Space-Time Characteristics of ALOHA Protocols in High-Speed Bidirectional Bus Networks

Whay Chiou Lee, *Member, IEEE,* and Pierre A. Humblet, *Fellow, IEEE*

Abstract—We study the space-time characteristics of ALOHA multiple-access protocols in bidirectional bus networks where transmissions are in the form of packets of constant length. For point-to-point communications, the maximum throughput of unslotted ALOHA is known to be $1/(2e)$, independent of station configuration. We show that, with a uniform probabilistic station configuration, the maximum throughput of slotted ALOHA tends to a nonzero constant that is less than $1/(2e)$, when a, the end-to-end propagation delay normalized with respect to the packet transmission time, tends to infinity. However, when N stations are evenly spaced on the bus, the maximum throughput of slotted ALOHA vanishes as a tends to infinity. For broadcast communications, the maximum throughput of slotted ALOHA is well known to be $1/\{e(1+a)\}$. For unslotted ALOHA, we show that, if the offered load intensity is constant along the bus, the maximum broadcast throughput achievable by a station varies along the bus and is maximized at its center. We also derive the optimal profile of the offered load intensity for achieving a constant throughput intensity. In both cases, the maximum broadcast throughput is greater than that derived by conventional analysis.

I. INTRODUCTION

THE ALOHA protocol, which is the first and simplest contention-based multiple-access protocol, may either be asynchronous or require packet transmissions starting only at the beginning of fixed-length time slots [1], [2]. The former version is known as unslotted ALOHA, and the latter slotted ALOHA. In this paper, we present the space-time characteristics of both slotted and unslotted ALOHA protocols in bidirectional bus networks where transmissions are in the form of packets of constant length. We assume that the bus is of unit length and has perfectly nonreflecting terminations at both ends. We consider both point-to-point and broadcast communications. For point-to-point communications, each transmission is designated for successful reception by exactly one station. For broadcast communications, each transmission must be successfully received by all stations.

A bus network is often specified by a parameter a, which denotes the end-to-end propagation delay normalized with respect to the packet transmission time [3]. The end-to-end propagation delay for a bus of length D meters is given by D/v, where v is the propagation velocity in meters per second.

Let W be the data rate of the channel in bits per second, and B the packet length in bits. Then, a $= DW/(vB)$. With the bus length normalized to 1, we focus our attention on the change of a due to variations in the packet length or data rate. The normalization does not affect the final results. Without loss of generality, we let positions on the bus be specified with respect to the center of the bus, so that any position must fall within the range $[-1/2, 1/2]$. We are particularly interested in high speed bidirectional bus networks where a $\gg 1$. We will not consider carrier sensing since it is known to be inefficient for contention-based multiple-access protocols for this case [4].

The bidirectional bus is a possible configuration for high-speed all-optical networks using optical amplifiers to compensate fiber and coupler losses. Although fiber is largely a unidirectional medium, one can implement a high-speed bidirectional bus network using two unidirectional buses with signals propagating in opposite directions, and each transmitter/receiver pair attached to both buses for cost saving (e.g., see [5]–[7]). In these networks, a can be significantly greater than 1.

When propagation delay is negligible, as conventionally assumed, there is no difference in the performance of the ALOHA protocols between point-to-point and broadcast communications. The vulnerability of a transmission is simply characterized by the time interval over which any other packet transmitted could cause a collision. During this time interval, the vulnerable period, the given transmission is vulnerable everywhere on the bus [1], [3]. There is also no difference in performance between point-to-point and broadcast communications when there is a central repeater, as in satellite and some packet radio networks.

The effect of propagation delay on the performance of ALOHA systems was not recognized initially, even though that on CSMA systems was investigated as early as in 1975 [3]. It was later shown in [8] that propagation delay may actually stabilize the ALOHA systems. In [9], Maxemchuk showed that for unidirectional bus networks, slotted protocols are always more efficient than unslotted protocols, even when propagation delay is taken into consideration. We show a different result for bidirectional bus networks.

The spatial properties of the ALOHA protocol were first studied by Abramson, who analyzed the spatial densities of traffic and throughput in a packet radio broadcasting network with capture [10]. The space-time behavior of the ALOHA protocol on bus networks was reported in [6], [7], [9], and [11]. Gonsalves and Tobagi conducted a simulation study of the effects of station locations on the broadcast performance

of Ethernet type bus networks using the CSMA/CD protocol [12]. They observed that, with stations uniformly distributed along the bus, those near the center of the bus obtain better performance than those near the ends. We confirm the above behavior analytically for unslotted ALOHA.

Due to propagation delay, packets transmitted in different directions may overlap nondestructively. When this occurs, we say that there is channel reuse [11], [13]. Allowing for channel reuse, point-to-point transmissions are generally less demanding on channel resources than broadcast transmissions, and one expects the former to have better throughput performance. Although such distinction was previously explored in [6] and [7], it is seldom emphasized in the literature because the difference is insignificant in networks with small propagation delays.

The analysis of multiple-access protocols in bus networks makes extensive use of space-time diagrams. They were used in [14] and [15] for deterministic analysis of a token bus network and of demand assignment multiple-access bus networks, respectively. They were also used in stochastic models for the analysis of CSMA protocols in bidirectional bus networks [16]–[20], and in star-like networks [21].

In [17], offered load was characterized by a load intensity function, which is an adaptation of the traffic density function originally defined by Abramson [10]. We also use the load intensity function in our space-time models.

In a bidirectional bus network where channel reuse is possible, vulnerable time periods do not adequately characterize the vulnerability of transmissions. We need to consider space-time vulnerable regions instead. A vulnerable region associated with a specific transmission is the space-time region over which any other packet arriving at the network could cause a collision. The size of the vulnerable regions is a limiting factor on the performance of a contention-based protocol. In general, the larger the size of the vulnerable regions, the smaller is the probability of success of each transmission. One notion of the vulnerable region was developed by Sohraby et al., where the region is defined with respect to "idle points" and a is constrained to be less than 0.5 [16]–[18]. We make use of a notion of the vulnerable region that is defined differently and that requires no upper limit on a. As the vulnerable region associated with a transmission depends on the location of the source and destination, the throughput of the system depends on the traffic pattern.

Levy and Kleinrock used a space-time model in [6] and [7] to study the behavior of a high-speed bidirectional bus network where there are N evenly spaced stations and $a = (N - 1)$. In their model, every transmission from a station is vulnerable to transmissions from every other station, as their definition of a slot is different from ours. We provide a throughput analysis that is valid for $a \geq (N - 1)$.

We discuss slotted ALOHA and unslotted ALOHA in Section II and Section III respectively. In each case, we first specify our ALOHA model, and review the basic results from conventional throughput analysis. We then examine the space-time characteristics of the protocol, and present some new results on maximum throughput. We show how maximum throughput depends on the offered load intensity. For slotted

ALOHA, we show that, with a uniform probabilistic station configuration, the maximum point-to-point throughput tends to a constant when a tends to infinity. Nevertheless, with a deterministic station configuration where N stations are evenly spaced on the bus, the point-to-point throughput of slotted ALOHA vanishes as a tends to infinity. For unslotted ALOHA, we show that if the offered load intensity is constant along the bus, the maximum broadcast throughput achievable by a station varies along the bus and is maximized at its center. We also derive the optimal profile of the offered load intensity for achieving a uniform throughput intensity.

II. SLOTTED ALOHA

In this section, we study the slotted ALOHA protocol in a bidirectional bus network. Time is divided into slots of length $(1 + a)$ units of packet transmission time. Our definition of a slot is different from that in [6], [7], and [9], but consistent with [1] and [2] where a slot is the duration of time during which the transmission of a packet is in progress in the network. While a was assumed to be small in [1] and [2], we allow it to be large. A packet arriving at a station during a slot is transmitted at the beginning of the following slot, and is completely received by the designated station before the end of the same slot. We summarize our slotted ALOHA model as follows:

- large population of users located along a bus of unit length;
- synchronous transmissions at discrete points in time with period $(1 + a)$ units of packet transmission time;
- offered traffic including retransmissions is a memoryless random process, and is characterized by an offered load intensity function;
- statistical equilibrium.

A. Conventional Analysis

Conventional analysis of slotted ALOHA without channel reuse is based on the assumption that a transmission in a given slot is successful only if there are no other transmissions within the same slot. Let N be the number of stations on the bus. Let G be the average offered traffic per slot, in packets per packet transmission time. The offered traffic is assumed to be uniform across all stations. By symmetry, each station is active during a slot with probability $(1 + a)G/N$. Thus, the probability of success is

$$P_a(G) = \left\{ 1 - (1 + a)\frac{G}{N} \right\}^{N-1} \quad \text{for } a \geq 0. \quad (1)$$

The throughput is then $S_a(G) = G \cdot P_a(G)$. For large N, $S_a(G) = Ge^{-(1+a)G}$, whose maximum with respect to G is

$$S_a^* = \left(\frac{1}{1 + a} \right)\frac{1}{e} \quad \text{for } a \geq 0. \quad (2)$$

The above analysis applies to both point-to-point and broadcast communications. Note that S_a^* vanishes as a tends to infinity. When channel reuse is taken into consideration, we reach a different conclusion, depending on the station configuration, for slotted ALOHA with point-to-point communications.
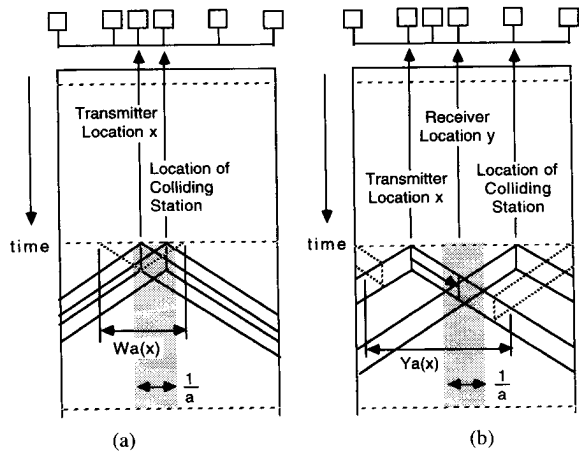
Fig. 1. Totally and potentially destructive collisions. (a) Totally destructive collisions. (b) Potentially destructive collisions.

Fig. 2. Vulnerable regions for slotted ALOHA.

B. Space-Time Characteristics

For broadcast communications with $a \geq 0$, and point-to-point communications with $0 \leq a \leq 1$, the entire previous slot is the vulnerable region. Hence, the maximum throughput is the same as that derived by conventional analysis. For point-to-point communications with $a > 1$, a vulnerable region may be considerably smaller than a whole slot. We show that the point-to-point throughput of slotted ALOHA does not necessarily degrade indefinitely as a becomes very large.

In Fig. 1, we show how two simultaneously transmitted packets may collide destructively in the same time slot. We call the inverted V-shaped space-time region covered by a transmission a transmission region.

In the case of Fig. 1(a), two transmitting stations are within $(1/a)^-$ units of distance from each other, where

$$(*)^- = \min(*, 1). \tag{3}$$

There is a totally destructive collision since no station can successfully receive the transmission. The spatial interval, in which no other transmission may originate without causing a totally destructive collision, is called a totally vulnerable interval. Let the length of the interval be $W_a(x)$ where x is the position of the transmitting station [see Fig. 1(a)]. Then, we have

$$W_a(x) = \left(\frac{2}{a}\right)^-. \tag{4}$$

In the case of Fig. 1(b), two transmission regions cross each other, and there is a potentially destructive collision. The collision is nondestructive if neither of the two designated receivers is located within the spatial interval where the transmission regions cross each other. This spatial interval is $(1/a)^-$ units long. The spatial interval in which no other transmission may originate without causing a potentially destructive collision is called a potentially vulnerable interval. It does not exist if y, the position of the receiving station, falls outside the following range [see Fig. 1(b)]:

$$Y_a(x) \equiv \left[\frac{1}{2}\left(x - \frac{1}{2} - \frac{1}{a}\right), \frac{1}{2}\left(x + \frac{1}{2} + \frac{1}{a}\right)\right] \cap \left[-\frac{1}{2}, \frac{1}{2}\right]. \tag{5}$$
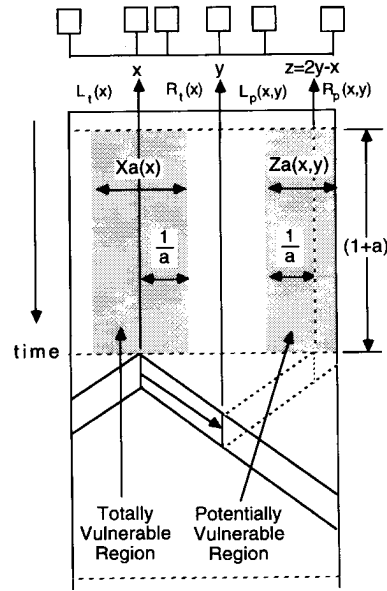
In Fig. 2, we show a typical transmission, and its corresponding totally and potentially vulnerable regions, which are, respectively, specified by the totally vulnerable interval, $[L_t(x), R_t(x)]$, and the potentially vulnerable interval, $[L_p(x, y), R_p(x, y)]$ where

$$R_t(x) = \min\left\{\left(x + \frac{1}{a}\right), \frac{1}{2}\right\} \tag{6}$$

$$L_t(x) = \max\left\{\left(x - \frac{1}{a}\right), -\frac{1}{2}\right\} \tag{7}$$

$$R_p(x, y) = \max\left\{\min\left\{\left(2y - x + \frac{1}{a}\right), \frac{1}{2}\right\}, -\frac{1}{2}\right\} \tag{8}$$

$$L_p(x, y) = \min\left\{\max\left\{\left(2y - x - \frac{1}{a}\right), -\frac{1}{2}\right\}, \frac{1}{2}\right\}. \tag{9}$$

Note that $R_p(x, y) = L_p(x, y)$ when y falls outside of $Y_a(x)$.

Let $X_a(x)$ denote the length of the totally vulnerable interval, and $Z_a(x, y)$ that of the potentially vulnerable interval. Then, it can be verified that

$$X_a(x) = R_t(x) - L_t(x) \leq \left(\frac{2}{a}\right)^- \tag{10}$$

and

$$Z_a(x, y) = R_p(x, y) - L_p(x, y) \leq \left(\frac{2}{a}\right)^- \quad \text{for } y \in Y_a(x). \tag{11}$$

Note that for $a \gg 1$, end effects are negligible, and the above relations hold with equality (see Fig. 3). From (5), we see that

$$\lim_{a \to \infty} Y_a(x) = \left[\frac{1}{2}\left(x - \frac{1}{2}\right), \frac{1}{2}\left(x + \frac{1}{2}\right)\right] \tag{12}$$

and the length of $Y_a(x)$ tends to $1/2$ as a tends to infinity.

We refer to the union of the vulnerable intervals as the vulnerable interval-set, and the union of the vulnerable regions as the vulnerable union. The larger the size of the vulnerable union, the smaller is the maximum throughput of slotted ALOHA.
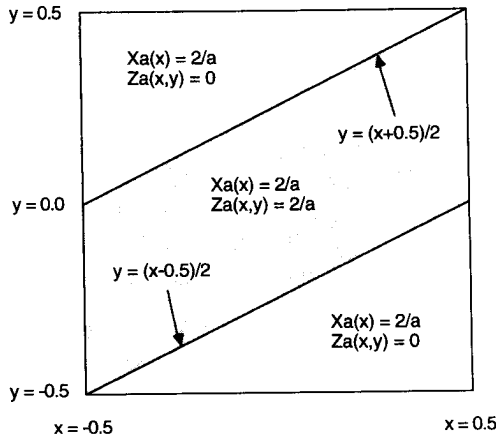
Fig. 3. Vulnerable intervals for $a \gg 1$.

C. Maximum Point-to-Point Throughput

We consider two kinds of station configurations. In the uniform probabilistic station configuration, each packet originates at a randomly chosen point, and the configuration represents a large number of stations uniformly distributed on the bus. In the deterministic station configuration, the location of each station on the bus is fixed. A special case of the deterministic configuration is the regular bus network, in which the distance between any two adjacent stations is a constant.

1) Uniform Probabilistic Station Configuration: We consider slotted ALOHA in a bidirectional bus network with a uniform probabilistic station configuration where $a \gg 1$, and show that it is less efficient than unslotted ALOHA for point-to-point communications.

Let $g(x, y)$ and $S_a(x, y)$ respectively denote the offered load intensity and the point-to-point throughput intensity for transmissions from position x to position y. Let $g(x)$ be the offered load intensity for transmissions originating from location x. By definition, we have

$$g(x) = \int_{-1/2}^{1/2} g(x, y) \, dy. \qquad (13)$$

Let the aggregate throughput that depends on $g(x, y)$ be

$$S = \lim_{a \to \infty} S_a = \lim_{a \to \infty} \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} S_a(x, y) \, dy \, dx. \qquad (14)$$

We characterize S^*, the maximum of S, in Theorem 1 for two different cases.

Theorem 1:

Case 1—Constant Offered Load Intensity: If $g(x, y)$ is constant for all x and y in $[-1/2, 1/2]$, then

$$S^* = 0.1304. \qquad (15)$$

Case 2—Constant Throughput Intensity: If $S_a(x, y)$ is constant for all x and y in $[-1/2, 1/2]$, then

$$S^* = 0.1167. \qquad (16)$$

Proof of Theorem 1: Let $B_a(x, y)$ represent the vulnerable interval-set. Then, the spatial intensity of the probability of success for transmissions from position x to position y is

$$P_a(x, y) = \exp \left\{ - \int_{z \in B_a(x, y)} (1 + a) g(z) \, dz \right\}. \qquad (17)$$

Consider $a \gg 1$. Using Figs. 2 and 3, it is easy to verify that

$$\lim_{a \to \infty} P_a(x, y)$$
$$= \begin{cases} \exp(-2g(x) - 2g(2y - x)) & \text{if } -\frac{1}{2} \le (2y - x) \le \frac{1}{2} \\ \exp(-2g(x)) & \text{otherwise.} \end{cases} \qquad (18)$$

The point-to-point throughput intensity for transmissions from position x to position y is

$$S_a(x, y) = g(x, y) P_a(x, y). \qquad (19)$$

From (18) and (19), we obtain $S_a(x, y)$ as follows, for $a \gg 1$:

$$S_a(x, y) =$$
$$\begin{cases} g(x, y) \exp(-2g(x) - 2g(2y - x)) & \text{if } \frac{x}{2} - \frac{1}{4} \le y \le \frac{x}{2} + \frac{1}{4} \\ g(x, y) \exp(-2g(x)) & \text{otherwise.} \end{cases} \qquad (20)$$

Case 1—Constant Offered Load Intensity: Suppose that $g(x, y) = G$. From (20), we obtain

$$\lim_{a \to \infty} S_a(x, y) = \begin{cases} G e^{-4G} & \text{if } \frac{x}{2} - \frac{1}{4} \le y \le \frac{x}{2} + \frac{1}{4} \\ G e^{-2G} & \text{otherwise.} \end{cases} \qquad (21)$$

From (14) and (21), we obtain the following aggregate throughput:

$$S = \frac{G}{2} \{ e^{-2G} + e^{-4G} \}. \qquad (22)$$

The maximum of S with respect to G is numerically determined to be 0.1304, and this is achieved with $G = 0.38$.

Case 2—Constant Throughput Intensity: Since $S_a(x, y)$ is constant, $g(x, y) P_a(x, y)$ is independent of x and y. We show that there exist two constants, α and β, such that the assignment

$$g(x, y) = \begin{cases} \alpha & \text{if } -\frac{1}{2} \le 2y - x \le \frac{1}{2} \\ \beta & \text{otherwise.} \end{cases} \qquad (23)$$

leads to a constant $S_a(x, y)$.

Equation (23) implies that $g(x)$ has a constant value G that relates to α and β as follows:

$$g(x) = \int_{-1/2}^{1/2} g(x, y) \, dy = \frac{1}{2} (\alpha + \beta) = G \qquad (24)$$

From (20), (23), and (24), and requiring $g(x, y) P_a(x, y)$ to be constant, we obtain

$$S_a(x, y) = \alpha e^{-4G} = \beta e^{-2G}. \qquad (25)$$

Hence,

$$\beta = \alpha e^{-2G}. \qquad (26)$$

We thus obtain two simultaneous equations, (24) and (26), from which we can solve for α and β. Using additionally (14) and (25), we obtain

$$S = \frac{2Ge^{-4G}}{1 + e^{-2G}}. \tag{27}$$

The maximum of S with respect to G is numerically determined to be 0.1167, and this is achieved with $G = 0.30$. □

Note that for $0 \leq a \leq 1$, there is no channel reuse, and the maximum throughput of slotted ALOHA is the same as in (2). For $a \geq 1$, $S_a^* \leq S_1^* = 1/(2e)$, which is the maximum throughput of unslotted ALOHA (see Section III–A2). We have thus shown that slotted ALOHA, as we have defined it, is less efficient than unslotted ALOHA for point-to-point communications in a very high speed bidirectional bus network. This interesting result is analogous to a result for CSMA protocols in Molle *et al.* [17].

2) Deterministic Station Configuration: For a bidirectional bus network with deterministic station configuration, each of the stations along the bus may transmit at most one packet in any period of length $(1 + a)$. Moreover, by considering only totally vulnerable intervals, one can show that the maximum possible number of successful synchronous transmissions cannot exceed $(1 + a)$. Hence, when there are N stations, an upper-bound on the maximum point-to-point throughput of slotted ALOHA is

$$S_a^* \leq \left(\frac{N}{1 + a}\right)^{-} \leq 1 \quad \text{for } a \geq 0. \tag{28}$$

The bound in (28) is valid for any station configuration. We don't observe the factor $N/(1 + a)$ in Theorem 1 because, in the case of uniform probabilistic station configuration, the number of synchronous transmissions that one may squeeze on the bus implicitly grows as a increases.

We now consider the case where N stations are evenly spaced on the bus and $a \geq (N - 1)$. We number the stations 1 through N from one end of the bus to the other. Let $G(m, n)$ and $S_a(m, n)$ respectively denote the offered load intensity and the point-to-point throughput intensity from station m to station n. We assume that $G(m, m) = 0$. Let $G(m)$ be the offered load intensity for transmissions from station m. By definition, we have

$$G(m) = \sum_{n=1}^{N} G(m, n). \tag{29}$$

Let the aggregate throughput that depends on $G(m, n)$ be

$$S_a = \sum_{m=1}^{N} \sum_{n=1}^{N} S_a(m, n). \tag{30}$$

We characterize S_a^*, the maximum of S_a, in Theorem 2 for two different cases.

Theorem 2:

Case 1—Constant Offered Load Intensity: If $G(m, n)$ is constant for any $m \neq n$, then

$$S_a^* = \frac{1}{2}\left(\frac{N}{1 + a}\right) \leq \frac{1}{2} \quad \text{for } a \geq (N - 1). \tag{31}$$
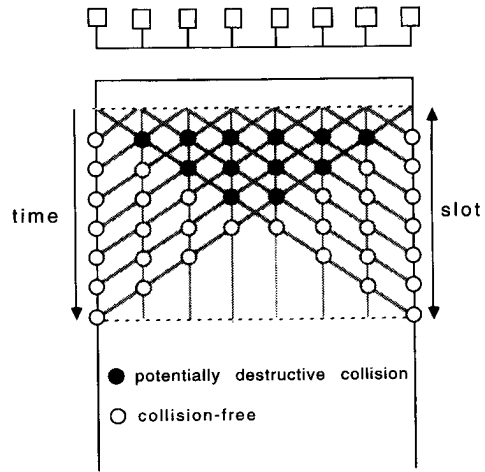

Fig. 4. A very high-speed regular bus network.

Case 2—Constant Throughput Intensity: If $S_a(m, n)$ is constant for any $m \neq n$, then

$$S_a^* = 0.3431\left(\frac{N}{1 + a}\right) \leq 0.3431 \quad \text{for } a \geq (N - 1). \tag{32}$$

Proof of Theorem 2: Let $C_a(m, n)$ be the number of stations located within the vulnerable interval-set associated with any transmission from station m to station n. For $a \geq (N - 1)$, no stations are sufficiently close to each other to allow any simultaneous transmissions that are totally destructive, and any transmission from station m to station n is always successful unless $L(m) \leq n \leq U(m)$, where $L(m)$ is the smallest integer greater than $(m + 1)/2$, and $U(m)$ is the smallest integer greater than $(N + m - 1)/2$ (e.g., see Fig. 4). Moreover, if $L(m) \leq n \leq U(m)$, a transmission from station m to station n is vulnerable only to transmissions that originate from station $(2n - m)$. Thus, for $a \geq (N - 1)$, we have

$$C_a(m, n) = \begin{cases} 1 & \text{if } L(m) \leq n \leq U(m) \\ 0 & \text{otherwise.} \end{cases} \tag{33}$$

The probability of success for a transmission from station m to station n is

$$P_a(m, n) = (1 - (1 + a)G(2n - m))^{C_a(m, n)}. \tag{34}$$

The throughput intensity from station m to station n is

$$S_a(m, n) = G(m, n)P_a(m, n). \tag{35}$$

It follows that

$$S_a(m, n) = \begin{cases} G(m, n)(1 - (1 + a)G(2n - m)) & \text{if } L(m) \leq n \leq U(m) \\ G(m, n) & \text{otherwise.} \end{cases} \tag{36}$$

Case 1—Constant Offered Load Intensity: Since $G(m, n)$ is constant, there is a constant G such that

$$G(m, n) = \left(\frac{1}{N - 1}\right)\frac{G}{N}. \tag{37}$$

It follows from (35) and (37) that

$$S_a(m, n) = \left(\frac{1}{N-1}\right)\frac{G}{N}P_a(m, n). \qquad (38)$$

The throughput intensity for transmissions originating from station m is

$$S_a(m) = \sum_{n=1}^{N} S_a(m, n). \qquad (39)$$

We thus obtain

$$\begin{aligned}
S_a(m) &= \left(\frac{1}{N-1}\right)\frac{G}{N}\Big\{(N-1) - K(m) \\
&\quad + K(m)\left(1 - (1 + a)\frac{G}{N}\right)\Big\} \\
&= \frac{G}{N}\left\{1 - \frac{K(m)}{N-1}(1 + a)\frac{G}{N}\right\} \qquad (40)
\end{aligned}$$

where

$$K(m) = \begin{cases} (N-2)/2 & \text{if } N \text{ is even} \\ (N-1)/2 & \text{if } N \text{ and } m \text{ are odd} \\ (N-3)/2 & \text{if } N \text{ is odd but } m \text{ is even.} \end{cases} \qquad (41)$$

Since $K(m) \leq (N-1)/2$, we obtain

$$S_a(m) \geq \frac{G}{N}\left\{1 - (1 + a)\frac{G}{2N}\right\}. \qquad (42)$$

For large N, $S_a(m)$ approaches the bound in (42) asymptotically. Hence, the aggregate throughput is

$$S_a = \sum_{m=1}^{N} S_a(m) = \left\{1 - (1 + a)\frac{G}{2N}\right\}G. \qquad (43)$$

Then, we can show that

$$S_a^* = \frac{1}{2}\left(\frac{N}{1 + a}\right) \qquad (44)$$

and S_a^* is achieved when $G = N/(1 + a)$. Since $a \geq (N - 1)$, $S_a^* \leq 1/2$.

Case 2—Constant Throughput Intensity: Since $S_a(m, n)$ is constant, $G(m, n)\, P_a(m, n)$ is independent of m and n. We show that there exist $\alpha(a)$ and $\beta(a)$ such that the assignment

$$G(m, n) = \begin{cases} \alpha(a) & \text{if } L(m) \leq n \leq U(m) \\ \beta(a) & \text{otherwise} \end{cases} \qquad (45)$$

leads to a constant $S_a(m, n)$.

Equation (45) implies that $G(m)$ has a constant value G_N that relates to $\alpha(a)$ and $\beta(a)$ as follows:

$$G(m) = \sum_{n=1}^{N} G(m, n) = \frac{N}{2}\alpha(a) + \frac{N}{2}\beta(a) = G_N. \qquad (46)$$

From (36), (45), (46), and requiring $G(m, n)P_a(m, n)$ to be constant, we obtain

$$S_a(m, n) = \beta(a) = \alpha(a)\{1 - (1 + a)G_N\}. \qquad (47)$$

We thus obtain two simultaneous equations, (46) and (47), from which we can solve for $\alpha(a)$ and $\beta(a)$. Using additionally

(30) and (47), and asuming $N \gg 1$ for simplicity, we obtain the following aggregate throughput.

$$S_a = 2\left(\frac{N}{1 + a}\right)((1 + a)G_N)\left(\frac{1 - (1 + a)G_N}{2 - (1 + a)G_n}\right). \qquad (48)$$

Maximizing with respect to G_N, we obtain

$$S_a^* = (6 - 4\sqrt{2})\left(\frac{N}{1 + a}\right) = 0.3431\left(\frac{N}{1 + a}\right). \qquad (49)$$

The above maximum occurs when $G_N = (2 - \sqrt{2})/(1 + a)$. Since $a \geq (N - 1)$, $S_a^* \leq 0.3431$. This maximum throughput was obtained independently by Sibal [22]. $\square$

As a tends to infinity, the maximum throughput for the case with uniform probabilistic station configuration tends to a nonzero constant, whereas the maximum throughput for the case with deterministic station configuration with evenly spaced stations vanishes. When the N stations are not evenly spaced, the maximum throughput for $a \geq (N - 1)$ may be higher because a smaller fraction of the potential receiving points are vulnerable to collisions. For a given N, the station configuration that offers the maximum throughput remains to be a subject for further study.

It is interesting to observe that, when a is large, the maximum point-to-point throughput of the ALOHA protocol is better for the unslotted version than the slotted version. For slotted ALOHA, the maximum throughput is worse in the case where stations are evenly spaced on the bus than the case where stations are uniformly distributed. In general, we conclude that slotting time or space, as done here, can do more harm than good to the performance of the ALOHA protocol in high-speed bidirectional bus networks since bunching of transmissions in space and time increases the chance of destructive collisions.

III. UNSLOTTED ALOHA

In this section, we study the following model of the unslotted ALOHA protocol in a bidirectional bus network:

- asynchronous transmissions;
- offered traffic including retransmissions is a Poisson process, and is characterized by an offered load intensity function;
- statistical equilibrium.

A. Conventional Analysis

We first review the conventional analysis of unslotted ALOHA for both broadcast and point-to-point communications.

1) Broadcast Communications: Conventional analysis of unslotted ALOHA without channel reuse is based on the assumption that a transmission is successful only if there are no other transmissions within a vulnerable period of $2(1 + a)$. This time interval is chosen for the worst case in which an end-station broadcasts a packet to every other station. The conventional vulnerable region for unslotted ALOHA with broadcast communications is shown in Fig. 5.

Let G be the constant offered traffic rate, in packets per second, including retransmissions. Then, the probability of
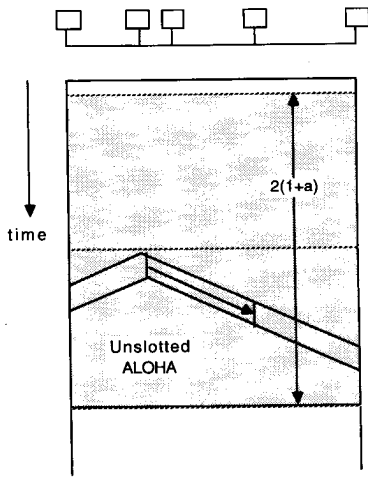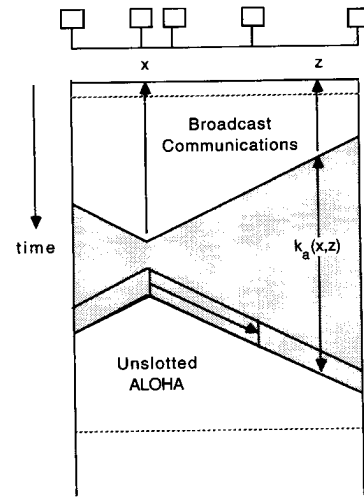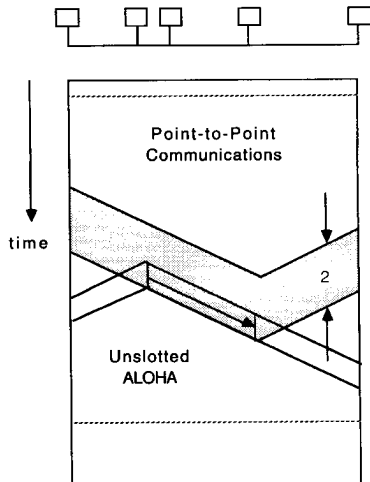
Fig. 5. Conventional vulnerable region.

Fig. 6. Vulnerable region for point-to-point communications.

success is

$$P_a = e^{-2(1+a)G} \qquad \text{for } a \geq 0. \tag{50}$$

The broadcast throughput is given by $S_a = Ge^{-2(1+a)G}$, whose maximum with respect to G is

$$S_a^* = \left(\frac{1}{1+a}\right)\frac{1}{2e} \qquad \text{for } a \geq 0. \tag{51}$$

Note that S_a^* vanishes as a increases to infinity, with a factor of 2 faster than that of unslotted ALOHA. When channel reuse is taken into consideration, we obtain different results.

2) Point-to-Point Communications: Allowing for channel reuse, the vulnerable region for point-to-point communications is actually smaller than that for broadcast communications. As shown in Fig. 6, the space-time area of a point-to-point vulnerable region is always equal to 2 (i.e., 2 units of packet transmission time × 1 unit of bus length), independently of a provided the bus length remains normalized to 1. It is well known that the throughput of unslotted ALOHA for point-to-point communications is Ge^{-2G}, whose maximum is $1/(2e)$ (see [1]).

Fig. 7. Vulnerable region for broadcast communications.

B. Space-Time Characteristics for Broadcast Communications

For broadcast communications, the vulnerable region for a transmission is shown in Fig. 7. Let $V_a(x)$ be its area. It is easy to verify that

$$V_a(x) = 2 + a/2 + 2ax^2 \tag{52}$$

$V_a(x)$ is symmetric about, and minimized at, $x = 0$. Hence, we could expect the throughput performance to be a function of x, and to be largest in the middle of the bus. Since $V_a(x)$ increases with a, and is less than $(2 + a) \leq 2(1 + a)$, the broadcast throughput of unslotted ALOHA indeed degrades as a increases, but more slowly than that under the conventional assumption.

C. Maximum Broadcast Throughput

We show that, if the offered load intensity is constant along the bus, the maximum throughput intensity depends on the location of the transmitting station. To achieve a constant throughput intensity, the offered load intensity has to vary along the bus. In any case, the maximum aggregate throughput degrades with a.

Let $g(x)$ be the offered load intensity at position x, in packets per second. We assume that $g(x)$ is symmetric about the center of the bus. The throughput intensity at position x for broadcast communications is

$$S_a(x) = g(x)P_a(x) \tag{53}$$

where $P_a(x)$ is the spatial intensity of the probability of success. In Theorem 3, we derive a differential equation relating the throughput intensity to the offered load intensity. We then obtain complete solutions for two special cases, as Abramson did in [10] for a packet radio broadcasting network with capture. Specifically, we determine bounds on the maximum aggregate throughput in each case.

Theorem 3: For $a \geq 0, S_a(x)$ is the solution to the following differential equation:

$$S_a'(x)g(x) = S_a(x)\{g'(x) - g(x)h_a(x)\} \tag{54}$$

where $f'(x)$ denotes the derivative of a function, $f(x)$, with respect to x, and

$$h_a(x) = 2a \left\{ \int_{-1/2}^{x} g(z)\,dz - \int_{x}^{1/2} g(z)\,dz \right\}. \quad (55)$$

Let the aggregate throughput be

$$S_a = \int_{-1/2}^{1/2} S_a(x)\,dx. \quad (56)$$

We characterize S_a^*, the maximum of S_a, for two different cases.

Case 1—Constant Offered Load Intensity: If $g(x)$ is constant, then S_a^* is bounded as follows:

$$\left(\frac{1}{1+a/2} \right) \frac{1}{2e} \le S_a^* \le \left(\frac{1}{1+a/4} \right) \frac{1}{2e} \quad \text{for } a \ge 0. \quad (57)$$

Case 2—Constant Throughput Intensity: If $S_a(x)$ is constant, then S_a^* is bounded as follows:

$$\left(\frac{1}{1+a} \right) \frac{1}{2e} \le S_a^* \le \left(\frac{1}{1+a/2} \right) \frac{1}{2e} \quad \text{for } a \ge 0. \quad (58)$$

Proof of Theorem 3: For $a \ge 0$, let $k_a(x, z)$ be the temporal length of the vulnerable region at location z when the transmission originates at location x. For broadcast communications, as shown in Fig. 7,

$$k_a(x, z) = 2(1 + a|z - x|). \quad (59)$$

The spatial intensity for the probability of success is

$$P_a(x) = \exp \left\{ - \int_{-1/2}^{1/2} k_a(x, z) g(z)\,dz \right\}. \quad (60)$$

Taking the derivative of (53), multiplying each side by $g(x)$, and using (60), we obtain

$$S_a'(x)g(x) - g'(x)S_a(x) = g(x)S_a(x) \int_{-1/2}^{1/2} k_a'(x, z)g(z)\,dz. \quad (61)$$

It is easy to verify that

$$k_a'(x, z) = \begin{cases} -2a & \text{if } x \le z \\ +2a & \text{if } x > z \end{cases}. \quad (62)$$

It follows that (54) holds with $h_a(x)$ defined below:

$$h_a(x) = \int_{-1/2}^{1/2} k_a'(x, z)g(z)\,dz$$
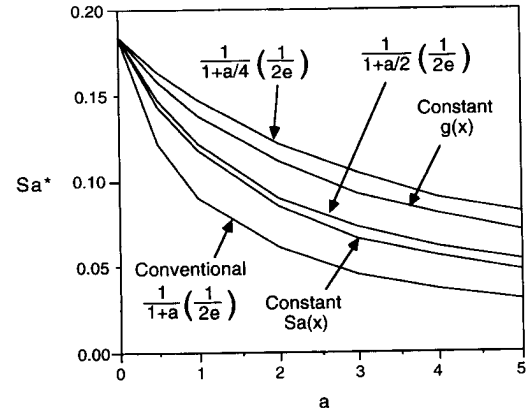$$= 2a \left\{ \int_{-1/2}^{x} g(z)\,dz - \int_{x}^{1/2} g(z)\,dz \right\}. \quad (63)$$

Fig. 8. Maximum broadcast throughput for unslotted ALOHA.

Case 1—Constant Offered Load Intensity: Suppose that

$$g(x) = G. \quad (64)$$

From (55), we have

$$h_a(x) = 4aGx. \quad (65)$$

From (54), (64), and (65), we obtain the following differential equation:

$$S_a'(x) = -4aGx S_a(x). \quad (66)$$

Solving (66), we obtain the following broadcast throughput intensity:

$$S_a(x) = Ge^{-(2+a)G} \{ e^{-2aG(x^2 - 1/4)} \}. \quad (67)$$

Note that for a given G, $S_a(x)$ is minimized at the ends and maximized at the center of the bus. It follows from (67) that

$$Ge^{-(2+a)G} \le S_a(x) \le Ge^{-(2+a/2)G}. \quad (68)$$

Using (56) and (67), we can write

$$S_a = Ge^{-(2+a/2)G} \left(\frac{\pi}{2aG} \right)^{1/2} \text{erf}\left(\left\{ \frac{aG}{2} \right\}^{1/2} \right) \quad (69)$$

where $\text{erf}(*)$ is the following standard error function:

$$\text{erf}(y) = \frac{2}{\sqrt{\pi}} \int_{0}^{y} e^{-w^2}\,dw. \quad (70)$$

We can now determine S_a^* by maximizing S_a with respect to G.

From (56) and (68), we obtain the bounds in (57). Note that for large a, the lower bound is twice the maximum broadcast throughput derived by conventional analysis. In Fig. 8, we show S_a^* and its bounds. We have included the result of Case 2 and that of conventional analysis for comparison.

Case 2—Constant Throughput Intensity: Suppose that $S_a(x)$ is constant. This corresponds to the interesting case where all stations have the same throughput. By symmetry, we have

$$g(x) = g(-x). \tag{71}$$

Taking the derivative of (55) and using (71), we obtain

$$h'_a(x) = 4ag(x). \tag{72}$$

Since $S'_a(x) = 0$, we obtain from (54) that

$$g'(x) = h_a(x)g(x). \tag{73}$$

Taking the derivative of (73), multiplying each side by $g(x)$, and using (72), we obtain

$$g''(x)g(x) = 4ag^3(x) + \{g'(x)\}^2. \tag{74}$$

Solving the above differential equation, using (71), we obtain

$$g(x) = \frac{b^2}{2a\cos^2(bx)} = \frac{b^2}{2a}\sec^2(bx) \tag{75}$$

for a given constant b. Define R as follows:

$$R = \pi/(2b). \tag{76}$$

Note that $g(x)$ is unbounded if

$$|x| \geq R. \tag{77}$$

If $b > \pi$, then $R < 1/2$, and (60) implies that $P_a(x) = 0$ for $x \in [-1/2, 1/2]$. It follows that S_a can only be zero. For a given b, [10] defines the Sisyphus Distance as the value of x with which $g(x)$ in (75) becomes unbounded. It does not appear to have any physical meaning in this case (without capture), as b is an arbitrary parameter. In the analysis below, b is always smaller than π, so that $R > 1/2 \geq x$.

To evaluate $P_a(x)$, we make use of the following indefinite integral [23]:

$$\int x\sec^2(bx)\,dx = \frac{x}{b}\tan(bx) + \frac{1}{b^2}\ln\{\cos(bx)\}. \tag{78}$$

Using (59), (60), (75), and (78), we obtain

$$P_a(x) = \exp\left\{-b\{\tan(b/2)\}\left(1 + \frac{2}{a}\right) - \ln\{\sec^2(bx)\}\right\}. \tag{79}$$

From (53), (75), and (79), we obtain the following aggregate throughput:

$$S_a = \frac{b^2}{2a}\exp\left\{-\frac{b}{a}\{\tan(b/2)\}(a + 2)\right\}. \tag{80}$$

We can now determine S_a^* by maximizing S_a with respect to b. Taking the derivative of the right-hand side of (80) with respect to b, and setting it to zero, we obtain

$$\left(\frac{b}{2}\right)^2\tan^2\left(\frac{b}{2}\right) + \left(\frac{b}{2}\right)\tan\left(\frac{b}{2}\right) + \left(\frac{b}{2}\right)^2 = \left(\frac{a}{a+2}\right). \tag{81}$$

Equation (81) can be solved numerically to determine the value of b that maximizes S_a. Making use of the fact that
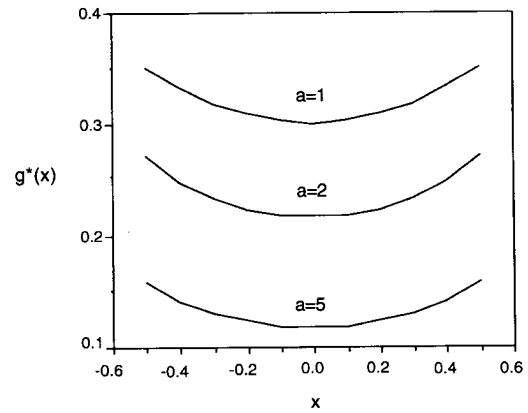

Fig. 9. Optimal offered load intensity.

$\tan(b/2) > (b/2)$, we obtain from (80) the upper bound in (58). The lower bound in (58) follows from the fact that the vulnerable region considered in the analysis is smaller than that assumed in the conventional analysis. □

We show in Fig. 8 the behavior of S_a^* as a function of a. The optimal offered load intensity, $g^*(x)$, which is obtained from (75) with the optimal value of b, is shown in Fig. 9. Note that $g^*(x)$ decreases with increasing value of a. As $g^*(x)$ is proportional to the number of retransmissions, this confirms the observation in [12].

IV. CONCLUSION

We have evaluated the throughput performance of slotted and unslotted ALOHA in a bidirectional bus network by giving special attention to the inherent channel reuse characteristics of the protocols. We have particularly examined the behavior of the ALOHA protocols when propagation delays are much larger than the packet transmission time. We have shown that conventional analysis sometimes overestimates the maximum throughput by neglecting the effect of propagation delay, and sometimes underestimates the maximum throughput by not considering channel reuse. We have investigated how the ALOHA protocols depend on offered load and throughput intensities. In both slotted and unslotted ALOHA, the maximum aggregate throughput is higher for the case with constant offered load intensity than for the case with constant throughput intensity.

For point-to-point communications in a bidirectional bus network with uniform probabilistic station configuration, the maximum throughput for slotted ALOHA degrades below that of unslotted ALOHA when propagation delay is large, but remains above zero. When the station configuration is deterministic with a fixed number of stations evenly spaced on the bus, the maximum throughput for slotted ALOHA vanishes as a tends to infinity. We learn that slotting time or space, as done here, can do more harm than good to the performance of the ALOHA protocol in high-speed bidirectional bus networks since bunching of transmissions in space and time increases the chance of destructive collisions.

For unslotted ALOHA with broadcast communications, we have shown that, if the offered load intensity is uniform along the bus, the maximum throughput achievable by a station

varies along the bus, and is maximized at its center. To achieve a uniform throughput intensity, the offered load intensity has to vary along the bus. We have derived the optimal profile of the offered load intensity.

In conclusion, this paper contributes to a better understanding of contention-based multiple-access protocols on high-speed bidirectional bus networks. The results reported in this paper are a significant generalization of some of those in [6] and [7]. They are also largely consistent with and complementary to those obtained for CSMA in [16], [17], and [18] where the range of a is limited to small values.

ACKNOWLEDGMENT

REFERENCES

[1] L. G. Roberts, "ALOHA packet system with and without slots and capture," *Comput. Commun. Rev.*, vol. 5, no. 2, Apr. 1975.
[2] N. Abramson, "The ALOHA system—Another alternative for computer communications," in *AFIPS Conference Proceedings*. Montvale, NJ: AFIPS Press, 1970, vol. 37.
[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, Dec. 1975.
[4] S. R. Sachs, "Alternative local area network access protocols," *IEEE Commun. Mag.*, vol. 26, Mar. 1988.
[5] M. A. Marsan, G. Albertengo, A. Francese, E. Leonardi and F. Neri, "All-optical bidirectional Manhattan networks," in *Proc. IEEE ICC*, 1992.
[6] H. Levy, "Non-uniform sturctures and synchronization patterns in shared-channel communication networks," Ph.D. Dissertation, Dep. Comput. Sci., UCLA, Aug. 1984.
[7] L. Kleinrock and H. Levy, "On the behavior of a very fast bidirectional bus network," *IEEE Trans. Commun.*, vol. 38, Oct. 1990. (Also in *ICC'81*)
[8] Y. Onozato, J. Liu, S. Shimamoto, and S. Noguchi, "Effect of propagation delays on ALOHA systems," in *Comput. Networks & ISDN Syst.*, vol. 12, 1986.
[9] N. Maxemchuk, "Twelve random access strategies for fiber optic networks," *IEEE Trans. Commun.*, vol. 36, Aug. 1988.
[10] N. Abramson, "The throughput of packet broadcasting channels," *IEEE Trans. Commun.*, vol. COM-25, Jan. 1977.
[11] W. Lee, "Channel reuse multiple access in bidirectional bus networks," Doctoral dissertation, Dep. Elec. Eng. Comput. Sci., Massachusetts Inst. Technol., Feb. 1989.
[12] T. A. Gonsalves and F. A. Tobagi, "On the performance effects of station locations and access protocol parameters in Ethernet networks," *IEEE Trans. Commun.*, vol. 36, Apr. 1988.
[13] W. Lee and P. Humblet, "Channel reuse multiple access protocol for bidirectional bus networks," in *Proc. IEEE ICC*, Apr. 1990.
[14] J. O. Limb, "Performance of local area networks at high speed," *IEEE Commun. Mag.*, vol. 22, Aug. 1984.
[15] M. Fine and F. A. Tobagi, "Demand assignment multiple access schemes in broadcast bus local area networks," *IEEE Trans. Comput.*, vol. C-33, Dec. 1984.
[16] K. Sohraby, M. L. Molle, and A. N. Venetsanopoulos, "Why analytical models of Ethernet-like local networks are so pessimistic," in *Proc. IEEE GLOBECOM*, Nov. 1984.
[17] M. L. Molle, K. Sohraby, and A. N. Venetsanopoulos, "Space-time models of asynchronous CSMA protocols for local area networks," *IEEE J. Select. Areas Commun.*, vol. SAC-5, July 1987.
[18] K. Sohraby, "Carrier sense multiple access on the bus topology," Doctoral dissertation, Dep. Elec. Eng., Comput. Syst. Res. Inst., Univ. of Toronto, Toronto, Ont., Canada, Oct. 1985.
[19] D. K. Choi and B. G. Kim, "The expected (not worst-case) throughput of the Ethernet protocol," *IEEE Trans. Comput.*, vol. 40, Mar. 1991.
[20] A. E. Kamal, "A discrete-time approach to the modeling of carrier-sense multiple access on the bus topology," *IEEE Trans. Commun.*, vol. 40, Mar. 1992.
[21] I. Vukovic and K. S. Vastola, "Throughput analysis of asynchronous CSMA protocols on star-like LAN topologies," in *Proc. IEEE ICC*, June 1992.
[22] S. Sibal, personal communication, Rensselaer Polytech. Inst., Troy, NY, Dec. 1993.
[23] I. N. Bronshtein and K. A. Semendyayev, *Handbook of Mathematics*. New York: Van Nostrand–Reinhold, 1985; English translation edited by K. A. Hirsch.
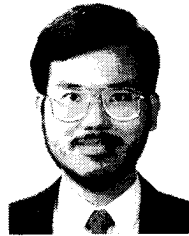
Whay Chiou Lee (S'83–M'88) received the B.S., M.S., E.E., and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge.

In 1982, he was an intern at COMSAT Laboratories, Clarksburg, MD, where he worked on coded modulation for satellite communications. In November 1988, he joined the Networking Research Department of Motorola, Mansfield, MA, where he conducted applied research in routing, connection management, and quality of service issues in high-speed integrated cell relay networks. Since January 1994, he has been representing Motorola in the PNNI Subworking Group of the ATM Forum.

Pierre A. Humblet (S'72–M'77–SM'91–F'93) received the E.E. degree from the University of Louvain, Belgium, in 1973, and the M.S.E.E. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge.

After graduating in 1978, he remained at M.I.T., where he became a Professor of Electrical Engineering. In 1993, he joined the Eurecom Institute, France. His teaching and research interests are in the area of communication systems, particularly digital mobile networks and optical networks. He is a consultant with a number of companies, most recently Motorola and IBM.