Throughput–Delay Characteristics of Some Slotted-ALOHA Multihop Packet Radio Networks

HIDEAKI TAKAGI, MEMBER, IEEE, AND LEONARD KLEINROCK, FELLOW, IEEE

Abstract—**A Markovian model is formulated to find the throughput-delay performance for slotted-ALOHA multihop packet radio networks with a fixed configuration of packet radio units (terminals and repeaters) and fixed source-to-link paths for packets. Improvements in performance which are obtained by the adjustment of transmission parameters (suppression/acceleration) according to the states of nearby units and/or by having repeaters equipped with multiple buffers are demonstrated.**

I. Introduction

The packet radio network considered in this paper is a ground-based minicomputer communication network using a shared multiple-access radio channel. One of its potential uses will be providing real-time computer-based communication for packet radio-equipped military users, both in garrison and in the battlefield. Another application is to replace regional wired packet-switching networks without the need for cable extension.

Although some intensive experimental research on packet radio networks has taken place at several locations during the last few years (e.g., PRNET in [7]), little theoretical work about their performance evaluation seems to have been published so far. Compared to the analysis of one-hop broadcast networks, for which extensive literature has appeared, one of the difficulties in dealing with multihop networks is inherent in the fact that the issue of routing comes into play as in the wire-based store-and-forward networks. However, because of colliding transmissions from multiple packet radio units, we have not found any exact solution—whether in a product form or not—for evaluating the mean packet delay of a general class of multihop packet radio networks. One of the reasons that a discrete-time queueing network (modeled on the slotted-ALOHA system) does not lend itself to a product-form solution is that more than one event can occur in a single slot [1].

As for the approximate evaluation of the average packet delay and the optimal routing with respect to it, some contributions may be noted. Leiner [11] showed an approximate way to get the delay at each link, given its traffic requirement, using Kleinrock's ZAP approximation [8] for the throughput-delay curves for a variety of channel access protocols in single-hop systems. Kung [10] speculates that the average delay is a convex function in the space over the traffic requirements on all links, on the basis of the ZAP approximation [8] of the throughput–delay curves. Thus, he adapts the flow deviation method, originally developed for wire-based store-and-forward networks in [4], to the multihop packet radio networks. Some other authors [2], [3], [13] create more or less idealistic assumptions (such as zero propagation delay and perfect delay capture) to inhibit interference of transmissions and discuss the resulting throughput and optimal routing. Some two-hop networks are analyzed by Tobagi [15].

In this paper, we take a Markov-chain approach to find the throughput–delay characteristics for a general class of slotted-ALOHA multihop packet radio networks which consist of a relatively small number of packet radio units. In Section II we describe our basic model of packet radio networks in detail. This is followed in Section III by the Markov-chain formulation to calculate the throughput and average end-to-end packet delay for a given network. The tradeoffs between them are shown for two example networks. In the following sections, we propose and analyze three ways (and their combinations) of reducing the average packet delay for a given throughput requirement and increasing the maximum supportable throughput.

II. Basic Model

In this section, we describe in detail our basic model of packet radio networks. Consider a network consisting of a fixed number of packet radio units, each having an omnidirectional antenna, thereby being capable of transmitting or receiving a packet, but not both simultaneously. We distinguish the two kinds of packet radio units: terminal and repeater. A terminal is defined to be a unit which can be a source and/or a sink of packets but does not relay any packets in transit. A repeater is defined to be a unit which neither generates nor absorbs any packets but only relays them.

We assume that every unit is within the transmission range of some other units but not necessarily of all others; this hearing topology is given and fixed. Let us represent the hearing configuration by a matrix (h_{ij}) defined by

$$h_{ij}=h_{ji}=\begin{cases}1 & \text{if units } i \text{ and } j \text{ hear each other}\\ 0 & \text{otherwise}\end{cases}$$

$$h_{ii}=1. \tag{1}$$

We also assume a given set of fixed paths for packets which connect pairs of specific source and sink terminals via a number of repeaters. Thus, packets originating at the source terminal of a particular path are sent (with specific destination ID for each link) in a store-and-forward manner through several repeaters along a unique path down to the sink terminal and absorbed there. Let these paths be numbered $k = 1, 2, 3, \cdots$. Fig. 1 shows two such networks, which we will call networks 1 and 2. In the (undirected) graph representation of hearing topology, nodes are packet radio units (either a terminal shown by a circle or a repeater shown by a square), and arcs are drawn between the units in the hearing range of each other. Each packet path is shown by a bold line, with arrows indicating the direction of the path. We will use these two examples to evaluate the throughput and average source-to-sink packet delay. Although these example network configurations look simple, they contain various cases of hearing topology and overlaid paths and already have a great number of Markovian states, as demonstrated below.

Note that the paths in our example networks are carefully

Paper approved by the Editor for Computer Communication of the IEEE Communications Society. Manuscript received February 14, 1983; revised January 10, 1985. This work was supported by the Defense Advanced Research Projects Agency under Contract MDA 903-82-C-0064.

H. Takagi is with IBM Japan Science Institute, Tokyo 102, Japan.

L. Kleinrock is with the Department of Computer Science, University of California, Los Angeles, CA 90024.

0090-6778/85/1100-1200$01.00 © 1985 IEEE

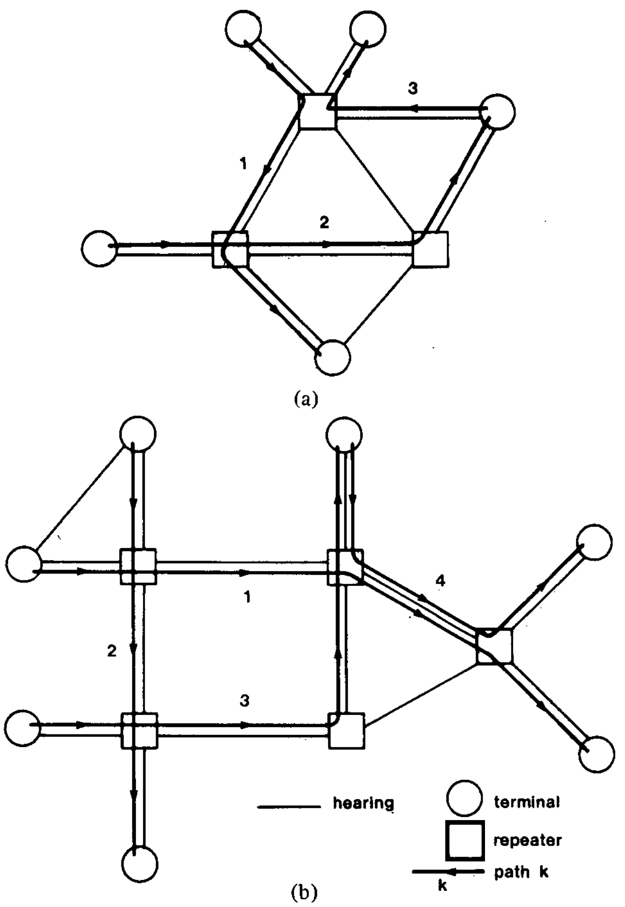

Fig. 1. Examples of networks. (a) Network 1—five terminals, three repeaters, and three paths. (b) Network 2—seven terminals, five repeaters, and four paths.

laid out in order to prevent the deadlock situations inherent in uncontrolled finite-buffer systems. For example, the reversal of direction of path 1 in network 2 would bring about an indirect store-and-forward deadlock. Any return path (connecting the same units as in a packet path in the opposite direction for end-to-end acknowledgment) or even a full-duplex link (a common link in the two paths in the opposite directions) would cause a direct store-and-forward deadlock. Note, further, that the direct store-and-forward deadlock can be avoided, for example, by reserving separate buffer spaces for the opposite directions and that the indirect store-and-forward deadlock can be prevented, e.g., by the structured buffer pool flow control [6]. We believe that these flow control schemes can also be incorporated within the framework of Markov-chain formulation as given below; however, we do not consider them in this paper.

All units in the network are assumed to use the common radio channel band. The reference time is slotted, and the slot size is such that it includes the transmission time of a packet, its propagation delay, and the time needed to notify the transmitter of the results of transmission (successful or not). This hop-by-hop acknowledgment is assumed to be given for free. We employ the slotted-ALOHA protocol, by which we mean that the channel slot is used or idle. Also, we neglect channel noise and assume no channel errors for single transmissions.

We now proceed to describe in more detail the properties of our terminals and repeaters. Let us begin with a terminal. In our model, a terminal can be a source of, at most, one path (for simplicity) and/or a sink of multiple paths, and it possesses buffer space for a single packet only. The packets received at proper sink terminals are consumed immediately so that they do not claim any buffer space. Let us represent the state of terminal i, s_i, by the state of its buffer; that is, it is in the "empty" state when the buffer is empty and in the "k-backlogged" state when the buffer contains a packet which belongs to path k. Thus,

$$s_i = \begin{cases} 0 & \text{empty} \\ k(\neq 0) & k\text{-backlogged.} \end{cases} \tag{2}$$

Notice that, since a terminal can be a source of, at most, one path, every source terminal has exactly two possible states. On the other hand, sink-only terminals are always in the empty state.

A source terminal of path k is the empty state generates a new packet at the beginning of a slot instantaneously with probability $\lambda(k)$, and, in such a case, it transmits the packet with probability 1 in the same slot. (Thus, $1 - \lambda(k)$ is the probability of no action in any given slot when in the empty state.) Suppose that the destination of the first transmission from a source terminal i is using j (repeater or sink terminal). The conditions for this transmission to be successful are: i) that all units which can be heard by j, including j and excluding i, do not transmit in the same slot, and ii) that unit j is in the empty state. (The states of a repeater are explained shortly.) If the first transmission is successful, the terminal remains the empty state for the next slot. If it is unsuccessful because i) and/or ii) are not met (we call them "collision" and "buffer blockage," respectively), the terminal goes into the k-backlogged state.

A k-backlogged terminal transmits its packet with probability $p(k)$ in any slot and delays action until the next slot with probability $1 - p(k)$. The conditions for successful transmission are the same as above in i) and ii). Immediately following its successful transmission, a backlogged terminal switches to the empty state. In the cases of no action and unsuccessful transmission, it remains in the same backlogged state for the next slot. We assume that a sink terminal receives a packet with success only on the condition of no collision, irrespective of its buffer state being empty or occupied by an outgoing packet. The successful reception at a sink terminal does not affect its state.

Next, we describe the operation of repeaters. As said before, a repeater neither generates nor absorbs any packets but simply relays them. Let each repeater be equipped with a single packet buffer. The state of repeater i, denoted by s_i, is identified by the contents of its buffer, just as for terminals. Thus, we have the same representation:

$$s_i = \begin{cases} 0 & \text{empty} \\ k(\neq 0) & k\text{-backlogged.} \end{cases} \tag{3}$$

However, since a repeater can be used by multiple paths, as exemplified by the networks depicted in Fig. 1, the number of distinct states for a given repeater is equal to the number of paths which pass that repeater plus 1 (the empty state).

A repeater in the empty state takes no action in any slot with probability 1. A repeater in the k-backlogged state behaves just like a terminal in the k-backlogged state; i.e., in any slot the probability of transmitting a packet is $p(k)$, and that of not transmitting is $1 - p(k)$. The conditions for successful transmission from a repeater are similar to those of the terminal mentioned above. We note that the successful transmission of a packet from a repeater in the k-backlogged state changes its state into the empty state. On the other hand, the successful reception of a packet belonging to path k at a repeater in the empty state changes its state into the k-backlogged state.

We may call the transmission protocol at a source terminal the immediate-first-transmission (IFT) with parameter $\lambda(k)$

and call that at a repeater the *delayed-first-transmission* (DFT) with parameter $p(k)$, where the terms IFT and DFT were introduced in [15] in the context of two-hop systems. Note that our protocol is a natural extension (to multihop networks) of the assumptions traditionally adopted for single-hop systems and then used in [15] for the two-hop system. Despite these simplifying assumptions, our model still captures major transmission characteristics of multihop networks, such as contention among units in the hearing range and coexistence of the successful transmissions occurring remotely.

III. Formulation

We now turn our attention to the formulation of the procedure for calculating the throughput and the average source-to-sink packet delay for networks such as those described in Section II. Note that the packet transmission process at each unit in any slot is based only on its current state and not on the past states. This memoryless property makes slot boundaries Markov points. Thus, we follow the usual formulation of discrete-time homogeneous Markovian systems.

First, let us represent the state of the whole network in a given slot, $\boldsymbol{s}$, by the Cartesian concatenation of the states of all units in the network: $\boldsymbol{s} = (s_1, s_2, \cdots, s_M)$, where M is the total number of units involved. Also, represent the behavior of the network for the slot, $\boldsymbol{e}$, by the Cartesian concatenation of the actions of all units: $\boldsymbol{e} = (e_1, e_2, \cdots, e_M)$, where

$$e_i = \begin{cases} 0 & \text{unit } i \text{ does not transmit} \\ k(\neq 0) & \text{unit } i \text{ transmits a packet of path } k. \end{cases} \tag{4}$$

The behavior of the network is a stochastic phenomenon, given the current state of the network. Since each unit behaves independently of others, we may write

$$p(\boldsymbol{e}|\boldsymbol{s}) \triangleq \text{Prob [behavior} = \boldsymbol{e} \,|\, \text{current state} = \boldsymbol{s}]$$
$$= \prod_{i=1}^{M} Q_i(s_i, e_i) \tag{5}$$

where each factor $Q_i(s_i, e_i)$ for unit i is given as follows. For terminal i which is the source of path k,

$$Q_i(s_i, e_i) = \begin{cases} 1-\lambda(k) & s_i = e_i = 0 \\ \lambda(k) & s_i = 0,\ e_i = k(\neq 0) \\ 1-p(k) & s_i = k(\neq 0),\ e_i = 0 \\ p(k) & s_i = e_i = k(\neq 0). \end{cases} \tag{6}$$

For terminal i where no paths originate, $Q_i(s_i, e_i) = 1$. For repeater i,

$$Q_i(s_i, e_i) = \begin{cases} 1 & s_i = e_i = 0 \\ 1-p(k) & s_i = k(\neq 0),\ e_i = 0 \\ p(k) & s_i = e_i = k(\neq 0). \end{cases} \tag{7}$$

Given the current state $\boldsymbol{s}$ and behavior $\boldsymbol{e}$ of the network, it is not difficult to determine the state of the system for the next slot, $\boldsymbol{s}' = (s_1', s_2', \cdots, s_M')$. For example, nontransmission at unit i does not affect the states of any other units. The successful transmission from unit i to unit j of a packet belonging to path k brings

$$s_j' = \begin{cases} s_j & j \text{ is a sink terminal of path } k \\ k & j \text{ is a repeater} \end{cases}$$
$$s_i' = 0. \tag{8}$$

The unsuccessful transmission from unit i simply gives

$$s_i' = s_i. \tag{9}$$

Thus, by examining all possible events for each state, we can construct the transition probabilities of our homogeneous Markovian system:

$$P(\boldsymbol{s}'|\boldsymbol{s}) \triangleq \text{Prob [next state} = \boldsymbol{s}' \,|\, \text{current state} = \boldsymbol{s}]$$
$$= \sum_{\boldsymbol{e} \text{ such that it gives } \boldsymbol{s}'} p(\boldsymbol{e}|\boldsymbol{s}). \tag{10}$$

Let us denote by $\pi(\boldsymbol{s})$ the equilibrium probability that the network is in state $\boldsymbol{s}$:

$$\pi(\boldsymbol{s}) \triangleq \text{Prob [state} = \boldsymbol{s}]. \tag{11}$$

Then we have the equilibrium state equations

$$\begin{cases} \pi(\boldsymbol{s}') = \sum_{\boldsymbol{s}} \pi(\boldsymbol{s}) P(\boldsymbol{s}'|\boldsymbol{s}) & \text{for all } \boldsymbol{s}' \\ \sum_{\boldsymbol{s}} \pi(\boldsymbol{s}) = 1. \end{cases} \tag{12}$$

This system of linear simultaneous equations may be solved numerically, given values of $\lambda(k)$ and $p(k)$.

Once the solution $\pi(\boldsymbol{s})$ is obtained, we can compute the following quantities of interest. First, the average backlog of packets along path k, $Q(k)$, is given by

$$Q(k) = \sum_{\boldsymbol{s}} Q(\boldsymbol{s};\ k)\pi(\boldsymbol{s}) \tag{13}$$

where $Q(\boldsymbol{s}; k)$ denotes the number of k-backlogged units when the network is in state $\boldsymbol{s}$. The total average backlog Q is given by

$$Q = \sum_{k} Q(k). \tag{14}$$

Second, the throughput of path k, $S(k)$, is defined as the average number of successfully delivered packets per slot from source to sink of path k. Since no packets disappear on their way, they can be counted at the sink terminal:

$$S(k) = \sum_{\boldsymbol{s}} \pi(\boldsymbol{s}) \sum_{\boldsymbol{e}} p(\boldsymbol{e}|\boldsymbol{s}) S(k|\boldsymbol{s},\ \boldsymbol{e}) \tag{15}$$

where

$$S(k|\boldsymbol{s}, \boldsymbol{e}) = \begin{cases} 1 & \text{successful transmission to the sink terminal of path } k, \text{ given } \boldsymbol{s} \text{ and } \boldsymbol{e} \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

Thus, the total throughput of the network, S, is given by

$$S = \sum_{k} S(k). \tag{17}$$

We call the maximum attainable throughput with respect to changing the values of $\lambda(k)$ and $p(k)$ the capacity of the network. Last, the average packet delay of path k, $D(k)$, is defined as the average time, in slots, that a packet of path k takes to go from its source to sink terminal. Applying Little's result [12] to each path, this is given by

$$D(k) = 1 + \frac{Q(k)}{S(k)}. \tag{18}$$

Note that the first term accounts for the first transmission from the source. Applying Little's result to the whole network, the overall average packet delay D is given by

$$D = 1 + \frac{Q}{S}. \tag{19}$$

This concludes the formulation of our basic model.

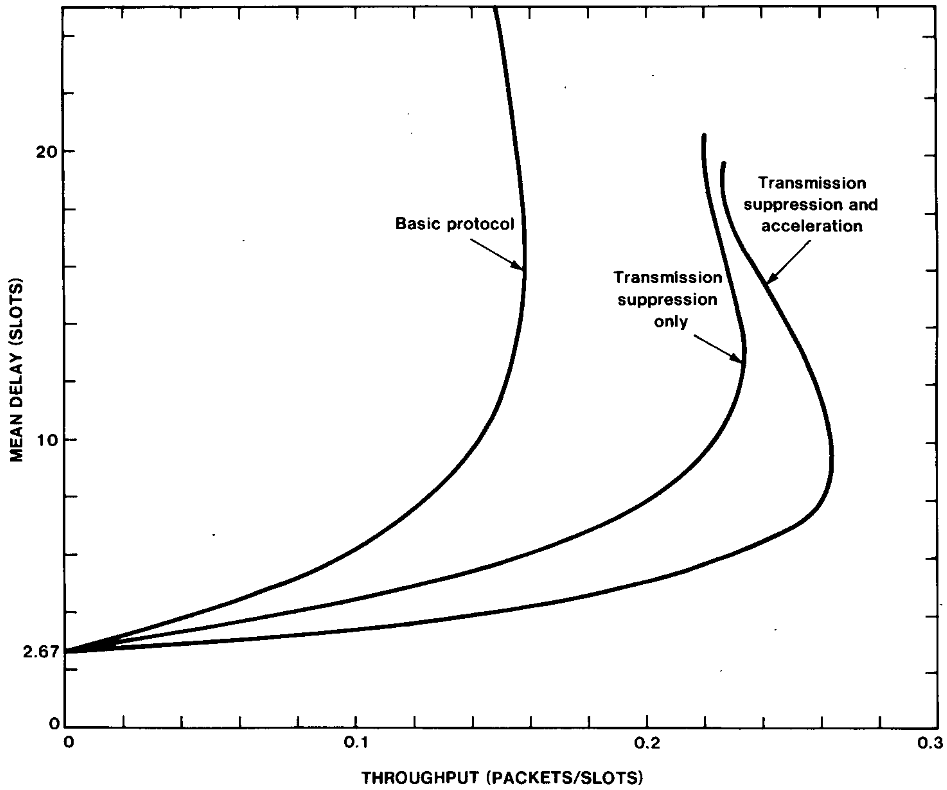

Fig. 2. Throughput–delay characteristics for network 1 with single-buffered repeaters.

We have calculated the above quantities for networks 1 and 2 depicted in Fig. 1, and the results are shown in Figs. 2 (for network 1) and 3 (for network 2) in the form of the overall average packet delay (D) (for the basic protocol) versus the total throughput (S). Throughout this paper we have assumed that $\lambda(k) = \lambda$ and $p(k) = p$ for all paths, not only for simplicity but also for fairness among paths. The displayed curves are actually optimum envelopes in the sense that, given λ, the value of p is adjusted in order to minimize D. (The optimization procedure is based on the Fibonacci search method for a unimodal function [9]. The unimodality of D in p has been assumed.) The curves in Fig. 4 show the throughput–delay relations for individual paths in network 2. Note that these curves have been obtained for the optimal values of p, where $p(k) = p$ for all paths is assumed.

Now let us look at the behavior for small values of λ. The throughput of each path nears λ, irrespective of the values of p, because collisions and buffer blockage are rare. Also, the average packet delays for paths 1, 2, and 3 of network 1 are given by $1 + 2/p$, $1 + 2/p$, and $1 + 1/p$, respectively, and thus, the overall average packet delay for network 1 is given by

$$\frac{1}{3}\left[\left(1+\frac{2}{p}\right)+\left(1+\frac{2}{p}\right)+\left(1+\frac{1}{p}\right)\right]=1+\frac{5}{3p}.$$

Similarly, the overall average packet delay for network 2 for small λ is given by

$$\frac{1}{4}\left[\left(1+\frac{3}{p}\right)+\left(1+\frac{2}{p}\right)+\left(1+\frac{3}{p}\right)+\left(1+\frac{2}{p}\right)\right]=1+\frac{5}{2p}.$$

Thus, it is clear that $p = 1$ evokes minimum delays of 8/3 and 7/2 for networks 1 and 2, respectively.

Under moderate values of λ, we see from Fig. 4 that the throughputs are rather fairly distributed among all paths while the delays are nearly proportional to the individual path lengths.

For larger values of λ, we reach the maximum in throughput, the capacity. The capacities of networks 1 and 2 are about 0.16 and 0.19 packets per slot, respectively, for the present transmission protocol. As the throughput approaches its maximum, the average packet delay increases rapidly for a marginal increase in throughput.

IV. Improvement by Transmission Suppression

In the basic protocol described in Section II, we have assumed that the behavior of each unit in any slot is based only on its own current state (isolated strategy). Thus, it may happen that a unit transmits a packet to a unit having no available buffer, resulting in certain failure. If information about the buffer state of the receiver were available to each transmitter, it could avoid this foreseeable wastage of channel capacity by suppressing the transmission, thus making it available for others. In this section, we exploit this possibility.

Suppose that repeaters with no available buffer broadcast a "busy tone," in a different channel, towards their neighbors at the beginning of any slot. (This busy tone should not be confused with the one introduced in [14] to solve the hidden terminal problem. The busy tone in [14] is emitted by a receiver when it is receiving a packet.) It is assumed that the hearing topologies on both channels are identical. Also, suppose (for the tractability of the problem) that the busy tones from multiple repeaters do not collide, and necessary information is always captured correctly by potential transmitters instantaneously. Knowing that the destination has no available buffer, a unit with a packet will suppress its transmission with probability 1 (i.e., p is set to 0) for the current slot. Otherwise, the value of p is unchanged. When the destination is a sink of a path (e.g., path 2 in network 1) and has a packet of another path (path 3 in the same example), the transmission to that destination is suppressed (despite no buffer blockage) since the destination unit still emits a busy tone.

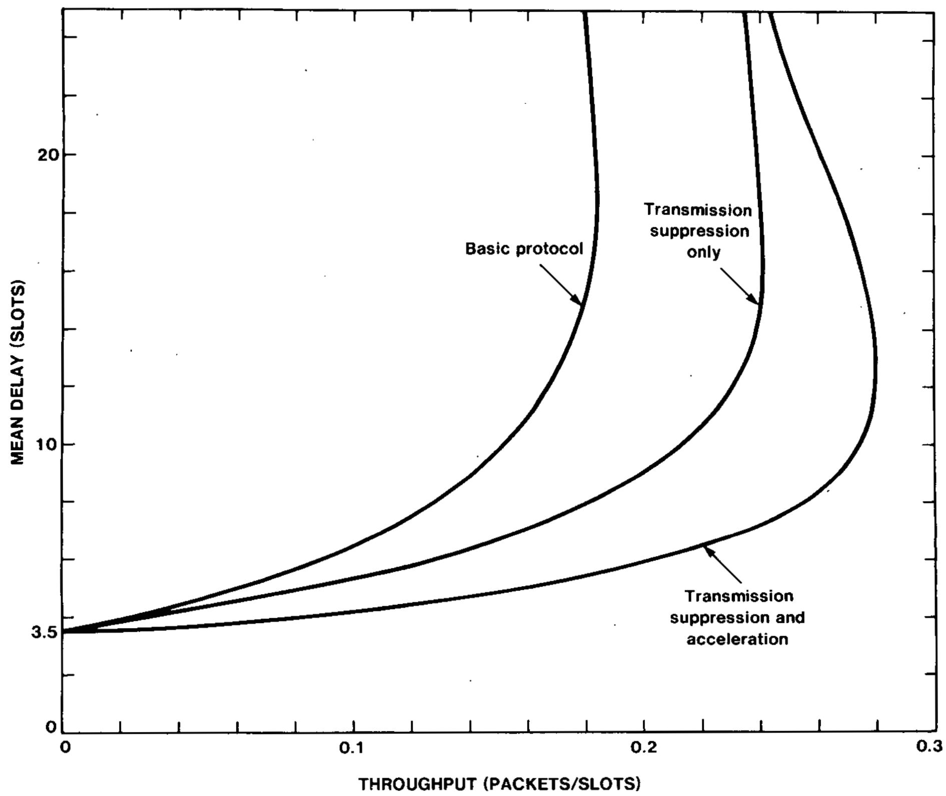

Fig. 3. Throughput-delay characteristics for network 2 with single-buffered repeaters.

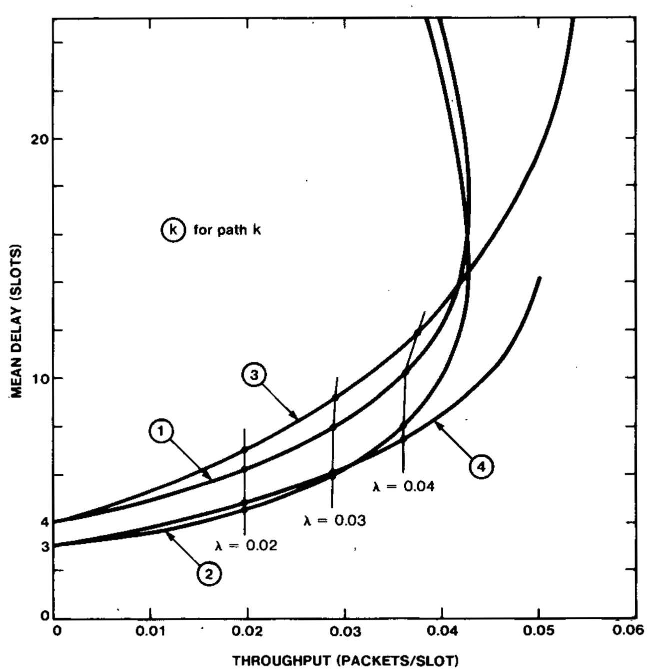

Fig. 4. Throughput-delay characteristics for individual paths in network 2 (basic protocol and single buffer).

Thus, the probability of behavior $\boldsymbol{e}$, given the current state $\boldsymbol{s}$, is expressed as

$$p(\boldsymbol{e}\,|\,\boldsymbol{s})=\prod_{i=1}^{M} Q_i(s_i,\ s_j,\ e_i) \tag{20}$$

where j is the destination ID of the transmission from unit i (which is unique given s_i) and each factor $Q_i(s_i, s_j, e_i)$ is given as follows. For terminal i which is a source of path k,

$$Q_i(s_i,\ s_j,\ e_i)=\begin{cases} 1-\lambda(k) & s_i=e_i=0,\ s_j=0 \\ \lambda(k) & s_i=0,\ e_i=k(\neq 0),\ s_j=0 \\ 1-p(k) & s_i=k(\neq 0),\ s_j=e_i=0 \\ p(k) & s_i=e_i=k(\neq 0),\ s_j=0 \\ 1 & s_j\neq 0,\ e_i=0. \end{cases} \tag{21}$$

For terminal i where no paths originate, $Q_i(s_i, s_j, e_i) = 1$. For repeater i,

$$Q_i(s_i,\ s_j,\ e_i)=\begin{cases} 1 & s_i=0 \text{ or } s_j\neq 0,\ e_i=0 \\ 1-p(k) & s_i=k(\neq 0),\ s_j=e_i=0 \\ p(k) & s_i=e_i=k(\neq 0),\ s_j=0. \end{cases} \tag{22}$$

Since the above modification of transmission parameters is again based on the current system state only, we still have the Markovian property for the network behavior. Thus, the formulation proceeds as in Section III. With respect to any empty state source terminal which generates a new packet but is forced to suppress its transmission, we can distinguish two models: i) it then goes into the backlogged state, retaining the packet (retransmission model); and ii) it remains in the empty state, dropping the packet (loss model). Since the delay due to buffer blockage at entry to the network must also be counted as much as one slot for a user, we opt for model i) in this paper. If we were to evaluate the delay for only those packets that are accepted in the network, we would employ model ii). Thus, in the case of transmission suppression as well as unsuccessful transmission from unit i of a packet of path k, the next state of unit i will be $s_i' = k$.

In Figs. 2 and 3 we show the throughput-delay characteristics improved by the transmission suppression scheme for networks 1 and 2, respectively. We see that the capacities of the networks are greatly increased. A close comparison of what is happening at each unit and along each path between the basic protocol and the transmission suppression protocol for the same values of λ and p has revealed that the present scheme gives rise to many fewer collisions and, thus, many fewer backlogs with only slightly higher throughput. This results in much lower average packet delay, due to Little's result. Also, it has turned out that the utilizations (fractions of time when the buffer is occupied) of units are much lower, especially for source terminals. It seems that, in case of congestion, newly entering packets are likely to be blocked at source terminals, which prevent them from entering the subnetwork consisting of the repeaters; therefore, we think that this transmission suppression scheme provides a natural flow control at the network access level [6] and speculate that the buffer-full condition at entry repeaters could be a good indication of network congestion. The fashion in which the buffer-full condition propagates to neighbors is analogous to the "backpressure bit" scheme for the virtual circuit (hop level) flow control employed in Tymnet [6].

V. Improvement by Transmission Acceleration

In this section we seek another way to take advantage of information about the network state. In addition to knowledge of the state of the immediate destination, let us assume that each unit with a packet knows the states of all hearing neighbors of the destination. Then, when the destination unit (which could be a sink terminal) and its neighbors (among which must not be a source terminal) are all in the empty state, it is foolish to flip a coin to decide whether to transmit or not. Since we know for sure there are no other transmissions around the destination, we simply raise the value of p to 1 and, with probability 1, transmit our packet. We call this operation, combined with the transmission suppression scheme described in Section IV, the transmission suppression/acceleration protocol. Of course, the above-mentioned information is not obtained for free; however, it is interesting to examine the resulting improvement in the throughput-delay performance as an ideal limit. The delay versus throughput curves based on this protocol are shown in Figs. 2 and 3 for networks 1 and 2, respectively. As expected, we see the increased capacity as well as the reduced delay for given throughput.

VI. Improvement by Multiple Buffers for Repeaters

So far, we have assumed that terminals and repeaters are equipped only with single packet buffers, and one of the conditions for successful transmission is that the buffer of the destination unit be empty (except for the final delivery to sink terminals). Therefore, some improvement in throughput-delay performance is expected by letting repeaters have more than one buffer. (Multiple buffers for terminals do not help performance in our model.) In this section, we examine this effect.

Let m be the number of buffers (each being capable of containing a single packet) provided for each repeater. The packets in buffers 1, 2, $\cdots$ form a queue waiting for transmission in this order. We express the state of repeater i, s_i, as

$$s_i = (s_i^{(1)}, s_i^{(2)}, \cdots, s_i^{(m)}) \tag{23}$$

where

$$s_i^{(l)} = \begin{cases} 0 & \text{buffer } l \text{ is empty} \\ k(\neq 0) & \text{buffer } l \text{ contains a packet of path } k \end{cases}$$
$$l = 1, 2, \cdots, m. \tag{24}$$

Note that the state $s_i^{(l)} = 0$ some l implies that $s_i^{(l')} = 0$ for those values of l' where $l < l' \leq m$.

Now a repeater having at least one packet transmits a packet in buffer 1 (say, of path k), with probability $p(k)$ and does not with probability $1 - p(k)$. In addition to "no collision," the condition for the successful transmission to a repeater is that there be at least one empty buffer at the receiver. If the transmission is successful, the transmitted packet joins on the tail of the queue (i.e., it is placed in the lowest numbered available buffer) at the receiver, while, at the transmitter (if it is a repeater), other packets, if any, are moved toward the head of the queue by one position. If the transmission is unsuccessful, the failed packet remains in buffer 1. In other aspects, the transmission protocols are the same as before.

For networks with multibuffer repeaters, we can still formulate a Markov chain problem and solve it numerically to obtain the delay versus throughput curves. Specifically, the formulation for the basic protocol parallels Section III with the following modification. For convenience, let $m = 1$ for terminals. First, in expressing the probability of network behavior $\boldsymbol{e}$ [defined by (4)] given the current state $\boldsymbol{s} = (s_1, s_2, \cdots, s_M)$ [now defined by (23) and (24)], we use $s_i^{(1)}$ (the state of buffer 1) instead of s_i in (5)–(7). Second, given the current state $\boldsymbol{s}$ and behavior $\boldsymbol{e}$ of the network, the next state $\boldsymbol{s}' = (s_1', s_2', \cdots, s_M')$ where $s_i' = (s_i'^{(1)}, s_i'^{(2)}, \cdots, s_i'^{(m)})$ is determined as follows: i) nontransmission by unit i does not affect the states of any other units, ii) the successful transmission from unit i to unit j of a packet of path k brings (let l_i be the lowest available buffer number at unit i given s_i)

$$s_j'^{(l_j)} = \begin{cases} 0, & j \text{ is a sink terminal of path } k \\ k, & j \text{ is a repeater} \end{cases}$$
$$s_j'^{(l)} = s_j^{(l)}, \qquad l = 1, 2, \cdots, l_j - 1$$
$$s_i'^{(l_i-1)} = 0, \quad s_i'^{(l)} = s_i'^{(l+1)}, \quad l = 1, 2, \cdots, l_i - 2 \tag{25}$$

and iii) the unsuccessful transmission by unit i results in $s_i' = s_i$. Last, given $P(\boldsymbol{s}'|\boldsymbol{s})$, the procedure to calculate the throughput and average delay is completely similar to (11)–(19) in Section III.

In Fig. 5 we plot the optimum delay curves for $m = 1, 2$, and 3 for network 1 without adjustment of transmission parameters. From this figure we see that the capacity of the network for $m = 2$ (0.21 packets per slot) is about a 30 percent increase over the single-buffer case. This is also accompanied by a reduced delay for a given throughput. However, the improvement by going from $m = 2$ to $m = 3$ is not as outstanding as increasing m from 1 to 2. A similar observation has been made by Tobagi [15] for some two-hop networks, where he comments that the lack of any important improvement experienced by increasing m is mainly explained by the fact that the system, at optimum, is mostly "channel bound" as opposed to "storage bound."

The effect of the transmission suppression/acceleration scheme described in Sections IV and V in the case of multibuffer repeaters is demonstrated in Fig. 6 for network 1 with $m = 2$. Here the transmission suppression is assumed to be in effect when all the buffers at the destination are occupied. The transmission acceleration takes place when all neighbors of the destination (which must not be source terminals) have empty buffers. In Fig. 6 we still see some improvement brought about by these schemes, although they are not as large as in the single-buffer case shown in Fig. 2.

VII. Conclusion

We have analyzed the throughput-delay characteristics for slotted-ALOHA multihop packet radio networks where the hearing configuration of packet radio units (terminals and repeaters) and source-to-sink paths of packets are given and fixed. The problems are formulated as discrete-time Markov chains and then solved numerically.

Besides the basic model—characterized by isolated trans-

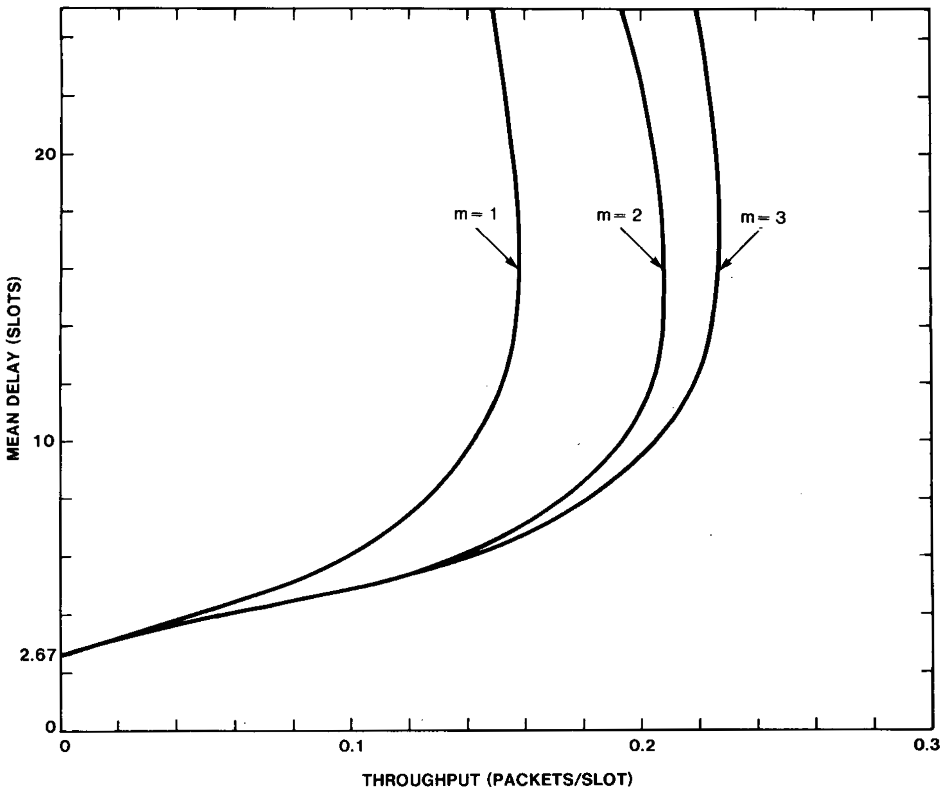

Fig. 5. Throughput–delay characteristics for network 1 (basic protocol) with m buffers for repeaters (m = 1, 2, 3).

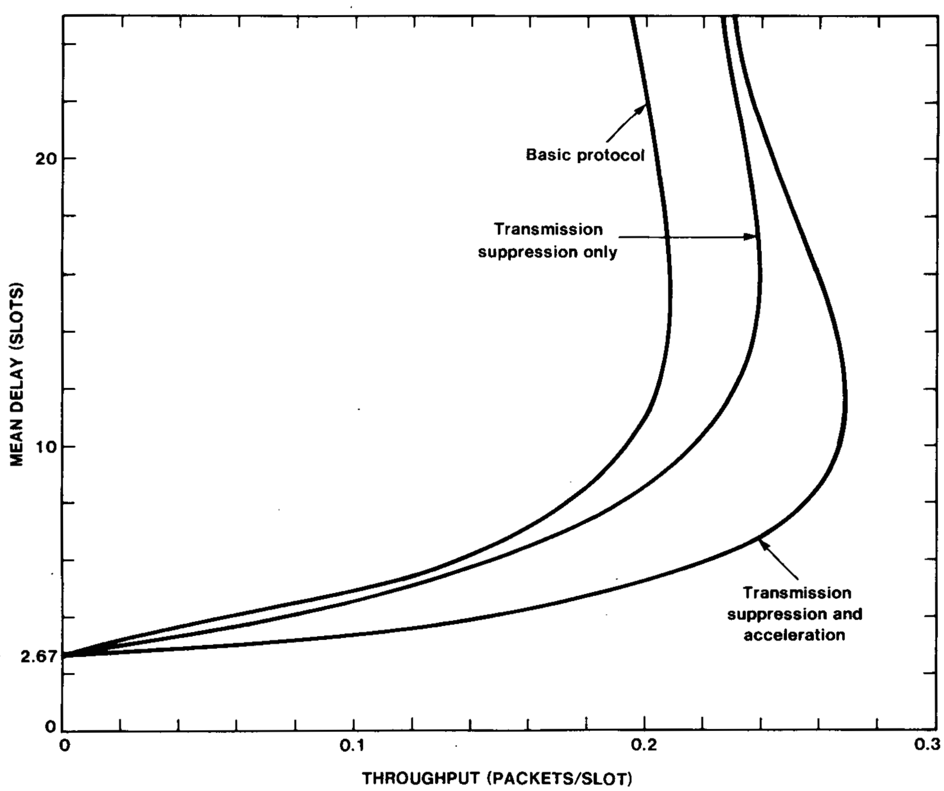

Fig. 6. Throughput–delay characteristics for network 1 with double-buffered repeaters.

TABLE I
THE NUMBER OF STATES AND THE SPARSENESS OF $P(s'|s)$ FOR SOME NETWORK MODELS DEPICTED IN FIG. 1. SUPPRESSION OR ACCELERATION OF TRANSMISSION IS NOT EMPLOYED. m = NUMBER OF BUFFERS IN EACH REPEATER.

cases		number	nonzero elements in $P(s'\|s)$	
network	m	of states	number	percentage
1	1	144	1190	5.74
1	2	1176	12424	0.89
1	3	7200	83412	0.16
2	1	3456	76377	0.64

mission behavior and single-buffered repeaters—three ways to improve the throughput-delay performance have been exploited. They include i) transmission suppression when the destination's buffer is occupied, ii) transmission acceleration when the buffers of all neighbors of the destination are empty, and iii) multiple buffers for repeaters.

It has been shown that the transmission suppression scheme provides natural flow control at the network access level to prevent packets from entering the "repeater subnet." This brings about significantly lower delay for a given throughput and achieves a much higher maximum throughput. The transmission accleration combined with appropriate suppression gives further improvement in the throughput-delay tradeoffs, at the cost of necessitating more information about the network state.

With more than one buffer for repeaters, we have fewer chances of failure of transmission due to buffer shortage at destinations. It has been shown that increasing the number of buffers from one to two offers more performance enhancement than going from two buffers to three. The effect of transmission suppression/acceleration in the multibuffer case has also been demonstrated.

Although the Markov chain approach used in this paper is not suitable for large-size networks (due to too much computational time and storage required), it may be useful for examining the effect of any particular heuristic protocol in prototype (small) network models. Also, it can provide a benchmark against which simulation models are validated.

Finally, we have some comments on the computational aspects involved in solving a large system of linear equations of the type as in (12). It turns out, in our problems, that most elements of the matrix are null; i.e., $P(s'|s)$ is a sparse matrix. In Table I we show the number of states and the percentage of nonzero elements in $P(s'|s)$ for some of the cases we have dealt with in this paper. We have used the Gauss-Seidel iteration method [5], which is time- and storage-efficient for the sparse matrix solution. As a matter of fact, in the last case of Table I (3456 states), it took less than 1 min to solve a system of equations of the form of (12) on a VAX-11/780 at the U.C.L.A. Department of Computer Science. So, it seems that a major time-consuming part in our calculation of throughputs and delays is now constructing the transition probability matrix $P(s'|s)$, which involves enumerating all possible events which can occur for every state of the network and determining the resulting next state for each of these events.

REFERENCES

[1] K. Bharath-Kumar, "Discrete-time queueing systems and their networks," *IEEE Trans. Commun.*, vol. COM-28, pp. 260-263, Feb. 1980.

[2] R. R. Boorstyn and A. Kershenbaum, "Throughput analysis of multihop packet radio," in *Proc. Int. Conf. Commun.*, Seattle, WA, June 8-12, 1980, 80CH1505-6, pp. 13.6.1-13.6.6.

[3] J. M. Brazio and F. A. Tobagi, "Theoretical results in throughput analysis of multihop packet radio networks," in *Proc. IEEE Int. Conf. Commun.*, Amsterdam, The Netherlands, May 14-17, 1984, 84CH2028-9, pp. 448-455.

[4] L. Fratta, M. Gerla, and L. Kleinrock, "The flow deviation method: An approach to store-and-forward communication network design," *Networks*, vol. 3, pp. 97-133, 1973.

[5] C. F. Gerald, *Applied Numerical Analysis*, 2nd ed. Reading, MA: Addison-Wesley, 1978.

[6] M. Gerla and L. Kleinrock, "Flow control: A comparative survey," *IEEE Trans. Commun.*, vol COM-28, pp. 553-574, Apr. 1980.

[7] R. E. Kahn, S. A. Gronemeyer, J. Burchfiel, and R. C. Kunzelman, "Advances in packet radio technology," *Proc. IEEE*, vol. 66, pp. 1468-1496, Nov. 1978.

[8] L. Kleinrock, "Performance of distributed multi-access computer-communication systems," in *Information Processing 77, Proc. IFIP Congr.* 77, Toronto, Ont., Canada, Aug. 8-12, 1977. New York: North-Holland, pp. 547-552.

[9] J. L. Kuester and J. H. Mize, *Optimization Techniques with Fortran*. New York: McGraw-Hill, 1973.

[10] K. C. Kung, "Flow deviation algorithm in a multihop packet radio network," Private commun., working paper 81006, ATS Group, Dep. Comput. Sci., Univ. California, Los Angeles, 1981.

[11] B. M. Leiner, "A simple model for computation of packet radio network communication performance," *IEEE Trans. Commun.*, vol. COM-28, pp. 2020-2023, Dec. 1980.

[12] J. D. C. Little, "A proof for the queueing formula: $L = \lambda W$," *Oper. Res.*, vol. 9, pp. 383-387, May-June 1961.

[13] M. Sidi and A. Segall, "A busy-tone-multiple-access-type scheme for packet-radio networks," in *Performance of Data Communication Systems and Their Applications*, G. Pujolle, Ed. New York: North-Holland, 1981, pp. 1-10.

[14] F. A. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part II—The hidden terminal problem in carrrier sense multiple-access and busy tone solution," *IEEE Trans. Commun.*, vol. COM-23, pp. 1417-1433, Dec. 1975.

[15] F. A. Tobagi, "Analysis of a two-hop centralized packet radio network—Part I: Slotted ALOHA," *IEEE. Trans. Commun.*, vol. COM-28, pp. 196-207, Feb. 1980.

★

Hideaki Takagi (S'80-M'83), for a photograph and biography, see p. 638 of the July 1985 issue of this TRANSACTIONS.

★

Leonard Kleinrock (S'55-M'64-SM'71-F'73), for a photograph and biography, see p. 638 of the July 1985 issue of this TRANSACTIONS.