Packet Switching in Radio Channels: Part II—The Hidden Terminal Problem in Carrier Sense Multiple-Access and the Busy-Tone Solution

FOUAD A. TOBAGI AND LEONARD KLEINROCK, FELLOW, IEEE

Abstract—**We consider a population of terminals communicating with a central station over a packet-switched multiple-access radio channel. The performance of carrier sense multiple access (CSMA) [1] used as a method for multiplexing these terminals is highly dependent on the ability of each terminal to sense the carrier of any other transmission on the channel. Many situations exist in which some terminals are "hidden" from each other (either because they are out-of-sight or out-of-range). In this paper we show that the existence of hidden terminals significantly degrades the performance of CSMA. Furthermore, we introduce and analyze the busy-tone multiple-access (BTMA) mode as a natural extension of CSMA to eliminate the hidden-terminal problem. Numerical results giving the bandwidth utilization and packet delays are shown, illustrating that BTMA with hidden terminals performs almost as well as CSMA without hidden terminals.**

I. INTRODUCTION

THE USE of packet switching in a multiple-access broadcast radio channel for communication between terminals and a central station was presented in Part I [1]. We also introduced and analyzed a new random-access mode, the carrier sense multiple-access mode (CSMA), as a means of multiplexing a large number of terminals communicating with the station over the shared radio channel.[1] Briefly, CSMA consists of reducing the level of interference (caused by overlapping packets) in the random multiaccess environment by allowing terminals to sense the carrier due to other users' transmissions; based on the information gained in this way about the state of the channel (busy or idle), the terminal takes an action prescribed by the particular CSMA protocol being used (in particular, a terminal never transmits when it senses the channel busy). In Part I we described and analyzed three protocols referred to as: 1-persistent; nonpersistent; and p-persistent CSMA. The evaluation of performance of the various protocols obtained there (Part I) was based on the assumption that all terminals are in line-of-sight (LOS) and within range of each other. However there are many situations in which this is not true, forcing us to relax the above assumption here. Two terminals can be within range of the station but out-of-range of each other; or they can be separated by some physical obstacle opaque to UHF signals. Two such terminals are said to be "hidden" from each other. It is evident that the existence of hidden elements in an environment affects (degrades) the performance of CSMA. In this paper we first attempt to gain some insight about this effect. This is the subject of Section II. (For simplicity, we restrict our study to 1-persistent and nonpersistent CSMA protocols only.)

Second, in this paper, we consider a solution to the hidden-terminal problem which we call the busy-tone multiple-access (BTMA) mode. This is the subject of Section III in which we give 1) a description of the operation of BTMA under a nonpersistent protocol, 2) an analysis to determine the throughput-delay characteristics along with the effect of various system parameters, and 3) a discussion of some numerical results.

II. THE HIDDEN-TERMINAL PROBLEM

Below, we shall describe the model and define an adequate representation for hidden-terminal configurations. Then we proceed with the analysis for throughput and delay. Finally we shall consider some examples to which we apply analytical and simulation techniques.

A. The Model

We assume an environment consisting of a large number of terminals communicating with a single station over a shared radio channel. All terminals are in line-of-sight and within range of the station but not necessarily with respect to each other. All other system assumptions introduced in Part I hold true. The total traffic source will be approximated by an independent Poisson source with an aggregate mean packet generation rate of λ packets/s.

We characterize a terminal configuration with hidden elements as follows. Let $i = 1,2,\cdots,M$ index the M terminals in the population. By definition, terminal i "hears" (is connected to) terminal j if i and j are within range and in line-of-sight of each other. To represent the connections among terminals, we use an $M \times M$ square matrix $\boldsymbol{M}$ such that the element m_{ij} is

$$m_{ij} = \begin{cases} 1, & \text{if } i \text{ hears } j \\ 0, & \text{otherwise.} \end{cases}$$

Paper approved by the Associate Editor for Computer Communication of the IEEE Communications Society for publication after presentation at the National Computer Conference, Anaheim, Calif., 1975. Manuscript received December 5, 1974; revised June 11, 1975. This work was supported in part by the Advanced Research Projects Agency, Department of Defense under Contract DAHC15-73-C-0368.

The authors are with the Computer Science Department, University of California, Los Angeles, Calif. 90024.

[1] Throughout the paper, the reader is assumed to be familiar with the results and terminology introduced in [1].

Since two terminals that hear the same subset of the population behave similarly, it is advantageous to partition the population into several groups (say $N, N \leq M$) such that all terminals within a group hear exactly the same subset of terminals in the population. This partitioning is easily formed by collecting all terminals with identical rows or columns in $\boldsymbol{M}$ into one group. We now define a "hearing" graph with N nodes and make a one-to-one correspondence between the nodes of the graph and the N groups just obtained. A link between two nodes k and l represents the fact that group k and group l hear each other, and this is easily determined by the fact that there exists a terminal i in group k that hears a terminal j in group l. This procedure provides us with the minimum number of groups describing the configuration. Let $h(i)$ be the set of groups that group i can hear. In the sequel, we shall isolate the case of independent groups from the general case of dependent groups. The former is characterized by the absence of links in the hearing graph.

We shall further assume that each group i consists of a large number of users who collectively form an independent Poisson source with an aggregate mean packet generation rate λ_i packets/s such that $\sum_{i=1}^{N} \lambda_i = \lambda$.

As in Part I, we characterize the traffic as follows. Let $S_i = \lambda_i T$. Under steady-state conditions, S_i is the throughput of group i. Let $S = \lambda T = \sum_{i=1}^{N} S_i$; S is the total throughput and utilization of the channel. Let $\mathcal{S} = (S_1, S_2, \cdots, S_N)$. Let $\mathcal{U} = (u_1, u_2, \cdots, u_N)$ where $u_i = (S_i/S)$; $\mathcal{U}$ describes a direction in N-dimensional space. The *capacity* of the channel along the direction $\mathcal{U}$ is defined as

$$C(\mathcal{U}) = \max_{0 \leq S \leq 1} S$$

such that the set of inputs determined by the vector $S\mathcal{U}$ is achievable. In other words, $C(\mathcal{U})$ is the maximum achievable throughput or maximum attainable channel utilization, when for all i, the input source of group i constitutes a fraction u_i of the total input source. In addition, let G_i denote the mean offered traffic rate (per T seconds) of group i. Let $\mathcal{G} = (G_1, G_2, \cdots, G_N)$ and $G = \sum_{i=1}^{N} G_i$.

We determined in Part I [1] the necessity of introducing a random retransmission delay X with mean $\bar{X}$ to avoid repeated conflicts. We shall further assume here that $\bar{X}$ is the same for all groups and that Assumptions 1 and 2 of Part I still hold true, as follows.

Assumption 1': $\bar{X}$ is large compared to the transmission time T, so that the interarrival times of the point processes defined by the start times of all the (new) packets plus retransmissions from group i are of independent increments and exponentially distributed, with mean interarrival time $1/G_i$.

B. Analysis

We wish to answer the following basic questions of interest.

Question 1: Given an input pattern $\mathcal{U}$, what is the channel capacity $C(\mathcal{U})$? An equivalent question is: Is a given set of input rates $\mathcal{S}(\mathcal{U})$ achievable or does it saturate the channel?

Question 2: For a given achievable set of input rates $\mathcal{S}(\mathcal{U})$, what is the relative performance of the various groups?

We shall first treat the simple case of independent groups for both the 1-persistent and nonpersistent CSMA, then we shall proceed with an approximate analysis for the dependent groups case under a nonpersistent protocol.

Independent Groups Case: We first recognize that S_i/G_i is merely the probability of success of an arbitrary packet from group i. This quantity is a function of the traffic vector $\mathcal{G}$. By expressing S_i/G_i for each i in terms of $\mathcal{G}$, we obtain a set of equations relating the components of $\mathcal{S}$ to the components of $\mathcal{G}$. For a given $\mathcal{G}$ and under the system and model assumptions stated above, the probability of success of an arbitrary packet from group i is given as follows.

1-Persistent CSMA:

$$P_{s_i} = \frac{S_i}{G_i} = \frac{[1 + G_i + aG_i(1 + G_i + aG_i/2)]}{(1 + aG_i) \exp\{-G_i(1 - 2a)\}} \prod_{j=1}^{N} \frac{(1 + aG_j) \exp\{-2G_j\}}{G_j(1 + 2a) - (1 - \exp\{-aG_j\}) + (1 + aG_j) \exp\{-G_j(1 + a)\}}. \quad (1)$$

Nonpersistent CSMA:

$$P_{s_i} = \frac{S_i}{G_i} = \exp\{G_i(1 - 2a)\} \cdot \prod_{j=1}^{N} \frac{\exp\{-G_j(1 - a)\}}{G_j(1 + 2a) + \exp\{-aG_j\}}. \quad (2)$$

Proof: By definition, a packet transmission is said to be *i-successful* if the packet is free from interference caused by packets from group i. A packet transmission is said to be *totally successful* if and only if it is i-successful for all i; $i = 1, 2, \cdots, N$.

Consider first the 1-persistent CSMA case. An arbitrary packet from group i is successful if the following two mutually independent conditions are satisfied.

$\mathcal{C}_1$: The packet transmission is i-successful.

$\mathcal{C}_2$: The packet is j-successful, for all $j \neq i$.

($\mathcal{C}_1$ and $\mathcal{C}_2$ are independent since we are dealing with independent groups.)

Consider for each group i a time line which exhibits packet transmissions from group i only (see Fig. 1). We observe on time line i an alternate sequence of busy and idle periods as defined in [1]. Moreover, because of the independence among groups (completely disconnected graph), this sequence is completely determined by the traffic rate G_i. Condition $\mathcal{C}_1$ is satisfied with a probability equal to the probability of success of a packet in 1-persistent CSMA without hidden terminals when the traffic rate is G_i. It is given by [1]

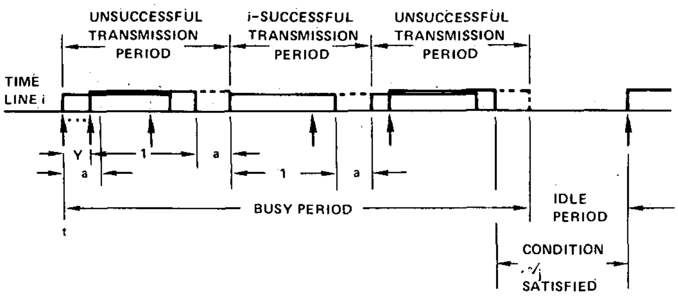

Fig. 1. 1-persistent CSMA: time line i.

$$\Pr\{\mathcal{C}_1\} = \frac{[1 + G_i + aG_i(1 + G_i + aG_i/2)] \exp\{-G_i(1 + 2a)\}}{G_i(1 + 2a) - (1 - \exp\{-aG_i\}) + (1 + aG_i) \exp\{-G_i(1 + a)\}}. \tag{3}$$

Consider now the time line j corresponding to group $j, j \neq i$. Here again, we observe an alternate sequence of busy and idle periods denoted by B_j and I_j, respectively, completely determined by the rate G_j. The average busy and idle periods are expressed as [1]

$$\bar{B}_j = \frac{1 + a + \bar{Y}_j}{q_{0,j}} \tag{4}$$

$$I_j = \frac{1}{G_j} \tag{5}$$

$$\Pr\{\mathcal{C}_2\} = \prod_{j \neq i} \frac{(1/G_j)[\exp\{-G_j(1 - a)\} - \exp\{-G_j\}] + \bar{I}_j \exp\{-G_j\}}{\bar{I}_j + \bar{B}_j}$$

$$= \prod_{j \neq i} \frac{(1 + aG_j) \exp\{-2G_j\}}{G_j(1 + 2a) - (1 - \exp\{-aG_j\}) + (1 + aG_j) \exp\{-G_j(1 + a)\}}. \tag{8}$$

where

$$\bar{Y}_j = a - \frac{1}{G_j}(1 - \exp\{-aG_j\}) \tag{6}$$

and

$$q_{0,j} = (1 + aG_j) \exp\{-G_j(1 + a)\}. \tag{7}$$

Our tagged packet is j-successful if and only if the following two conditions are satisfied.

$\mathcal{A}_j$: The start of transmission of the tagged packet does not occur during any transmission period (with the exception of the last a seconds of the last transmission period of the busy period [see Fig. 1]).

$\mathcal{B}_j$: No packet from group j starts transmission during the transmission time of the tagged packet.

We know that, by assumption 1′, the arrival of an arbitrary packet represents a random look in time. In 1-persistent CSMA, the start of transmission of the packet may not correspond to the arrival of the packet since the packet may incur a pretransmission delay in case the channel is sensed busy at its arrival time. However, by assuming that the start of transmission possesses the same property,[2] we have:

$\Pr\{$the tagged packet starts transmission during the last a seconds of a busy period on time line $j\} = a/(\bar{I}_j + \bar{B}_j)$.

In this event, the probability that no packets from group j start transmission during the transmission time of the tagged packet is the probability of no transmission from group j in an interval $(1 + x - a)$ with x uniformly distributed over the last a seconds of the busy period and is given by

$$\int_0^a \exp\{-G_j(1 + x - a)\} \frac{dx}{a}$$

$$= \frac{1}{aG_j}\left[\exp\{-G_j(1 - a)\} - \exp\{-G_j\}\right].$$

On the other hand, the probability that the tagged packet starts transmission during an idle period is $\bar{I}_j/(\bar{I}_j + \bar{B}_j)$ and the probability that no packets from group j start transmission during its transmission time is $\exp\{-G_j\}$. Since the groups are independent, we then have

Conditions $\mathcal{C}_1$ and $\mathcal{C}_2$ being mutually independent, we have

$$P_{s_i} = \frac{S_i}{G_i} = \Pr\{\mathcal{C}_1\} \Pr\{\mathcal{C}_2\}$$

in which we substitute the expressions found above to get (1).

The proof for (2) is exactly identical to the one above, in which we have the following expressions for the various quantities [1]:

$$\bar{B}_j = 1 + a + \bar{Y}_j$$

$$\bar{I}_j = \frac{1}{G_j}$$

$$\bar{Y}_j = a - \frac{1}{G_j}(1 - \exp\{-aG_j\})$$

[2] Such an assumption is not needed in the nonpersistent CSMA case since packets will not incur pretransmission delays. In that case the analysis will be exact. The comparison between results obtained from simulation of the "Poisson" model, to be discussed in the following section, and those obtained by this analysis (with $a = 0.01$) shows that the effect of this assumption is not noticeable.

$$\Pr\{\mathcal{C}_1\} = \frac{\exp\{-aG_j\}}{G_i(1+2a)+\exp\{-aG_i\}} \tag{9}$$

and

$$\Pr\{\mathcal{C}_2\} = \prod_{j \neq i} \frac{\exp\{-G_j(1-a)\}}{G_j(1+2a)+\exp\{-aG_j\}}. \tag{10}$$

Therefore

$$P_{si} = \frac{S_i}{G_i} = \exp\{-G_i(1-2a)\}$$

$$\cdot \prod_{j=1}^{N} \frac{\exp\{-G_j(1-a)\}}{G_j(1+2a)+\exp\{-aG_j\}}. \quad \text{Q.E.D.}$$

Thus we obtain a set of equations relating the components of the input vector $\mathcal{S}$ to the components of the traffic vector $\mathcal{G}$ of the form

$$\frac{S_j}{G_i} = f_i(G_1,G_2,\cdots,G_N). \tag{11}$$

For a given input vector $\mathcal{S}$, we can numerically solve for G_i, $i = 1,\cdots,N$. This we do by writing (11) in the form

$$G_i = S_i/f_i(G_1,\cdots,G_N) \tag{12}$$

and by solving the set of equations iteratively, starting with the initial values $\mathcal{G} = \mathcal{S}$. If the iterative procedure results in a (finite) traffic vector $\mathcal{G}$ then the input vector is feasible. (We do not claim we can prove existence and unicity of the solution but it has been our experience that $\mathcal{G} = \mathcal{S}$ is a good starting solution and simulation results always agreed with the results obtained by the iterative procedure whenever a finite solution could be reached. Lack of convergence is assumed whenever a certain preset maximum number of iterations is exceeded.) Thus, the convergence of the iterative procedure determines the feasibility of the input vector $\mathcal{S}$ and the final values G_i/S_i; $i = 1,2,\cdots,N$ give the average number of transmissions and schedulings a packet from group i undertakes before success. This will be our measure of relative performance of the various groups. Some simple examples are treated in the following section.

Before we proceed with the case of dependent groups, we consider here the particular case in which the N independent groups are identical, i.e.,

$$S_i = s, \qquad \forall i$$

$$G_i = g, \qquad \forall i.$$

Equations (1) and (2) reduce to, respectively,

$$\frac{S}{G} = \frac{s}{g} = \frac{1+g+ag(1+g+ag/2)}{(1+ag)\exp\{-g(1-2a)\}}\left[\frac{(1+ag)\exp\{-g\}}{g(1+2a)-(1-\exp\{-ag\})+(1+ag)\exp\{-g(1+a)\}}\right]^N \tag{13}$$

$$\frac{S}{G} = \frac{s}{g} = \exp\{g(1-2a)\}\left[\frac{\exp\{-g(1-a)\}}{g(1+2a)+\exp\{-ag\}}\right]^N. \tag{14}$$

Naturally, if we let $N \to \infty$, $s \to 0$, and $g \to 0$ such that $Ns = S$ and $Ng = G$, we expect the CSMA mode to reduce to the ALOHA access mode. Indeed it is easy to see from (13) and (14) that for both protocols

$$\frac{S}{G} \underset{N\to\infty}{\longrightarrow} e^{-2G}$$

the probability of success of a packet in ALOHA mode!

The Case of Dependent Groups: The dependence among the groups renders the determination of the sequence of idle and busy periods relative to a group, say i, rather difficult. Moreover this sequence is a function of the entire set $\{G_j, j \in h(i)\}$. Thus, no tractable analysis is yet available for the 1-persistent CSMA mode. However, an *approximate* model is presented here for the nonpersistent CSMA protocol. For this we make some fairly strong assumptions of statistical independence among the groups as well as exponential distributions for the interpoint times of various processes. (We claim that such assumptions are particularly valid when the load on the channel is low). Simulation techniques are considered in the next section allowing us to verify the validity of the approximate models.

Consider again the time axis on which we represent packet arrivals and packet transmissions. Time line i, relative to group i, is obtained by deleting from the time axis all packet transmissions belonging to groups other than i; that is, time line i exhibits packets from group i only. As discussed above, we can observe on time line i an alternate sequence of idle and busy periods (see Fig. 2). The simplicity in studying this protocol is mainly due to the fact that any busy period consists of a single transmission period [1]. In nonpersistent CSMA, when a terminal becomes ready it senses the channel. An arrival corresponds then to a *sense point*. A sense point will result in an actual transmission if the channel is sensed idle, otherwise the sense point (or arrival) is said to be *blocked*. Furthermore, by definition a sense point is said to be *j-unblocked* if the sense point is not blocked by packet transmissions from group j.

Consider time line i for example and let G_i, as before, denote the *total* rate of sense points generated by group i. The point process defined by these sense points is assumed to be of independent increments and Poisson (see Section II-A). Let G_i' be the rate of sense points which are j-unblocked for all j; $j \neq i$; i.e., G_i' is the rate of sense points which did not find the channel busy because of the transmission of a packet from group j; $j \neq i$. Obviously, j must then belong to the subset $h(i)$. The independence assumption that we make at this point can be stated as follows.

Assumption 2': The point process defined by the unblocked sense points relative to group i is independent of

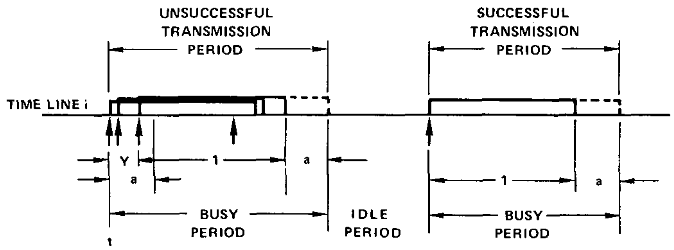

Fig. 2. Nonpersistent CSMA: busy and idle periods on time line i.

the state (busy or idle) of any time line j; $j \neq i$, is of independent increments, and is Poisson. That is, the point process is completely determined by the rate G_i'.

We recognize that this statement would be true only if all groups were independent. Nevertheless it is valid when the system is lightly loaded (at any rate, simulation is used to check the validity of results obtained under these assumptions); if $P_b{}^i$ is the probability that a sense point on time line i is blocked by packet transmissions from and group j; $j \neq i$, then we can write

$$G_i' = G_i(1 - P_b{}^i).$$

The introduction of Assumption 2′ simplifies the problem yielding *approximate* relationships between the various quantities defined so far.

Under the model assumptions and the additional Assumption 1′, the relationship between the components of $\mathcal{S}$ and the components of $\mathcal{G}$ is given by the following system of equations:

$$S_i = G_i \frac{\prod_{j \epsilon h(i)} \exp\{-aG_j'\} \prod_{k \notin h(i)} \exp\{-G_k'(1-a)\}}{\prod_{l=1}^{N} [G_l'(1+2a) + \exp\{-aG_l'\}]} \tag{15}$$

where

$$G_i' = G_i \prod_{j \epsilon h(i);\, j \neq i} \frac{1 + aG_j'}{G_j'(1+2a) + \exp\{-aG_j'\}}. \tag{16}$$

Proof: Consider time line i on which we observe an alternate sequence of busy and idle periods (Fig. 2). By Assumption 2′, this sequence is completely determined by the rate G_i'. The average busy and idle periods can then be expressed as [1]

$$\bar{B}_i = 1 + \bar{Y}_i' + a$$

$$\bar{I}_i = \frac{1}{G_i'}$$

where

$$\bar{Y}_i' = a - \frac{1}{G_i'}(1 - \exp\{-aG_i'\}).$$

By the Poisson assumption, an arbitrary sense point represents a random look in time. The probability that an arbitrary sense point from group i is j-unblocked, $j \neq i$, is the probability that a random look at time line j falls either in an idle period or during the first a seconds of a busy period of time line j and is expressed as

$$\frac{\bar{I}_j + a}{\bar{B}_j + \bar{I}_j} = \frac{1 + aG_j'}{G_j'(1+2a) + \exp\{-aG_j'\}}.$$

By the independence assumption we then establish (16).

$$G_i' = G_i \prod_{j \epsilon h(i);\, j \neq i} \frac{1 + aG_j'}{G_j'(1+2a) + \exp\{-aG_j'\}}.$$

Consider now an arbitrary (unblocked) sense point from group i. For this to result in a totally successful transmission, the following conditions must be simultaneously satisfied:

$\mathcal{A}$: The sense-point corresponds to the start of an i-successful transmission.

$\mathcal{B}$: The tagged sense point (which is j-unblocked, $\forall_j \in h(i)$; $j \neq i$) occurs during an idle period of time line j, $\forall j \in h(i)$; $j \neq i$.

$\mathcal{C}$: There are no arrivals from any group $j \in h(i)$, $j \neq i$ during the first a seconds of the tagged transmission.

$\mathcal{D}$: The tagged packet is k-successful, for all $k \notin h(i)$.

It is easy to see that, as in (9), we have:

$$\Pr\{\mathcal{A}\} = \frac{\exp\{-aG_i'\}}{G_i'(1+2a) + \exp\{-aG_i'\}}. \tag{17}$$

On the other hand, knowing that the tagged packet is j-unblocked, $j \in h(i)$; $j \neq i$, we have

$$\Pr\{\mathcal{B}/\text{packet is } j\text{-unblocked}, j \in h(i), j \neq i\}$$

$$= \prod_{j \epsilon h(i);\, j \neq i} \frac{1/G_j'}{1/G_j' + a}$$

$$= \prod_{j \epsilon h(i);\, j \neq i} \frac{1}{1 + aG_j'} \tag{18}$$

and

$$\Pr\{\mathcal{C}\} = \prod_{j \epsilon h(i);\, j \neq i} \exp\{-aG_j'\}. \tag{19}$$

Moreover, similarly to (10), we have

$$\Pr\{\mathcal{D}\} = \prod_{k \notin h(i)} \frac{\exp\{-G_k'(1-a)\}}{G_k'(1+2a) + \exp\{-G_k'\}}. \tag{20}$$

By the independence assumption, we have

$$S_i = G_i' \Pr\{\mathcal{A}\} \cdot \Pr\{\mathcal{B}\} \cdot \Pr\{\mathcal{C}\} \cdot \Pr\{\mathcal{D}\}.$$

Using the expressions found in (17), (18), and (20) and the expression for G_i' given by (16), we get (15). Q.E.D.

Before we proceed with the examples, let us consider again the case of N independent and identical groups to which we apply (15) and (16). In this case, $h(i) = \{i\}$; $\forall i = 1,2,\cdots,N$. Denoting as before $G_i = g$, $S_i = s$; $\forall i$, and $Ns = S$, $Ng = G$, (16) reduces to $G_i' = G_i = g$; $\forall i$, and (15) reduces to

$$\frac{s}{g} = \frac{S}{G} = \exp\{g(1-2a)\} \left[\frac{\exp\{-g(1-a)\}}{g(1+2a) + \exp\{-ag\}}\right]^N$$

which is identical to (14).

C. Examples

In the present section, we consider some examples to which we apply the analytical results found in Section II-B. Simulation techniques are also used 1) whenever the analysis is intractable, and 2) to check the validity of the assumptions on which the *analysis* was based. The simulation model is based on the same system assumptions as in Section II-A. Among these, in particular, we assume that the input processes for the various groups are Poisson. However the assumptions pertaining to the characterization of the offered traffic (see Assumption 1′) and the independence assumptions introduced for analytic tractability are all relaxed. For the various examples simulated, the comparison of results obtained from simulation and the results obtained from the analytic model match very well. In the present section, we also draw various conclusions about the effect hidden terminals have on the performance of CSMA. For the following examples and numerical results, we restrict ourselves to $a = 0.01$.

Independent Groups Case: A Symmetric Configuration: We have already considered the example in which the population is partitioned into N groups of equal size. The (S,G) relationship is given by (13) for 1-persistent CSMA and by (14) for the nonpersistent protocol.

In this example, for each terminal there exists a fraction β of the population which is hidden, namely $\beta = [(N-1)/N]$ (>0.5). The channel capacity for various values of N is plotted in Fig. 3. Note that the channel capacity experiences a drastic decrease between the two cases: $N = 1$ (no hidden terminals, $\beta = 0$) and $N = 2$ $(\beta = 0.5)$. For $N \geq 2$, slotted ALOHA performs better than CSMA.[3] This decrease is more critical for the nonpersistent CSMA than for the 1-persistent CSMA as shown in the figure. For $N > 2$, the channel capacity is rather insensitive to N and approaches pure ALOHA for large N, as was shown in Section II-B.

Independent Groups Case: Complementary Couple Configuration: The previous example did not show the effect of a small fraction of the population being hidden from the rest. In this example the population consists of two independent groups $(N = 2)$ of unequal sizes such that $\mathfrak{U} = (\alpha, 1-\alpha)$; that is,

$$S_1 = \alpha S$$

$$S_2 = (1-\alpha) S.$$

Equations (1) and (2) are readily applicable. The channel capacity is plotted versus α for both CSMA protocols in Fig. 4. Here again we note that the capacity decreases *rapidly* as α increases from 0. This decrease is much more critical for the nonpersistent than for the 1-persistent.

In answering Question 2 we note that a good measure of delay is given by G_i/S_i, the average number of transmissions and scheduling of a packet until success. We note that the larger group (i.e., the group with the higher aggregate input rate) performs much better than the smaller one. (See [2] for further details.)

[3] Recall that a is assumed constant and equal to 0.01.

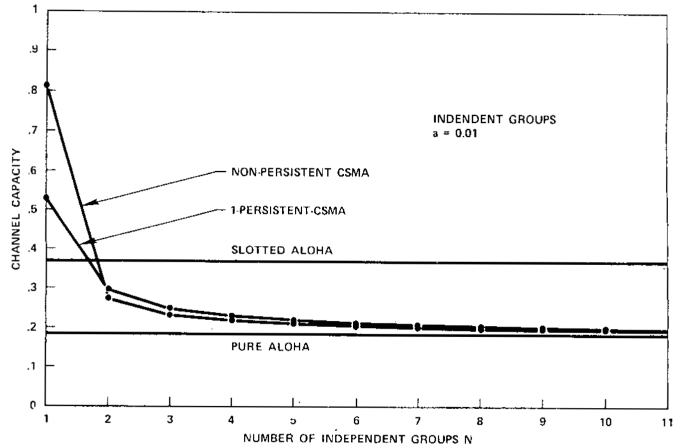

Fig. 3. Independent groups case—channel capacity versus the number of groups.

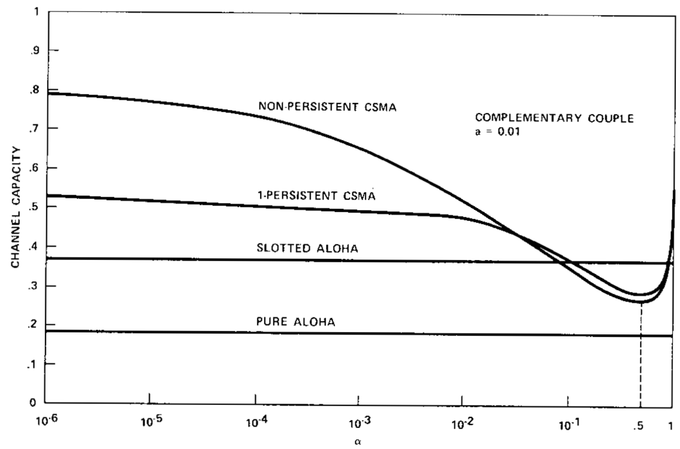

Fig. 4. Complementary couple configuration—channel capacity versus α.

Dependent Groups Case: A Symmetric Configuration: Let us now consider the situation in which the population is partitioned into N groups of equal size such that for each group all but one group are within sight. The graph representation of such a configuration is shown in Fig. 5. Obviously, this situation falls within the case of dependent groups, and corresponds to the instance in which, for each terminal, there is a fraction β of hidden terminals such that $\beta < 0.5$, namely $\beta = (1/N)$. Simulation techniques have been employed to study this instance under a 1-persistent CSMA protocol. In Fig. 6, we show the relationship between the total throughput S and the total traffic G for various values of β. We note again that the higher the value of β is, the smaller is the channel capacity. We further note that for a given achievable throughput S, the larger β is, the larger the traffic rate G is and hence the smaller is the probability of success of a packet, and the larger its average number of transmissions is.

Dependent Groups Case: The "Wall" Configuration: Consider a uniform distribution of terminals over a circular area, the station being located at the center. All terminals are within range of each other, but the presence of a "wall" (hill) as displayed in Fig. 7(a) causes some terminals to be hidden from others. A terminal T_0 at an angle α_0 ($\alpha_0 <$ 180°) from the wall can only hear terminals in the region

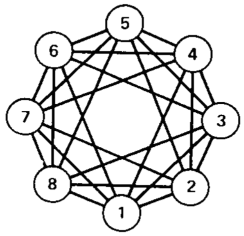

Fig. 5. The hearing graph of a symmetric dependent groups configuration.

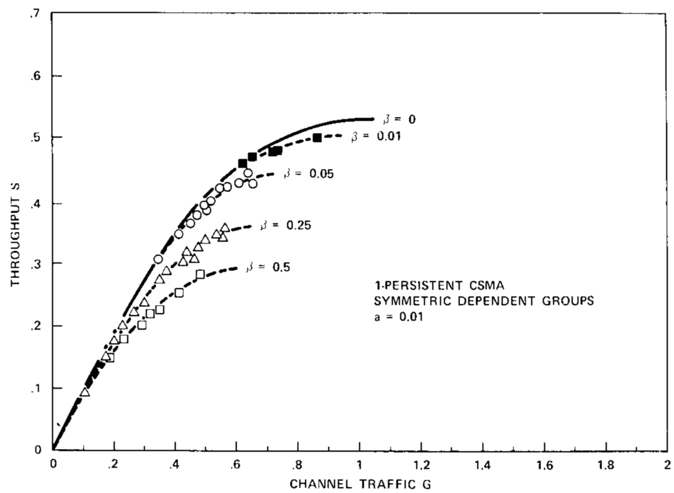

Fig. 6. Symmetric dependent groups configuration—throughput versus channel traffic.

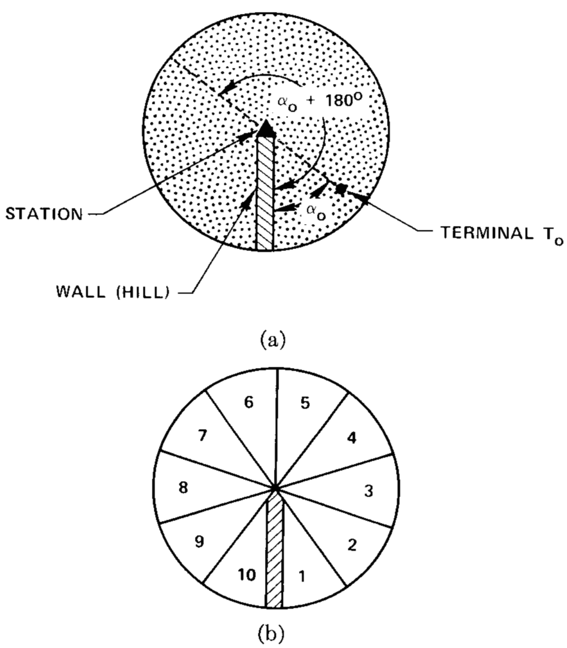

Fig. 7. The wall configuration.

$(0,\alpha_0 + 180°)$. To study a continuous problem such as this, we examine the discrete approximation obtained by breaking up the uniform population into small sectors, each sector considered then as a group. The smaller the sectors are, the more precise the approximation becomes. The problem is solved numerically below for $N = 10$; that is, we partition the population into 10 equal sectors as shown in Fig. 7(b). In the latter, we note that, for each group i, there exists a group $\bar{\imath}$ diametrically opposite such that a terminal in group i can only hear a fraction of the terminals in group $\bar{\imath}$. Therefore, the solution to the continuous problem can be bounded by the following two cases.

Case 1—$\bar{\imath} \notin h(i)$: This gives a lower bound on CSMA performance; that is a lower bound on channel capacity or upper bound on the channel traffic and on the average number of transmissions and schedulings.

Case 2—$\bar{\imath} \in h(i)$: This gives an upper (i.e., optimistic) bound on the performance of CSMA.

The hearing graph representations corresponding to the two cases are shown in Fig. 8. The analytic results and simulation results were compared for the case $\bar{\imath} \notin h(i)$ for values of $a = 0.01$ and 0.1 (see [2]). The close match again validates the analytical model. The results obtained by the latter are shown in Figs. 9 and 10. In Fig. 9 we plot the total throughput S versus the aggregate rate G. The channel capacity is shown to be bounded by $0.37 \leq C \leq 0.44$ $(a = 0.01)$. The existence of the wall decreased the channel capacity by a factor approximately equal to 2. Similarly, in Fig. 10 we plot the upper and lower bounds on the overall average number of transmissions and schedulings per packet.

Remark: The set of equations relative to the dependent groups case given in Section II-2 provides us with an approximate solution based on the independence Assumption 2′. Simulation results agree with those obtained by the model for the examples considered so far. However, there are cases where the independence assumption is not satisfied and the model inapplicable. Consider, for example, the symmetric configuration depicted in Fig. 5. Assume $a = 0$ for simplicity and let $g' = G_i'$, $g = G_i$, and $s = S_i$; $\forall i$. From (15) and (16) we have

$$g' = g\left(\frac{1}{1+g'}\right)^{N-2}$$

$$s = g\,\frac{e^{-g'}}{(1+g')^N}\,.$$

If we now let $N \rightarrow \infty$ $s \rightarrow 0$, $g \rightarrow 0$, $g' \rightarrow 0$ such that $Ns = S$, $Ng = G$, and $Ng' = G'$, then we get

$$S = G' = Ge^{-S}. \tag{21}$$

Such a result is certainly wrong in the limit since, for this particular case, we expect to reach the nonpersistent CSMA result with no hidden terminals, namely [1]

$$S = \frac{G}{1+G}\,. \tag{22}$$

This is basically due to the independence assumption introduced above. However, (21) and (22) are equivalent in the case of low channel utilization, since then approximating e^{-S} by $1 - S$, (21) reduces to (22).

In summary, for the various particular configurations that were considered, it is to be noted that the performance of carrier sense is badly degraded by the existence

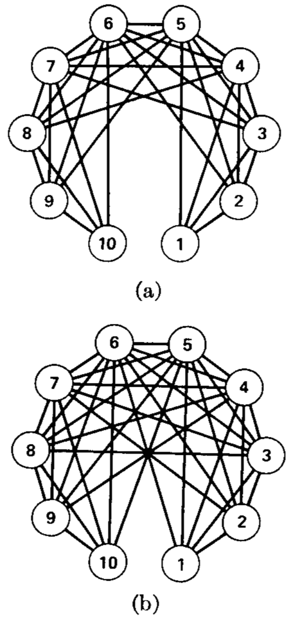

Fig. 8. The wall configuration—the hearing graphs for the ten sectors discrete approximation. (a) $\bar{\imath} \notin h(i)$. (b) $\bar{\imath} \in h(i)$.

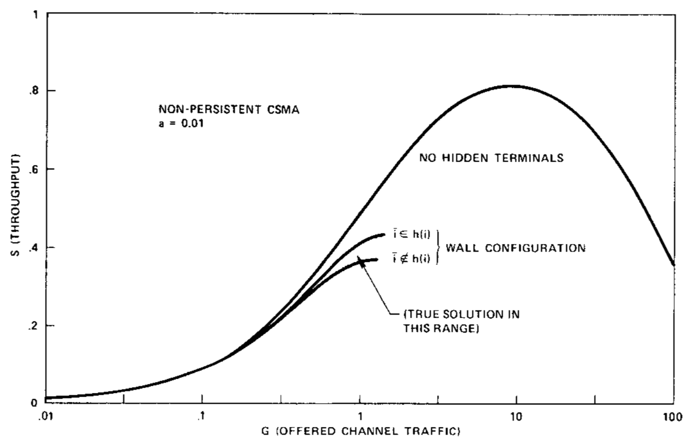

Fig. 9. The wall configuration—throughput versus offered channel traffic.

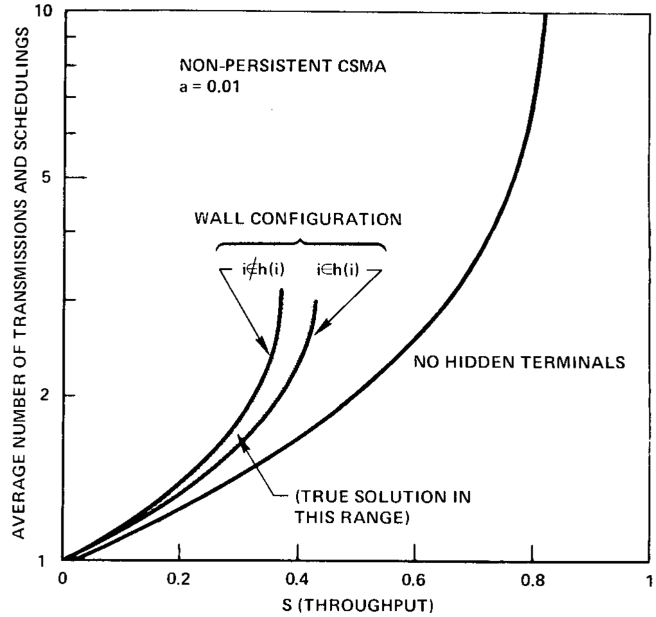

Fig. 10. The wall configuration—overall average number of schedulings and transmissions.

of hidden elements. Fortunately, in a single-station environment, the hidden-terminal problem can be eliminated by dividing the available bandwidth into two separate channels: a busy-tone channel and a message channel. This solution is the subject of the following section.

III. BUSY-TONE MULTIPLE-ACCESS (BTMA)

A. System Operation and Protocol.

The operation of BTMA rests on the assumption that the station is, by definition, within range and in line-of-of-sight of all terminals. The total available bandwidth is to be divided into two channels: a message channel and a busy-tone (BT) channel.[4] As long as the station senses (terminal) carrier on the incoming message channel it transmits a (sine wave) busy-tone signal on the busy-tone channel. It is by sensing carrier on the busy-tone channel that terminals determine the state of the message channel. The action pertaining to the transmission of the packet that a terminal takes (again) is prescribed by the particular protocol being used. In this section, we shall restrict ourselves to the nonpersistent protocol because of its simplicity in analysis and in implementation, as well as its relatively high efficiency as shown in Part I. In CSMA [1], the difficulty of detecting the presence of a signal on the message channel when this message used the entire bandwidth was minor and therefore was neglected. It is not so when we are concerned with the (statistical) detection of the (sine wave) busy-tone signal on a narrow-band channel. The detection time, denoted by t_d, is no longer negligible and must be accounted for. The nonpersistent BTMA protocol is similar to the nonpersistent CSMA protocol [1] and corresponds to the following. Whenever a terminal has a packet ready for transmission, it senses the busy-tone channel for t_d seconds (the detection time) at the end of which it decides whether the busy-tone signal is absent (in which case it transmits the packet); otherwise it reschedules the packet for transmission at some later time incurring a random rescheduling delay. At this new point in time, it senses the busy-tone channel and repeats the algorithm. In the event of a conflict, which the terminal learns about by failing to receive an acknowledgment from the station, the terminal again reschedules the transmission of the packet for some later time, and repeats the above process.

Of interest to this paper is first, the determination of the channel capacity under a nonpersistent BTMA protocol and second, the throughput delay characteristics of the latter. The total available bandwidth being the limiting resource, the problem then reduces to selecting the system parameters in order to achieve the best system performance. For this analysis we make the same assumptions as in Part I. While the effect of noise is assumed to

[4] The busy-tone concept in the context of packet radio was first suggested by Fralick [3].

be negligible on the message channel, we do account for it in the (narrow-band) busy-tone channel (See Section III-B). τ is the one-way propagation delay to (and from) the station. Each packet is of constant length requiring T_m seconds for transmission on the message channel. We let $S_m = \lambda T_m$. Let ψ be the fraction of the bandwidth assigned to the busy-tone channel. The overall channel utilization S is

$$S = (1 - \psi) S_m.$$

Let γ denote the mean offered traffic rate. (This is the rate of sense points since each arrival corresponds in this protocol to sensing the busy-tone channel before taking an action.) In Section III-B the problem of detecting a sine wave signal on a narrow-band channel is examined and the effect of various system parameters is characterized.

B. Signal Detection

The detection of the busy-tone signal is the problem of detecting a signal of known form in the presence of noise. The useful signal is a given function with some unknown parameters, namely, phase and amplitude.[5] However the observation (detection) time is usually small compared to the "fluctuation time" of these parameters, and the unknown phase and amplitude can be regarded as constant.

The problem of detecting a signal in a background of white Gaussian noise is a classical statistical problem involving the choice of one hypothesis from two mutually exclusive hypotheses. This has been extensively studied in the literature [4]. The quality of the decision can be characterized by the following two probabilities:

- D Probability of correct detection (in presence of the signal)
- F Probability of incorrect detection or false alarm.

Again let t_d be the observation time, i.e., the width of the window over which the channel is observed in making the decision. If μ_{t_d} is the signal-to-noise ratio (SNR)[6] when the signal is present over the entire window t_d, and $r*$ is the optimum threshold of the statistical detector, then for the usually assumed Rayleigh channel, it can be shown that [4]

$$\begin{aligned} F &= \exp\{-r*^2/2\mu_{t_d}\} \\ r* &= (-2\mu_{t_d} \log F)^{1/2}. \end{aligned} \tag{23}$$

That is, for a given observation time t_d and a desired false alarm probability F, the optimum threshold $r*$ can be determined from (23).

In this case (i.e., when the *useful* signal is present over the entire window), the probability of correct detection is given by [4]

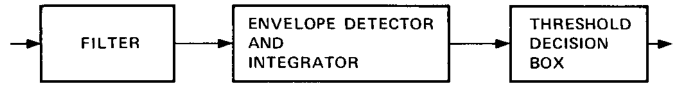

Fig. 11. Block diagram of the busy tone signal detector.

$$\begin{aligned} D_{t_d} &= \exp\left(-\frac{r*^2}{2\mu_{t_d}(1+\mu_{t_d})}\right) \\ &= F^{1/(1+\mu_{t_d})}. \end{aligned} \tag{24}$$

Equation (24) rests on the fact that the SNR of the received useful signal during the observation time is actually μ_{t_d}. In the following paragraph we investigate the (transient) situations where the signal may be present during only a fraction of the observation window.

Transient Behavior: The detector at the receiver consists mainly of a filter, an integrator, and a threshold decision box (see Fig. 11). Assume the step response of the busy-tone detect filter is exponential [3]; the amplitude at time t for the output of the busy-tone filter is then[7]

$$A_{BT}(t) = A_{\max}(1 - e^{t/k}) \tag{25}$$

where $A_{\max}$ is the maximum amplitude and k is the filter time constant. If we assume [3] that the same peak power is used for the busy tone as for the message on the message channel, then

$$A_{\max} = A_m$$

where A_m is the amplitude of the message signal on the message channel. Since the energy of a signal is the integral of its squared amplitude (which equals $(A^2t/2)$ for a sinusoid of amplitude A and duration t), we define the SNR as the ratio of the signal energy to the noise-energy and express it as

$$\mu = \frac{A^2}{2N_0W}$$

where W is the bandwidth of the channel under consideration and N_0 is the (assumed white) noise power density.

Let μ_m be the SNR of the message on the message channel required for suitable operation (typically $\mu_m = 10$). Then from the last equation, it is clear that

$$A_{\max} = A_m = (2N_0W_m\mu_m)^{1/2}$$

and

$$A_{BT}(t) = [2N_0W_{BT}\mu(t)]^{1/2} = (2N_0W_m\mu_m)^{1/2}(1 - e^{-t/k}).$$

Taking the time constant k to be one half of the inverse of the busy-tone channel bandwidth and recalling that ψ is the fraction of the total bandwidth W assigned to the busy-tone channel, we have the following function defining the SNR $\mu(t)$ on the busy-tone channel:

$$\mu(t) = \mu_m \frac{1-\psi}{\psi} (1 - \exp\{-2\psi Wt\})^2. \tag{26}$$

[5] Because of the mobility of terminals, the signal fluctuates. Thus we assume it to be of unknown amplitude. In the case of fixed terminals, we may idealize the problem to be that of detecting a signal with known amplitude but unknown phase.

[6] Ratio of the signal energy to the noise energy in the time-frequency window.

[7] At time $t = 0$, the filter is just turned on, and the signal is assumed to be present.

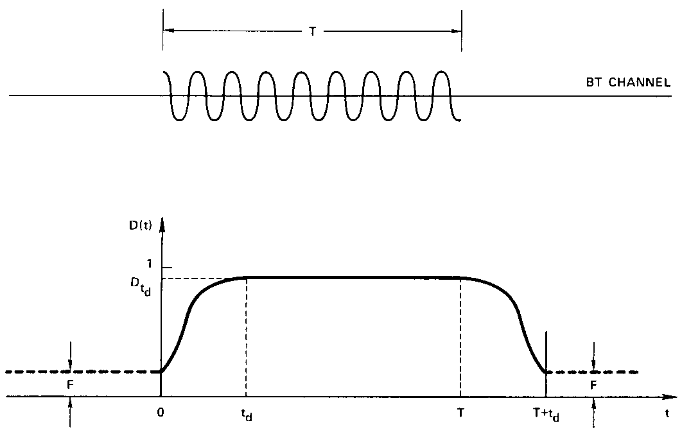

Fig. 12. $D(t)$ for an isolated busy tone signal of length T.

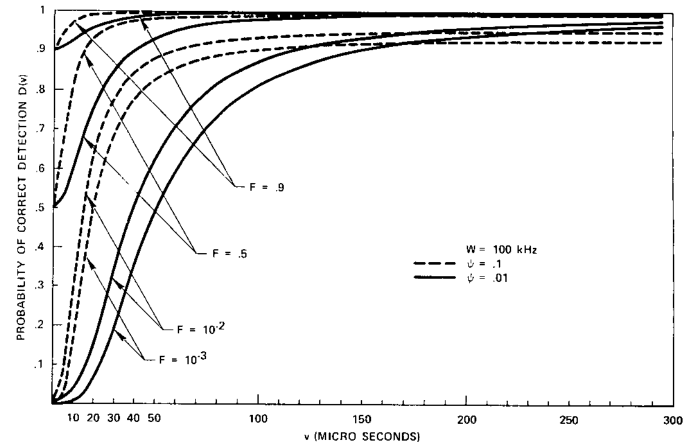

Fig. 13. Probability of correct detection $D(v)$.

Particularly

$$\mu_{t_d} = \mu(t_d) = \mu_m \frac{1-\psi}{\psi} (1 - \exp\{-2\psi W t_d\})^2. \quad (27)$$

As the starting time of a busy tone is unpredictable, the busy tone may not be present during the entire window of t_d seconds. The resulting SNR is a function of the time u over which the busy-tone signal is present [see (26)]. By the same token, the probability of proper detection is also a function of the time over which the busy-tone signal is present.

Consider now a signal starting at $t = 0$ and terminating at $t = T$. Let $D(t)$ be the probability of correct detection at time t after having observed the channel over t_d seconds (t is the time at which the decision is made). $D(t)$ is determined by [4]

$$D(t) = F^{\{1/[1+\mu(u)]\}} \quad (28)$$

where

$$u = \begin{cases} t, & \text{if } 0 \le t \le t_d \\ t_d, & \text{if } t_d \le t \le T \\ T - t + t_d, & \text{if } T < t \le T + t_d. \end{cases}$$

For $t > T + t_d$, the probability of false alarm is F. $D(t)$ is sketched in Fig. 12. A detailed graph of $D(t)$ for $0 < t < t_d$ can be seen in Fig. 13 in which we plot, for various values of F and ψ, the function

$$D(v) = F^{\{1/[1+\mu(v)]\}}, \qquad v \in [0,\infty]$$

where $\mu(v)$ is given in (26). ($D(v)$ here is the probability of correct detection if the useful signal is present over v seconds.) For very large v, namely when $v \to \infty$, the probability of correct detection reaches an asymptotic value equal to

$$D(v) \underset{v\to\infty}{\longrightarrow} F^{(1+[(1-\psi)/\psi]\mu_m)^{-1}}.$$

As usual, the larger F is, the better is the probability of correct detection. It is interesting to note that when v increases from 0, $D(v)$ increases quite rapidly reaching its asymptotic value for relatively small v. Moreover, this increase is faster for larger busy-tone bandwidth (larger ψ). A more complex situation occurs when two busy-tone signals are separated by a gap shorter than t_d. The window now can overlap over two busy-tone signals. Let $t = 0$ be the time at which the first signal terminates and t_g be the gap between the two signals. $D(t)$ for various values of t_g is sketched in Fig. 14. In the following section further approximations are introduced to avoid dealing with these complex situations. The approximations are checked to have a negligible effect on the evaluation of the system performance.

C. Throughput Analysis

Let us first summarize the notation in use:

τ	One-way propagation delay.
b_m	Number of bits per packet.
W	Total bandwidth available.
ψ	Fraction of W assigned to the busy tone channel.
T_m	Transmission time of a packet on the message channel; $T_m = \dfrac{b_m}{(1-\psi)W}$.
γ	Rate of the offered channel traffic (in packets/s).
G	Normalized rate of offered channel traffic (i.e., $G = \gamma T_m$).
t_d	Detection time.
F	The false alarm probability.
$D(t)$	Probability of correct detection given the signal is present for t seconds.

We wish to solve for the channel capacity, given the system parameters F,ψ,W,b_m,τ,t_d. This we do by solving for S in terms of γ and other system parameters. The channel capacity is then found by maximizing S in respect to γ.

Contrary to the CSMA modes the fraction of the population which decides to transmit is a function of time. The analytical approach consists (as in Part I [1]) of identifying the busy and idle periods and of determining the

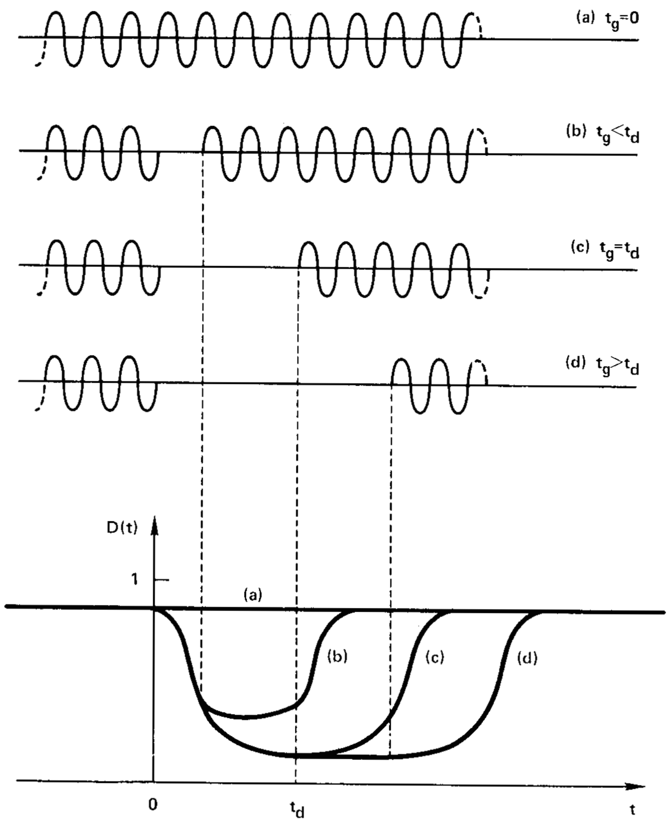

Fig. 14. $D(t)$ for various values of the gap t_g between two consecutive busy tone signals.

condition for a successful transmission over the busy period.

As stated above, to keep the analysis simple, some approximations will be made yielding a lower bound on throughput. Corresponding numerical results are presented and discussed in the following section.

Under stationary conditions and the model assumptions, a lower bound on the channel utilization S_l is given by

$$S \geq S_l = \frac{b_m}{W} \frac{\exp\left(-\gamma m(0,T_m)\right)}{\bar{B} + \bar{I}} \tag{29}$$

where the quantities $\bar{B},\bar{I},m(0,T_m)$ are defined in the following proof in (42), (46), and (31), respectively.

Proof: Since a terminal senses the busy tone channel for t_d seconds before deciding to transmit, an arrival is not effective, as far as the channel operation is concerned, until t_d seconds following its arrival time. To view the occurrence of events in this access mode more easily, we consider the Poisson arrival process to be shifted t_d seconds in time, bearing in mind that the terminal has already observed the channel for t_d seconds. The busy-tone signal emitted by the station is, as seen by the terminals, shifted in time by 2τ, a round-trip delay to the station (time until the station hears the transmission) and back (time until the busy tone from the station is heard by the other terminals).

If we let γ (packets/s) denote the offered traffic, the rate of transmission of packets at any time t is $\alpha(t)\gamma$ such that

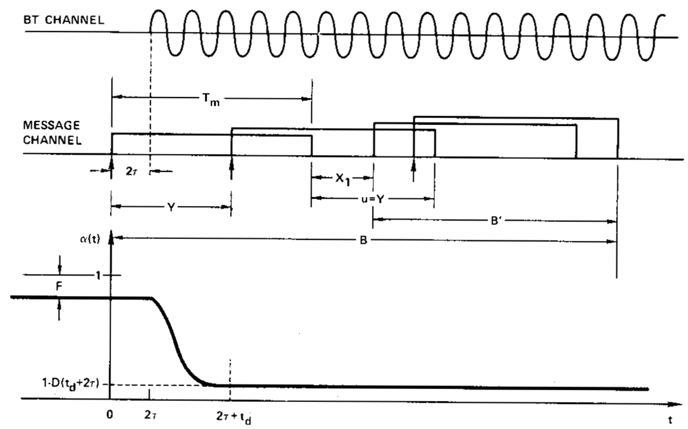

Fig. 15. BTMA—$\alpha(t)$ at the start of a busy period.

$$\alpha(t) = \begin{cases} 1 - D_{t_d}, & \text{if the busy-tone signal is present during the entire observation window} \\ 1 - F, & \text{if the busy-tone signal is absent during the entire window} \\ 1 - D(v), & \text{for the case when the busy-tone signal is only present over } v \text{ s, } v \leq t_d \end{cases}$$

where $D(t)$ is the probability of correct detection at time t and F is the false alarm probability. That is, at any time t there is a fraction $\alpha(t)$ of the population which decides to transmit.

Contrary to CSMA, an arrival in the middle of an ongoing transmission has a nonzero probability $(1 - D(t))$ of transmitting; thus, a busy period (period of time over which the channel is continuously used) can now exceed $T_m + 2\tau$.

Let t_{P1} be the time at which the transmission of the first packet of a busy period starts. The busy-tone signal of such a busy period starts 2τ seconds later, at time $t_{P1} + 2\tau$. The time dependent *packet transmission rate* $\alpha(t)\gamma$ depends on the length of the gap between the start of the busy-tone signal corresponding to the current busy period and the end of the busy-tone signal corresponding to the previous busy period. Let us assume $t_g > t_d + 2\tau$ for the present time. (This corresponds to the worst yet "cleanest" case.) Without loss of generality, let $t_{P1} = 0$. From the previous section, we can easily derive $\alpha(t)$ as follows (see Fig. 15):

$$\alpha(t) = \begin{cases} 1 - F & 0 < t \leq 2\tau \\ 1 - D(t), & 2\tau < t \leq 2\tau + t_d \\ 1 - D(t_d + 2\tau), & 2\tau + t_d < t \leq T_m \end{cases} \tag{30}$$

where

$$D(t) = F^{\{1/[1+\mu(t)]\}}$$

$$\mu(t) = \mu_m \frac{1-\psi}{\psi} [1 - \exp\{-2\psi W(t-2\tau)\}]^2.$$

Let

$$m(x_1,x_2) = \int_{x_1}^{x_2} \alpha(t)\, dt. \tag{31}$$

The first packet in the busy period is successful if there is no arrival from the nonhomogeneous Poisson arrival process with rate $\alpha(t)\gamma$ during the entire transmission time of the packet. The probability of success is given by

$$P_s = \exp(-\gamma m(0,T_m)). \tag{32}$$

To find the channel throughput, we use a "cycle" analysis [5]. For this, we must calculate the average busy and idle periods.

Let us first find the average busy period. Let Y be the random variable defining the time until the last arrival during the packet transmission time T_m (see Fig. 15). It is easy to see that:

$$\begin{aligned} \Pr\{Y \le y\} &= \Pr\{\text{no arrival in } (y,T_m)\} \\ &= \exp[-\gamma m(y,T_m)] \\ \Pr\{Y = 0\} &= \exp[-\gamma m(0,T_m)]. \end{aligned} \tag{33}$$

The Laplace transform $Y^*(s)$ is defined by

$$Y^*(s) \triangleq \int_0^{T_m} \exp\{-sy\}\, d\Pr\{Y \le y\}.$$

From this we obtain

$$\begin{aligned} Y^*(s) &= \Pr\{Y = 0\} \\ &\quad + \int_0^{T_m} \gamma \exp\{-sy\} \exp\{-\gamma m(y,T_m)\}\, dm(y,T_m) \\ &= \exp\{-\gamma m(0,T_m)\} \\ &\quad + \int_0^{T_m} \gamma\alpha(y) \exp\{-sy\} \exp\{-\gamma m(y,T_m)\}\, dy. \end{aligned} \tag{34}$$

Let

$$D \equiv D(t_d + 2\tau) = D_{t_d} \tag{35}$$

$$\Delta = 1 - D \tag{36}$$

$$\Phi = 1 - F. \tag{37}$$

The busy period is equal to:

$$\begin{cases} T_m + Y, & \text{if no arrival occurs in the period of time } (T_m, T_m + Y) \\ T_m + X_1 + B', & \text{otherwise} \end{cases}$$

where (see Fig. 15) 1) X_1 is the random variable defining the time elapsed from $t = T_m$ up to the first arrival in $(T_m, T_m + Y)$, and 2) B' is the length of a busy period when the Poisson arrival process is homogeneous with rate $\Delta\gamma$. (B' is the "sub-busy" period (see [5]) created by the first arrival in $(T_m, T_m + Y)$. Beyond $t = T_m$, $\alpha(t) = \Delta$.)

Conditioning on $Y = y$, the average busy period is given by

$$\begin{aligned} \bar{B}_y &= (T_m + y)e^{-\Delta\gamma y} + (T_m + \bar{X}_{1y} + \bar{B}')(1 - e^{-\Delta\gamma y}) \\ &= T_m + ye^{-\Delta\gamma y} + (\bar{X}_{1y} + \bar{B}')(1 - e^{-\Delta\gamma y}). \end{aligned}$$

The expected value $\bar{X}_{1y}$ (defined as the expected value of X_1 conditioned on $Y = y$) is derived as follows. First, we have

$$\Pr\{X_1 \le x/\text{some arrival occurred}, Y = y\} = \frac{1 - e^{-\Delta\gamma x}}{1 - e^{-\Delta\gamma y}}.$$

Therefore,

$$\begin{aligned} \bar{X}_{1y} &= \frac{1}{1 - e^{-\Delta\gamma y}} \int_0^y (e^{-\Delta\gamma x} - e^{-\Delta\gamma y})\, dx \\ &= \frac{1}{1 - e^{-\Delta\gamma y}} \left[\frac{1}{\gamma\Delta}(1 - e^{-\Delta\gamma y}) - ye^{-\Delta\gamma y} \right] \\ &= \frac{1}{\Delta\gamma} - \frac{ye^{-\Delta\gamma y}}{1 - e^{-\Delta\gamma y}}. \end{aligned} \tag{38}$$

The expected busy period $\bar{B}'$ is given by

$$\bar{B}' = \frac{1}{\Delta\gamma}\left(e^{\Delta\gamma T_m} - 1\right). \tag{39}$$

Indeed, we have

$$B' = \begin{cases} T_m, & \text{with probability } e^{-\Delta\gamma T_m} \\ X_2 + B', & \text{with probability } 1 - e^{-\Delta\gamma T_m} \end{cases}$$

where X_2 is the time elapsed from the start of transmission of a packet until the start of transmission of the first overlapping packet. $\bar{X}_2$ is obtained from $\bar{X}_{1y}$ (37) by simply setting $y = T_m$.

$$\bar{X}_2 = \frac{1}{\Delta\gamma} - \frac{T_m e^{-\Delta\gamma T_m}}{1 - e^{-\Delta\gamma T_m}}. \tag{40}$$

Therefore,

$$\begin{aligned} \bar{B}' &= T_m e^{-\Delta\gamma T_m} + (\bar{X}_2 + \bar{B}')(1 - e^{-\Delta\gamma T_m}) \\ &= T_m e^{-\Delta\gamma T_m} + \frac{1 - e^{-\Delta\gamma T_m}}{\Delta\gamma} \\ &\quad - T_m e^{-\Delta\gamma T_m} + \bar{B}'(1 - e^{-\Delta\gamma T_m}) \\ &= \frac{1}{\Delta\gamma}(e^{\Delta\gamma T_m} - 1). \end{aligned}$$

The average busy period $\bar{B}_y$, conditioned on $Y = y$, is given by

$$\bar{B}_y = T_m + ye^{-\Delta\gamma y} + \frac{1 - e^{-\Delta\gamma y}}{\Delta\gamma} - ye^{-\Delta\gamma y}$$

$$+ \frac{1}{\Delta\gamma}(e^{\Delta\gamma T_m} - 1)(1 - e^{-\Delta\gamma T_m})$$

$$= T_m + \frac{1}{\Delta\gamma} e^{\Delta\gamma T_m}(1 - e^{-\Delta\gamma y}). \quad (41)$$

Removing the condition on Y, we get

$$\bar{B} = \int_0^{T_m} \bar{B}_y \, d\Pr\{Y \le y\}$$

$$= T_m + \frac{1}{\Delta\gamma} e^{\Delta\gamma T_m}(1 - Y^*(\Delta\gamma)). \quad (42)$$

Let us now calculate the average idle period. Let $t = 0$ now denote the *end* of the busy period. It is easy to see that the transmission rate, which we denote by $\alpha'(t)\gamma$, is defined by (see Fig. 16)

$$\alpha'(t) = \begin{cases} 1 - D(t_d + 2\tau) = \Delta, & 0 \le t \le 2\tau \\ 1 - D'(t), & 2\tau \le t \le 2\tau + t_d \\ 1 - F = \Phi, & t \ge 2\tau + t_d \end{cases} \quad (43)$$

where

$$D'(t) = F^{\{1/[1+\mu'(t)]\}}.$$

Denoting by I the idle period, we have

$$\Pr\{I > z\} = \Pr\{\text{no arrival occurs in } (0,z)\}$$
$$= \exp[-\gamma m'(0,z)] \quad (44)$$

where

$$m'(z_1,z_2) = \int_{z_1}^{z_2} \alpha'(t)\, dt. \quad (45)$$

The average idle period is given by

$$\bar{I} = \int_0^\infty \exp\{-\gamma m'(0,z)\}\, dz. \quad (46)$$

The exact expression[8] for the channel throughput depends on the distribution of the gap t_g. Indeed, $D(t)$ as well as $\alpha(t)\gamma = [1 - D(t)]\gamma$ depends on t_g as shown in Fig. 14. However, to keep the analysis fairly tractable, we choose to give a lower bound on channel throughput by considering $\alpha(t)\gamma$ as determined in the worst case ($t_g \gg t_d$), that is as given in (30). Under this condition, the first transmission in the busy period is successful with probability

$$P_s = \exp(-\gamma m(0,T_m))$$

and the expected period of time $\tilde{U}$ that the channel is

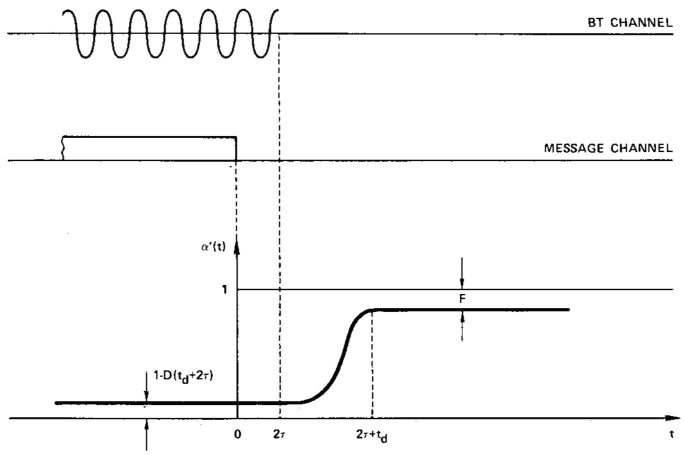

Fig. 16. BTMA—$\alpha'(t)$ at the start of an idle period.

transmitting packets without interference during a cycle (defined as a busy period plus an idle period) is

$$\tilde{U} = T_m \exp[-\gamma m(0,T_m)].$$

Therefore, the lower bound on throughput is given by

$$S_l = (1 - \psi)\frac{T_m \exp[-\gamma m(0,T_m)]}{\bar{B} + \bar{I}}$$

$$= \frac{b_m}{W}\frac{\exp[-\gamma m(0,T_m)]}{\bar{B} + \bar{I}}. \qquad \text{Q.E.D.}$$

"Upper bound" on S: To check the validity of the lower bound provided by (29), we consider the fraction f of busy periods for which the gap t_g is less than $t_d + 2\tau$. Indeed, it is for these cases that $\alpha(t)$ is overestimated by the expression given in (30), and could be underestimated by $1 - D_{t_d} = \Delta$.

The probability that the gap t_g is less than $t_d + 2\tau$ is readily given by

$$f = \Pr\{t_g \le t_d + 2\tau\} = 1 - \exp\{-\gamma m'(0,t_d + 2\tau)\}. \quad (47)$$

The smaller f is, the closer is the lower bound on S to the exact expression. On the other hand, an "upper bound"[9] on S is obtained by underestimating $\alpha(t)$ by Δ for a fraction f of the busy periods. That is, a lower bound on the expected busy period is obtained by

$$\bar{B}_l = T_m + \frac{1}{\Delta\gamma} e^{\Delta\gamma T_m}[1 - fY_1^*(\Delta\gamma) - (1 - f)Y^*(\Delta\gamma)] \quad (48)$$

where $Y_1^*(s)$ is the Laplace transform of Y when $\alpha(t) = \Delta\gamma$, and is given by

$$Y_1^*(s) = \gamma\Delta\frac{e^{-sT_m} - e^{-\gamma\Delta T_m}}{\gamma\Delta - s} + e^{-\gamma\Delta T_m} \quad (49)$$

[8] Exact under the provision that a is constant.

[9] With respect to the approximation concerning the gap t_g.

so that

$$Y_1^*(\Delta\gamma) = e^{-\gamma\Delta T_m}[1 + \gamma\Delta].$$

An upper bound on the probability of success of the first transmission in a busy period is obtained by

$$P_{s_u} = f \exp\{-\gamma\Delta T_m\} + (1-f)\exp\{-\gamma m(0,T_m)\}.$$

The upper bound on throughput is then given by

$$S_u = \frac{B_m}{W} \frac{f \exp\{-\gamma\Delta T_m\} + (1-f)\exp\{-\gamma m(0,T_m)\}}{\bar{B}_l + \bar{I}}. \tag{50}$$

Limit when $t_d \to 0$: When $t_d \to 0$, the channel capacity reduces to

$$S = (1-\psi)\frac{1}{2e}. \tag{51}$$

Proof: We realize that in the limit ($t_d \to 0$), the problems caused by the transient behavior are insignificant (nonexistent when $t_d = 0$). We see then from (24) that

$$D_{t_d} = F^{\{1/[1+\mu_{td}]\}} \to F$$

since $\mu_{t_d} \to 0$. Therefore, $\alpha(t)$ is constant and equal to $\Phi = \Delta = 1 - F$; (34) becomes:

$$Y^*(s) = \exp\{-\Phi\gamma T_m\} + \Phi\gamma \exp\{-\Phi\gamma T_m\}\left(\frac{\exp\{(-s+\Phi\gamma)T_m\} - 1}{\Phi\gamma - s}\right)$$

yielding

$$Y^*(\Delta\gamma) = \exp\{-\Phi\gamma T_m\}(1 + \Phi\gamma T_m).$$

Therefore, $\bar{B}$ as given by (42) becomes

$$\bar{B} = \frac{1}{\Phi\gamma}\exp\{\Phi\gamma T_m\} - \frac{1}{\Phi\gamma}.$$

Similarly, we can see that

$$\bar{I} = \frac{1}{\Phi\gamma}.$$

Substituting in (29) we get

$$S = (1-\psi)\Phi\gamma T_m \exp\{-2\Phi\gamma T_m\}$$

yielding a channel capacity equal to $(1-\psi)(1/2e)$ under the optimum condition $\Phi\gamma T_m = 1$. Q.E.D.

D. Numerical Results and Discussion

The expressions given in (29) and (50) relate the throughput S to the offered channel traffic γ. When all the system parameters have fixed values, the *information capacity* of the channel is defined as the maximum achievable throughput. This throughput is obtained at an optimum value of the traffic γ and results in infinite packet delays. To obtain finite delays, we must reduce the throughput below the capacity.

The design problem in BTMA consists of maximizing the channel capacity (under the nonpersistent protocol) by properly selecting the design variables ψ, F, and t_d when the number of bits per packet, b_m, and the total available bandwidth W are assumed to be given. Because of the complicated form of the expressions for S, numerical optimization techniques are used. Below, we first discuss the restrictions on the input data and the accuracy of the approximations. Then we give the numerical results in the form of curves. The subsequent curves depict the changes in system performance due to variations in the design parameters. The various tradeoffs which influence the performance of BTMA are also discussed.

Let us first discuss some restrictions and approximations. To reduce the dimensionality of the problem, and to provide an easy comparison with the previously analyzed CSMA protocols we restrict ourselves to the following:

τ (maximum propagation delay) $= 100\ \mu s$[10]

$\mu_m \doteq 10$

$\dfrac{b_m}{W} = 10^{-2}$ seconds.[11]

We consider two cases for b_m and W.

Case I: $b_m = 1000$ bits; $W = 10^5$ Hz.
Case II: $b_m = 10\,000$ bits; $W = 10^6$ Hz.

It is important to note that, for the same μ_m, Case II requires higher transmitting power than does Case I; the following curves also show that Case II offers a channel capacity higher than that offered by Case I.

Along with the numerical computation of S_l and S_u we computed the probability f that the idle period is smaller than $t_d + 2\tau$. The probability f never exceeded a few percent (less than 0.04) and the two estimates on throughput are *very close* to each other. (As an example, in Table I we give the values of f encountered for various values of the system parameters.) Therefore all numerical results given in the sequel will only correspond to the lower bound S_l.

For $F = 10^{-3}$ and various values of ψ we plot in Fig. 17 the channel capacity versus the observation window t_d. Similar curves can be plotted for other values of F. For each couple (F,ψ), the channel capacity reaches its maximum at some optimum value of t_d. This optimum is explained by the fact that the larger t_d is, the better is the probability of correct detection D_{t_d} when the signal is present during the entire window. However, the larger t_d is, the longer the idle period will be, as it can be seen from Fig. 16. The effect is reversed as t_d gets smaller.

Note that when the observation window shrinks to 0, the capacity of the channel decreases to $(1-\psi)(1/2e)$

[10] This corresponds to a maximum distance of about 20 miles. The ratio of propagation delay to transmission time of a packet, denoted by a, is in all cases less than (but very close to) or equal to 0.01.
[11] The bandwidth is assumed to be modulated at 1 bit/Hz·s.

TABLE I
ACCURACY OF THE APPROXIMATIONS

	γ	S_l	S_u	f
	10	0.0897	0.0897	0.0008
$F = 10^{-3}$	100	0.4580	0.4581	0.0084
$\psi = 10^{-2}$	400	0.6245	0.6269	0.0333
$t_d = 7 \times 10^{-4}$	500	0.6201	0.6238	0.0414
	600	0.6084	0.6137	0.0495
	10	0.0890	0.0890	0.0006
$F = 10^{-2}$	100	0.4586	0.4587	0.0067
$\psi = 10^{-2}$	500	0.6411	0.6440	0.0334
$t_d = 7 \times 10^{-4}$	600	0.6338	0.6380	0.0400
	700	0.6222	0.6279	0.0465
	10	0.0818	0.0818	0.0004
$F = 10^{-1}$	100	0.4443	0.0444	0.0045
$\psi = 10^{-2}$	500	0.6635	0.6653	0.0226
$t_d = 5 \times 10^{-4}$	600	0.6625	0.6651	0.0271
	700	0.6565	0.6599	0.0315
	10	0.0473	0.0473	0.0002
$F = 0.5$	100	0.3206	0.3206	0.0021
$\psi = 10^{-2}$	500	0.6318	0.6322	0.0106
$t_d = 5 \times 10^{-4}$	1000	0.6807	0.6825	0.0210
	2000	0.6406	0.6476	0.0417
	100	0.2249	0.2249	0.0012
$F = 0.7$	500	0.5511	0.5512	0.0062
$\psi = 10^{-2}$	1000	0.6540	0.6546	0.0125
$t_d = 5 \times 10^{-4}$	2000	0.6830	0.6855	0.0248
	3000	0.6583	0.6639	0.0370
	1000	0.4692	0.4692	0.0041
$F = 0.9$	2000	0.6018	0.6020	0.0083
$\psi = 10^{-2}$	3000	0.6552	0.6558	0.0124
$t_d = 5 \times 10^{-4}$	4000	0.6779	0.6790	0.0165
	5000	0.6858	0.6876	0.0206

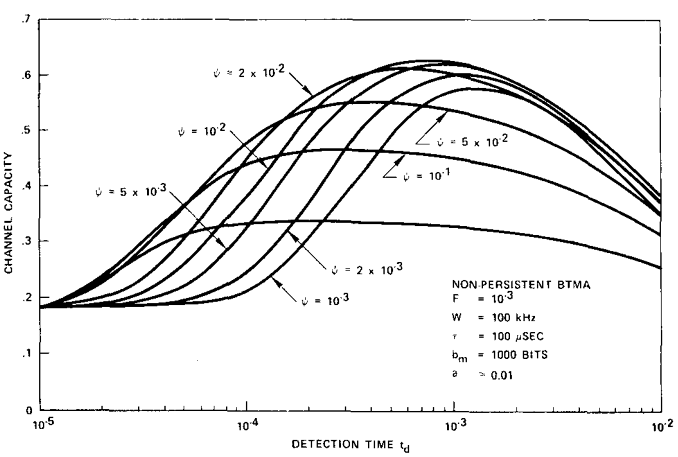

Fig. 17. BTMA—channel capacity versus observation window t_d.

(capacity provided by the pure ALOHA access mode [1], [6]) as shown in the previous section. Qualitatively speaking, $t_d \to 0$ reduces to very bad detection (i.e., $D \to 0$), and terminals behave in a pure ALOHA mode.

In Fig. 18, we plot for various F, the maximum capacity of the channel (maximized over t_d) versus ψ. We note here that the maximum capacity is not very sensitive to *small* variations of ψ. However there is a certain range of ψ which yields the best performance. For those values of F considered in the graph ($F = 10^{-3}, 10^{-2}, 10^{-1}, 0.5$), the optimum ψ is in the range $[10^{-2}, 2 \times 10^{-2}]$.

In Fig. 19, we plot the capacity (maximized over ψ and t_d) versus F. Note that for both Case I and Case II,

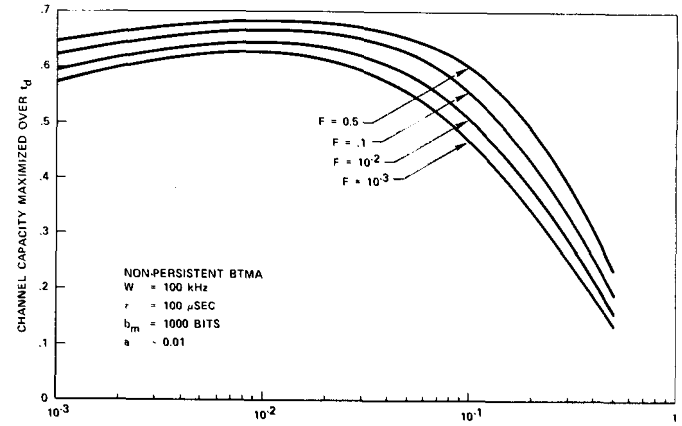

Fig. 18. BTMA—channel capacity (maximized over t_d) versus ψ.

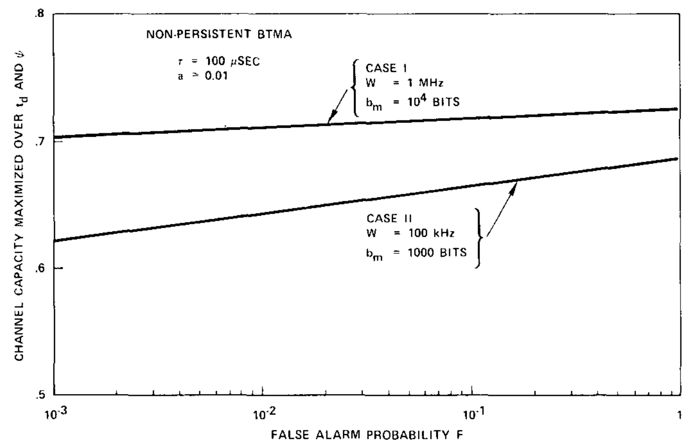

Fig. 19. BTMA—channel capacity (maximized over t_d and ψ) versus F.

the capacity of the channel is a logarithmic function of F. The ultimate performance ($\simeq$0.68 for Case I and $\simeq$0.72 for Case II) is obtained for $F \to 1$. However, the channel capacity is not very sensitive to variations of F. The effect of F can be explained as follows: For fixed values of ψ and t_d the larger F is, the better is the quality of correct detection $D(t)$ (see, for example, Fig. 13). Thus, for a particular value of γ, the channel time wasted by interference is smaller; the channel idle time, however, which is a function of $(1 - F)\gamma$ is larger; thus a tradeoff exists between idle time and time wasted by interference. The overall performance is not easily expressed. The plot corresponding to Case II exhibits the same linear behavior on the semilogarithmic graph, but acheives a larger capacity.[12]

To compare the delay performance of BTMA for various values of the system parameters, we first consider the quantity G/S, the average number of transmissions and schedulings that a packet incurs before successful transmission.

[12] The larger the bandwidth is, the better is the correct detection $D(v)$ [see (26)]. Thus larger W provides larger channel capacity. However, we note from Figs. 15 and 16 that the channel capacity is always bounded from above by the capacity of CSMA with propagation delay equal to 2τ.

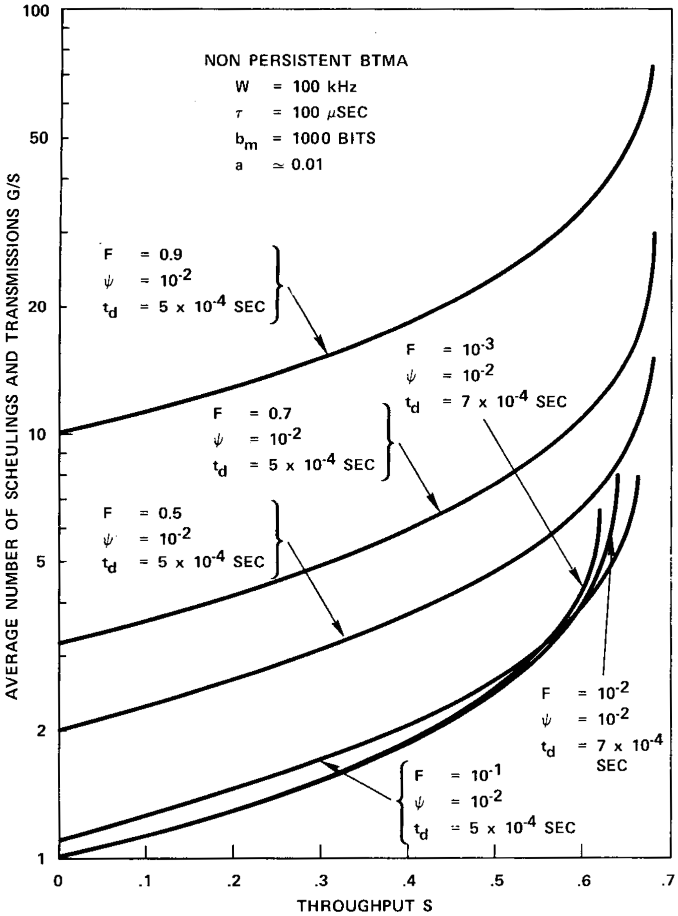

Fig. 20. BTMA—average number of schedulings and transmissions.

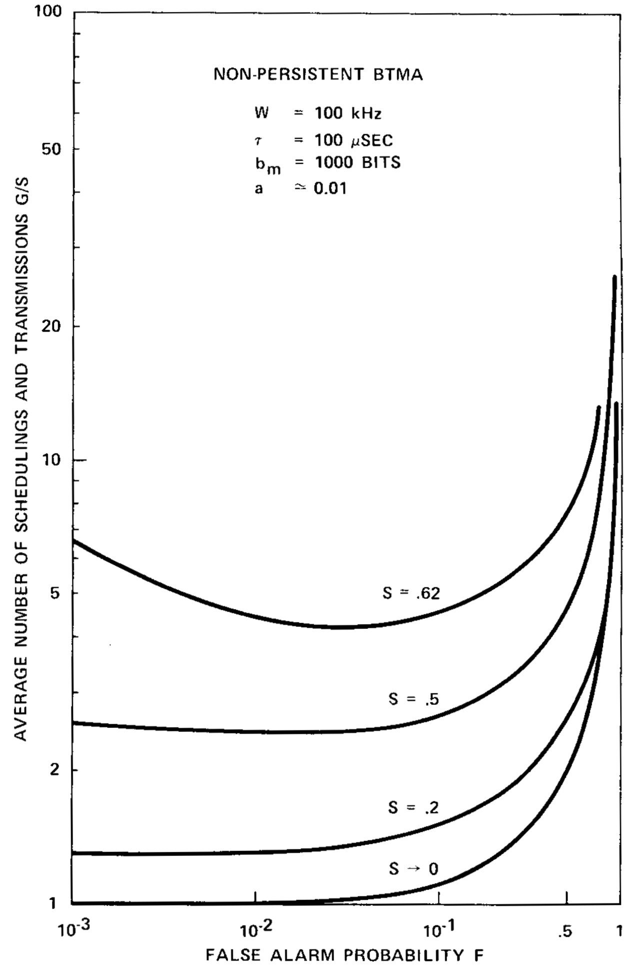

Fig. 21. BTMA—G/S versus F for constant S.

In Fig. 20 we plot, for each value of F, G/S versus S for those values of ψ and t_d yielding the maximum channel capacity. (Strictly speaking, we should plot G/S versus S for all pairs (ψ,t_d) and then draw their lower envelope. However, the difference between this lower envelope and the plotted curve corresponding to the optimum ψ and t_d for maximum capacity is so minor that we restrict ourselves to the latter.)

Note that for each value of S there exists a value of F minimizing G/S. However for relatively small values of S (not too close to the saturation point of the channel) we note that the higher the probability of false alarm F is, the larger is G/S. An explanation can be given by the following fact: when $G \to 0$ and $S \to 0$, the terminal incurs an average number of schedulings and transmissions equal to $(1/1 - F)$. This is shown on Fig. 20 at $S = 0$. (In some cases such as in the curve corresponding to $F = 0.9$, we have $\lim_{S\to 0} (G/S)$ slightly larger than $(1/1 - F)$; this is due to the fact that the curve does not correspond to the lower envelope.) To best compare the effect F has on G/S, we plot in Fig. 21 G/S versus F for constant S. Thus, we show that for $S < 0.5$ and $10^{-3} < F < 10^{-1}$, G/S is small and fairly insensitive to F, and that for $F > 10^{-1}$ it increases rapidly with increasing values of F. For larger values of throughput, the choice of F is more critical. A good operating point should then be in the flat part of the curves.

G/S, as a measure of delay, can be of importance since the complexity of the equipment and the implementation of the protocol can be directly related to the number of schedulings and transmissions that a packet incurs. For example, at each scheduling the terminal has to generate a random number determining the scheduling delay. Of even more importance in evaluating the performance of such a system is the determination of the actual packet delay, defined as the time lapse since the packet is first generated, until the time it is successful. As discussed in Part I [1], the mathematical determination of packet delays is fairly complex, and simulation techniques are employed. For various values of F ($F = 10^{-3}$ and $F = 0.5$), by selecting the optimum system parameters (ψ,t_d) with respect to channel capacity, we simulated the BTMA mode. In Fig. 22, we plot the throughput-minimum-delay[13] curve for these values of F. It is to be noted that, even though G/S can be significantly affected by F, the minimum delay is insensitive to F. However, for each value of S there exists a value of F which provides the lowest

[13] Delay is minimized with respect to $\bar{X}$. (See [1].) In BTMA, the larger F is, the larger is G/S. The minimum delay is obtained for very small values of $\bar{X}$.

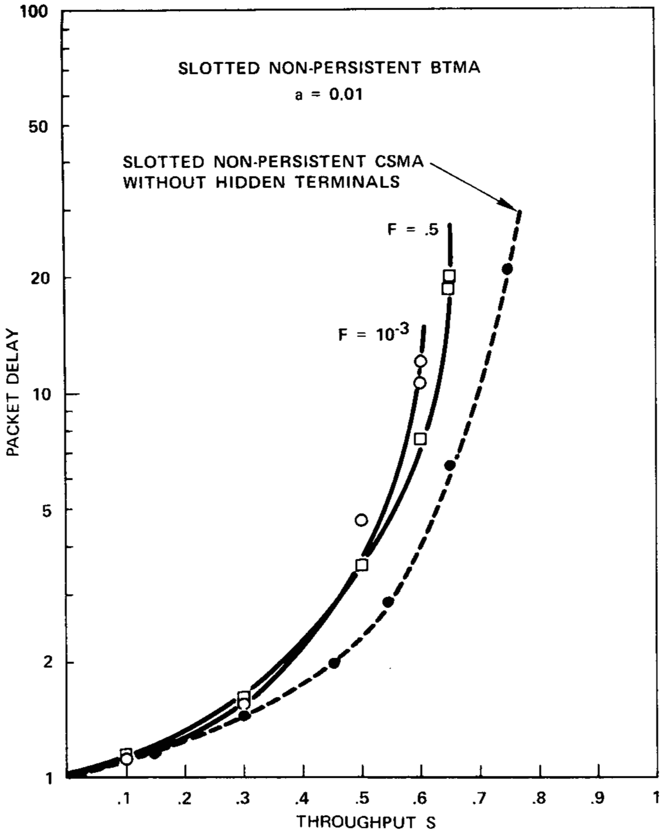

Fig. 22. BTMA—throughput-delay tradeoffs (a = 0.01).

delay. By comparing the lower envelope of these throughput-delay curves to the curve corresponding to the nonpersistent CSMA without hidden terminals, we note the relatively good performance of BTMA.

IV. CONCLUSION

The hidden-terminal problem seriously degrades the performance of CSMA. To eliminate this problem in single-station environments, the use of a busy-tone channel has been considered. In this paper, the nonpersistent BTMA mode has been analyzed and the approximations made in the analysis were shown to have a very minor effect in evaluating the channel performance. The channel capacity (optimized over ψ and t_d) and the packet delays are not very sensitive to F; but when G/S is considered, we see that (unreasonably) large values of F may have a significant effect even when $S < 0.5$ and that the choice of F is more critical for $S > 0.5$. For W = 100 kHz and $b_m = 10^3$ bits (Case 1), and $a \lesssim 0.01$, in order to keep G/S low, F should lie in the range $[10^{-3}, 10^{-1}]$. For this range of F the channel capacity lies in the range [0.62, 0.66]. These capacities are obtained for the optimum values of ψ lying in the range $[10^{-2}, 2 \times 10^{-2}]$ and the corresponding optimum values of t_d. Similar results are readily obtainable for other values of W.

TABLE II
CAPACITY C FOR THE VARIOUS PROTOCOLS CONSIDERED (a = 0.01)

Protocol	Capacity C
Pure ALOHA	0.184
Slotted ALOHA	0.368
1-persistent CSMA	0.529
Slotted 1-persistent CSMA	0.531
Nonpersistent BTMA	
W = 100 kHz	0.680
W = 1000 kHz	0.720
0.1-persistent CSMA	0.791
Nonpersistent CSMA	0.815
0.03-persistent CSMA	0.827
Slotted nonpersistent CSMA	0.857
Perfect scheduling	1.000

Thus the nonpersistent BTMA constitutes a fairly good solution to the "hidden terminal" problem, providing a maximum channel capacity of approximately 0.68 when $a = 0.01$ and W = 100 kHz, and 0.72 when $a = 0.01$ and W = 1 MHz, as compared to 0.82 for nonpersistent CSMA with no hidden terminals (0.85 for slotted nonpersistent CSMA with no hidden terminals). It should be noted that while in CSMA sensing the presence of a transmission involves a one-way propagation delay $a = 0.01$, BTMA requires a round-trip delay $2a = 0.02$ to perform the same operation. Furthermore, the performance of the nonpersistent BTMA is insensitive to the precise setting of the various system parameters ψ, t_d, and F.

We summarize the above results in Table II (CSMA capacities assume no hidden terminals) for $a = 0.01$.

REFERENCES

[1] L. Kleinrock and F. Tobagi, "Packet switching in radio channels: Part I—Carrier sense multiple access modes and their throughput delay characteristics," *IEEE Trans. Commun.*, this issue, pp. 1400–1416.

[2] F. Tobagi, "Random access techniques for data transmission over packet switched radio networks," Comput. Sci. Dep., School Eng. Appl. Sci., Univ. Calif., Los Angeles, UCLA-ENG 7499, Dec. 1974.

[3] S. Fralick, Stanford Res. Inst., Menlo Park, Calif., private communication.

[4] L. A. Wainstein and V. D. Zubakov, *Extraction of Signals from Noise*. Englewood Cliffs, N. J.: Prentice-Hall, 1962.

[5] L. Kleinrock, *Queueing Systems, Vol. I, Theory, Vol. II, Computer Applications*. New York: Wiley Interscience, 1975.

[6] N. Abramson, "The ALOHA system—Another alternative for computer communications," in *1970 Fall Joint Computer Conf., AFIPS Conf. Proc.*, vol. 37. Montvale, N. J.: AFIPS Press, 1970, pp. 281–285.

Fouad A. Tobagi, for a photograph and biography see page 1416 of this issue.

Leonard Kleinrock (S'55–M'64–SM'71–F'73), for a photograph and biography see page 423 of the April 1975 issue of this TRANSACTIONS.