean system time for this frame in comparison to the previous set of frames. This demonstrates the dependency of the mean system time upon the size of the intervals. If we adjust the frame to minimize the mean system time, as shown in Figure 10 (where for illustrative clarity we have only shown a portion of the frame), the resultant delay is approximately equal to 1 time unit throughout the entire range of ρ. For such a frame the majority of the arrivals to the system are immediately serviced in the following service interval.

4. CONCLUSIONS

In this paper we have presented a method to approximate the mean system time for packets in a network using spatial-TDMA. The approximation has been compared to simulation

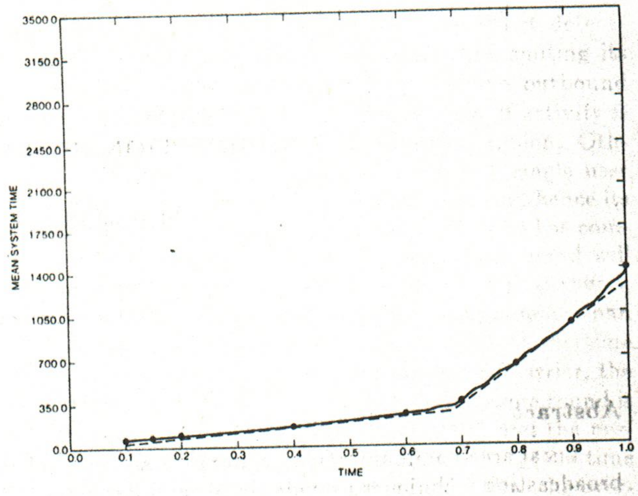


FIGURE 9

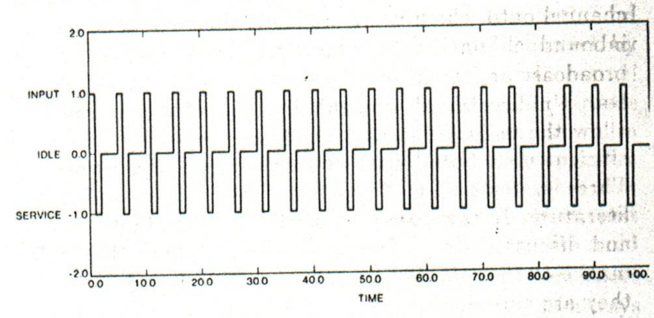


FIGURE 10

results and has been shown to be very accurate. The dependencies of this mean system time on the ordering and size of time assignments has been discussed and the property that the mean system time does not explode at $\rho = 1$ has been explained.

References

[1] L. G. Roberts, "ALOHA Packet System With and Without Slots and Capture," ASS Note 8 (NIC 11290), ARPA Network Information Center, Stanford Res. Inst., Menlo Park, Ca. (June 1972). reprinted in *Computer Communication Review*, Vol. 5, pp. 28-42 (April 1975).

[2] R. Nelson, *Channel Access Protocols for Multi-hop Broadcast Packet Radio Networks*, UCLA Computer Science Department, Los Angeles, Ca., Ph.D. Thesis.

[3] L. Kleinrock, *Queueing Systems, Vol II., Computer Applications*, Wiley-Interscience, New York (1976).

[4] J. Little, "A Proof of the Queueing Formula $L = \lambda W$," *Operations Research* 9(2), pp.383-387 (March 1961).