AT&T BELL LABORATORIES TECHNICAL JOURNAL
Vol. 63, No. 1, January 1984
Printed in U.S.A.

Analysis of a Demand Assignment TDMA Blocking System

By S. M. BARTA* and M. L. HONIG†

(Manuscript received May 6, 1983)

This paper presents an analysis of a multichannel Time Division Multiple Access (TDMA) blocking system. Such a system is of interest for real-time voice-traffic applications. The effects of different traffic-assignment algorithms, traffic loads, number of channels, number of time slots, and number of traffic nodes on system performance are studied, where performance is measured by the probability that an incoming message will be blocked. An approximate analytical solution is found, the results of which compare exceedingly well with results obtained from computer simulation. Also derived is a rigorous lower bound on the blocking probability. Collectively, these results indicate that, for most systems of interest, blocking probability is insensitive to the assignment algorithm used. The performance of an assignment algorithm that is simplest to implement is therefore nearly optimal.

I. INTRODUCTION

A multichannel Time Division Multiple Access (TDMA) protocol provides an efficient means of sharing a high-capacity communication channel among a network of users. In a multichannel TDMA system, the aggregate channel capacity is partitioned in both the frequency and time domains. Each of several channels has some fraction of total bandwidth and consists of a series of time slots. A fixed number of channel time slots are combined to form the TDMA frame. There is an extensive literature describing and analyzing this protocol, espe-

* AT&T Information Systems.
† AT&T Bell Laboratories. Now a member of the Central Services Organization of the Regional Bell Operating Companies.

cially in the context of scanning beam communication satellite systems.[1-6]

The sequence of slot-by-slot switching configurations, which describes the origin and destination nodes of the traffic links assigned to each channel and time slot, is called a traffic assignment. Much attention has been focused on the problem of designing efficient traffic-assignment algorithms for the static case, i.e., where the assignment schedule does not change from frame to frame. However, because messages originate at random times and are of random duration, such a static assignment can be wasteful of bandwidth, since a time slot is unused during idle periods.

To overcome the inefficiency of a static assignment, a network controller can allocate channels and time slots to the traffic nodes according to instantaneous traffic needs. In this scheme, the switching configuration may change from frame to frame. This dynamic assignment of channel capacity is called Demand Assignment TDMA (DA/TDMA).

This paper presents an analysis of a multichannel DA/TDMA protocol. We consider a blocking system in which incoming traffic that cannot be immediately serviced is blocked (i.e., turned away). Only one type of incoming traffic is considered. Thus, this model is appropriate for a voice-traffic system.

We compare the blocking probability obtained using an optimal-assignment algorithm, which allows a complete reconfiguration of the switching pattern in each frame, with the blocking probability obtained using a fixed-assignment algorithm, which allows no rearrangement of existing traffic; i.e., in the fixed-assignment case, a message that requires more than one frame for transmission occupies the same channel and time slot in each frame. Notice that both the optimal- and fixed-assignment algorithms are more general than a static reservation scheme in which the switching configuration is the same in each frame. A tight lower bound on blocking probability obtained using an optimal-assignment algorithm is derived in addition to an accurate approximation for the blocking probability resulting from a type of fixed assignment called random assignment. Based upon our analytical results, we conclude that, for systems of moderate size, the blocking probabilities obtained using optimal- and fixed-assignment schemes are nearly identical. This is a significant finding since it implies that the complexity of optimal assignment is usually unnecessary.

The next section describes the multichannel DA/TDMA protocol and the traffic assignment problem. Section III contains a description of the network mathematical model used for the analysis, and a derivation of the equilibrium-state equations for the associated Markov process. Section IV computes the probability that an incoming-traffic request is blocked, and Section V presents comparisons of analytical with computer simulation results.

II. THE MULTICHANNEL ASSIGNMENT PROBLEM

This section describes the multichannel traffic-assignment problem. We start with a network consisting of communicating traffic nodes. The channel capacity in this case is partitioned both in the frequency and time domains. In particular, the bandwidth, B, of the transmission medium is divided into m channels, each having bandwidth B/m, and each channel consists of a series of time slots. A prespecified number of channel time slots are combined to form a transmission frame that continually repeats itself. The reservation multichannel TDMA protocol under consideration assumes a network controller that assigns time slots to incoming-traffic requests on a noninterfering basis.

Figure 1 shows one frame of a multichannel TDMA scheme with three channels and four time slots per frame. Each slot shows a traffic link assigned to that interval. The configuration shown in Fig. 1 will henceforth be referred to as a channel-time slot matrix. Denoting the number of channels by m and the number of time slots by c, this matrix will in general have m rows and c columns, and each entry will be a two-dimensional vector consisting of the transmitting and receiving nodes. Notice that this multichannel technique assumes that each node can transmit and/or receive on any (single) channel during a given time slot, and that the channel on which a node is transmitting or receiving can change over successive time slots. Furthermore, we also assume that the assigned channel and time slot for a given traffic link may change from frame to frame. In particular, the traffic link 1–2 shown in Fig. 1 may be reassigned from channel 1/time slot 2 to some other channel and time slot in subsequent frames.

The assignment of incoming traffic to available time slots is gov-

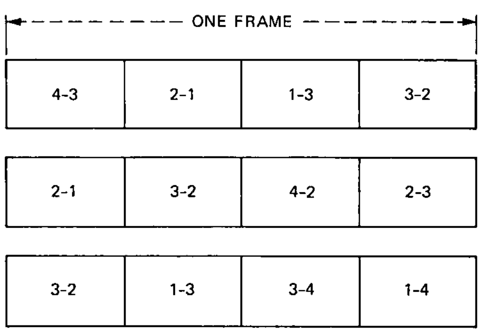

Fig. 1—One three-channel time division multiple access frame.

erned by the following *fundamental constraint*: one node may not transmit or receive on two different channels during the same time slot. In contrast to single-channel TDMA, in the multichannel case the network controller must not only determine whether a slot is being used, but must also determine whether a given traffic request can be assigned to that slot without violating the fundamental constraint. It therefore may not be possible to assign a given traffic request even though empty time slots exist.

Given unlimited computational capability, it may be desirable to rearrange traffic already assigned to the channel-time slot matrix in order to accommodate a new traffic arrival. It is therefore of interest to know under what conditions a set of traffic requests can be assigned. Denoting the number of traffic nodes by n, we define the $n \times n$ traffic matrix $\mathbf{T}$ as the matrix whose (i, j)th element contains the number of traffic-units node i is transmitting to node j (each traffic unit represents one packet and is assigned to one time slot). Given a traffic matrix, $\mathbf{T}$, and an empty channel-time slot matrix, all of the traffic can be assigned without violating the fundamental constraint if and only if the following *matrix constraints* are satisfied:[7]

$$\mathbf{T}\underline{1} \leq c\underline{1}, \tag{1a}$$

$$\underline{1}'\mathbf{T} \leq c\underline{1}' \tag{1b}$$

and

$$\underline{1}'\mathbf{T}\underline{1} \leq mc, \tag{1c}$$

where $\underline{1}$ is an n-dimensional vector whose elements are all unity, m denotes the number of channels, c denotes the number of time slots, and prime denotes transpose. These equations imply that no transmitting node requires more than c time slots, no receiving node requires more than c time slots, and that the total traffic demand is less than the total number (mc) of available time slots. To check whether it is possible to assign a new traffic request, one therefore need only check the matrix constraint set (1) where $\mathbf{T}$ contains traffic already assigned in addition to the new traffic request.

Because traffic requests are time varying, of ultimate interest is how to assign incoming traffic dynamically so as to minimize the number of times a new traffic request cannot be assigned without rearranging assigned traffic. This problem is quite difficult and has not yet been addressed in the open literature. The difficulty arises from the fact that future traffic requests are unknown, and hence, given any assignment rule, it is always possible to receive a sequence of traffic requests such that one can be assigned only if existing traffic is rearranged. Given a matrix $\mathbf{T}$ such that the constraint set (1) is satisfied, however, the static problem of efficiently assigning all of the traffic in $\mathbf{T}$ to an empty channel-time slot matrix has been addressed in Refs. 6 and 8 through 10. We also point out that these methods do not result in a unique assignment.

If we assume that a total rearrangement of existing traffic is allowed at the time each traffic request is made, then an optimal traffic assignment scheme can be inferred from the matrix constraint set (1). ("Optimal" in this case means that the probability of not being able to assign a new traffic request is minimized.) In particular, each time a new traffic request is made, the traffic matrix constraint set (1) is checked. If they are satisfied, then a "brute force" method for assigning the new traffic request would be to "empty out" the existing channel-time slot matrix and reassign all of this traffic along with the new traffic request via one of the methods suggested in Refs. 6 and 8 through 10. Certainly, this scheme requires much more computational power than necessary. If a new traffic request cannot be assigned to existing empty time slots, it is likely that very few (i.e., one or two) existing traffic assignments would have to be rearranged in order to assign the new traffic request.

If traffic assignments are to be made real time, as in a satellite system, the complexity of the assignment scheme becomes a crucial issue. The brute force optimal assignment scheme previously described would optimize system performance; however, under moderate to heavy loads, this scheme is likely to be impractical so that simpler assignment schemes, which yield suboptimal performance (i.e., a higher blocking probability), must be used. This raises the question of how much performance degradation is caused by using simpler assignment schemes rather than an optimal assignment scheme that permits an unlimited number of rearrangements.

The approach taken in this paper is to compare analytically the blocking probability obtained for a given system using an optimal assignment algorithm with the blocking probability obtained using fixed-assignment algorithms, which allow no rearrangement. Under a fixed-assignment algorithm, if a new traffic request cannot be assigned to a given channel-time slot matrix, it is blocked. Notice that to determine whether incoming traffic can be assigned when using an optimal-assignment algorithm, the traffic matrix must be examined, whereas when using a fixed-assignment algorithm, the channel-time slot matrix must be examined.

III. MATHEMATICAL FORMULATION

3.1 Traffic model

The purpose of this subsection is to specify the mathematical model used to generate the analytical results in the following sections. The incoming traffic is modeled as the sum of independent Poisson processes flowing between each pair of traffic nodes. We therefore have an arrival rate matrix, Λ, whose (i, j)th element is the Poisson rate at which messages are transmitted from node i to node j. The flow rates for traffic between each pair of nodes are assumed to be identical, i.e., $\Lambda = \lambda' \underline{1}\underline{1}'$ where λ' is a constant. The total traffic arrival rate, λ, is therefore the sum of the rates between each pair of nodes.

Because the traffic out of each node is assumed to be independent of the traffic out of all other nodes, given a new traffic request, the probability that the request originated from a specific node i is $1/n$ and, similarly, the probability that the destination is node j is also $1/n$. Because the transmission processes between each pair of nodes are independent, the probability that a given traffic request originates from node i *and* is sent to node j is $1/n^2$.

If a traffic request can be assigned, it occupies one slot of the channel-time slot matrix for a random amount of time, depending on the message length. This "service" time is assumed to be exponentially distributed. Associated with the incoming traffic is therefore the exponential service rate, μ. For analytical convenience, we assume that this distribution is continuous in the sense that departures from the channel-time slot matrix can occur at any time instant. Notice, however, that for a real system, departures occur only at the end of a given frame. The corresponding service time distributions must therefore be "discretized" since service times must be an integer numbers of frames. This effect becomes negligible, however, if the average service time consists of a large number of frames (i.e., >100).

3.2 Derivation of equilibrium-state equations for optimal assignment

The derivations of the state equations for the Markov processes associated with optimal and fixed assignments are virtually identical. We present, therefore, only the details of the derivation for optimal assignment and then indicate how to modify that derivation to obtain the state equations for fixed assignment.

When an optimal assignment algorithm is used, the traffic matrix $\mathbf{T}$ determines whether an incoming traffic request is blocked. Because the traffic arrival process is Poisson and the service times are exponential, the evolution of the traffic matrix $\mathbf{T}$ is described by a Markov process. The set of states for the Markov process associated with optimal assignment is, therefore, the set of $n \times n$ matrices $\mathbf{T}$ with integer elements that satisfy the constraint set (1). The following notation is helpful in defining the transition rates for the process:

$$S \equiv \{\mathbf{T} \mid \mathbf{T} \text{ satisfies (1a) through (1c)}\} \tag{2a}$$

$$S_k \equiv \{\mathbf{T} \mid \mathbf{T} \in S,\ \underline{1}'\mathbf{T}\underline{1} = k\} \tag{2b}$$

$$S_{\mathbf{T}}^- \equiv \{\mathbf{T}^* \mid \mathbf{T}^* \in S,\ \exists(i, j) \text{ such that } \mathbf{T}^* = \mathbf{T} - \underline{e}_i\underline{e}_j'\} \tag{2c}$$

$$S_{\mathbf{T}}^{+} \equiv \{\mathbf{T}^* \mid \mathbf{T}^* \in S,\ \exists(i, j) \text{ such that } \mathbf{T}^* = \mathbf{T} + \underline{e}_i \underline{e}_j'\}, \tag{2d}$$

where $\underline{e}_i$ is the $n \times 1$ vector with 1 in the ith place and 0 elsewhere. The set S_k is the set of all states such that there are k messages in the system; $S_{\mathbf{T}}^{-}$ and $S_{\mathbf{T}}^{+}$ are the sets of states into which the process makes transitions via departures and arrivals, respectively, given that the process is in state $\mathbf{T}$. Notice that if $\mathbf{T}^* \in S_{\mathbf{T}}^{-}$, then $\mathbf{T} \in S_{\mathbf{T}^*}^{+}$ and, similarly, if $\mathbf{T}^* \in S_{\mathbf{T}}^{+}$, then $\mathbf{T} \in S_{\mathbf{T}^*}^{-}$.

Let $r(\mathbf{T}^1, \mathbf{T}^2)$ be the transition rate† between any two states $\mathbf{T}^1$ and $\mathbf{T}^2$. Notice that a transition from state $\mathbf{T}^1$ to a state $\mathbf{T}^2$ can occur only via a single arrival or a single departure. The rate $r(\mathbf{T}^1, \mathbf{T}^2)$ is, therefore, nonzero if and only if $\mathbf{T}^1 = \mathbf{T}^2$, or $\mathbf{T}^1$ and $\mathbf{T}^2$ differ from each other in exactly one element and that difference is one. Thus,

$$r(\mathbf{T}^1, \mathbf{T}^2) = \begin{cases} t_{ij}^1 \mu, & \text{if } \mathbf{T}^2 \in S_{\mathbf{T}^1}^{-} \\ \lambda/n^2, & \text{if } \mathbf{T}^2 \in S_{\mathbf{T}^1}^{+} \\ 1 - \sum_{i,j} t_{ij}^1 \mu - \dfrac{\lambda}{n^2} \, |S_{\mathbf{T}^1}^{+}|, & \text{if } \mathbf{T}^1 = \mathbf{T}^2 \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where $\mathbf{T}^1$ and $\mathbf{T}^2$ differ in the (i, j)th element, t_{ij}^1 is the (i, j)th element of $\mathbf{T}^1$, $1/\mu$ is the mean service time for messages, λ/n^2 is the expected rate at which messages between nodes i and j are generated, and $|A|$ denotes the number of elements in the set A.

Because the performance of a TDMA assignment algorithm is measured in terms of steady-state network behavior, we focus on the limiting steady-state distribution of the state variable $\mathbf{T}$. To prove that the steady-state distribution exists, note that the state space consists of matrices with integer elements that satisfy the constraint set (1), and it is therefore finite. The transition rates given in (3) do not depend explicitly on time, so the process is time homogeneous. Finally, it can be shown the process is irreducible and contains no periodicities. Therefore, the steady-state distribution exists. Although we obtain a closed-form expression for the steady-state distribution in Appendix A, it can be explicitly computed only in very few cases. We therefore develop an alternative method of analyzing the system, which does not require the explicit formula. This approach, moreover, provides valuable insight into our problem and allows us to provide a unified treatment of both optimal- and fixed-assignment algorithms.

The next step is to derive the equilibrium equations that the steady-state probability distribution $p(\mathbf{T})$ must satisfy. Using the notation introduced earlier, we have the flow equation[11]

---

†The function $r(\cdot\,, \cdot)$ is technically the infinitesimal generator of the Markov process associated with optimal assignment.

$$\sum_{\substack{\mathbf{T}'\in S\\ \mathbf{T}'\neq T}} p(\mathbf{T})r(\mathbf{T}, \mathbf{T}') = \sum_{\substack{\mathbf{T}'\in S\\ \mathbf{T}'\neq \mathbf{T}}} p(\mathbf{T}')r(\mathbf{T}', \mathbf{T}), \qquad \mathbf{T} \in S. \tag{4}$$

The intuitive meaning of (4) is that for every state **T**, the probability flow rate out of **T** must equal the probability flow rate into **T**.

The following lemma may be used to derive an alternative form for (4). Let $|A|$ represent the number of elements in the set A and let $p(f\,|\,\mathbf{T})$ denote the probability, conditioned on state **T**, that an arrival, chosen from the uniform distribution of origin-destination pairs, does not violate the set of matrix constraints; i.e., $p(f\,|\,\mathbf{T})$ is the probability that an arrival "fits" (can be assigned). Then we have

Lemma 1: The probability flow out of state $\mathbf{T} \in S$ *satisfies*

$$\sum_{\substack{\mathbf{T}'\neq \mathbf{T}\\ \mathbf{T}'\in S}} p(\mathbf{T})r(\mathbf{T}, \mathbf{T}') = \sum_{\mathbf{T}'\in S_{\mathbf{T}}^{-}} p(\mathbf{T})r(\mathbf{T}, \mathbf{T}') + \sum_{\mathbf{T}'\in S_{\mathbf{T}}^{+}} p(\mathbf{T})r(\mathbf{T}, \mathbf{T}') \tag{5a}$$

$$= \mu k p(\mathbf{T}) + \lambda p(f\,|\,\mathbf{T})p(\mathbf{T}), \tag{5b}$$

where $\underline{1}'\mathbf{T}\underline{1} = k$, *which equals the number of occupied time slots, and*

$$p(f\,|\,\mathbf{T}) = \frac{|S_{\mathbf{T}}^{+}|}{n^2}. \tag{6}$$

The proof of this lemma appears in Appendix B.

This lemma shows that the flow out of any state **T** is composed of two terms: the first term in (5b), $\mu k p(\mathbf{T})$, reflects transitions that occur because there is a departure, while the second term, $\lambda p(f\,|\,\mathbf{T})p(\mathbf{T})$, reflects transitions that occur because an admissible request is generated.

Applying Lemma 1 to the left side of (4) and noting that, if $\mathbf{T}' \neq \mathbf{T}$, $r(\mathbf{T}', \mathbf{T}) \neq 0$ if and only if $\mathbf{T}' \in S_{\mathbf{T}}^{-}$ or $\mathbf{T}' \in S_{\mathbf{T}}^{+}$, we obtain the flow equation

$$\mu k p(\mathbf{T}) + \lambda p(f\,|\,\mathbf{T})p(\mathbf{T})$$
$$= \sum_{\mathbf{T}'\in S_{\mathbf{T}}^{-}} p(\mathbf{T}')r(\mathbf{T}', \mathbf{T}) + \sum_{\mathbf{T}'\in S_{\mathbf{T}}^{+}} p(\mathbf{T}')r(\mathbf{T}', \mathbf{T}), \qquad \mathbf{T} \in S_k. \tag{7}$$

Equation (7) describes the probabilistic flows into and out of each state. In principle, (7) can be solved to yield a closed form solution for the steady-state distribution of the state variable, $p(\mathbf{T})$ (see Appendix A). Unfortunately, it is very difficult to evaluate this expression for most cases of interest. In order to calculate the system blocking probability, P_B, however, it is unnecessary to evaluate the steady-state distribution of **T**. It is shown in the next section that it is enough to know only certain aggregate quantities depending only on the number of messages in the system, k. Notice that the "state" k is an aggregate state composed of all traffic matrices $\mathbf{T} \in S_k$. We therefore are interested in deriving an analogous equation to (7) that describes the steady-state flows between the sets of states S_k, $k = 0, 1, 2, \cdots$. A direct technique for obtaining this equation is to sum both sides of (7) over the set $\mathbf{T} \in S_k$.

Let $p_O(k)$ be the steady-state probability that the state of the system is contained in S_k, which is the probability that there are k occupied time slots; and let $p_O(f\,|\,k)$ be, in analogy with $p(f\,|\,\mathbf{T})$, the probability of a fit given that the system contains k messages, i.e.,

$$p_O(k) \equiv \sum_{\mathbf{T} \in S_k} p(\mathbf{T}) \tag{8a}$$

$$p_O(f\,|\,k) \equiv \frac{\sum_{\mathbf{T} \in S_k} p(f\,|\,\mathbf{T})p(\mathbf{T})}{p_O(k)}, \tag{8b}$$

where the O subscript indicates that an optimal assignment algorithm is assumed. The following theorem gives the flow equations for the distribution $p_O(k)$.

Theorem 1: The probabilities $\{p_O(k)\}$ *satisfy the equations*

$$[\mu k + \lambda p_O(f\,|\,k)]p_O(k) = \mu(k+1)p_O(k+1) + \lambda p_O(f\,|\,k-1)p_O(k-1), \tag{9}$$

where $p_O(k) = 0$ for $k < 0$ or $k > mc$.

The proof of this theorem is given in Appendix C.

Equation (9) expresses the equality of probability flows between the sets S_k. Given that the system is in state $\mathbf{T}$, the next transition can only be into a state $\mathbf{T}' \in S_{\mathbf{T}}^-$ or $\mathbf{T}' \in S_{\mathbf{T}}^+$. Suppose, for example, that $\mathbf{T} \in S_k$; then $\mathbf{T}' \in S_{k-1}$ or $\mathbf{T}' \in S_{k+1}$; i.e., transitions out of any state in S_k are always into a state in S_{k-1} or S_{k+1}. Note that (9) is a set of "birth-death" equations, which we will solve in Section IV to give an expression for $\{p_O(k)\}$ in terms of λ, μ, and $\{p_O(f\,|\,k)\}$. (These equations do not, however, imply that the k process is Markov.) In contrast to the solution of (7) we show in Section IV that it is possible to obtain a close approximation to the solution of (9) that can be easily evaluated.

3.3 Equilibrium state equations for fixed assignment

To derive the flow equations [corresponding to (5), (7), and (9)] for the Markov process defined by a fixed-assignment algorithm, we redefine the set of states as the set of channel-time slot matrices $\mathbf{M}$ defined in Section II. The following notation is analogous to (2):

$$L \equiv \{\mathbf{M}\,|\,\mathbf{M} \text{ is an admissible channel-time slot matrix}\} \tag{10a}$$

$$L_k \equiv \{\mathbf{M}\,|\,\mathbf{M} \in L, \mathbf{M} \text{ contains } k \text{ units of traffic}\} \tag{10b}$$

$L_{\mathbf{M}}^{-} \equiv \{\mathbf{M}^* \mid \mathbf{M}^* \in L, \exists (i-j) \text{ and } (m^*, c^*) \text{ such that } \mathbf{M}^* = \mathbf{M} - (i-j)\underline{e}_{m^*}\underline{e}_{c^*}'\}$ (10c)

$L_{\mathbf{M}}^{+} \equiv \{\mathbf{M}^* \mid \mathbf{M}^* \in L, \exists (i-j) \text{ and } (m^*, c^*) \text{ such that } \mathbf{M}^* = \mathbf{M} + (i-j)\underline{e}_{m^*}\underline{e}_{c^*}'\},$ (10d)

where $(i-j)$ represents one unit of traffic from node i to node j, $\underline{e}_{m^*}$ is the $m \times 1$ vector with 1 in the m^*th place and 0 elsewhere, e_{c^*} is the $c \times 1$ vector with 1 in the c^*th place and 0 elsewhere, and the notation $\mathbf{M}^* = \mathbf{M} + (i-j)\underline{e}_{m^*}\underline{e}_{c^*}'$ means that the channel-time slot matrices $\mathbf{M}^*$ and $\mathbf{M}$ differ from each other only in the (m^*, c^*) position. A + sign indicates that $\mathbf{M}^*$ contains the traffic pair $(i-j)$ in the (m^*, c^*) position, whereas the (m^*, c^*) position in $\mathbf{M}$ is empty; the − sign indicates that $\mathbf{M}$ contains the traffic pair $(i-j)$ in the (m^*, c^*) position, whereas the (m^*, c^*) position in $\mathbf{M}^*$ is empty.

Let $r(\mathbf{M}^1, \mathbf{M}^2)$ be the transition rate between any two states $\mathbf{M}^1$ and $\mathbf{M}^2$.[†] Then we have

$$\mathrm{r}(\mathbf{M}^1, \mathbf{M}^2) = \begin{cases} \mu, & \text{if } \mathbf{M}^2 \in L_{\mathbf{M}^1}^{-} \\ \lambda(\mathbf{M}^1, \mathbf{M}^2), & \text{if } \mathbf{M}^2 \in L_{\mathbf{M}^1}^{+} \\ 1 - |\mathbf{M}^1|\mu - \sum_{\mathbf{M}^2 \in L_{\mathbf{M}^1}^{+}} \lambda(\mathbf{M}^1, \mathbf{M}^2), & \text{if } \mathbf{M}^1 = \mathbf{M}^2 \\ 0, & \text{otherwise,} \end{cases} \quad (11)$$

where $|\mathbf{M}^1|$ is the total number of traffic pairs in $\mathbf{M}^1$. The transition rate $\lambda(\mathbf{M}^1, \mathbf{M}^2)$ is the rate at which a new arrival $(i-j)$ is assigned to $\mathbf{M}^1$, thereby changing $\mathbf{M}^1$ to $\mathbf{M}^2$. Let $a_{\mathbf{M}^1}^{(i-j)}$ denote the number of slots in $\mathbf{M}^1$ into which the arrival $(i-j)$ can be assigned. If assignments are made by selecting one out of the $a_{\mathbf{M}^1}^{(i-j)}$ available slots from a uniform distribution, then

$$\lambda(\mathbf{M}^1, \mathbf{M}^2) = \frac{\lambda}{n^2 a_{\mathbf{M}^1}^{(i-j)}},$$

where λ is the total traffic-arrival rate. This fixed-assignment scheme, where traffic arrivals are randomly assigned to available time slots according to a uniform distribution, will henceforth be referred to as random assignment. It is not necessary, however, to assume random assignment in order to derive the flow equation that follows.

We denote the steady-state probability distribution of the Markov process for fixed assignment as $p(\mathbf{M})$. The existence proof for $p(\mathbf{M})$ is the same as that for $p(\mathbf{T})$. Also, let $p(f \mid \mathbf{M})$ denote the conditional probability, given that the state is $\mathbf{M}$, that an arrival chosen from the

[†] The function $r(\cdot\,,\cdot)$ is the infinitesimal generator for the associated fixed assignment process.

uniform distribution of origin-destination pairs can be assigned without reassigning any existing traffic, i.e., the probability of a fit under the fixed-assignment rule. Finally, we define

$$A_{\mathbf{M}} \equiv \{(i - j) \,|\, a_{\mathbf{M}}^{(i-j)} > 0\} \qquad (12)$$

as the set of all traffic pairs which can be assigned to $\mathbf{M}$.

Proceeding as in the case of optimal assignment, we can derive the flow equations for the fixed-assignment case by merely changing notation. The probability of a fit, $p(f\,|\,\mathbf{M})$, in this case is simply the number of traffic pair arrivals that can be assigned under the fixed assignment rule divided by the total number of possible pairs. Thus, in analogy with (6) we have

$$p(f\,|\,\mathbf{M}) = \frac{|A_{\mathbf{M}}|}{n^2}, \qquad (13)$$

where $A_{\mathbf{M}}$ is defined in (12). In analogy with the flow equations (7) and (9) for optimal assignment, the fixed-assignment flow equations are

$$\mu k p(\mathbf{M}) + \lambda p(f\,|\,\mathbf{M})p(\mathbf{M}) = \sum_{\mathbf{M}' \in L_{\mathbf{M}}^-} p(\mathbf{M}')r(\mathbf{M}', \mathbf{M})$$
$$+ \sum_{\mathbf{M}' \in L_{\mathbf{M}}^+} p(\mathbf{M}')r(\mathbf{M}', \mathbf{M}), \qquad \mathbf{M} \in L_k, \quad k = 1, \cdots, mc, \qquad (14)$$

and

$$[\mu k + \lambda p_F(f\,|\,k)]p_F(k) = \mu(k+1)p_F(k+1)$$
$$+ \lambda p_F(f\,|\,k-1)p_F(k-1), \qquad k = 1, \cdots, mc, \qquad (15)$$

where

$$p_F(k) \equiv \sum_{\mathbf{M} \in L_k} p(\mathbf{M}) \qquad (16a)$$

$$p_F(f\,|\,k) \equiv \frac{\sum_{\mathbf{M} \in L_k} p(f\,|\,\mathbf{M})p(\mathbf{M})}{p_F(k)}, \qquad (16b)$$

and the F subscript indicates that a fixed-assignment algorithm is assumed. Note that (15) is, like (9), a system of birth-death equations.

IV. PERFORMANCE ANALYSIS

4.1 Properties of blocking probabilities

The performance of a multichannel TDMA blocking system is measured in terms of the steady-state probability that an arrival is blocked or lost. Of course, this blocking probability depends on the assignment algorithm. Conditioned on the state of the system, the blocking probability is simply $1 - p(f|\mathbf{T})$ or $1 - p(f|\mathbf{M})$, for optimal or fixed assignment, respectively. The unconditional blocking probability is given by

$$1 - P_B^O = \sum_{\mathbf{T}\in S} p(f|\mathbf{T})p(\mathbf{T}) \tag{17}$$

or

$$1 - P_B^F = \sum_{\mathbf{M}\in L} p(f|\mathbf{M})p(\mathbf{M}), \tag{18}$$

where P_B^O and P_B^F are blocking probabilities using respectively optimal- and fixed-assignment algorithms. As discussed in Section III and Appendix A, it is extremely difficult to compute $p(\mathbf{T})$ for most systems of interest, and hence (17) and (18) cannot be used directly to compute P_B^O and P_B^F. In this section, however, we derive bounds and approximations for (17) and (18) by using the aggregate-state equations (9) and (15). This provides a means of quantitively comparing assignment algorithms and estimating the performance of the system as parameters such as n, m, and c vary.

We obtain equivalent expressions for the blocking probabilities as follows:

$$\begin{aligned} 1 - P_B^O &= \sum_{\mathbf{T}\in S} p(f|\mathbf{T})p(\mathbf{T}) \\ &= \sum_{k=1}^{mc} \sum_{\mathbf{T}\in S_k} p(f|\mathbf{T})p(\mathbf{T}) \\ &= \sum_{k=1}^{mc} p_O(f|k)p_O(k), \end{aligned} \tag{19}$$

where $p_O(f|k)$ and $p_O(k)$ satisfy (8); similarly P_B^F satisfies

$$1 - P_B^F = \sum_{k=1}^{mc} p_F(f|k)p_F(k), \tag{20}$$

where $p_F(f|k)$ and $p_F(f)$ satisfy (16).

Equations (19) and (20) are particular cases of the following general expression for blocking probability;

$$1 - P_B = \sum_{k=1}^{mc} \phi_k p(k), \tag{21}$$

where ϕ_k and $p(k)$ satisfy the birth-death system of equations

$$[\mu k + \lambda\phi_k]p(k) = \mu(k+1)p(k+1) + \lambda\phi_{k-1}p(k-1), \tag{22}$$

subject to the boundary conditions $\phi_{mc} = 0$ and $p(k) = 0$ for $k < 0$ or $k > mc$; and $\Sigma_k\, p(k) = 1$. Solving (22) and substituting into (21) yields an expression for the blocking probability in terms of the ϕ_k's; i.e.,

$$p(k) = \frac{\frac{1}{k!} \rho^k \prod_{j=0}^{k-1} \phi_j}{\sum_{i=0}^{mc} \frac{1}{i!} \rho^i \prod_{j=0}^{i-1} \phi_j}, \tag{23}$$

and from (21)

$$1 - P_B = \frac{\sum_{k=0}^{mc-1} \frac{1}{k!} \rho^k \prod_{j=0}^{k} \phi_j}{\sum_{i=0}^{mc} \frac{1}{i!} \rho^i \prod_{j=0}^{i-1} \phi_j}, \tag{24}$$

where $\rho \equiv \lambda/\mu$. If $\phi_k = p_O(f \mid k)$ or $\phi_k = p_F(f \mid k)$, then $P_B = P_B^O$ or $P_B = P_B^F$, respectively. Another case of particular interest is where $\phi_k = 1$ if $0 \leqslant k < mc$ and $\phi_k = 0$ otherwise. We denote this set of probabilities, which corresponds to the mc server Erlang-loss system, as $p_E(f \mid k)$. The expression for blocking probability, P_B^E, is called Erlang's loss or B formula. The Erlang formula applies to a single-channel TDMA system where only the constraint (1c) in (1) must hold. An important property of P_B stated in the following lemma, which may be used to derive bounds on assignment algorithm performance, is that P_B is a monotonically nonincreasing function of ϕ_k for $k = 0, 1, \cdots, mc$.

Lemma 2: The blocking probability $P_B \equiv P_B(\phi_0, \cdots, \phi_{mc})$ satisfies

$$\frac{\partial P_B}{\partial \phi_k} \leqslant 0, \qquad k = 0, \cdots, mc. \tag{25}$$

The proof of this lemma appears in Appendix D.

It is intuitively reasonable that the blocking probability should increase as more system constraints are added. This is indeed the case, as the next theorem shows.

Theorem 2: The optimal assignment, fixed assignment, and Erlang blocking probabilities satisfy

$$P_B^E \leqslant P_B^O \leqslant P_B^F. \tag{26}$$

Proof: If the system is in state k, then an arrival can be assigned by a fixed-assignment algorithm only if it can be assigned by an optimal algorithm and so, for all k, we have $p_F(f \mid k) \leqslant p_O(f \mid k)$. It is clear that $p_O(f \mid k) \leqslant p_E(f \mid k)$, for all k. Therefore, (25) implies (26). Q.E.D.

As the ratio of the number of nodes n to the number of channels m increases, it becomes less likely that an arrival will match any of the traffic pairs already assigned to a given time slot. Consequently, for large n/m, we expect that if there is an open slot, it should be relatively easy to assign a new traffic request. To be more precise, as n/m increases, the system performance approaches that of an Erlang system.

Theorem 3:
$$\lim_{\frac{n}{m}\to\infty} P_B^F = \lim_{\frac{n}{m}\to\infty} P_B^O = P_B^E. \tag{27}$$

Proof: Theorem 2 implies that it is sufficient to show that $P_B^F \to P_B^E$. For all $k < (mc - 1)$ we have $p_F(f|k) \geqslant p_F(f|mc - 1)$. Because the nodes are symmetric, the probability of a fit given one empty slot is:

$$p_F(f|mc-1) = \left(\frac{n - m + 1}{n}\right)^2,$$

which is the probability that the origin and destination nodes of an arrival chosen from a uniform distribution do not match $(m - 1)$ randomly chosen origin-destination pairs. Thus, we have

$$\lim_{\frac{n}{m}\to\infty} p_F(f|k) = \begin{bmatrix} 1, & k < mc \\ 0, & k = mc, \end{bmatrix}$$

which is $p_E(f|k)$. Because (24) is a continuous function of the ϕ_k's, we have $\lim_{n/m\to\infty} P_B^F = P_B^E$. Q.E.D.

Although it is easy to compute the Erlang lower bound P_B^E, it is extremely difficult to calculate P_B^O or P_B^F. The difficulty lies in the calculation of $p_F(f|k)$ and $p_O(f|k)$ as functions of k. In particular, notice from (8b) and (16b) that the state probabilities $p(\mathbf{T} \in S_k)$ and $p(\mathbf{M} \in L_k)$ must be known. The aggregate flow eqs. (9) and (15) have therefore not made the exact computation of P_B^O and P_B^F any easier. However, the advantage in using these equations is that we can derive an accurate approximation for $p_F(f|k)$, thereby enabling the calculation of an approximation for the corresponding blocking probability.

4.2 Approximate calculation of blocking probability with random assignment

In order to approximate the blocking probability resulting from fixed assignment, we first approximate $p_F(f|k)$ and subsequently substitute the resulting expression for ϕ_k in (24). Let $\underline{k} \equiv (k_1, \cdots, k_c)$ be the vector of time slot occupancy; i.e., there are k_i units of traffic in the ith column of the channel-time slot matrix. The probability of a fit can be expressed as

$$p_F(f|k) = \sum_{\underline{k}\in\Omega_k} p_F(f|\underline{k})p_F(\underline{k}|k), \tag{28}$$

where Ω_k is the set of all occupancy vectors $\underline{k}$ such that $\Sigma_i k_i = k$, $p_F(f|\underline{k})$ is the conditional fit probability given that the occupancy vector is $\underline{k}$, and $p_F(\underline{k}\,|\,k)$ is the conditional probability that the occupancy vector is $\underline{k}$ given there are k units of traffic in the system.

To calculate $p_F(f\,|\,\underline{k})$, note that: (1) the k_i traffic units in time slot i are characterized by k_i pairs of integers between 1 and n, which satisfy the fundamental constraint, and because the traffic between nodes is assumed to be symmetric, these integer pairs are equally likely; (2) if $k_i = m$, then any new requests cannot be assigned to the ith time slot. We also use the following assumption; and (3) information about traffic in time slot i provides no information about traffic in a different time slot j (traffic in different slots is independent). This assumption is certainly very accurate for a large number of nodes; however, it is quite difficult to prove or disprove in general. An arriving unit of traffic can be assigned to time slot i if and only if its origin does not match any of the k_i origins and its destination does not match any of the k_i destinations of traffic already assigned to column i. Observation (1) implies that this event has probability $[(n - k_i)/n]^2$. A unit of traffic cannot be assigned, i.e., does not fit, if and only if it does not fit into any time slot. Thus, using observations (2) and (3), we have that the probability of no fit is the product

$$\prod_{\{i\,|\,k_i<m\}} \left[1 - \left(\frac{n - k_i}{n}\right)^2\right],$$

and hence, the probability of a fit, given $\underline{k}$, is

$$p_F(f\,|\,\underline{k}) = 1 - \prod_{\{i\,|\,k_i<m\}} \left[1 - \left(\frac{n - k_i}{n}\right)^2\right], \tag{29}$$

where the product is assumed to be one if $\{i\,|\,k_i < m\}$ is the empty set.

Intuition suggests that given k units of traffic in the system, the occupied slots, when random assignment is used, are uniformly distributed throughout the channel-time slot matrix. That is,

$$p_F(\underline{k}\,|\,k) = \frac{\prod_{i=1}^{c} \begin{bmatrix} m \\ k_i \end{bmatrix}}{\begin{bmatrix} mc \\ k \end{bmatrix}}, \tag{30}$$

where the numerator is the number of ways to have k_i units of traffic in slot i ($i = 1, \cdots, c$) and the denominator is the total number of ways to have k units of traffic in the system. The following example shows, however, that this intuition is misleading and (30) is not the correct distribution.

Consider a system with $n = 3$, $m = 2$, $c = 2$. Figure 2 is a transition diagram illustrating transition rates into the states corresponding to $k = 2$. The ordered pairs, e.g., (1, 1), represent the occupancy vectors

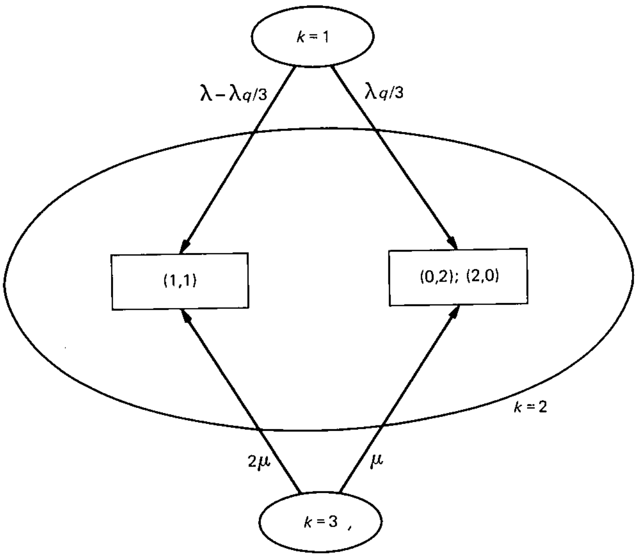

Fig. 2—Transition diagram for example with two channels, two time slots, and three nodes.

$\underline{k}$ for channel-time slot matrices containing two units of traffic, and q is the probability that an arriving unit of traffic can be assigned to the column containing the single unit already assigned. Using (30), it is easily verified that the uniformly distributed assumption implies that

$$p_F[(1, 1) \mid k = 2] = 2\{p_F[(0, 2) \mid k = 2] + p_F[(2, 0) \mid k = 2]\}. \quad (31)$$

It is easy to show by writing the flow equation and using the fact $p_F(\underline{k}) = p_F(\underline{k} \mid k)p(k)$ that (31) can be satisfied only if the rate from $k = 1$ into (1, 1) is twice the rate into the pair of states (0, 2) and (2, 0), i.e., only if $\lambda - \lambda q/3 = 2\lambda q/3$. But the preceding equation can be satisfied only if $q = 1$, which is clearly not true for this system. Therefore, (30) is not correct and the assumption of uniformly distributed occupancy is not strictly true. We may, however, approximate $p_F(\underline{k} \mid k)$ by (30), which when combined with (28) and the expression for $p_F(f \mid \underline{k})$, (29), gives an approximation to $p_F(f \mid k)$ when random assignment is used. That approximation may, in turn, be used in (24) to provide an approximation to P_B^F.

Intuitively, one would expect that the approximation for $p_F(f \mid k)$ given by (28) through (30) is accurate for the case of random assignment. In this case (30) becomes more accurate as the ratio of the number of nodes to the number of channels n/m increases. This is because the probability that a new arrival can be assigned to a randomly picked empty time slot increases with n/m. The selection of empty slots in the channel-time slot matrix to which new arrivals are assigned therefore becomes less biased. The simulations in Section V indicate that in fact our approximation is extremely accurate for relatively small systems (i.e., four channels, five time slots, and ten nodes). Fixed-assignment schemes other than random assignment are possible, however, where the approximation (30) may not be accurate. For example, it may be desirable to pack the assigned traffic as closely as possible to the left or to the right of the channel-time slot matrix to increase the probability that some column has a relatively large number of empty slots. For this case, the distribution of slot patterns $p(\underline{k}\,|\,k)$ may be significantly different from the distribution resulting from random assignment.

An upper bound on the blocking probability obtained using any fixed-assignment scheme can be derived in principle by calculating a lower bound on the fit probability, $p_F(f\,|\,k)$. From (28),

$$p_F(f\,|\,k) = \sum_{\underline{k}\in\Omega_k} p_F(f\,|\,\underline{k})p_F(\underline{k}\,|\,k)$$

$$\geq \min_{\underline{k}\in\Omega_k} p_F(f\,|\,\underline{k}) \equiv p_F^{MIN}(f\,|\,k). \tag{32}$$

A lower bound on $p_F(f\,|\,k)$ is therefore obtained by assuming that traffic already assigned is arranged in the configuration which minimizes the probability that a new arrival can be assigned. From Lemma 2 and Theorem 2 we have that

$$P_B^O \leq P_B^F \leq P_B[p_F^{MIN}(f\,|\,1), \cdots, p_F^{MIN}(f\,|\,mc)], \tag{33}$$

where the blocking probability, P_B, as a function of the fit probabilities is given by (24). Unfortunately, the expression for $p_F(f\,|\,\underline{k})$, given by (29), relies upon an independence assumption which has not been proven. Consequently, combining (24), (29), and (33) may not constitute a rigorous upper bound on the blocking probability. The derivation of a tight upper bound on $p_F(f\,|\,\underline{k})$, and hence on the blocking probability, P_B^F, therefore remains an open problem.

This completes the presentation of analytical results that can be used to evaluate multichannel TDMA performance using either optimal- or fixed-assignment schemes. To summarize, we have obtained an approximation for the blocking probability resulting from random assignment [given by (24), (28), (29), and (30)], and a lower bound on the blocking probability using optimal assignment (i.e., the Erlang blocking probability, P_B^E). These quantities can be easily evaluated with the aid of a computer. In the next section we compare these analytical results with computer simulation results.

V. NUMERICAL RESULTS

The analytical results of the last two sections are now illustrated via some specific examples. Figure 3 shows plots of the probability

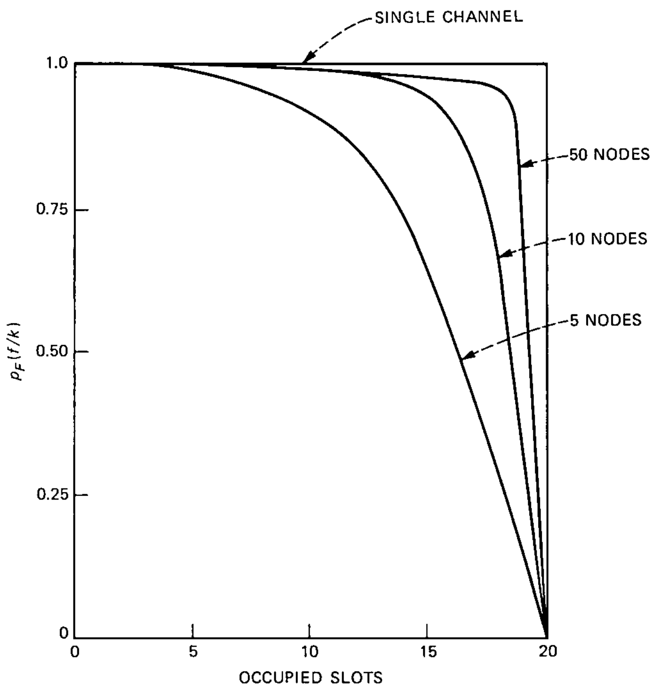

Fig. 3—Probability of a fit vs. number of occupied slots using random assignment for a system with four channels and five time slots.

that a new traffic arrival can be assigned, given that there are k units of traffic already present in the channel-time slot matrix [$p_F(f|k)$ given by (28) through (30)], vs. k for a system with four channels and five time slots per channel. Curves are shown for a system with 5 nodes, 10 nodes, and 50 nodes. These curves approximate the probability that an incoming traffic arrival can be assigned using random assignment. As the number of nodes increases, the curves converge rapidly to the single-channel (step function) case with nm time slots. The same set of curves computed for a system with 4 channels and 10 time slots per channel were nearly identical to those shown in Fig. 3 and are therefore omitted.

It is reasonable to expect that if the fit probabilities shown in Fig. 3 are close to the single-channel case, then the corresponding system blocking probabilities should also be close to the analogous single-channel blocking probability. This is indeed the case as illustrated in Figs. 4 through 7. In each case plots of blocking probabilities vs. normalized load for the single-channel case (Erlang B formula with nm servers), and for the multichannel case using random and optimal assignment are shown. The optimal-assignment curves were obtained by computer simulation. The random-assignment curves were ob-

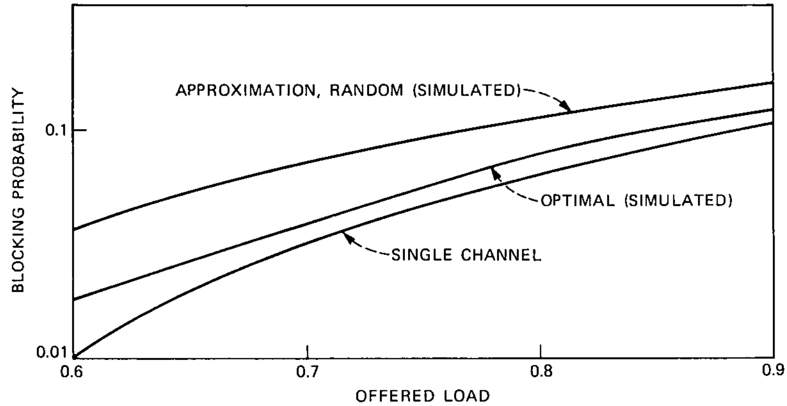

Fig. 4—Blocking probability vs. offered load for a system with 4 channels, 5 time slots, and 10 nodes.

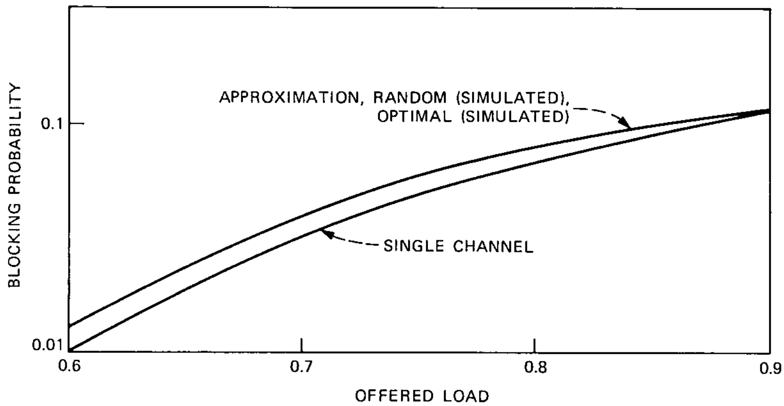

Fig. 5—Blocking probability vs. offered load for a system with 4 channels, 5 time slots, and 50 nodes.

tained both analytically via the approximation described in the last section [(24) and (28) through (30)], and by computer simulation. In all cases the approximate analytical curves are nearly identical to the corresponding simulated curves. Figures 4 and 5 show plots for a system with 4 channels, 5 time slots per channel, and 10 nodes and 50 nodes, respectively. Figures 6 and 7 show analogous plots for systems with 4 channels and 10 time slots per channel.

Figures 4 through 7 indicate that the differences between the simulation results, the analytical approximation, and the lower (single-channel) bound on multichannel blocking probabilities are significant only for systems with a relatively small number of nodes. For the cases shown here, the single-channel system exhibits at most moderate

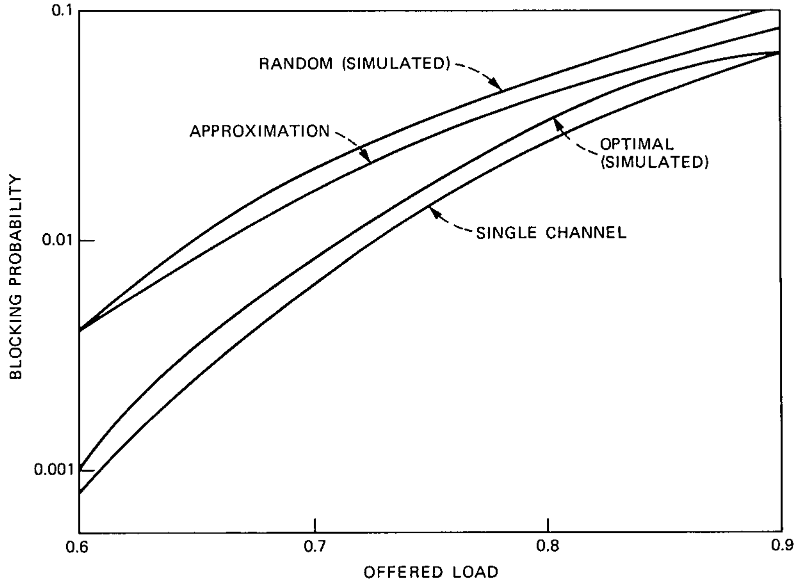

Fig. 6—Blocking probability vs. offered load for a system with 4 channels, 10 time slots, and 10 nodes.

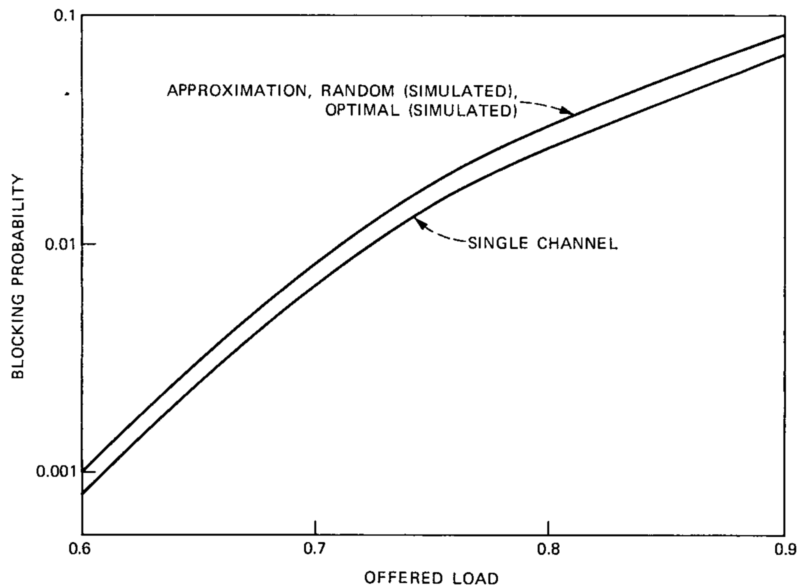

Fig. 7—Blocking probability vs. offered load for a system with 4 channels, 10 time slots, and 50 nodes.

performance improvements over the multichannel random-assignment case. Results obtained for additional cases indicate that if the ratio of nodes to channels (n/m) is greater than 10, the difference between blocking probabilities obtained using a multichannel system with a fixed-assignment algorithm and an analogous single-channel system is negligible. This condition is likely to be satisfied in many satellite systems, and hence we reach the important conclusion that, for practical systems, the simplest of assignment schemes will perform nearly as well as an optimal-assignment scheme.

We point out that the traffic model used here must be modified in order to study duplex voice traffic. In this case each traffic request from node i to node j also generates a simultaneous traffic request from node j to node i. This alternative traffic model does not require major changes in any of our previous arguments, and hence results obtained from using this model should correspond with those given here.

VI. CONCLUSIONS

This paper has provided tools with which to evaluate the performance of multichannel TDMA blocking systems. For any multichannel assignment scheme (fixed or optimal), a lower bound on system blocking probability has been obtained along with an accurate approximation for the blocking probability resulting from random assignment.

The numerical results in Section V indicate that multichannel blocking probability is relatively insensitive to the assignment algorithm used when a moderate number of nodes are present. If the ratio of the number of nodes to number of channels is 10 or greater, the difference between blocking probabilities obtained using a multichannel system with a fixed-assignment algorithm and an analogous single-channel system is negligible. This conclusion is fortunate since it implies that the performance of an assignment algorithm, which is simplest to implement, will be nearly optimal.

The results in this paper pertain to networks which handle voice traffic only. Of equal interest are analogous results which apply to networks handling data traffic. Specifically, it would be useful to know whether the performance of a multichannel TDMA queueing system is also insensitive to the particular assignment algorithm used. This issue requires further investigation.

VII. ACKNOWLEDGMENTS

The authors thank A. S. Acampora and B. E. Simon for many valuable discussions and comments concerning this work.

REFERENCES

1. R. Cooperman and W. G. Schmidt, "A Satellite Switched SDMA/TDMA System for Wideband Multibeam Satellites," ICC Conf. Rec., June 1973.
2. T. Muratani, "Satellite-Switched Time Domain Multiplex Access," EASCON, 1974.
3. D. O. Reudink and Y. S. Yeh, "A Scanning Spot Beam Satellite System," B.S.T.J., *56,* No. 8 (October 1977), pp. 1549–60.
4. D. O. Reudink, "Spot Beams Promise Satellite Communications Breakthrough," IEEE Spectrum, *15,* No. 9 (September 1978), pp. 36–42.
5. A. S. Acampora, D. O. Reudink, and Y. S. Yeh, "The Transmission Capacity of Multibeam Communication Satellites," Proc. IEEE, *69,* No. 2 (February 1981), pp. 209–25.
6. I. Gopal, D. Coppersmith, and C. K. Wong, "Minimizing Packet Waiting Time in a Multibeam Satellite System," IEEE Trans. Commun., *COM-30,* No. 2 (February 1982), pp. 305–6.
7. A. S. Acampora and B. R. Davis, "Efficient Utilization of Satellite Transponders via Time-Division Multibeam Scanning," B.S.T.J., *57,* No. 8 (October 1978), pp. 2901–14.
8. G. Bongiovanni et al., "An Optimum Time Slot Assignment Algorithm for an SS/TDMA System with Variable Number of Transponders," IEEE Trans. Commun., *COM-29,* No. 5 (May 1981), pp. 721–6.
9. Y. Ito, "Analysis of a Switch Matrix for an SS/TDMA System," Proc. IEEE, *65,* No. 3 (March 1977), pp. 411–9.
10. T. Inukai, "An Efficient SS/TDMA Time Slot Assignment Algorithm," IEEE Trans. Commun., *COM-27,* No. 10 (October 1979), pp. 1449–55.
11. L. Kleinrock, *Queueing Systems,* Vol. 1, New York: John Wiley, 1975, Chapters 2 and 4.
12. F. P. Kelly, *Reversibility and Stochastic Networks,* New York: John Wiley, 1979, Chapter 1.
13. A. S. Acampora, unpublished work.
14. E. J. Messerli, "Proof of a Convexity Property of the Erlang B Formula," B.S.T.J., *51,* No. 4 (April 1972), pp. 951–3.

APPENDIX A

Steady-State Distribution of the Markov Process Associated with Optimal Assignment

To derive the steady-state distribution of the Markov process associated with optimal assignment, consider a set of n^2 independent $M/M/\infty$ queues, labeled by the indices (i, j), with arrival rates λ_{ij} and mean service times μ_{ij}^{-1}. Define the Markov process q_{ij} as the number of customers in the queue (i, j). Then each process q_{ij} is a birth-death process with a steady-state distribution given by

$$p_{ij}(q_{ij}) = \frac{1}{q_{ij}!}\left(\frac{\lambda_{ij}}{\mu_{ij}}\right)^{q_{ij}} e^{-\lambda_{ij}/\mu_{ij}}, \qquad q_{ij} \geqslant 0.$$

Thus, the matrix process $\mathbf{Q} \equiv [q_{ij}]$ has the steady-state distribution given by

$$p(\mathbf{Q}) = \prod_{i,j} \frac{1}{q_{ij}!}\left(\frac{\lambda_{ij}}{\mu_{ij}}\right)^{q_{ij}} e^{-\lambda_{ij}/\mu_{ij}}, \qquad q_{ij} \geqslant 0.$$

Now restrict the process $\mathbf{Q}$ to the set of states such that the matrix constraints (1) are satisfied, and set $\lambda_{ij} = \lambda/n^2$ and $\mu_{ij} = \mu$. The resulting process is the Markov process $\mathbf{T}$ defined in Section III. By

Corollary 1.10 in Kelly,[12] the steady-state distribution of $\mathbf{T}$ is

$$p(\mathbf{T}) = \frac{\prod_{i,j} \frac{1}{t_{ij}!}\left(\frac{\lambda}{n^2\mu}\right)^{t_{ij}} e^{-\lambda/n^2\mu}}{\sum_{\mathbf{T}\in S} \prod_{i,j} \frac{1}{t_{ij}!}\left(\frac{\lambda}{n^2\mu}\right)^{t_{ij}} e^{-\lambda/n^2\mu}}, \qquad \mathbf{T} \in S, \tag{34}$$

where S is defined by (2a).† Note that (34) is the conditional probability distribution of the $\mathbf{Q}$ process, given that the state is contained in S, i.e., $p(\mathbf{Q} \mid \mathbf{Q} \in S)$. Evaluating the denominator in (34) requires an enumeration off all the states in S, a formidable task for moderately sized systems.

APPENDIX B

Proof of Lemma 1

This appendix contains the proof of Lemma 1. From (4), we have

$$\sum_{\mathbf{T}'\neq\mathbf{T}} p(\mathbf{T})r(\mathbf{T}, \mathbf{T}') = \sum_{\mathbf{T}'\in S_{\mathbf{T}}^-} p(\mathbf{T})r(\mathbf{T}, \mathbf{T}') + \sum_{\mathbf{T}'\in S_{\mathbf{T}}^+} p(\mathbf{T})r(\mathbf{T}, \mathbf{T}'). \tag{35}$$

Substituting (3) for $r(\mathbf{T}, \mathbf{T}')$ in the first term on the right yields

$$\begin{aligned} \sum_{\mathbf{T}'\in S_{\mathbf{T}}^-} p(\mathbf{T})r(\mathbf{T}, \mathbf{T}') &= p(\mathbf{T}) \sum_{\mathbf{T}'\in S_{\mathbf{T}}^-} r(\mathbf{T}, \mathbf{T}') \\ &= p(\mathbf{T}) \sum_{i,j} t_{ij}\mu = p(\mathbf{T})\mu k, \end{aligned} \tag{36}$$

where $\sum_{i,j} t_{ij} = k$ is the total traffic in $\mathbf{T}$. Similarly, for the second term on the right,

$$\begin{aligned} \sum_{\mathbf{T}'\in S_{\mathbf{T}^+}} p(\mathbf{T})r(\mathbf{T}, \mathbf{T}') &= p(\mathbf{T}) \sum_{\mathbf{T}'\in S_{\mathbf{T}^+}} r(\mathbf{T}, \mathbf{T}') \\ &= p(\mathbf{T}) \sum_{\mathbf{T}'\in S_{\mathbf{T}^+}} \lambda/n^2 \\ &= p(\mathbf{T})\lambda \mid S_{\mathbf{T}^+} \mid /n^2. \end{aligned} \tag{37}$$

An arrival is modeled as a selection from n^2 equally likely possibilities, of which $\mid S_{\mathbf{T}^+} \mid$ can be assigned, given that the state is $\mathbf{T}$. Therefore, the conditional probability that an arrival can be assigned is

$$p(f \mid \mathbf{T}) = \mid S_{\mathbf{T}}^+ \mid /n^2. \tag{38}$$

Combining (35) through (38) gives (5).

---

†This distribution was also derived in Ref. 13.

APPENDIX C

Proof of Theorem 1

To prove Theorem 1, we sum the flow equations in (7) over the set of states with k units of traffic, i.e.,

$$\sum_{\mathbf{T}\in S_k} [\mu k p(\mathbf{T}) + \lambda p(f|\mathbf{T})p(\mathbf{T})]$$
$$= \sum_{\mathbf{T}\in S_k} \sum_{\mathbf{T}'\in S_{\mathbf{T}}^{+}} p(\mathbf{T}')r(\mathbf{T}', \mathbf{T}) + \sum_{\mathbf{T}\in S_k} \sum_{\mathbf{T}'\in S_{\mathbf{T}}^{-}} p(\mathbf{T}')r(\mathbf{T}', \mathbf{T}). \quad (39)$$

The left-hand term is

$$\sum_{\mathbf{T}\in S_k} [\mu k p(\mathbf{T}) + \lambda p(f|\mathbf{T})p(\mathbf{T})] = \mu k \sum_{\mathbf{T}\in S_k} p(\mathbf{T}) + \lambda \sum_{\mathbf{T}\in S_k} p(f|\mathbf{T}) \cdot p(\mathbf{T})$$
$$= \mu k p_O(k) + \lambda p_O(f|k)p_O(k). \quad (40)$$

We evaluate the terms on the right side of (39) by interchanging the order of summation, which is permissible since the sums are always finite. Because $S_{k+1} = \cup_{\mathbf{T}\in S_k} S_{\mathbf{T}}^{+}$, and for $\mathbf{T}' \in S_{k+1}$, $r(\mathbf{T}', \mathbf{T}) \neq 0$ only if $\mathbf{T}' \in S_{\mathbf{T}}^{+}$, interchanging the sums in the first term on the right yields

$$\sum_{\mathbf{T}\in S_k} \sum_{\mathbf{T}'\in S_{\mathbf{T}}^{+}} p(\mathbf{T}')r(\mathbf{T}', \mathbf{T}) = \sum_{\mathbf{T}'\in S_{k+1}} \sum_{\mathbf{T}\in S_k} p(\mathbf{T}')r(\mathbf{T}', \mathbf{T}).$$

But if $\mathbf{T}' \in S_{k+1}$ and $\mathbf{T} \in S_k$, then $r(\mathbf{T}', \mathbf{T}) \neq 0$ only if $\mathbf{T} \in S_{\mathbf{T}'}^{-}$. Consequently,

$$\sum_{\mathbf{T}'\in S_{k+1}} \sum_{\mathbf{T}\in S_k} p(\mathbf{T}')r(\mathbf{T}', \mathbf{T}) = \sum_{\mathbf{T}'\in S_{k+1}} \sum_{\mathbf{T}\in S_{\mathbf{T}'}^{-}} p(\mathbf{T}')r(\mathbf{T}', \mathbf{T})$$
$$= \sum_{\mathbf{T}'\in S_{k+1}} p(\mathbf{T}')\mu(k+1)$$
$$= \mu(k+1)p_O(k+1), \quad (41)$$

where the next to the last step follows from Lemma 1. A similar argument shows that the second term on the right is

$$\sum_{\mathbf{T}\in S_k} \sum_{\mathbf{T}'\in S_{\mathbf{T}}^{-}} p(\mathbf{T}')r(\mathbf{T}', \mathbf{T}) = \sum_{\mathbf{T}'\in S_{k-1}} \sum_{\mathbf{T}\in S_k} p(\mathbf{T}')r(\mathbf{T}', \mathbf{T})$$
$$= \sum_{\mathbf{T}'\in S_{k-1}} \sum_{\mathbf{T}\in S_{\mathbf{T}'}^{+}} p(\mathbf{T}')r(\mathbf{T}', \mathbf{T})$$
$$= \sum_{\mathbf{T}'\in S_{k-1}} p(\mathbf{T}')\lambda p(f|\mathbf{T}')$$
$$= \lambda p_O(f|k-1)p_O(k-1). \quad (42)$$

Combining (40) through (42) yields (9).

APPENDIX D

Proof of Lemma 2

Lemma 2 is stated in the literature and although some unpublished proofs are referenced in Ref. 14, there does not seem to be a published proof.

We prove this theorem by calculating the partial derivatives and showing that they are nonnegative in the region of interest. Differentiating (24) gives

$$\frac{\partial}{\partial \phi_l}\left[\frac{\sum_{k=0}^{mc-1} \frac{1}{k!} \rho^k \prod_{j=0}^{k} \phi_j}{\sum_{k=0}^{mc} \frac{1}{k!} \rho^k \prod_{j=0}^{k-1} \phi_j}\right] = \frac{b_1(l)a_2(l) - a_1(l)b_2(l)}{[b_1(l) + b_2(l)\phi_l]^2}, \tag{43}$$

where

$$a_1(l) \equiv \sum_{k=0}^{l-1} \frac{1}{k!} \rho^k \alpha(k) \qquad b_1(l) \equiv \sum_{k=0}^{l} \frac{1}{k!} \rho^k \alpha(k-1)$$

$$a_2(l) \equiv \sum_{k=l}^{mc-1} \frac{1}{k!} \rho^k \beta(k) \qquad b_2(l) \equiv \sum_{k=l+1}^{mc} \frac{1}{k!} \rho^k \beta(k-1)$$

and

$$\alpha(k) \equiv \prod_{j=0}^{k} \phi_j \qquad \beta(k) \equiv \prod_{\substack{j=0 \\ j \neq l}}^{k} \phi_j$$

$$\alpha(-1) = \beta(-1) = 1.$$

The derivative (43) is nonnegative if and only if

$$b_1(l)a_2(l) - a_1(l)b_2(l) \geqslant 0. \tag{44}$$

The expression (44) is a polynomial in ρ with powers ranging from l to $mc + l - 1$. We will show that the coefficient of each power is nonnegative. Let $mc + l - 1 \geqslant s \geqslant l$ and define the sets

$$\Omega_1 \equiv \{(k_1, k_2) \mid k_1 + k_2 = s,\ 0 \leqslant k_1 \leqslant l,\ l \leqslant k_2 \leqslant mc - 1\}$$

$$\Omega_2 \equiv \{(k_3, k_4) \mid k_3 + k_4 = s,\ 0 \leqslant k_3 \leqslant l - 1,\ l + 1 \leqslant k_4 \leqslant mc\}.$$

Now the coefficient of ρ^s in (44) may be written as

$$\sum_{\Omega_1} \frac{1}{k_1! k_2!} \alpha(k_1 - 1)\beta(k_2) - \sum_{\Omega_2} \frac{1}{k_3! k_4!} \alpha(k_3)\beta(k_4 - 1). \tag{45}$$

Define the set

$$\Omega_{12} = \{(k_1, k_2) \mid (k_1, k_2) \in \Omega_1,\ k_1 \geqslant 1\}$$

and note that $(k_3, k_4) \in \Omega_2$ if and only if $k_3 = k_1 - 1$, $k_4 = k_2 + 1$, where $(k_1, k_2) \in \Omega_{12} \subset \Omega_1$. This implies that (45) is equal to $\omega_1 + \omega_2$, where

$$\omega_1 \equiv \sum_{\Omega_1 - \Omega_{12}} \frac{1}{k_1!k_2!} \alpha(k_1 - 1)\beta(k_2) \geq 0 \tag{46a}$$

$$\omega_2 \equiv \sum_{\Omega_{12}} \left[\frac{1}{k_1!k_2!} - \frac{1}{(k_1 - 1)!(k_2 + 1)!}\right] \alpha(k_1 - 1)\beta(k_2). \tag{46b}$$

A term in the sum ω_2 is negative if and only if $(k_2 + 1)/k_1 < 1$. But this inequality and the definition of Ω_{12} imply that $l + 1 \leq k_2 + 1 < k_1$, which is impossible because $k_1 \leq l$ for $(k_1, k_2) \in \Omega_{12}$. Consequently, we have $\omega_2 \geq 0$. This $\omega_1 \geq 0$ implies that (44) and (45) are nonnegative; therefore, each partial derivative (43) is nonnegative. Q.E.D.

AUTHORS

Michael L. Honig, B.S. (Electrical Engineering), 1977, Stanford University; M.S., Ph.D. (Electrical Engineering), University of California, Berkeley, in 1978 and 1981, respectively; AT&T Bell Laboratories, 1981–1982; AT&T Information Systems, 1983—. At AT&T Bell Laboratories and AT&T Information Systems, Mr. Honig has worked on echo cancellation of voiceband data signals, performance evaluation of convolutional codes, and performance analysis of local area networks. He is currently working on office information networks. Member, IEEE, Tau Beta Pi, Phi Beta Kappa.

Steven M. Barta, B.S. (Mathematics and Physics), 1973, Yale University; S.M., Ph.D. (Electrical Engineering), Massachusetts Institute of Technology, in 1976 and 1978, respectively; AT&T Bell Laboratories, 1978–1982; AT&T Information Systems, 1983—. As a Member of the Technical Staff at AT&T Bell Laboratories and AT&T Information Systems, Mr. Barta worked on economic studies, performance evaluation of convolutional codes, and analysis and simulation of traffic flow in data communication networks. He is currently developing a real-time traffic simulator for local area networks. Member, Phi Beta Kappa, Sigma Xi, IEEE.