

Space-Time Models of Asynchronous CSMA Protocols for Local Area Networks

MART L. MOLLE, MEMBER, IEEE, KHOSROW SOHRABY, AND ANASTASIOS N. VENETSANOPOULOS, SENIOR MEMBER, IEEE

Abstract—Carrier sense multiple-access (CSMA) protocols are widely used in local area networks (LAN's) to control access to a shared communications channel such as a coaxial cable or radio frequency band. CSMA protocols are designed to exploit the property that the signal propagation time across the LAN is much smaller than the packet transmission time. Consequently, their performance depends on the exact timing of a sequence of asynchronous events originating at different points in the network. Previous models of CSMA protocols have not captured this essential space-time characteristic. Instead, the system has been reduced to a 1-dimensional model (i.e., a sequence of events on a single time line) by assuming either that events are synchronous (i.e., "slotted" operation of the protocol) or that all stations are mutually equidistant (so it can be assumed that each station's time line is identical). The novelty in our work is to describe the system state in terms of an $(N + 1)$ -dimensional "ribbon" of space-time, allowing us to faithfully model the exact timings of events for LAN's where the stations are distributed over a fully connected region in N -space. For example, an Ethernet-like "bus" network can be viewed as a 1-dimensional LAN, a terrestrial packet radio network can be viewed as a 2-dimensional LAN, and so on. First we highlight some of the key properties exhibited by such a model in the general case, including the notion of an embedded Markovian sequence of "idle points," and the decomposition of the space-time ribbon into cycles (each of which can be further decomposed into a number of "regions"). Then we present extensive results for the interesting special case of a 1-dimensional LAN, i.e., an Ethernet-like "bus" network. Our results show that previous models of bus LAN's can significantly underestimate their performance.

I. INTRODUCTION

CARRIER sense multiple-access (CSMA) protocols are distributed algorithms for serializing the transmission of packets in a local area network (LAN). All the stations are connected to a common channel, such as a coaxial cable, radio frequency band, etc., where every transmission originating from each station is *broadcast* to all the other stations. The key properties that enable CSMA protocols to be used on this type of LAN are that i) we require each station to be able to track changes in the "local state" (i.e., idle/busy/. . .) of the common channel, and ii) the network diameter is sufficiently small for different

stations to observe the same "evolutionary history" of the local state, except for small perturbations in the state transition times.

In the case of CSMA, the first property is satisfied using *carrier sensing*, which allows each station to distinguish between an "idle" channel (where no packet transmissions are taking place) and a "busy" channel (where one or more concurrent packet transmissions are taking place). And in the case of CSMA/CD, we further assume that each station is capable of *collision detection*, and hence can determine within a few bit times whether the channel is currently "busy" because of a *single* packet transmission or *multiple, mutually interfering* packet transmissions (i.e., a *collision*), even if the station is among the active transmitters. Since the probability that a packet is corrupted because of random channel noise is typically orders of magnitude smaller than the probability that it is involved in a collision, it is customary to equate the event that a packet transmission is not part of a collision with the event that the packet transmission is "successful," i.e., that a correct copy of the packet is received by all stations, including the intended (set of) recipient(s).

The second property is obtained by limiting the distance between the stations in the LAN relative to the transmission time for a packet. More precisely, we require that in the worst case, the signal propagation time between any pair of transmitters is smaller than the transmission time for a packet.

Unlike more conventional models of CSMA, we allow the network to be a region in N -space where the position of each station, i for instance, is given by a point $\vec{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,N})$, and the distance between two stations i and j is given by the Euclidean distance between \vec{x}_i and \vec{x}_j in N -space, namely

$$a_{ij} \triangleq \sqrt{(x_{i,1} - x_{j,1})^2 + (x_{i,2} - x_{j,2})^2 + \dots + (x_{i,N} - x_{j,N})^2}.$$

Notice that $a_{ij} = a_{ji}$ holds (although many of our results in Sections II and III do not depend on this result), and so does the triangle inequality, i.e., $a_{ik} \leq a_{ij} + a_{jk}$. Without loss of generality, we shall normalize time so that the expected transmission time for a packet is unity (in which case the previous requirement may be written as $\max_{i,j} a_{ij} \ll 1$), and we normalize distance so that an average

Manuscript received February 3, 1986; revised December 22, 1986. This work was supported by the Natural Sciences and Engineering Research Council of Canada under Grants A5517 and G0669.

M. L. Molle is with the Computer Systems Research Institute, University of Toronto, Toronto, Ont. M5S 1A4, Canada.

K. Sohraby is with AT&T Bell Laboratories, Holmdel, NJ 07733.

A. N. Venetsanopoulos is with the Department of Electrical Engineering, University of Toronto, Toronto, Ont. M5S 1A4, Canada.

IEEE Log Number 8714857.

packet travels one unit of distance during its transmission time (which, of course is unity).

Although a variety of different CSMA protocols exist, we will restrict our attention in this work to the asynchronous (“unslotted”) version of nonpersistent CSMA. Here a station transmits its packet immediately if the channel was sensed idle when the packet was generated, and simply drops the packet if it was generated when the channel was busy. We consider only nonpersistent CSMA because it is the simplest one to analyse. Since our primary goal is to consider the effect of topology on performance, attempting to use another, more complicated protocol will only make our models become intractable sooner! And we restrict our attention to asynchronous operation because the performance of a synchronous (“slotted”) protocol is independent of the topology and thus uninteresting for our study.

II. ESTABLISHING A CONSISTENT VIEW OF THE CHANNEL ACTIVITY

Before we can proceed with the analysis, we must first establish that it is possible to discuss “the performance” of a CSMA protocol without reference to a particular station’s view of the network. To do this, we now show that if we enforce a (possibly station dependent) lower bound on the transmission times, then we can guarantee that the local views of the channel activity by each station (and indeed, with respect to any sensible frame of reference) will be consistent, in the following sense. Whenever two (or more) packet transmissions are observed to collide by some station, then it must also be the case that they are observed to collide by every other station (including the transmitters). Note that it follows from this property that every time a packet transmission is successfully received by one station, it must also be successfully received by every other station, and hence that the probability of success for a given packet transmission is independent of its destination.

Lemma 1: Suppose two distinct stations i and j are permitted to transmit packets according to the rules of carrier sensing, beginning at times t_i and t_j , respectively. Then if we enforce on station i a *minimum* transmission time of at least

$$\tau_{ij} \triangleq \max_k \{a_{ij} + a_{jk} - a_{ik}\} \quad (1)$$

it must either be the case that *all* stations (including i and j) observe these transmissions to collide or that *none* of the stations do.

Proof: Let t_i and t_j be the times that stations i and j , respectively begin transmitting their packets. Without loss of generality, assume $t_i \leq t_j$; otherwise simply interchange the roles of the two stations. Clearly, if $t_j \geq t_i + a_{ij}$, then these two packet transmissions cannot collide because carrier sensing will have forced station j to defer its transmission until station i was finished. And because the triangle inequality (i.e., $a_{ik} \leq a_{ij} + a_{jk}$) holds in N -space, the end of the transmission by station i must arrive at k

before the start of the transmission by j does. Thus the lemma is true if we can show that whenever $t_i \leq t_j < t_i + a_{ij}$, every station must have observed a collision between these two packets.

Suppose some station k (possibly i or j) does *not* observe a collision in this case. Then clearly the end of the earliest packet to reach k must have gotten there before the start of the latest packet to reach there. But this is clearly impossible if and only if

$$\begin{aligned} & \max \{t_i + a_{ik}, t_j + a_{jk}\} \\ & \leq \min \{t_i + a_{ik} + \tau_{ij}, t_j + a_{jk} + \tau_{ji}\}. \end{aligned} \quad (2)$$

Let $t_j = t_i + \delta$, $0 \leq \delta \leq a_{ij}$. Then (2) may be simplified to yield

$$\max \{a_{ik}, \delta + a_{jk}\} \leq \min \{a_{ik} + \tau_{ij}, \delta + a_{jk} + \tau_{ji}\} \quad (3)$$

from which it follows that

$$\begin{aligned} \tau_{ij} & \geq \max_{0 \leq \delta \leq a_{ij}} \{0, \delta + a_{jk} - a_{ik}\} \\ \tau_{ji} & \geq \max_{0 \leq \delta \leq a_{ij}} \{0, a_{ik} - a_{jk} - \delta\}. \end{aligned} \quad (4)$$

Substituting a_{ij} for δ , the maximum possible value, yields the term due to station k that appears above in (1); the term is clearly nonnegative because of the triangle inequality. Note that we may interchange the roles of stations i and j in (4) (so that i is the *last* rather than the first to begin transmitting) to give us another lower bound, namely $\tau_{ij} \geq a_{jk} - a_{ik}$, but this is clearly weaker than the previous one.

Lemma 2: The minimum transmission time that must be enforced on station i to ensure that all stations observe a consistent view of the channel history is exactly

$$\tau_i \triangleq \max_j \{a_{ij} + a_{ji}\}. \quad (5)$$

Proof: Necessity follows immediately from Lemma 1 when we let station i take on the role of station k , and optimise with respect to j . To see that this worst case round-trip time between i and each of the other transmitters is sufficient, we proceed as follows. Suppose there were some other station, j , and observation point z such that an observer z would *not* see a collision if station j were to begin a transmission (of duration τ_j) δ time units after station i began a transmission of duration τ_i , $0 \leq \delta \leq a_{ij}$. Suppose this situation arose because the time at which the end of the packet from station i arrived at z , namely $\tau_i + a_{iz}$, is less than the time at which the start of the packet from j arrived at z , namely $\delta + a_{jz}$. But $\delta \leq a_{ij}$, and because of the triangle inequality we have that $a_{jz} \leq a_{ji} + a_{iz}$, so to *avoid* seeing a collision at z we must have

$$\tau_i < a_{ij} + a_{ji}$$

which contradicts the definition of τ_i . Thus the situation must have arisen because the time at which the end of the

packet from station j arrived at z , namely $\delta + a_{jz} + \tau_j$, is less than the time at which the start of the packet from i arrived at z , namely a_{iz} . But $\delta \geq 0$ and we have that $a_{iz} \leq a_{ij} + a_{jz}$ from the triangle inequality, so to avoid seeing a collision at z we must have

$$\tau_j < a_{ij}$$

which contradicts the definition of τ_j . This completes the proof of the Lemma 2.

For the case of 1-dimensional "bus" networks, Field and Wong [1] first showed that a slightly stronger condition, namely $\tau_i \geq 2 \max_{j,k} \{a_{jk}\}$, is sufficient to ensure consistency among the stations. It is worth noting that in the proof of Lemma 2, we placed no restriction on the position of the observation point z . That is, it could be one of the transmitters, an arbitrary point "inside" the network (i.e., within the convex hull of the set of transmitters $\{\bar{x}_i\}$), or indeed *anywhere* in N -space! This means that we can freely add passive extensions to the channel (where there may be additional receivers but not transmitters) without affecting the operation of the CSMA protocol in any way, and we can be sure that these external receivers see every successful transmission as a success, and every collision as a collision. Thus, as shown by the following theorem, it makes no difference what frame of reference we use to calculate the throughput of the protocol in equilibrium.

Theorem 1: Let y and z be any two observation points. Then, providing each station satisfies its minimum transmission time requirement, namely τ_i for station i , the limiting value of the observed throughput will be the same at y and z .

Proof: Let $I_k^{(z)}$ and $U_k^{(z)}$ be the time at which the k th transition from busy to idle on the channel takes place and the duration of any successful transmissions that took place during the k th channel busy time, respectively, according to an observer situated at z . Then we define

$$S_{M,N}^{(z)} \triangleq \frac{\sum_{k=M+1}^N U_k^{(z)}}{I_N^{(z)} - I_M^{(z)}} \quad (6)$$

as the normalized throughput between $I_M^{(z)}$ and $I_N^{(z)}$ as observed from z . Clearly, if a limiting value exists for the throughput observed from z , $S^{(z)}$, then we must have

$$S^{(z)} = \lim_{N \rightarrow \infty} S_{M,N}^{(z)} \quad (7)$$

for any M . Similarly, we can define the analogous quantities with respect to an observer at y to obtain

$$S_{M,N}^{(y)} \triangleq \frac{\sum_{k=M+1}^N U_k^{(y)}}{I_N^{(y)} - I_M^{(y)}} \quad (8)$$

$$S^{(y)} = \lim_{N \rightarrow \infty} S_{M,N}^{(y)} \quad (9)$$

Let us now fix M and N and compare $S_{M,N}^{(y)}$ with $S_{M,N}^{(z)}$. From Lemmas 1 and 2 (and the obvious fact that the du-

ration of a given station's transmission is the same from any stationary observation point), we have that

$$U_k^{(y)} = U_k^{(z)} \quad (10)$$

must hold for all k , and hence the numerators of (6) and (8) are identical. Now consider $I_N^{(y)}$ and $I_N^{(z)}$. First, suppose that $I_N^{(z)} \geq I_N^{(y)}$ and that the transmission from station i is the last one to finish at z . Since the channel must have remained busy at y at least until the transmission from i finished, we know that

$$I_N^{(y)} \geq I_N^{(z)} - a_{iz} + a_{iy}$$

But from the triangle inequality we have that $a_{iz} \leq a_{iy} + a_{yz}$, and hence that $I_N^{(y)} \geq I_N^{(z)} - a_{yz}$. Conversely, if $I_N^{(z)} \leq I_N^{(y)}$, we can easily show that $I_N^{(z)} \geq I_N^{(y)} - a_{zy}$, and hence that

$$I_N^{(z)} - a_{yz} \leq I_N^{(y)} \leq I_N^{(z)} + a_{zy} \quad (11)$$

must hold for any N . Notice that (11) depends only on the geometry of the network and has nothing to do with the evolution of the channel state up to I_N . Thus, it should be clear that in the worst case, the greatest difference that can arise between the denominators of (6) and (8) is when one of the observers, say z , is maximally ahead of the other at I_M , and then maximally behind the other at I_N , in which case

$$\left| (I_N^{(y)} - I_M^{(y)}) - (I_N^{(z)} - I_M^{(z)}) \right| \leq a_{yz} + a_{zy}$$

must hold for all M and N . Now since $I_N \rightarrow \infty$ as $N \rightarrow \infty$, the result follows.

Notice that the proof of Theorem 1 did not depend on the duration of any station's transmission, as long as Lemma 2 is satisfied, so the result is true whether or not we allow variable packet lengths and/or collision detection. Thus, in the remainder of this work, we shall insist that each station always satisfies its minimum transmission time requirement defined by Lemma 2. In this way, we can freely choose any observation point(s) from which to monitor the operation of the protocol.

III. PARTITIONING THE SPACE-TIME EXECUTION TRACE INTO CYCLES

Suppose $\{a_{ij}\}$ is such that the LAN can be embedded in N -space. Then we can construct an $(N+1)$ -dimensional *execution trace* for the protocol in which the first N dimensions represent position in the LAN and the $(N+1)$ st dimension represents the passage of time. For example, an Ethernet-like network [4] consisting of a single unbranched coaxial cable may be viewed as a linear (one-dimensional) LAN—where the location of a station is given by its distance from, say, the left-hand end of the cable—and its execution trace consists of a two dimensional "ribbon" of space-time. Similarly, a ground radio packet-switched network [2] may be viewed as a planar (two dimensional) LAN where the execution trace is a three dimensional "cylinder" of space-time, and so on.

Now, in view of Lemma 2 and Theorem 1, we can partition the execution trace into an alternating (in time) se-

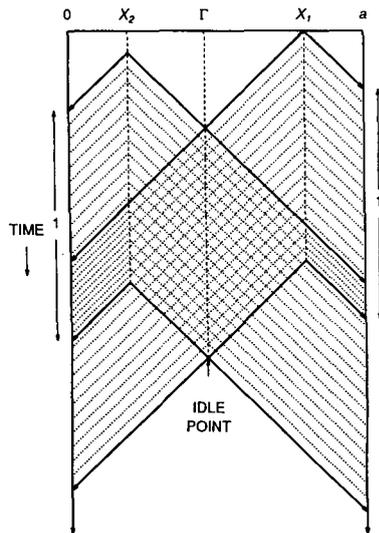


Fig. 1. The busy band that results when transmitters at X_1 and X_2 collide.

quence of busy and idle “bands.” Note that the form of these “bands” is, in general, quite complicated because of the spatial geometry of collisions. Whenever some station, i for instance, transmits a packet of duration T_i , beginning at the time t_i , it causes the channel to be busy in a cone-shaped region in space-time, centered at $(\vec{x}_i; t_i)$, namely

$$B_i(t_i, T_i) \triangleq \left\{ (\vec{x}_j; t) \mid \vec{x}_j \in \mathbb{R}^N, t_i + |\vec{x}_i - \vec{x}_j| \leq t \leq t_i + T_i + |\vec{x}_i - \vec{x}_j| \right\}.$$

Thus, whenever stations i, j, \dots collide, the resulting busy band is the union of $B_i(\cdot), B_j(\cdot) \dots$. In Fig. 1, for example, we show the busy band that results from a collision between two transmitters on a linear (bus-like) LAN. Notice that the “initial” surface of the band (i.e., earliest in time) contains a local minimum at each transmitter—as it must because carrier sensing only permits a station to begin transmitting if it is the “first” one according to its own frame of reference. Conversely, the “final” surface of the band (i.e., latest in time) contains a single, global minimum that we shall call an *idle point* Γ the earliest point in the LAN to observe the transition from busy to idle channel.¹ The remainder of the LAN senses the transition as the leading edge of a convex *idle cone* reaches them.

It is not hard to see that even in N -space, the final surface of each busy band must be a convex idle cone, starting from a unique idle point. To see this, we recall that the idle band is actually the *intersection* of one or more convex cones—bounded by the final surfaces of $B_i(\cdot), B_j(\cdot) \dots$ —and the intersection of convex sets is itself convex.

¹The existence of idle points on 1-dimensional LAN's was first shown by Field and Wong [1] where they were called *time origin points*. However, their work did not consider the stochastic structure induced by the idle points. Instead, they calculated throughput under the approximation that both the traffic and the location of the idle points are uniformly distributed along the bus.

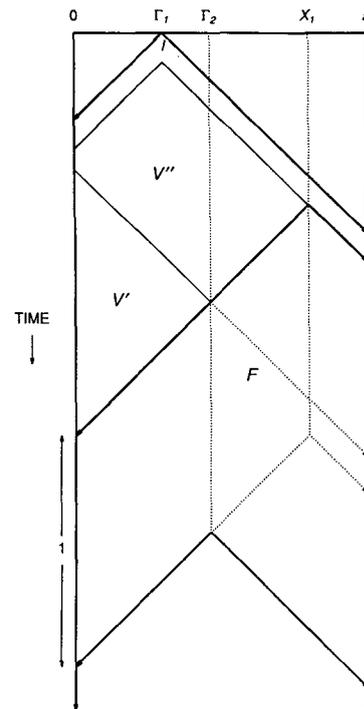


Fig. 2. The components I, V , and U for a cycle with a successful transmission.

Following [3], the throughput analysis of asynchronous CSMA protocols is usually carried out with the aid of renewal theory in terms of a sequence of *transmission cycles*, each consisting of an “idle period” followed by a “busy period.” It is worth noting that in previous work (where, for simplicity, the worst case “star” topology was assumed) it made no difference whether we defined a cycle to be an idle period followed by a busy period, or a busy period followed by an idle period. However, in our work (where we are explicitly modelling the network topology) it is important to define a cycle to begin with the idle period so the boundary between successive cycles in space-time is as clean as possible, namely a convex idle cone.

In Figs. 2–4, we illustrate the various components of a cycle that arise in a successful transmission, a collision without collision detection, and a collision with collision detection, respectively, in the case of a 1-dimensional (linear) LAN. Notice that unlike the worst case “star” analysis (where because of symmetry and the memoryless property of Poisson arrivals, successive cycles are i.i.d.), there is a dependence between successive transmission cycles through the geometry of the separating idle cone. But because of the Poisson traffic assumption, it should be clear that the sequence of idle cones forms an embedded Markov chain.

Each transmission cycle begins with an idle volume I obtained by “sweeping” the initiating idle cone forward along the time axis until the first transmission is encountered. Note that the station, say j at position \vec{x}_j , that terminates I is the one that begins its transmission at time t_j after sensing the channel idle for the *minimum* time. Be-

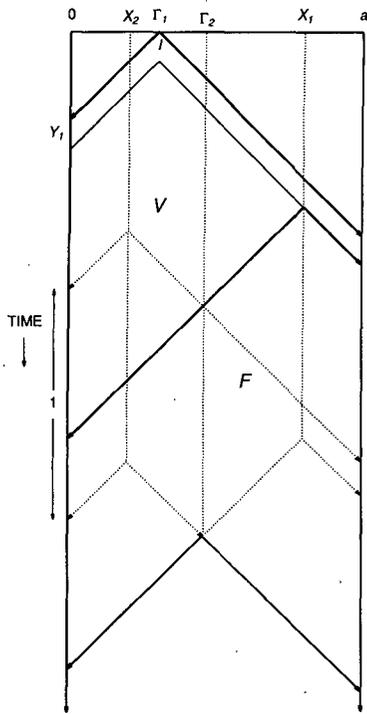


Fig. 3. The components I , V , and F for a cycle with a collision, assuming no collision detection.

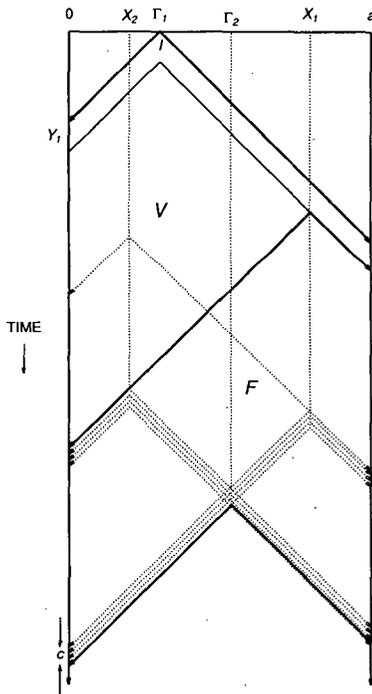


Fig. 4. The components I , V , and F for a cycle with a collision, assuming each transmitter jams the channel for c time units after detecting a collision.

forms the idle cone that terminates this cycle. If the transmission from j is a failure (i.e., a collision occurs), then we have several cases, depending on the durations of each conflicting transmission (which, in turn, may depend on how collision detection operates, if it exists).

First of all, it is easy to see that if one of the transmitting stations, j for instance, observes its own transmission to be the *last* one to terminate from its *own* point of view (i.e., at \bar{x}_j), then we say that its transmission *dominates* the rest in the sense that the trailing edge of $B_j(\cdot)$ completely defines the idle cone. To see this, suppose that stations i and j stop transmitting their packets at times t'_i and t'_j , respectively, and that the end of station i 's transmission arrives at j before j stops transmitting, i.e.,

$$t'_i + a_{ij} < t'_j. \tag{12}$$

Now if there is some other observation point, say z , where the order of termination is reversed, then we must have that

$$t'_i + a_{iz} > t'_j + a_{jz}. \tag{13}$$

But $a_{ij} + a_{jz} \geq a_{iz}$ from the triangle inequality, so that if (13) holds then so must

$$t'_i + a_{ij} + a_{jz} > t'_j + a_{jz}$$

which contradicts (12).²

In general, if none of the conflicting transmissions dominates the rest, then it is difficult to say much about the structure of the busy band without further assumptions. However, if we restrict our attention to linear (bus-like) networks with constant length packets [6], then it is easy to see that in a collision, the trailing edge of the busy band is completely determined by the *two* transmitters that are closest to the opposite extremes of the network, regardless of the total number of stations involved in the conflict.

Lemma 3: Suppose a conflict arises on a "bus" network among $K \geq 2$ stations, such that the *positions* of the transmitters (with respect to the left-hand end of the bus, say) satisfy $\bar{x}_1 \leq \bar{x}_2 \leq \dots \leq \bar{x}_K$. Assume that t_1, t_2, \dots, t_K are the respective times that each station began its transmission. Then, if the *durations* of all transmissions are equal to some constant T the idle point will be completely determined by the two extreme transmitters, i.e., the left-hand ray leaving the idle point will be the end of the transmission from \bar{x}_K and the right-hand ray leaving the idle point will be the end of the transmission from \bar{x}_1 .

Proof: First, we note that the last transmission to arrive at any point z must also be the last one to terminate there because of the constant packet length assumption. Now suppose that $\bar{x}_1, \dots, \bar{x}_j$ all fall to the left of z . In this case, because of the linearity of the network, we have

²It is worth noting that if station j were to retransmit its packet immediately in this case, then this new attempt would be successful with probability one under the Poisson traffic model since the vulnerable volume would be zero.

cause of carrier sensing, this transmission will be successful if and only if no other transmissions arise within the corresponding vulnerable volume V that separates I from $B_j(\cdot)$, the contribution to the busy band due to j . If the transmission from j is successful (Fig. 2), then the busy band is exactly $B_j(\cdot)$, and hence its trailing edge

that for the m th such transmitter, $1 < m \leq j$,

$$a_{1z} = a_{1m} + a_{mz}. \quad (14)$$

Since adding t_1 to both sides of (14) cannot affect the equality, and we must have $t_m \leq t_1 + a_{1m}$ (or else carrier sensing would have prevented station m from transmitting in the first place), we see immediately that the transmission from station m could not have arrived at z after the one from station 1. Thus it follows that we need only consider the left-most transmitter to the left of z to determine the time at which the channel goes idle at z . Similarly, if there are transmitters to the right of z , we need only consider the right-most one. Thus, to complete the proof of the Lemma, we need only show that if at some point z the transmission from station 1 arrives after the transmission from station K , then the same must be true for all points $z' > z$. But this is obviously true, since the difference in arrival times of the packets from the two extreme transmitters, according to an observer at z , is given by $t_1 + a_{1z} - t_K - a_{Kz}$, and $a_{1z} - a_{Kz}$ is a nondecreasing function of z . This completes the proof of Lemma 3.

It is worth noting that the new idle point Γ_2 will be located at the point z^* , such that $t_1 + a_{1z^*} = t_K + a_{Kz^*}$.

Unfortunately, if we now consider a slightly more general system, then we find that many of these nice results no longer hold. For example, consider a generalization of Lemma 3 to a 2-dimensional (i.e., planar) network where we would like to show that the idle cone is once again determined by the extreme transmitters. Suppose we once again restrict all the transmissions to be of constant duration, and even restrict our attention to the special case of a collision among three *collinear* transmitters, denoted by ij and k , as shown in Fig. 5. To simplify the problem (after all, we are just providing a counterexample!) assume that $a_{ij} = a_{jk} = a_{ik}/2$, and that their respective transmission times satisfy $t_i = t_k$, $t_j = t_i + C$ for some $0 < C < a_{ik}$. Now consider the difference Δt between the times that the ends of the transmissions from i and j , respectively, arrive at some point z on the right bisector to ik . (Note that from symmetry, we need not consider the transmission from k .) Clearly, we have

$$\begin{aligned} \Delta t &= (t_i + a_{iz} + T) - (t_j + a_{jz} + T) \\ &= a_{iz} - (C + a_{jz}). \end{aligned} \quad (15)$$

Now since a_{iz} is the hypotenuse of a right triangle, and, obviously, $\sqrt{x^2} = x$, (15) may be rewritten as

$$\Delta t = \sqrt{a_{ij}^2 + a_{jz}^2} - \sqrt{C^2 + 2Ca_{jz} + a_{jz}^2}. \quad (16)$$

Thus, since \sqrt{x} is monotonic in x , Δt is clearly positive whenever

$$a_{jz} \leq \frac{a_{ij}^2 - C^2}{2C} \quad (17)$$

holds. Equation (17) is clearly true for small values of a_{jz} because we have already imposed $0 < C < a_{ij}$. However, since the right-hand side of (17) is independent of z , and the left-hand side depends on the distance between j and

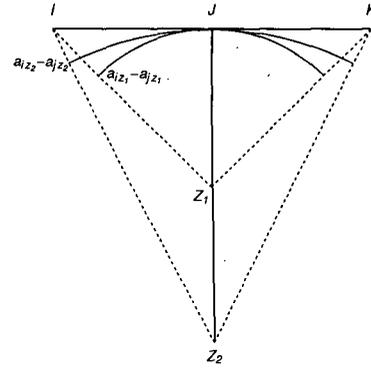


Fig. 5. The transmitter directly between two others can affect the idle cone in 2-space.

z , the transmission from j will be the last one to terminate at z for all points sufficiently far from j .

Thus, we have found that even with constant packet lengths, a transmitter situated on the line joining two other transmitters may contribute to the definition of the idle cone in 2-space. The problem of finding the idle cone—or even just the set of transmitters contributing to its boundary—is even more difficult when we consider variable packet lengths, such as might arise with collision detection. (As we will see in Section V, the problem of finding the idle cone—or, equivalently, the failure area in a collision cycle—is the main stumbling block in this type of the analysis.)

IV. TOWARDS THE TRAFFIC-DEPENDENT ANALYSIS: FINDING THE PROBABILITY OF SUCCESS FOR THE “FIRST” TRANSMITTER

In previous sections, we derived some properties that are independent of the traffic generation model. In this section, we begin the task of evaluating some traffic-dependent properties of the protocol with the ultimate aim of finding the throughput as a function of load and the network topology.

In the remainder of this work, we will consider the traffic (including the retransmission of previously collided packets) to be a time homogeneous, space inhomogeneous $(N + 1)$ -dimensional Poisson process at an aggregate rate of G packet transmission attempts per packet transmission time. That is, we allow the distribution of this traffic over the spatial extent of the network to be governed by an arbitrary nonnegative *load intensity function* $g(\vec{x})$ satisfying

$$\iiint_{\mathbb{R}^N} g(x_1, x_2, \dots, x_N) dx_1 dx_2 \dots dx_N = G. \quad (18)$$

Thus, for example, to model a large number of stations uniformly distributed over some region R_0 , we let

$$g(\vec{x}) = G \int \int \int_{R_0} dx_1 dx_2 \dots dx_N$$

for all $\vec{x} \in R_0$; to represent a set of discrete stations at \vec{x}_1 ,

$\vec{x}_2, \vec{x}_3, \dots$ generating Poisson traffic at rates G_1, G_2, G_3, \dots , respectively, we let

$$g(\vec{x}) = \sum_i G_i \delta(\vec{x} - \vec{x}_i) \quad , \sum_i G_i = G$$

and so on.

From renewal theory, the throughput over all time can be found from the expected behaviour of the protocol over a transmission cycle as the ratio of the expected amount of useful work performed in a cycle $E[U]$ to the expected duration of a cycle, $E[L]$. The main result of this section is to derive a simple lower bound on the probability of success for the 'first' transmitter in a cycle, in terms of a "natural" measure of the dispersion of the traffic we call the *average vulnerable period* (\bar{a}). This bound holds for the general case of an arbitrary load intensity function in N -space. Note that because of carrier sensing, the success of a given transmission attempt is independent of the packet length. Thus, since we have already normalized time so that the average packet length is unity, $E[U]$ is equal to the probability of success for the "first" transmitter in a cycle.

Before proceeding with the derivation of the bound, it is instructive to consider the notion of a *frame of reference* F with respect to which we say that a given station was the first one to transmit. We say that two events that took place at points \vec{x}_i and \vec{x}_j in the network and at times t_i and t_j , respectively, are simultaneous with respect to F if

$$t_i - t_j = f_{ij} \quad (19)$$

where f_{ij} is a constant that depends on \vec{x}_i, \vec{x}_j and satisfies $f_{ij} = -f_{ji}$. We say that F is a consistent frame if

$$|f_{ij}| \leq |\vec{x}_i - \vec{x}_j| \equiv a_{ij}$$

holds for all \vec{x}_i, \vec{x}_j . In other words, were we to send observers to both points, neither one could disprove our claim of simultaneity because he observed the distant event before the local one. Some examples of consistent frames of reference include "absolute" time ($f_{ij} \equiv 0$), each station's local measurement of elapsed time since the start of the last idle period (f_{ij} is defined by the idle cone), and a single station's measurement of the time at which it observed every event (f_{ij} is defined by a right circular hypercone, centered at the observer, that opens toward the negative time axis).

Lemma 4: Assume that enough time has elapsed in the current idle period for the entire network to be "simultaneously" idle with respect to a consistent frame of reference, F . Then the pdf for the location of the first station (with respect to F) to transmit in the idle period is given by $g(\vec{x})/G$, independent of our choice of F .

Proof: Consider two contours of simultaneity with respect to F , namely C , upon which the entire network is seen to be idle, and C' , obtained from C by advancing it forward in time by Δt . Clearly C must exist under the assumptions of the Lemma. Since the channel is completely idle prior to C , no station will be forbidden from

transmitting between C and C' because it senses an "earlier" (with respect to F) transmission. However, should one station begin transmitting after C , it may prevent other stations from transmitting due to carrier sensing. (And since F is a consistent frame, no station will be prevented from transmitting between C and C' because it senses a transmission that began after C' .) Thus, the probability that *exactly* one transmission occurs between contours C and C' , given that *at least* one occurs, is lower bounded by the probability that *exactly* one packet is generated between contours C and C' , given *at least* one. But since the total traffic is Poisson, it is well known that as $\Delta t \rightarrow 0$, this lower bound converges to unity. Furthermore, since we may interpret the total traffic G as the superposition of many Poisson streams—say at rate $g(\vec{x}) d\vec{x}$ from an infinitesimal volume $d\vec{x}$ at position \vec{x} —it should be clear that the pdf for the position of the first transmitter is given by the normalized load intensity function, $g(\vec{x})/G$, i.e., the proportion of the total traffic originating from each point.

Before continuing on with the main theorem of this section, we note that in view of Lemma 4 and the memoryless property of the Poisson distribution, it should be clear that we must "sweep" the contour C forward along the time axis for an exponentially distributed distance with mean $1/G$ before we encounter the first transmission.

Theorem 2: If the channel is completely idle (with respect to a consistent frame of reference F) when the "first" station begins its transmission, then its probability of success, averaged over the pdf for the location of the first station, satisfies

$$P[\text{success}] \geq e^{-\bar{a}G} \quad (20)$$

where

$$\bar{a} \triangleq \int_{\vec{x}_j \in \mathbb{R}^N} d\vec{x}_j \int_{\vec{x}_k \in \mathbb{R}^N} d\vec{x}_k \frac{g(\vec{x}_j)}{G} \frac{g(\vec{x}_k)}{G} |\vec{x}_j - \vec{x}_k| \quad (21)$$

is the average distance between two points drawn at random according to the normalized load intensity function $g(\vec{x})/G$ and is independent of the total traffic G and of F .

Proof: Clearly the probability of success, conditioned on the location of the first transmitter \vec{x}_1 is given by the probability that no other traffic is generated within the vulnerable volume V that lies between the contour of simultaneity C and the busy cone due to \vec{x}_1 , namely $B_1(\cdot)$. Hence, as illustrated in Fig. 6, we have

$$P[\text{success} | \vec{x}_1] = \exp \left(- \int_{\vec{x}_2 \in V} d\vec{x}_2 [a_{12} - f_{12}] g(\vec{x}_2) \right) \quad (22)$$

where

$$a_{12} = |\vec{x}_1 - \vec{x}_2|.$$

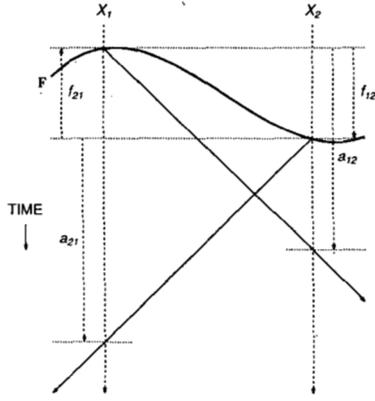


Fig. 6. The interference between X_1 and X_2 when one of them is the “first” transmitter.

Unconditioning on \vec{x}_1 , (22) gives us

$$P[\text{success}] = \frac{1}{G} \int_{\vec{x}_1 \in \mathbb{R}^N} d\vec{x}_1 g(\vec{x}_1) \cdot \exp \left(- \int_{\vec{x}_2 \in \mathbb{R}^N} d\vec{x}_2 [a_{12} - f_{12}] g(\vec{x}_2) \right). \quad (23)$$

But since $\exp(\cdot)$ is a concave \cup function, we can apply Jensen’s inequality to (23) to obtain

$$P[\text{success}] \geq \exp \left((-1/G) \int_{\vec{x}_1 \in \mathbb{R}^N} d\vec{x}_1 \cdot \int_{\vec{x}_2 \in \mathbb{R}^N} d\vec{x}_2 g(\vec{x}_1) g(\vec{x}_2) [a_{12} - f_{12}] \right). \quad (24)$$

But since $f_{12} = -f_{21}$, we see that by reversing the roles of \vec{x}_1 and \vec{x}_2 in (24) the effect of F cancels, leaving

$$P[\text{success}] \geq \exp \left((1/G) \int_{\vec{x}_1 \in \mathbb{R}^N} d\vec{x}_1 \cdot \int_{\vec{x}_2 \in \mathbb{R}^N} d\vec{x}_2 g(\vec{x}_1) g(\vec{x}_2) [a_{12}] \right). \quad (25)$$

Substituting (21) into (25) gives us (20), thus completing the proof of the theorem.

Before ending this section, it is worth considering the probability of success in the special case of a 1-dimensional (bus) network in greater detail. First, we have been able to show [7] that $\bar{a} \leq a/2$ must hold for every load intensity function, with equality in the case of two stations with equal loadings of $G/2$ at opposite extremes of the bus. Thus, we have immediately from Theorem 2 that the probability of success in the bus is at least $e^{-aG/2}$. This result is significantly stronger than the well-known result, obtained using a worst case “star” approximation, namely e^{-aG} , especially when a and/or G is large. And, in addition, the relatively simple structure of the 1-dimensional case has allowed us to obtain even stronger results on the probability of success as a function of the idle point Γ . The next two theorems are proven in [7].

Theorem 3: Consider a 1-dimensional network having a (strictly) continuous load intensity function $g(x)$, $0 \leq x \leq a$. Then the probability of success, conditioned on the position of the idle point Γ has continuous first and second derivatives with respect to Γ in $[0, a]$, and attains (possibly local) maxima at $\Gamma = 0$ and $\Gamma = a$. Furthermore, if $g(x)$ is symmetric (i.e., $g(x) = g(a - x)$ for all $0 \leq x \leq a$), then the conditional probability of success as a function of Γ (and thus also the *unconditional* probability of success) is lower bounded by

$$P[\text{success} | \Gamma] \geq \frac{g(a/2)}{g(a/2) + G^2/2}. \quad (26)$$

If we rewrite (26) in terms of $g^*(x) \triangleq (a/G) g(ax)$ instead of $g(x)$, i.e.,

$$P[\text{success} | \Gamma] \geq \frac{g^*(1/2)}{g^*(1/2) + aG/2} \quad (27)$$

then we can separate the influences on the probability of success from the “shape” of the load intensity function (given by $g^*(\cdot)$), the “scaling” parameter a and the load G . It is interesting to note that these last two factors only enter (27) through the product aG , and that the lower bound is actually *hyperbolic* in aG , and thus inversely proportional to aG when $aG \gg 1$. Indeed, this type of asymptotic result is generally true for *any continuous* $g(x)$ where we have

$$P[\text{success} | \Gamma = \gamma, aG \gg 1] \approx$$

$$\begin{cases} \frac{g(\gamma)}{2 \left(\int_0^\gamma g(x) dx \right) \left(\int_\gamma^a g(x) dx \right)} & 0 < \gamma < a \\ \frac{\sqrt{g(\gamma)} \pi}{2G} & \gamma \in \{0, a\}, g(\gamma) \neq 0. \end{cases}$$

Theorem 4: Consider a 1-dimensional network having a discrete load intensity function

$$g(x) = \sum_{i=1}^M G_i \delta(x - a_i)$$

where $0 = a_1 \leq a_2 < \dots < a_M = a$, $G_i > 0$ for all i and $\sum_{i=1}^M G_i = G$. Then the probability of success, conditioned on the position of the idle point Γ is continuous in $[0, a]$; its derivative is continuous in the open interval (a_i, a_{i+1}) and does not exist at $\{a_i\}$ (each of which is a local maximum). Furthermore, we have that

$$P[\text{success} | \Gamma = a_i] \geq \frac{G_i}{G}.$$

It is also possible to obtain the *complete* solution to the two-station case, including the pdf for the idle point and the exact probability of success, which in its symmetric form attains the worst possible value of \bar{a} on the 1-dimen-

sional (bus) network [7]. In this case, we have

$$P[\text{success}] = \frac{1}{1 + aG/2}$$

which, surprisingly, is equal to the lower bound for uniform traffic (i.e., $g(x) = G/a$ for all x) implied by Theorem 3.

V. RESULTS ON THROUGHPUT

To find throughput as a function of G and the network topology, it remains to find the expected duration of a transmission cycle. Since the expectation of a sum is equal to the sum of the expectations, we may consider each of the components of a cycle separately. Recall that in Section III, we decomposed each cycle into the idle volume, I , followed by the vulnerable volume V and finally either the useful volume U or the failure volume F in the case of a successful transmission or a collision, respectively.

The idle volume I is constructed by translating the idle cone forward along the time axis until the first transmission is encountered. Thus, for each cycle, the measured duration of I must be *exactly* the same, no matter what observation point we use to make the measurement. Since the idle cone is a consistent frame of reference, it is clear that I is unaffected by the outcome in this (or, for that matter, the previous) cycle. Furthermore, since the generation of traffic is Poisson (recall that our traffic model is a time homogeneous, space inhomogeneous $N + 1$ -dimensional Poisson process), it is clear that from any observation point, say z , the measured duration of I must be exponentially distributed with mean $E_t^{(z)}[I] = 1/G$. (For region R , we use the notation $E_t^{(z)}[R]$ to mean the expected value of the length of R along the time axis at observation point z .)

Similarly, we note that the length of a successful transmission is (obviously) independent of our observation point. Thus, since we have defined the average packet length to be unity, we have immediately that $E_t^{(z)}[U] = P[\text{success}]$.

The remaining two terms, namely V and F , are more difficult to characterize, since for a given cycle it is clear that the measured duration of these volumes can vary drastically as a function of our observation point. In particular, the shapes of V and F depend on both the geometry of the idle cone and the position of the first transmitter. In addition, the shape of F also depends on the position(s), differences in starting time(s), and differences in transmission time(s) between the first transmitter and a (random number of) conflicting transmitter(s). Thus we face a far more challenging problem in evaluating $E_t^{(z)}[V]$ and $E_t^{(z)}[F]$.

In view of the discussion above, it seems quite remarkable that with the aid of the following "trick" we can obtain a simple expression for $E_t[V]$, which is *independent* of the idle cone. The main idea is that instead of a single fixed observation point, we now propose to make our measurements of the duration of a cycle with respect to a generalized observation point z^* consisting of a

weighted average of the measurements obtained from many different observation points.

First of all, since

$$\begin{aligned} E[\alpha_1 X_1 + \alpha_2 X_2 + \dots] \\ = \alpha_1 E[X_1] + \alpha_2 E[X_2] + \dots \end{aligned} \quad (28)$$

holds for any random variables X_1, X_2, \dots and real numbers $\alpha_1, \alpha_2, \dots$, we may replace $\{X_i\}$ by the respective sets of measurements of I and U at z_i and replace $\{\alpha_i\}$ by any set of nonnegative "weights" summing to unity to see that both

$$E_t^{(z^*)}[I] = \sum_i \alpha_i E_t^{(z_i)}[I] = E_t[I]$$

and

$$E_t^{(z^*)}[U] = \sum_i \alpha_i E_t^{(z_i)}[U] = E_t[U]$$

must hold for all z^* . Thus, from now on we can use z^* as our observation point instead of some fixed point z without affecting our measurements of I or U .

Returning now to the derivation of $E_t^{(z^*)}[V]$, we now show that the key property of the independence from the idle cone holds when the weight assigned to the observation at \vec{x}_i , is given by $g(\vec{x}_i)/G$, the pdf for the normalized load intensity function. To see this, we take the idle cone as our frame of reference and condition of the position of the first transmitter X_1 to obtain

$$E_t[V|X_1 = \vec{x}, F] = \int_{\vec{z} \in \mathbb{R}^N} d\vec{z} [a_{xz} - f_{xz}] g(\vec{z})/G.$$

Unconditioning on X_1 , we have

$$\begin{aligned} E_t[V|F] = \int_{\vec{x} \in \mathbb{R}^N} d\vec{x} g(\vec{x})/G \int_{\vec{z} \in \mathbb{R}^N} d\vec{z} \\ \cdot [a_{xz} - f_{xz}] g(\vec{z})/G. \end{aligned} \quad (29)$$

But since $f_{xz} = -f_{zx}$, (29) simplifies to (21). Thus, if the measurements of V at the positions of each transmitter are weighted in proportion to their load generation rate, we see that the resulting estimate of $E_t[V]$ is *always* equal to \bar{a} , independent of the frame of reference. Thus, even if without the steady-state distribution of the idle cone (which we usually cannot find, anyway), we know that $E_t^{(z^*)}[V] = \bar{a}$.

At this point, the reader may recall that in the worst-case "star" analysis without collision detection, the average duration of a cycle is either $1/G + a + 1$ if it is successful (with probability e^{-aG}), or $1/G + a + \bar{Y}_a + 1$ if it is a collision, [3]. The term

$$\begin{aligned} \bar{Y}_a &\triangleq \frac{1}{1 - e^{-aG}} \int_0^a (a - y) G e^{-yG} dy \\ &= \frac{a - (1 - e^{-aG})/G}{1 - e^{-aG}} \end{aligned}$$

is the expectation of the difference between the starting times of the first and last transmitters in a collision, which

depends on both the total traffic G and the end-to-end propagation time, a . In the case of general loading in N -space, we have now shown that the average duration of a cycle (with respect to the generalized observation point, z^* , defined above) is either $1/G + \bar{a} + 1$ if it is successful (for which the probability is at least $e^{-\bar{a}G}$), or $1/G + \bar{a} + E_i^{(z^*)}[F]$ if it is a collision. Thus, we see that if the inequality

$$E_i^{(z^*)}[F] \stackrel{?}{\leq} 1 + \bar{Y}_{\bar{a}} \quad (30)$$

were to hold, then we could obtain a lower bound to the throughput as a function of $g(\cdot)$ in the general case by replacing the worst case propagation time a by its "average" value \bar{a} and then simply pretending that we have a "star" network. Unfortunately, as tempting as it is to jump to this conclusion, we are not so lucky. This is because F depends on the *worst* of the interfering stations if there is more than one, and hence on the distribution of traffic as a function of distance from the first transmitter \bar{x}_1 , on the separation between the interfering transmitters (if there is more than one), on our frame of reference, and on the observation point. Furthermore, in the 1-dimensional symmetric two-station case (for which we have an exact solution available [7]), we have found that the expected length of a busy period is given by $1/G + a/2 + 1 + P[\text{collision}]a/2$, from which it follows that $E_i^{(z^*)}[F] = 1 + a/2$. But since $\bar{Y}_{a/2} < a/2$ for all finite values of G , we have found a case where the conjectured inequality in (30) does *not* hold. Thus, the use of this heuristic substitution is questionable, even in the case of 1-dimensional networks! (However, it is interesting to note that although (30) is not satisfied in the two station case, the heuristic substitution does appear to give a lower bound on S , presumably due to its underestimation of $P[\text{success}]$.)

Before returning to the problem of finding $E_i[F]$, we

$$E_i^{(z^*)}[F] = 1 + \frac{(1 - e^{-\Gamma G})[a - \Gamma - 1/G] + (1 - e^{-(a-\Gamma)G})[\Gamma - 1/G]}{(1 - e^{-\Gamma G}) + (1 - e^{-(a-\Gamma)G})}$$

note that as a consequence of Theorem 1, the expected duration of a cycle *must* be independent of the observation point. And thus, since I and U are constant, the sum $E_i[V] + E_i[F]$ must also be independent of the observation point. We note, however, that in order to find these expectations with respect to some specific observation point, we must in general find the steady state distribution governing the geometry of the idle cone (so we can take expectations with respect to it). Unfortunately, we have found that even for 1-dimensional networks, it is difficult to find the steady-state distribution of the idle point. Furthermore, as shown by the following example, the fact that $E_i^{(z^*)}[V]$ is independent of the idle point does *not* imply that this same independence also holds for $E_i^{(z^*)}[F]$.

Consider a cycle in the symmetric two station case, as shown in Fig. 7. Clearly we have from (21) that $\bar{a} = a/2$ in this case, since half the time we will choose the same

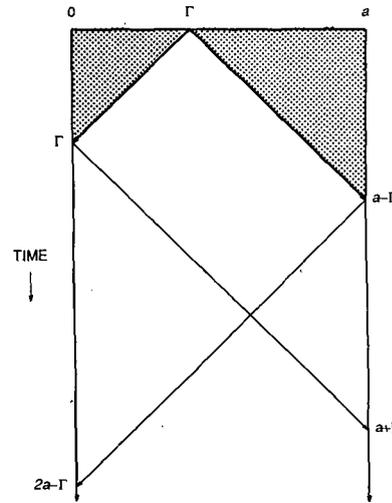


Fig. 7. The timing of events in a cycle in the two station case, letting $t = 0$ at the idle point, Γ .

end of the bus for x_j and x_k and the rest of the time we will pick opposite ends, which yields $a_{jk} = 0$ and $a_{jk} = a$, respectively. To find $E_i^{(z^*)}[V]$, we first condition on the idle point being at position Γ on the bus, for some $0 \leq \Gamma \leq a$, and then take the average of the measurements of V at the two endpoints of the bus for the two equally likely cases, namely $X_1 = 0$ and $X_1 = a$. Thus, we have

$$E_i^{(z^*)}[V] = \frac{1}{2} \left[\frac{[0] + [2\Gamma]}{2} + \frac{[2(a - \Gamma)] + [0]}{2} \right] = \frac{a}{2}$$

which, of course, is independent of Γ . $E_i^{(z^*)}[F]$ can be obtained in a similar way, except the probability that $X_1 = 0$ given that a collision occurred is given by

$$\frac{P[\text{collision} | X_1 = 0]}{P[\text{collision} | X_1 = 0] + P[\text{collision} | X_1 = a]}$$

and not by $1/2$. Thus

which is clearly *not* independent of Γ . Thus, Theorem 1 is of little help to us, and the estimation of F must be done using other means.

Because of the special structure of 1-dimensional networks, we can make further progress in the estimation of $E_i[F]$, and hence S . We now briefly outline two methods that appear in [7]. First, in the *partitioning method*, we find upper and lower bounds to $E_i[F]$ by partitioning the vulnerable area V into two components V' and V'' , as shown in Fig. 8. If we assume that all transmissions are of constant length equal to unity, then we have a direct relation between the regions V' and F . That is, in the event that no transmissions originate from within the region V' , we can use a simple geometrical argument to establish an upper bound on $E_i^{(z^*)}[F]$ (no transmissions in V') for any observation point z . But since a ready station would be forbidden from transmitting in V' if it sensed carrier from another transmission that originated within region V'' , it

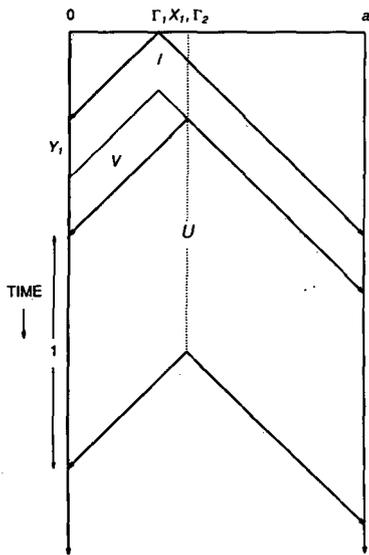


Fig. 8. The partitioning of V into regions V' and V'' , from which we can bound F .

is easy to see that

$$P[\text{no transmission in } V'] \geq \exp\left(-\iint_{V'} g(x) dx dt\right)$$

must hold, and hence how to upper bound $E_i^{(z)}[F]$. To find a matching lower bound, we note that the probability that at least one transmission occurs in region V' given that a collision has occurred may be viewed as the union of two events A and B where A is the event that at least one transmission occurs in region V' and none occur in region V'' , and B is the event that at least one transmission occurs in region V' and at least one occurs in region V'' . Since $P[A \cup B] \geq P[A]$, we have that

$$P[\text{no transmissions in } V'] \leq \exp\left(-\iint_{V'} g(x) dx dt\right) - \exp\left(-\iint_{V' \cup V''} g(x) dx dt\right)$$

from which the lower bound on $E_i^{(z)}[F]$ follows.

The bounds on F obtained with the partitioning method may be sharpened using the *layering method*, as shown in Fig. 9. In this case, we use the idle cone to partition the vulnerable volume V into successive "layers" V_1 and V_2 where layer V_1 terminates when the *second* transmission in the cycle is encountered. (Recall that X_1 has already been defined to mean the first transmitter to break the idle period.) If X_1 and X_2 are on opposite sides of Γ , then we know from Lemma 3 that these two transmitters are sufficient to define F , whether or not any more transmissions take place in this cycle. Thus the interesting case is when X_1 and X_2 are on the same side of Γ , as shown in the figure. First, we note that in the region $V_2 \cap B_2$, no fur-

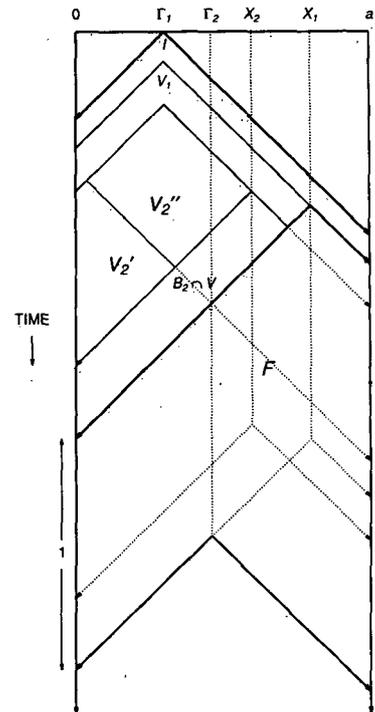


Fig. 9. The layering method applied to a cycle where a collision has occurred.

ther transmissions can possibly arise because the stations would already have sensed the transmission from X_2 . Thus, to bound F we need only examine the remainder of V_2 for further transmissions. If there are none, then X_1 and X_2 define F . Otherwise, we can bound F by dividing this region into two parts, V_2' and V_2'' , and applying the partitioning method.

VI. DISCUSSION AND CONCLUSIONS

The methods outlined here have been successfully applied to linear (bus-like) networks [6]–[8] where we have obtained bounds on the throughput for nonpersistent CSMA with and without collision detection in the case of general intensity functions. A sampling of these results are shown in Figs. 10 and 11.

In Fig. 10, we consider nonpersistent CSMA without collision detection, while in Fig. 11, we assume that there is collision detection with a jam time of $c = 2a$. In all cases, we have assumed a moderately large value for the normalized propagation time, namely $a = 0.1$, which is quite appropriate for cable-based LAN's operating at high data rates. Furthermore, to illustrate the role of the load intensity function on performance, we have included results for three different ones, namely a) $g^*(x) = 6x(1-x)$, which concentrates the load near the middle of the bus, giving $\bar{a} = 9a/35$; b) $g^*(x) = 1$, which distributes the load uniformly over the bus, giving $\bar{a} = a/3$; and c) $g^*(x) = 12(x - 1/2)^2$, which concentrates the load near the two ends of the bus, giving $\bar{a} = 3a/7$. The corresponding results under the worst case "star" topology assumption and for the slotted version of the protocol are also included. Notice that in all cases, the performance of

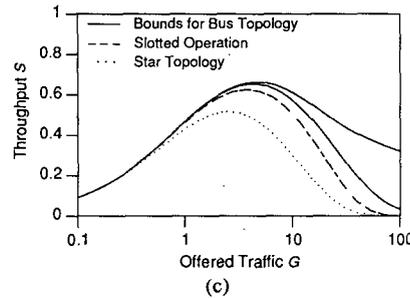
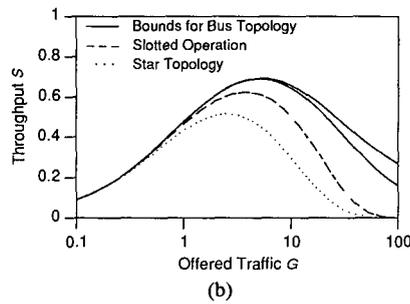
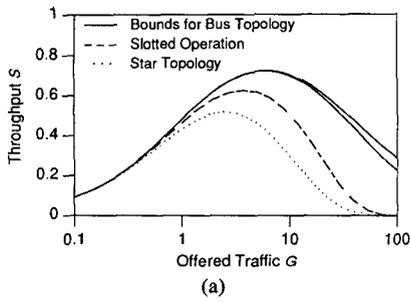


Fig. 10. The effect of the load intensity function on throughput for asynchronous nonpersistent CSMA (without collision detection) on a bus network ($a = 0.1$). (a) $g^*(x) = 6x(1 - x)$, $\bar{a} = 9a/35$. (b) $g^*(x) = 1$, $\bar{a} = a/3$. (c) $g^*(x) = 12(x - 1/2)^2$, $\bar{a} = 3a/7$.

the asynchronous protocol on a bus network is grossly underestimated by the worst case “star” topology assumption, as one might expect. What is more interesting, however, is that the results for the slotted protocol *also* underestimate its throughput—even though it has been widely believed since the discovery of slotted Aloha [5] that slotting improves the efficiency of a random access protocol.

Our efforts to extend this approach to a more general setting, where we allow stations (with inhomogeneous traffic generation rates) to be placed arbitrarily within a region in N -space, have met with limited success. We have been able to determine some of the properties of such a LAN, including a lower bound on the probability of success. However, we found the problem of bounding the throughput as a function of load *and* topology in the general case to be intractable. The main reason for this was the complexity of the geometry of collisions in higher dimensions, which is partially due to the correlations between the actions of the stations that result from carrier sensing. As a result, the problem of estimating the expected duration of a collision, as it would be measured by

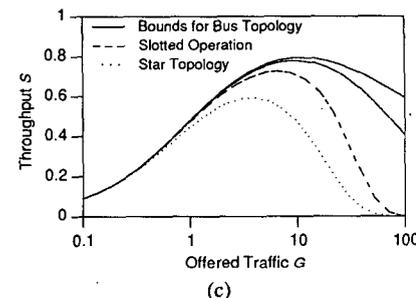
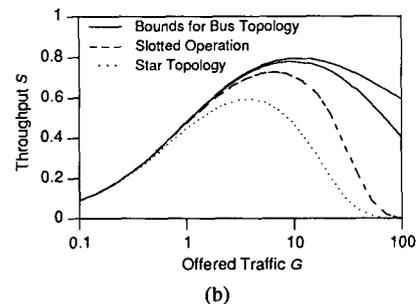
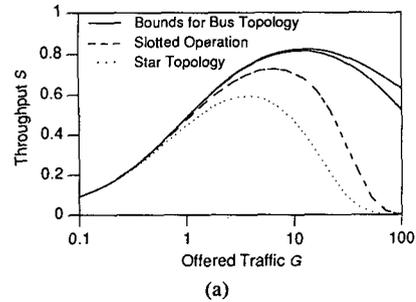


Fig. 11. The effect of the load intensity function on throughput for asynchronous nonpersistent CSMA/CD (with collision detection) on a bus network ($a = 0.1$). (a) $g^*(x) = 6x(1 - x)$, $\bar{a} = 9a/35$. (b) $g^*(x) = 1$, $\bar{a} = a/3$. (c) $g^*(x) = 12(x - 1/2)^2$, $\bar{a} = 3a/7$.

an observer positioned somewhere within the network, was found to be intractable in the general case.

Although we were unable to generalize the results on throughput from one dimension to the general case, we wish to point out that this limitation is not as severe as it might first appear. First, we note that unlike the linear case (where the medium could be a cable that supports very high data rates), planar and 3-dimensional LAN's are typically packet radio networks [2] that must operate in an environment where bandwidth is a scarce resource. Thus, we expect the worst case normalized propagation time a to be much smaller in 2- and 3-dimensional space than it is for 1-space. Since the details of the topology are of little consequence when $a \ll 1$, we expect little difference in the performance predictions for a packet radio network between a worst case “star” approximation and a more elaborate two- or three-dimensional model. Second, we note that even if the density of traffic per unit area/volume remains constant while we increase the dimensionality of the network, the difference between the average case and worst case separation between stations decreases rapidly.

REFERENCES

- [1] J. A. Field and J. W. Wong, "An analysis of a carrier sense multiple access system with collision detection," *Comput. Commun. Networks Group, Univ. Waterloo, Waterloo, Canada, May 1981, CCNG E-Rep. E-95.*
- [2] R. E. Kahn, S. A. Gronemeyer, J. Burchfiel, and R. C. Kunzelman, "Advances in packet radio technology," *Proc. IEEE*, vol. 66, pp. 1468-1496, Nov. 1978.
- [3] L. Kleinrock and F. A. Tobagi, "Packet switching in radio channels: Part I—Carrier sense multiple-access modes and their throughput-delay characteristics," *IEEE Trans. Commun.*, vol. COM-23, pp. 1400-1416, December 1975.
- [4] R. M. Metcalfe and D. R. Boggs, "Ethernet: Distributed packet switching for local computer networks," *Commun. Ass. Comput. Mach.*, vol. 19, July 1976.
- [5] L. G. Roberts, "ALOHA packet system with and without slots and capture," ARPA Network Information Center, Stanford Res. Inst., Menlo Park, CA, ASS Note 8 (NIC 11290), June 1972; reprinted in *Comput. Commun. Rev.*, vol. 5, pp. 28-42, April 1975.
- [6] K. Sohraby, M. L. Molle, and A. N. Venetsanopoulos, "Why analytical models of ethernet-like local networks are so pessimistic," in *Proc. IEEE Global Telecommun. Conf.*, pp. 19.4.1-19.4.6, November 1984.
- [7] K. Sohraby, "Carrier sense multiple access on the bus topology," Ph.D. dissertation, Dep. Elec. Eng., Comput. Syst. Res. Inst., Univ. Toronto, Toronto, Canada, October 1985, Tech. Rep. CSRI-174.
- [8] K. Sohraby, M. L. Molle, and A. N. Venetsanopoulos, "On the stochastic structure of asynchronous CSMA protocols applied to linear (bus-like) networks," in *Proc. IEEE Int. Conf. Commun.*, pp. 58.4.1-58.4.5, June 1986.



Mart L. Molle (S'80-M'82) was born in Toronto, Ont., Canada on November 2, 1953. He received the B.Sc. (Hons.) degree in mathematics/computing science from Queen's University at Kingston, Kingston, Ont., Canada in 1976, and the M.S. and Ph.D. degrees in computer science from the University of California, Los Angeles in 1978 and 1981, respectively.

Since September 1981 he has been on the faculty at the University of Toronto, where he is now an Associate Professor in the Department of Computer Science, with a cross-appointment in the Department of Electrical Engineering, and is a member of the Computer Systems Research Institute. His research interests include the modeling and analysis of protocols for computer networks and distributed systems. He has published extensively on the performance evaluation of various local area network access protocols and distributed conflict resolution algorithms, and won the 1986 ICC Best Paper Award.

Dr. Molle is a member of the Association for Computing Machinery.



Khosrow Sohraby received the B.Eng. and M.Eng. degrees from McGill University, Montreal, P.Q., Canada, in 1979 and 1981, respectively and the Ph.D. degree from University of Toronto, Toronto, Ont., Canada, in 1984, all in electrical engineering.

From December 1984 to July 1986 he was a Research Associate at INRS Telecommunications, Montreal, P.Q., Canada, where he worked on the performance modeling and analysis of new hybrid switching strategies. Since August 1986 he has been with AT&T Bell Laboratories, Holmdel, NJ. He is currently a Member of Technical Staff in the Department of Teletraffic Theory and System Performance. His research interests are in applied stochastic processes, modeling and performance analysis of computer and communication networks, integrated voice/data systems, and multiple-access protocols.



Anastasios N. Venetsanopoulos (S'66-M'69-SM'79) received the B.S. degree in electrical and mechanical engineering from the National Technical University of Athens, Greece (1965) and the M.S. (1966), the M.Phil. (1968) and the Ph.D. (1969) degrees in electrical engineering, all from Yale University, New Haven, CT.

He joined the University of Toronto, Canada in September 1968, where he is now a Professor in the Department of Electrical Engineering. He also served as Chairman of the Communications Group of the same department between 1974-1978 and 1981-1986, and as Associate Chairman between 1978 and 1979 and held visiting posts at the National Technical University of Athens, Athens, Greece, the Federal University of Rio de Janeiro, Brazil, and the University of Florence, Florence, Italy. He was on research leave at the Imperial College of Science and Technology and the University of Grenoble, Grenoble, France, and was an Adjunct Professor at Concordia University, Montreal, P.Q., Canada. He has been lecturer of numerous short courses to industry and continuing education programs; contributor to eight books and over 200 papers in digital communications, digital filters and image processing, and consultant to many organizations. He served as the Assistant Editor (1979-1980) and Editor (1981-1983) of the *Canadian Electrical Engineering Journal*.

Dr. Venetsanopoulos was the President of the Canadian Society for Electrical Engineering and Vice President of the Engineering Institute of Canada (1983-1986). He was a Fulbright Scholar, A. F. Schmitt Scholar, and recipient of the J. Vakis Award. He is a member of the New York Academy of Sciences, Sigma Xi, the International Society for Optical Engineering, and the Technical Chamber of Greece, he is a Fellow of the Engineering Institute of Canada, and is a Registered Professional Engineer in Ontario and Greece. He was Program Chairman of the International Communications Conference (ICC'78 and ICC'86) which were held in Toronto. He has also served as Chairman of the IEEE Toronto Section and Central Canada Council of the IEEE, is presently Associate Editor for Digital Signal Processing of the *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS*, and the Guest Editor of a special issue of the same *TRANSACTIONS* on digital image processing, which will be published in November 1987.