COLLEGE OF ENGINEERING AND APPLIED SCIENCES
STATE UNIVERSITY OF NEW YORK
STONY BROOK, NY 11794

CEAS Technical Report #659

January 7, 1993

ANALYSIS OF MULTIMEDIA PCN MEDIUM ACCESS CONTROL IN A MIXED TRAFFIC ENVIRONMENT

Nisha P. Newman and Stephen S. Rappaport
Department of Electrical Engineering

Analysis of Multimedia PCN Medium Access Control in a Mixed Traffic Environment

Nisha P. Newman and Stephen S. Rappaport
Department of Electrical Engineering
State University of New York
Stony Brook, NY 11794-2350, USA

ABSTRACT

Future wireless PCNs will be required to support a variety of services, including digitized voice, interactive data and personal computer based multi-media. The resulting mixed traffic environment can contain real-time traffic, with its associated maximum delay (expiration time) constraints. We consider a simple medium access control technique for a slotted multiple access channel in a mixed traffic environment. The focus is to optimize the multiple access techniques so that the maximum delay constraints are satisfied. We consider a slotted Aloha type protocol for a uniform traffic scenario, where all packets have the same maximum delay constraint. Formulas that characterize steady state performance for fixed and adaptive retransmission techniques (based on mean retransmission delay) are derived. Expressions for channel throughput, average packet delay, and average packet loss rate are developed. Trade-offs involved in the selection of system parameters are discussed.

A generalized multiple access environment, consisting of two distinct classes of traffic, each with its own maximum delay constraints is also considered. The analysis methodology used, can be easily extended to include additional classes of traffic. Using adaptive retransmission techniques, significant improvement in performance can be obtained in the high throughput operating region. For a special case of a mix of real-time and non-real time traffic (no maximum delay constraints), the system parameters can be chosen to satisfy a low packet loss (≤0.1%) requirement. However, this can result in system instability. The performance does not change significantly with the actual mix (%) of real-time and non real-time traffic. The access control techniques discussed are applicable to a variety of communications applications, including low earth orbit satellite communication systems.

January 1993

The research reported in this paper was supported in part by the U. S. National Science Foundation under Grant No. NCR-90-25131 and in part by IST/SDIO under Grant No. N00014-91-J-4063 administered by the U. S. Office of Naval Research.

I. INTRODUCTION

In first generation cellular systems, multiple access techniques are used by a ready user to gain access to the radio channel for transmissions to/from a base station. The transmissions on the radio channel are primarily the voice calls, and analog FM/FDM techniques are used on these voice channels. Second generation cellular systems use digital transmission standards such as GSM, IS54, etc. These standards address primarily the voice traffic scenario, but can be extended to include data traffic. A major thrust in cellular systems architecture is towards personal communications which will provide ubiquitous radio communication access in an integrated traffic environment [CO87, CO90, CO91].

If voice codecs used with digital cellular systems are constrained to produce a constant bit-rate output, there will be variations of reconstructed voice quality and/or delays due to buffering. It is desirable to employ variable bit-rate (VBR) voice codecs to maintain voice quality. Since wireline (B-ISDN/ATM) systems will support integrated traffic consisting of voice, data, video and intelligent network services, wireless PCN's would be required to support some of these applications within the radio spectrum limitations. With the trend towards allocation of separate spectrum worldwide for wireless PCN applications (WARC '92), it will be possible to support integrated traffic consisting of packet voice, data and other personal computer (lap-top portables) based multi-media applications. In this mixed traffic environment, the real-time traffic, can have stringent delay constraints.

Packet multiple access techniques [BO81, RA79, RA81, WI85] can provide efficient trunking for applications requiring bandwidth on-demand. In general, the multiple access channel throughput can be increased at the cost of additional delay. Packet retransmission delay can be changed using back-off strategies [ME76] to allow low average packet delay under light traffic conditions and avoid congestion on the channel in heavy traffic. This reduces the chances of operating in the low throughput/high delay region. In computer communication applications, the packets may not have stringent maximum delay constraints. However, in a multimedia PCN environment with real-time applications, some data is unusable beyond a certain delay (expiration time). These data packets are not available to generate the reconstructed signal and cannot be used later. The focus in this paper will be to optimize the multiple access techniques so that the maximum delay constraints are satisfied and the packet loss rate is minimized. Maximum delay (expiration time) constraints in a queueing environment have been addressed in [CA89, CH89].

Many real time applications such as digitized voice and video, must accommodate packet loss due to channel errors, since there is no time available to request retransmission of the lost data. Error concealment techniques [WA91] help to approximate this lost data. These can be extended to partially compensate for packets lost because of time-of-expiration constraints. Concealment techniques increase receiver complexity. Furthermore, the reconstructed (voice/ video) signal quality deteriorates compared to the no loss situation. Therefore, it is important to maintain the packet loss probability below some acceptable level.

The slotted Aloha [KL73] multiple access technique requires little coordination among users and has a maximum throughput of ~0.36. Therefore, it is well suited for early PCN applications in a cellular environment with integrated traffic. If improved throughput is required, a time division multiple access (TDMA) [WI92] or PRMA [GO91] scheme may be used. Some slots in each TDMA frame can be assigned for stream traffic and remaining slots can be accessed in slotted Aloha mode for data (real-time and non-real time) traffic. The Aloha traffic can include stream channel assignment requests. A VBR voice codec may be treated as a stream traffic source with its own TDMA channel assignment and the excess capacity required for VBR operation can be accommodated as real-time data. The slotted Aloha type medium access control can also be applied to low earth orbit satellite based communication systems, where the congestion in the radio spectrum is less severe.

In this paper, the focus will be to optimize the multiple access techniques so that the maximum delay constraints are satisfied and the packet loss rate is minimized. First, we consider the multiple access environment where slotted Aloha is employed for packet radio transmission. All the traffic (packets) into the system have the same maximum delay (expiration time) constraint. Analytical expressions are derived for the channel throughput, average packet delay and packet loss probability. Analysis includes both fixed and adaptive (back-off) retransmission rates. Numerical results show the improvement in stability and reduction in packet loss rates that result from using adaptive retransmission techniques.

Next, a mixed traffic scenario is considered where a certain fraction of the incoming traffic has stringent maximum delay constraint (real-time traffic) relative to the remaining traffic. Again, the performance expressions are derived for both fixed and adaptive retransmission rate [ME76] scenarios. The analysis methodology used is general and can be extended easily to include multiple classes of traffic. Numerical results applicable to terrestrial and satellite environments are included. The results show that suitable selection of adaptive retransmission parameters can help improve the stable operation of the system (in the presence of small input traffic fluctuations), while keeping the average packet loss rates and average packet delays acceptably small.

In section II, a brief description of the system model is given along with the analysis leading to the system performance expressions for channel throughput, average packet delay and packet loss rate. The numerical results obtained for selected system parameters are discussed in section III. This is followed by the concluding remarks in section IV and references in section V. The probability distribution function computations are included in the Appendix.

II. SYSTEM DESCRIPTION AND ANALYSIS

In this section, the multiple access environment is described and the steady-state performance of slotted Aloha type medium access control schemes is analyzed. Expressions for the system performance measures such as throughput, average packet delay and average packet loss are obtained. Without loss of generality, the variables in the expressions are normalized to a slot duration on the multiple access channel. First, we consider a uniform traffic environment with fixed and adaptive packet retransmission delay (back-off). Next, a mixed traffic environment, with two types of traffic, each with a different maximum-delay requirement, is considered. Fixed and adaptive packet retransmission delay strategies are analyzed. A special case of mixed traffic consisting of real-time (maximum-delay constraint) and non-real time traffic is also analyzed. This type of integrated traffic scenario can occur in future wireless PCN systems.

2.1. Multiple access environment

The multiple access environment consists of a population of users accessing a single slotted channel, for packet transmissions to a central destination. We assume that all packets are of equal length (bits). The destination can be a base station of a cellular PCN system or a low earth orbit satellite (LEOS). The users access the channel following a slotted Aloha type protocol [KL73]. Here, a ready user, with a packet arrival in the current slot, transmits the packet in the following slot. If there is(are) no simultaneous transmission(s) from other users, the packet is assumed to be received correctly at the central destination. The user receives an acknowledgement from the central destination indicating successful packet reception. This acknowledgement is sent on a separate channel which is assumed to be very reliable. On receiving the acknowledgement, the user becomes idle, until its next packet arrival. The successful packet is considered to have left the multiple access environment.

If there are simultaneous transmissions on the channel by other users, the packets are assumed to be unsuccessful. No acknowledgment can be sent for any of the packets involved in this collision. After transmitting a packet, a user waits for a fixed interval, X (equal to the sum of (a) packet transmission time, (b) the round trip delay and (c) the processing time at the destination, expressed in slots) for an acknowledgement of its successful transmission. If it does not receive an acknowledgement (as in this event of collision(s)), it schedules its packet for retransmission after an exponentially distributed random delay. The random delay has a pre-specified mean.

For real-time traffic, a newly arrived packet is associated with a maximum delay (expiration time) constraint. If a packet is not successful in its previous attempts, it may be rescheduled for another transmission after a random retransmission delay. Retransmission of the packet occurs only if the total time delay from the arrival time to the next retransmission time is less than the maximum delay (expiration time) allowed. Otherwise, the user flushes out the packet and becomes idle. This packet is assumed to be lost from the multiple access environment. For non-real time traffic, by definition there is no maximum delay constraint, and retransmissions take place until the packets are successfully transmitted.

2.2. Uniform traffic scenario

We assume that the new packet transmissions are Poisson distributed, with a normalized packet origination rate of N packets per slot (fig. 1). The normalized retransmitted packet arrival rate is R packets per slot. The total offered traffic on the multiple access channel, which is the sum of the new and retransmitted packets, is also assumed to be Poisson distributed, with a normalized offered traffic rate, G (=N+R) packets per slot duration. Without loss of generality, the packet transmission time or slot interval is assumed to be of unit time duration. In an actual system, the normalized packet origination rate, N may be made up of the new packet arrivals from a population of users, each with its own mean packet origination rate.

The average number of packets that leave the multiple access environment after being successfully transmitted to the central destination, is denoted by the normalized channel throughput, S packets per slot. For real-time traffic with maximum time delay constraints, the average number of packets that exit the system (due to maximum delay constraints) before successful transmission to the central destination, is denoted by the normalized packet loss rate, L packets per slot. Therefore, in steady state,

$$N = S + L \quad . \qquad (1)$$

For non-real time traffic, the packet loss rate is zero, and S is equal to N.

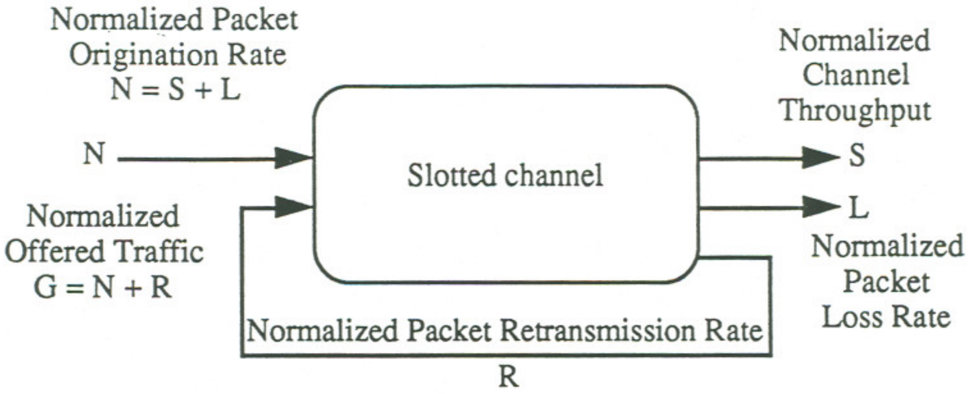

Figure 1: Uniform Traffic - Steady State System Parameters

Although a (new or rescheduled) packet arrival can occur at any time within the slot duration, it is considered to have arrived at the end of the slot interval, for the delay calculations. The elapsed time from the new packet arrival time to the completion of its successful transmission (which can involve many retransmissions) is taken to be the total time delay for successful transmission of this packet. The normalized average packet delay, D (slots) denotes the average delay for packet success.

The maximum delay (expiration time), E (slots) for a newly arrived packet denotes the maximum time interval (in slots) available for the packet to make a final transmission attempt on the channel, before it exits the system as a lost packet. Note that a maximum of M retransmissions can take place, where M is equal to the largest integer that does not exceed the value of (E/X). In the uniform traffic environment, all newly arrived packets are assumed to have the same expiration time, E. For a very large E (relative to X and the mean packet retransmission delay) the performance expressions can be approximated by the slotted Aloha protocol results.

2.2.1. Analysis for fixed mean retransmission delay

Let us assume that the average packet retransmission delay, T is fixed. The system performance expressions are derived below. We begin with the expression for the probability of success, P_S for a newly arrived packet.

The probability of a packet success in a given slot, p is the probability that there is no other simultaneous packet transmission in that slot. From [KL73],

$$p = \exp(-G) \quad . \tag{2}$$

The corresponding packet failure probability, q is (1 - p).

Let R_i denote the retransmission delay for i th retransmission attempt of a packet. The total delay, Z_i (starting from the arrival time) until the beginning of i th retransmission is

$$Z_i = (R_1 + X) + (R_2 + X) + ... + (R_i + X) = \sum_{k=1}^{i} R_k + i\,X = Y_i + i\,X \quad . \tag{3}$$

where Y_i is the sum of the retransmission delays R_i 's. The probability that the total delay, Z_i is less than or equal to the expiration time, E, is given by

$$P_c(i) = P\left\{Z_i \le E\right\} = P\left\{Y_i \le E - (iX)\right\} = P\left\{Y_i \le C(i)\right\} \quad , \tag{4}$$

where C(i) is (E - i X).

For a packet to make the m th retransmission, it must have failed in all the previous m attempts (probability = q^m) and the total delay Z_m should be less than or equal to the expiration time, E (probability denoted by $P_c(m)$). Then it can succeed with probability, p in this m th retransmission attempt. Therefore, the probability of a packet success in the m th retransmission, $P_r(m)$ is

$$P_r(m) = p\,(q^m)\,P_c(m) \quad . \tag{5}$$

The probability of a packet success, P_s is the probability that it succeeds in the first transmission or it succeeds in the 1st, 2nd, etc. up to M retransmissions. Thus,

$$P_s = p + \sum_{m=1}^{M} P_r(m) = p\left\{1 + \sum_{m=1}^{M} (q^m)\,P_c(m)\right\} \quad . \tag{6}$$

The probability, $P_c(m)$ that the sum of m retransmission delays, Y_m is less than or equal to C(m) must be found. Since R_i's are i.i.d. exponential random variables, the sum of m retransmission delays , Y_m has an Erlangian type distribution [GR85] (see appendix). Hence, the probability of $(Y_m \le C(m))$ is given by

$$P_c(m) = \int_0^{C(m)} \frac{(A^m)}{(m-1)!}\left(y^{m-1}\right)\left(e^{-Ay}\right) dy \quad , \tag{7}$$

in which (by definition), $\frac{1}{A}$ is the mean of retransmission delay random variable, R_i. That is, $\frac{1}{A}$ = T. Evaluating the integral (7) by parts, one finds that

$$P_c(m) = 1 - (\exp(-AC(m))) \sum_{k=1}^{m} \frac{(AC(m))^{m-k}}{(m-k)!} \quad . \tag{8}$$

Using the expressions (2), (6) and (8), the probability of a newly originated packet success, P_s can be computed for a given offered traffic, G. The normalized packet loss rate, L due to maximum delay (expiration time) constraint is

$$L = N\,(1 - P_s) \quad . \tag{9}$$

The loss rate may also be expressed as the percentage of the new arrivals that are lost (due to expiration time constraints), that is, $(L/N) = 100\,(1 - P_s)$ (%).

The normalized throughput, S defined as the average number of successful packets per slot, is

$$S = p\,G = G \exp(-G) \tag{10}$$

from (2). This is also expressed as the product of the normalized new packet arrival rate, N (packets per slot) and the probability of a new packet success, P_s. That is,

$$S = N\,P_s \quad . \tag{11}$$

The total delay, Z_m until the beginning of m th retransmission has a mean of $\left(X+\frac{1}{A}\right)$ m. The probability of a packet success in its m th retransmission is $P_r(m)$. Therefore, the average delay for a packet success, D (slots) is

$$D = 1 + \frac{1}{P_s} \sum_{m=1}^{M} P_r(m) \left(X+\frac{1}{A}\right) m \quad . \tag{12}$$

The performance curves are obtained with G as independent variable, and computing the P_s, L, S, N, and D using the equations given above.

2.2.2. Analysis for adaptive retransmission delay

In the low offered traffic region, the packet throughput increases with an increase in the offered traffic. But beyond a certain value of the offered traffic (close to the maximum channel throughput region), the channel becomes congested and the increased collisions cause a further drop in the throughput. It is desirable to operate the system close to the maximum channel throughput region -

but the chances of the system drifting into unstable region increases. One way to avoid the low throughput high delay operation is to increase the retransmission delay. However, this technique increases the delay even when the channel is lightly loaded. Hence, adaptive schemes (back-off) have been proposed, where the mean retransmission delay for k th retransmission, a_k is adaptively changed [ME76]. The performance expressions derived here are applicable to a variety of adaptive retransmission delay (definition of a_k's) strategies.

In the numerical evaluations, we computed the performance of adaptively changing the retransmission delays, a_k in two ways:

$$a_k = A/(B(k-1)+1)$$

and $$a_k = A/b^{(k-1)} \qquad (13)$$

where $\frac{1}{A}$ is the mean retransmission delay for the first retransmission (k=1); and B, b are constants. The latter (13) gave better performance for a wide range of offered traffic.

The procedure used to obtain the analytical expressions for steady state system performance with fixed retransmission delay is applicable here also. Therefore, only the few exceptions, for which new expressions are derived, are discussed below.

The total delay, Z_i until the beginning of the i th retransmission is given in (3). Note that the random variables R_k have a mean of $(1/a_k)$. The probability that the total delay, Z_i is less than or equal to the expiration time , E is given by

$$P_z(i) = P\{Z_i \le E\} = P\{Y_i \le C(i)\} \quad ,$$

where C(i) is (E - iX) and $Y_i = \sum_{k=1}^{i} R_k$. The expression for the probability, $P_z(m)$ is obtained using the characteristic functions of probability distributions and partial fractions. Details are included in the appendix. The result is that $P_z(m)$ can be written as a sum of products as shown below.

$$P_z(m) = \sum_{n=1}^{m}(1-\exp(-a_n C(m))) \prod_{\substack{k=1 \\ k\neq n}}^{m} \frac{a_k}{(a_k-a_n)} \quad . \qquad (14)$$

The probability of success in the m th retransmission, $P_R(m)$ is given by (5), where $P_c(m)$ is replaced by $P_z(m)$. Therefore,

$$P_R(m) = p\,(q^m)\,P_z(m)\ . \tag{15}$$

The probability of a packet success, P_s is the probability that it succeeds in the first transmission or it succeeds in the 1st, 2nd, etc. up to M retransmissions.

$$P_s = p + \sum_{m=1}^{M} P_R(m) = p\left\{1 + \sum_{m=1}^{M} (q^m)\,P_z(m)\right\}\ . \tag{16}$$

The total delay, Z_m until the beginning of m th retransmission has a mean of $\left(mX+\sum_{k=1}^{m} \frac{1}{a_k}\right)$. The probability of a packet success in its m th retransmission is $P_R(m)$. Therefore, the average delay for a packet success, D (slots) is

$$D = 1 + \frac{1}{P_s} \sum_{m=1}^{M} P_R(m) \left(mX+\sum_{k=1}^{m} \frac{1}{a_k}\right)\ . \tag{17}$$

The normalized throughput, S, the normalized new packet arrival rate, N and the normalized packet loss rate, L are given by equations (10), (11) and (9), respectively.

The performance curves may be obtained with G as independent variable and computing P_S, D, S, N and L, as described above.

2.3. Mixed traffic scenario

We consider a generalized traffic environment where there are two types (j=1,2) of traffic input to the system, each with its own arrival rates and expiration times. The system design parameters, such as the mean or adaptive retransmission delay parameters, can be different for these two types of traffic. We assume that the new packet transmissions are Poisson distributed, with a normalized packet origination rates of N_1 and N_2 packets per slot (fig. 2), for the type 1 and type 2 traffic, respectively. The total offered traffic on the multiple access channel, which is the sum of the new and retransmitted packets, is also assumed to be Poisson distributed, with normalized offered traffic rates, G_1 and G_2, packets per slot duration.

Since the sum of two Poisson distributed random variables is Poisson distributed, the total normalized input traffic, N and the total normalized offered traffic, G are also Poisson distributed. The ratio of type 1 packet origination rate to the total new packet origination rate is denoted by α ,

$$\alpha = (N_1/N) \tag{18}$$

Therefore, $N_2 = (1-\alpha)N$. The total offered traffic, G is equal to $(G_1 + G_2)$.

The average number of packets of each type, that leave the multiple access environment after successful transmission to the central destination, are denoted by the normalized channel throughputs, S_1 and S_2 packets per slot. For real-time traffic with maximum time delay constraints, the packets of each type that exit the system (due to maximum delay constraints) before successful transmission to the central destination, are denoted by the normalized packet loss rates, L_1 and L_2 packets per slot. Therefore, in steady state,

$$N_1 = S_1 + L_1; \tag{19}$$

and

$$N_2 = S_2 + L_2 \quad .$$

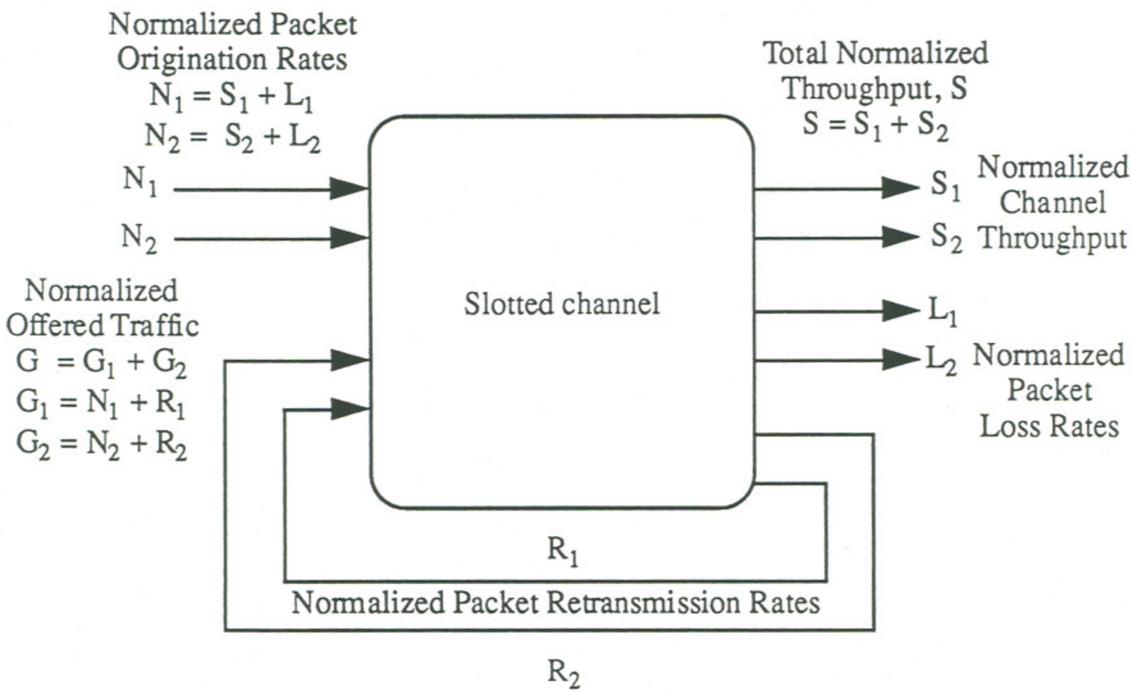

Figure 2: Mixed Traffic - Steady State System Parameters

As before, a (new or rescheduled) packet arrival can occur at any time within the slot duration, it is considered to have arrived at the end of the slot interval for delay calculations. The elapsed time from the new packet arrival time to the completion of its successful transmission (which can involve many retransmissions) is taken to be the total time delay for successful transmission of this packet. The normalized average packet delay variables, D_1 and D_2 denote the average delay for packet success, in slots.

The maximum delays (expiration times) are E_1 and E_2 (slots) for the two types of traffic.

2.3.1. Analysis for fixed mean retransmission delay

Let us assume that the average retransmission delays, T_1 and T_2 are fixed for both types of traffic, given by $T_1 = (1/A_1)$ and $T_2 = (1/A_2)$. The steady-state system performance expressions are given below.

The probability of a packet success in a given slot, p is the probability that there is no other simultaneous packet transmission in that slot. From [KL73],

$$p = \exp(-G) \quad .$$

The corresponding packet failure probability, q is (1 - p).

The probability of a packet success for the two types of traffic, $P_s(j)$, j=1 and 2, are derived using the same procedure that was employed for the uniform traffic scenario with fixed retransmission delay. The expressions for the success probabilities are:

$$P_s(j) = p + \sum_{m=1}^{M(j)} P_r(m,j) = p\left\{1 + \sum_{m=1}^{M(j)} (q^m)\, P_c(m,j)\right\}, \qquad j=1,2 \qquad (20)$$

where, M(j) is equal to the maximum integer $[E_j/X]$, X is the wait time for acknowledgement, $P_r(m,j)$ is the probability of success of a type j packet, in its m th retransmission and $P_c(m,j)$ is the probability that the total delay till the beginning of the m th retransmission of type j packet is less than or equal to its expiration time, E_j. The expressions for $P_r(m,j)$ and $P_c(m,j)$ are given by:

$$P_r(m,j) = p\,(q^m)\, P_c(m,j) \quad ; \qquad (21)$$

$$P_c(m,j) = 1 - \left(e^{-A_jC(m,j)}\right) \sum_{k=1}^{m} \frac{(A_jC(m,j))^{m-k}}{(m-k)!} \; ; \quad C(m,j) = E_j - mX \; ; \quad j=1,2. \quad (22)$$

The normalized packet loss rates, L_1 and L_2 for the two types of traffic are

$$L_1 = N_1\,(1 - P_s(1)) \quad \text{and} \quad L_2 = N_2\,(1 - P_s(2)) \quad . \qquad (23)$$

The loss rates may also be expressed as the percentage of the new arrival of the respective traffic types, that is, $(L_1/N_1) = 100\,(1 - P_s(1))\ (\%)$ and $(L_2/N_2) = 100\,(1 - P_s(2))\ (\%)$.

The normalized throughput for the two types of traffic, S_1 and S_2 are

$$S_1 = p\,G_1 = N_1\,P_s(1) \tag{24}$$

and $$S_2 = p\,G_2 = N_2\,P_s(2) \quad .$$

The total normalized throughput, S defined as the average number of successful packets per slot, is

$$S = p\,G = G\exp(-G) = S_1 + S_2 \ . \tag{25}$$

The average delay for a packet success, D_j (slots) of type j traffic is

$$D_j = 1 + \frac{1}{P_s(j)} \sum_{m=1}^{M(j)} P_r(m,j)\left(X + \frac{1}{A_j}\right) m \qquad ; \qquad j = 1\,,\,2. \tag{26}$$

Given the normalized packet origination rates, N_1 and N_2 the performance curves are obtained as follows:

From (24 and 25) we find that

$$pG = N_1\,P_s(1) + N_2\,P_s(2) \qquad .$$

Substituting the expressions for G, $P_s(1)$ and $P_s(2)$, and after some algebraic manipulation, we obtain

$$-\ln(1-q) = N_1 \left\{1 + \sum_{m=1}^{M(1)} (q^m)\,P_c(m,1)\right\} + N_2 \left\{1 + \sum_{m=1}^{M(2)} (q^m)\,P_c(m,2)\right\} \quad .$$

This equation is numerically solved for q, to obtain the steady state operation parameter, G (using equation (2)). At this value of G, the system performance expressions are evaluated to obtain the numerical results.

2.3.2. Analysis for adaptive retransmission delay

The analytical expressions for the the steady-state system performance for a mixed traffic scenario (two traffic types, j=1 and 2) with adaptive retransmission delays, are same as that for the mixed traffic scenario with fixed retransmission delay, with the following exceptions:

The probability, $P_z(m,j)$ that the total delay until the beginning of the m th retransmission is less than or equal to the expiration time, E_j of type j traffic is given by

$$P_z(m,j) = \sum_{n=1}^{m}\left(1-\exp(-a_{n,j}C(m,j))\right)\prod_{\substack{k=1\\k\neq n}}^{m}\left(\frac{(a_{k,j})}{(a_{k,j}-a_{n,j})}\right) \quad ; \qquad j=1, 2. \tag{27}$$

In numerical results, for packets of type j traffic, the mean retransmission delays for k th retransmission was computed using the adaptive retransmission expression:

$$\frac{1}{(a_{k,j})} = \frac{b_j^{(k-1)}}{A} \qquad j=1,2. \tag{28}$$

where $\frac{1}{A}$ is the mean retransmission delay for first retransmission of type j packets and b_j is a constant for type j traffic.

In the derivation of the performance expressions for the mixed traffic scenario with fixed retransmission delay, the probability, $P_c(m,j)$ is replaced by the above expression for $P_z(m,j)$, to obtain the system performance expressions for the mixed traffic scenario with adaptive retransmissions.

The probability of success of type j packet in the m th retransmission, $P_R(m,j)$ is given by (21), where $P_c(m,j)$ is replaced by $P_z(m,j)$. Therefore,

$$P_R(m,j) = p\,(q^m)\,P_z(m,j) \;, \qquad j = 1, 2. \tag{29}$$

Also, the probability of a type j (j=1,2) traffic packet success is, $P_s(j)$

$$P_s(j) = p + \sum_{m=1}^{M(j)} P_R(m,j) = p\left\{1 + \sum_{m=1}^{M(j)} (q^m)\,P_z(m,j)\right\} , \qquad j=1,2\;. \tag{30}$$

The total delay until the beginning of m th retransmission has a mean of $\left(mX+\sum_{k=1}^{m}\frac{1}{(a_{k,j})}\right)$. The probability of a type j packet success in its m th retransmission is $P_R(m,j)$. Therefore, the average delay for a packet success, D_j (slots) is

$$D_j = 1 + \frac{1}{P_s(j)}\sum_{m=1}^{M(j)} P_R(m,j)\left(mX+\sum_{k=1}^{m}\frac{1}{(a_{k,j})}\right) \quad ; \quad j=1\,,2\,. \tag{31}$$

Given the normalized packet origination rates, N_1 and N_2 the performance curves are obtained as follows:

From (24) and (25) we find that

$$pG = N_1 P_s(1) + N_2 P_s(2) \quad .$$

Substituting the expressions for G, $P_s(1)$ and $P_s(2)$, and after some algebraic manipulation, we obtain

$$-\ln(1-q) = N_1 \left\{ 1 + \sum_{m=1}^{M(1)} (q^m)\, P_z(m,1) \right\} + N_2 \left\{ 1 + \sum_{m=1}^{M(2)} (q^m)\, P_z(m,2) \right\}$$

This equation is numerically solved for q, to give the steady state operation parameter, G (using equation (2)). At this value of offered traffic, G the system performance expressions are evaluated to obtain the numerical results.

III. DISCUSSION OF NUMERICAL RESULTS

In the numerical results presented here, the channel rate is assumed to be 200 Kbps. For a packet size of 1000 bits, the packet transmission time or slot duration is 5 ms. The time delay for acknowledgement, X is taken to be 3 slots (15 ms). This includes one packet transmission time slot and two time slots for the round trip delay (up to 1000 Km for LEOS) and the processing delays at the satellite/ base station. The results do not change significantly for X equal to 2 or 3, corresponding to terrestrial or satellite application.

For a uniform traffic environment where all the incoming packets have the same maximum delay constraint of E equal to 50 slots (250 ms), the normalized channel throughput, S (packets per slot) versus the normalized packet origination rate, N (packets per slot) is shown in fig. 3 for the average retransmission delay, T equal to 3, 5, 10 and 20 slots. As the average retransmission delay, T is increased, the throughput, S becomes less than that given by S=N straight line, in the peak throughput region. (The deviation gives the normalized packet loss rate, L in packets per slot). While the higher T can reduce the congestion (especially during peak traffic) and increase the packet success rate, the increased retransmission delay allows fewer attempts to be made on the channel before a packet reaches its expiration time, E. Together, these effects result in the overall reduction in the throughput.

The operating point (and system parameters such as mean retransmission delay, T) on the throughput curve should be such that the system can tolerate small fluctuations in the incoming traffic. With a small T, the system throughput follows S=N but it is also more likely to enter the low throughput/ high delay region during peak throughput operation. In order to have more stable high throughput condition with low delay during light traffic, the mean retransmission delay is adaptively increased (back-off) as the traffic, N increases (actually the increase in N results in more packet failures due to collisions and the retransmission delay is increased by the concerned user with each failure). We considered linear as well as exponential increase in mean retransmission delay of a packet (back-off, equation 13). The exponential back-off gave better performance compared to the linear back-off case for the system parameters chosen here. The throughput curve with adaptive retransmission parameter b equal to 2 (fig. 3) and initial retransmission delay (=1/A) of 3 slots, has nearly the same throughput compared to a constant T equal to 10. However, the corresponding packet loss rates and packet delays are less over a wide range of input traffic.

The loss rate (as %) versus the packet origination rate is shown in fig. 4. As the input traffic, N is increased, the packet loss rate increases due to the increased congestion on the channel. Depending on the robustness of the voice codec, losses of up to 1% of the packets may be tolerated. The normalized throughput, S is 0.36, 0.26 and 0.07, respectively, for T equal to 3, 5 and 10 slots. The adaptive retransmission with A=1/3 and b=2.0 gives a throughput of 0.27 at 1% loss. In order to further decrease the packet loss rate, one has to either increase the packet retransmission rate (reduce T) or operate at a lower throughput.

The average packet delay, D for the successful packets versus normalized packet origination rate, N is shown in fig. 5. As expected, the average packet delay increases with the retransmission delay, T for normal operation (origination rate at or below that corresponding to the maximum throughput). Further increase (even small fluctuation) will drive the T equal to 3 operation into the high delay region. Beyond that the delay curve flattens as the packet loss (due to expiration time constraint) increases. The adaptive retransmission curve has nearly the same packet delay as the T equal to 3 curve throughout the operating region (less than 5 slots or 25 ms for S equal to 0.27) and has relatively better delay performance beyond the maximum throughput region. Hence, the adaptive retransmission scheme is expected to result in a more stable operation.

The discussion so far can be related to most real-time traffic which typically have a maximum delay constraint of 250 ms. The numerical results are useful for the selection of the system parameters in a mixed traffic environment.

The performance curves were repeated for uniform traffic environment where the expiration time, E was chosen to be 200 slots (1 sec). The relaxed delay constraint can be considered to represent the non-real time traffic, as the interactive users are getting accustomed to faster modems and higher transmission rates for the inter-computer communications.

The normalized throughput, S versus normalized origination rate, N is shown in fig. 6 for various mean retransmission delay, T . It should be noted that the mean retransmission delay, T had to be increased to 20 slots, in order to achieve the maximum (possible slotted Aloha) normalized throughput of 0.36. The adaptive retransmission employed has a initial (k=1) mean retransmission delay (=1/A) of 3 slots and b equal to 2.5, in order to achieve the stability and throughput performance better than the T equal to 35 curve. The corresponding average packet delay, D (fig. 7) and average packet loss, L (fig. 8) are better than those for T in the range of 20 to 50. Of course, for a truly non-real time traffic that can tolerate higher delays, the back-off parameter b has to be increased further (as we will see in the mixed traffic scenario) in order to improve stability. Theoretically, the packet loss rate is zero for non-real time traffic; however, due to the large delay the users will be discouraged from using the system.

For a general mixed traffic environment consisting of two types of traffic, the performance curves are shown in fig. 9 to 11. The packet expiration time for type 1 traffic is E_1 equal to 50 slots (250 ms), representing the real-time traffic with maximum delay requirement. The type 2 traffic has a much relaxed time delay constraint of E_2 equal to 200 slots (1 sec), some what like interactive data where the users have come to expect faster response time. Based on the earlier results for uniform traffic, the fixed mean retransmission delay for type 1 and type 2 traffic are T_1=10 and T_2 =35, in order to have a relatively stable operating region with a normalized throughput greater than 0.3. It is assumed that the total normalized origination rate, N consists of N_1 (=αN) of type 1 traffic and the rest is type 2 traffic. The normalized throughputs, S, S_1 and S_2 versus the normalized origination rate, N is shown in fig. 9 for α equal to 0.1 and 0.3. There is no significant change in performance with 10 or 30 % of type 1 traffic. In order to achieve the same throughput performance, but better delay and packet loss characteristics, the adaptive retransmission rate parameters were chosen to be A=1/3 and b_1 = 2.0, b_2 = 2.5. The packet loss rates (fig. 10) for adaptive retransmission is nearly half the loss rate of the fixed retransmission scheme in the high throughput region. The average packet delay, particularly for type 2 traffic, decreases to less than 50% of that obtained by the fixed retransmission scheme, over a wide range of input traffic (N between 0.2 and 0.4).

The packet delay, D_1 of about 10 slots (50 ms) is acceptable for real-time traffic, but the packet loss rates for a satisfactory operation of say, a vocoder should be below 1%. For good reconstructed signal quality, the loss rate is required to be 0.1% or less. In a mixed traffic environment, where part of the incoming traffic has expiration time constraints and the remaining traffic has practically no time constraint, the real-time traffic can be given priority by suitably reducing its retransmission delay. The remaining traffic has to have adaptive retransmission (backoff) so that a stable operating point may be chosen close to its peak throughput region. Such a selection of parameters with E_1=50 slots, E_2=1500 slots, b_1=1.1 and b_2=3.5 is considered, for traffic mix consisting of 30% real-time data. The large expiration time, E_2 (7.5 sec) makes it a close approximation to non-real time traffic (higher expiration time values could not be considered due to computation accuracy constraints). Both the throughput curves, S_1 and S_2 (fig. 12) have a near constant slope for a wide range of input traffic. The packet delay, D_1 is less than 10 slots (50 ms) for normalized input traffic, N up to 0.36 (fig. 13).

The packet loss rate (fig. 14) is about 0.1% for a normalized packet origination rate of 0.3 or less. Because of the stringent loss constraints, the retransmission delay can not be increased beyond a certain value and hence, the throughput curve has a narrow peak compared to the results presented for a generalized mixed traffic (fig. 9). But, it is apparent that a careful selection of system parameters will help achieve acceptably low loss rates in the high throughput region.

IV. CONCLUSION

It is expected that the demand for additional capacity for voice calls will drive the evolution of the next generation digital cellular system. With the flexibility of digital transmission and the growing interest in personal communications services, the future PCN systems will be required to provide support for a variety of services, including digitized voice, interactive data and personal computer based multi-media. The resulting mixed traffic environment can contain real-time traffic, with its associated maximum delay (expiration time) constraints. This can result in the loss of transmitted packets. The lost packets can be compensated to a limited extent by error concealment procedures.

In this paper, we have considered a simple medium access control technique for a multiple access channel in a mixed traffic environment. The real-time traffic can make up the whole or part of the offered traffic on the channel. The remaining traffic is considered to have relatively less stringent maximum delay constraints. First, the slotted Aloha protocol is analyzed for a uniform traffic where all packets have the same maximum delay constraint, and the steady state system performance expressions for the slotted channel throughput, average packet delay and average packet loss rate are derived. The numerical examples show that the average retransmission delay must be higher than a certain minimum value in order to have a stable operation close to the maximum possible slotted Aloha throughput of 0.36. The adaptive retransmission schemes (back-off) are analyzed and numerical results for the improvement in performance are obtained for (a) real-time traffic constraints and (b) much relaxed time delay constraints. These results are used to select the parameters for the mixed traffic environment.

A generalized case consisting of two distinct classes of traffic, each with its own maximum delay time constraints, is analyzed for fixed and adaptive retransmission strategies. The performance curves show that the system parameters can be chosen to provide high channel efficiency, low average delay and low average packet loss rates. The performance curves do not change significantly with the actual mix (%) of real-time and non-real time traffic.

For the special case, of a mixture of real-time and non-real time traffic, the numerical results are obtained for a choice of system parameters such that the packet loss rate is less than 0.1%. This leads to a relatively sharper peak for the throughput versus new packet origination rate curve, indicating a possible problem with the stability of the system even for small fluctuations in input traffic (near high throughput region). Compared to the slotted Aloha developed for computer communications, the system with a mixed traffic is likely to recover from the low throughput high delay undesirable region, as packets expire at a higher rate (causing temporary channel outage) with the increased congestion on the channel. The access control techniques discussed are applicable to a variety of communications applications, including low earth orbit satellite communication systems.

V. REFERENCES

[BO81] S. Bose and S. S. Rappaport, "Demand Assigned Multiple Access Systems Using Collision Type Request Channels: Stability and Delay Considerations," IEE (British) Proc. on Computers and Digital Techniques, January 1981, vol. 128, Part E, no. 1, pp. 37-43.

[CA89] C. G. Cassandras, M. H. Kallmes, D. Towsley, "Optimal routing and flow control in networks with real-time traffic," IEEE Infocom 1989, pp. 784-790.

[CH89] R. Chipalkatti, J. F. Kurose, D. Towsley, "Scheduling policies for real-time and non-real-time traffic in a statistical multiplexer," IEEE Infocom 89, pp. 774-783.

[CO87] D. C. Cox, "Universal portable radio communication," Proc. IEEE, April 1987, pp. 436-477.

[CO90] D. C. Cox, "Personal communications - A viewpoint," IEEE Communications Magazine, Nov. 1990, pp. 8-20.

[CO91] *IEEE Communications Magazine*, February 1991.

[GO91] D. J. Goodman, "Trends in cellular and cordless communications," IEEE Commn. Mag., vol. 24, no. 6, June 1991, pp. 31-41.

[GR85] D. Gross and C. M. Harris, *Fundamentals of Queuing Theory*, II edition, 1985, pp. 170-172.

[KL73] L. Kleinrock and S. S. Lam, "Packet-switching in a slotted satellite channel," AFIPS Conf. Proc., National Computer Conference, vol. 42, 1973, pp. 703-10.

[ME76] R. M. Metcalfe, D. R. Boggs, "Ethernet: Distributed packet switching for local computer networks," Commun. of ACN, 1976, pp. 395-404.

[PA65] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 1965, pp. 159.

[RA79] S. S. Rappaport, "Demand Assigned Multiple Access Systems Using Collision Type Request Channels: Traffic Capacity Comparisons," IEEE Trans. on Commn., Sept. 1979, vol. COM-27, no. 9, pp. 1325-1331.

[RA81] D. Raychaudhuri, "Performance analysis of random access packet switched Code Division Multiple Access systems," IEEE Trans. on Commun., June 1981, pp. 1895-1901.

[WA91] Y. Wang and Q. -F. Zhu, "Signal loss recovery in DCT-based image and video codecs," Proc. of SPIE on Visual Commn. and Image Processing, Boston, Nov. 91, pp. 667-678

[WI85] N. D. Wilson and S. S. Rappaport, "Cellular Mobile Packet Radio Using Multiple Channel CSMA," IEE (British) Proc., Part F, Communications, Radar and Signal Processing, Oct. 1985, vol. 132, no. 6, pp. 517-526.

[WI92] N. Wilson, R. Ganesh, K. Joseph and D. Raychaudhuri, "CDMA versus Dynamic TDMA for access control in an integrated voice/data PCN," Proc. of the 1st Int'l. Conference on Universal Personal Communications, Dallas, Sept. 1992.

Acknowledgement

The authors acknowledge helpful discussions with Newman Wilson in the preparation of this manuscript.

Appendix

The probability distribution function for the sum of several independent exponential random variables is obtained here. The results are used to derive the performance expressions for systems with fixed or adaptive retransmission delay strategies.

Let R_i be independent exponentially distributed random variable. The probability density function, $f_{R_i}(r_i)$ of R_i with a mean of $\frac{1}{a_i}$ is

$$f_{R_i}(r_i) = a_i \exp(-a_i r_i) \quad . \tag{A1}$$

Let $\Phi_{R_i}(\omega)$ be the characteristic function of the exponential random variables, R_i. Then [PA65],

$$\Phi_{R_i}(\omega) = \int_0^{\infty} a_i \exp(-a_i r_i) \exp(jwr_i)\, dr_i = \frac{a_i}{a_i - jw} \quad . \tag{A2}$$

Let Y_m be the sum of m independent exponential random variables, R_i. That is, $Y_m = \sum_{i=1}^{m} R_i$.

Then, the characteristic function, $\Phi_{Y_m}(\omega)$ is

$$\Phi_{Y_m}(\omega) = \prod_{i=1}^{m} \Phi_{R_i}(\omega) \quad . \tag{A3}$$

A.1. Fixed mean retransmission delay

Here, all the R_i's are i.i.d. exponential random variables with the same mean, that is, $\frac{1}{a_i} = \frac{1}{A}$.

From (A2) and (A3), the characteristic function of the sum random variable, Y_m is

$$\Phi_{Y_m}(\omega) = \left(\frac{A}{A-jw}\right)^m \quad . \tag{A4}$$

The probability distribution of Y_m is (using tables of transforms in [GR85])

$$f_{Y_m}(y_m) = \frac{A^m}{(m-1)!}\, y_m^{\,m-1}\, e^{-Ay_m} \qquad (0 < y_m < \infty) \ . \tag{A5}$$

A. 2. Adaptive retransmission delay

For the general case where R_i's are independent exponential random variables, with their respective means, $\frac{1}{a_i}$ the characteristic function of Y_m is

$$\Phi_{Y_m}(\omega) = \prod_{k=1}^{m} \frac{a_k}{(a_k\text{-}jw)} \quad . \tag{A6}$$

The expression for $\Phi_{Y_m}(\omega)$ can be split into partial fractions resulting in

$$\Phi_{Y_m}(\omega) = \sum_{n=1}^{m} \frac{H_n}{(a_n\text{-}jw)} \quad , \tag{A7}$$

where

$$H_n = a_n \prod_{\substack{k=1 \\ k\neq n}}^{m} \left(\frac{a_k}{a_k\text{-}a_n}\right) \quad .$$

Using the result in (A2), the probability distribution of Y_m is

$$f_{Y_m}(y_m) = \sum_{n=1}^{m} H_n \exp(-a_n\, y_m) = \sum_{n=1}^{m} a_n \exp(-a_n\, y_m) \prod_{\substack{k=1 \\ k\neq n}}^{m} \left(\frac{a_k}{a_k\text{-}a_n}\right) \quad , \tag{A8}$$

and

$$\int_0^{C(m)} f_{Y_m}(y)\, dy = \sum_{n=1}^{m} (1 - \exp(-a_n C(m))) \prod_{\substack{k=1 \\ k\neq n}}^{m} \left(\frac{a_k}{a_k\text{-}a_n}\right) \quad . \tag{A9}$$

Fig. 3: Uniform traffic - fixed & adaptive retransmission delays

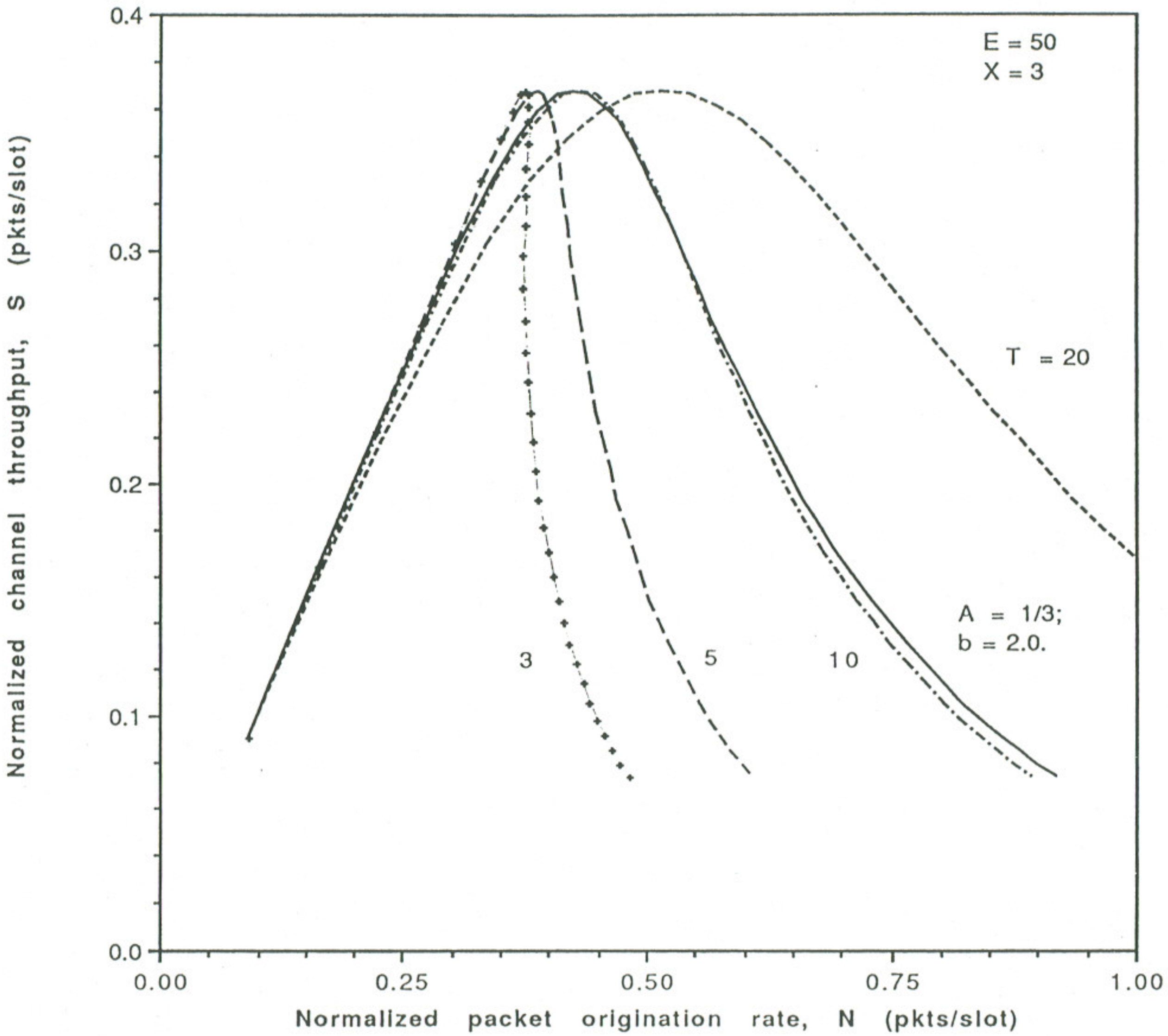

Fig. 4: Uniform traffic - fixed & adaptive retransmission delays

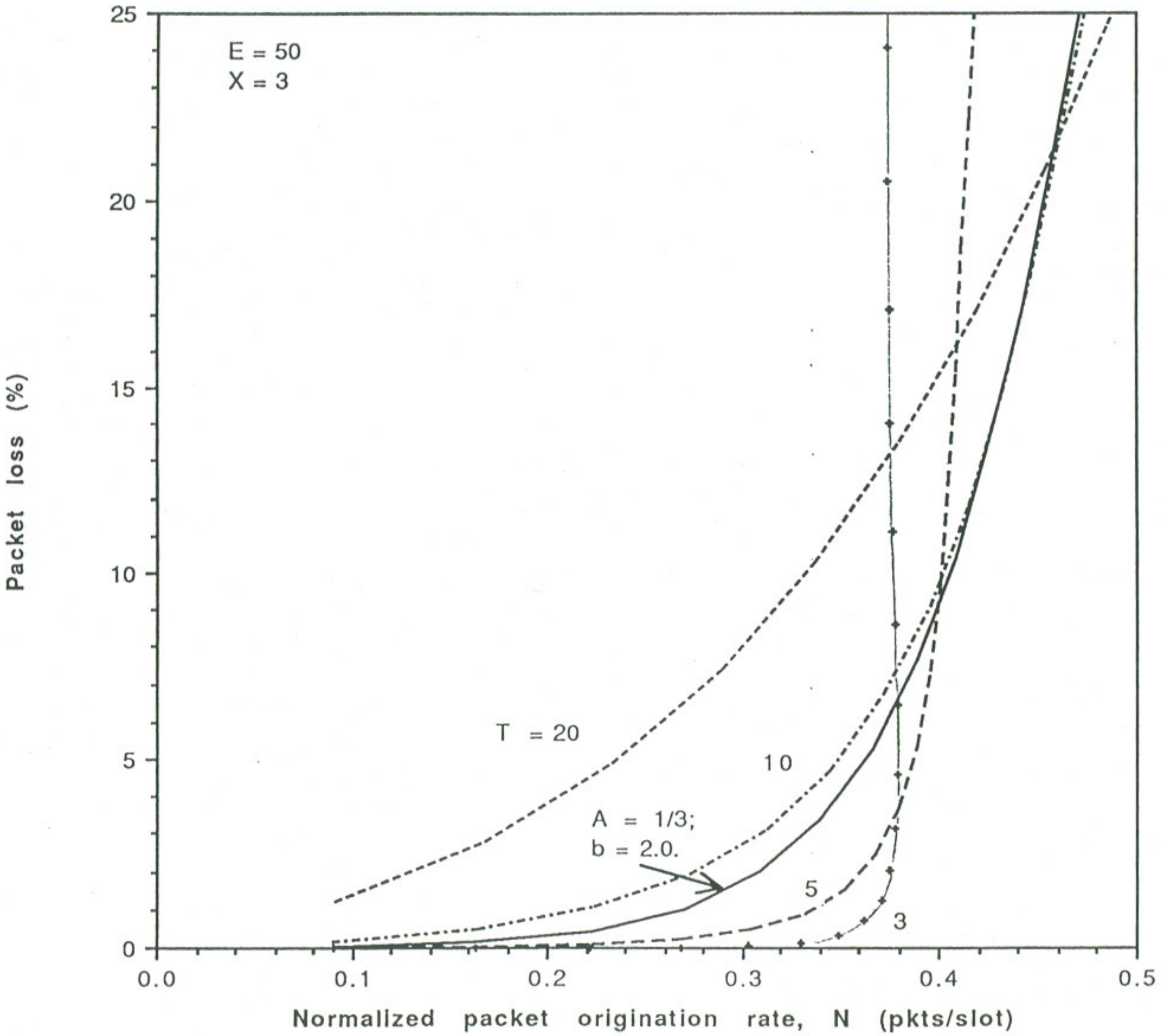

Fig. 5: Uniform traffic - fixed & adaptive retransmission delays

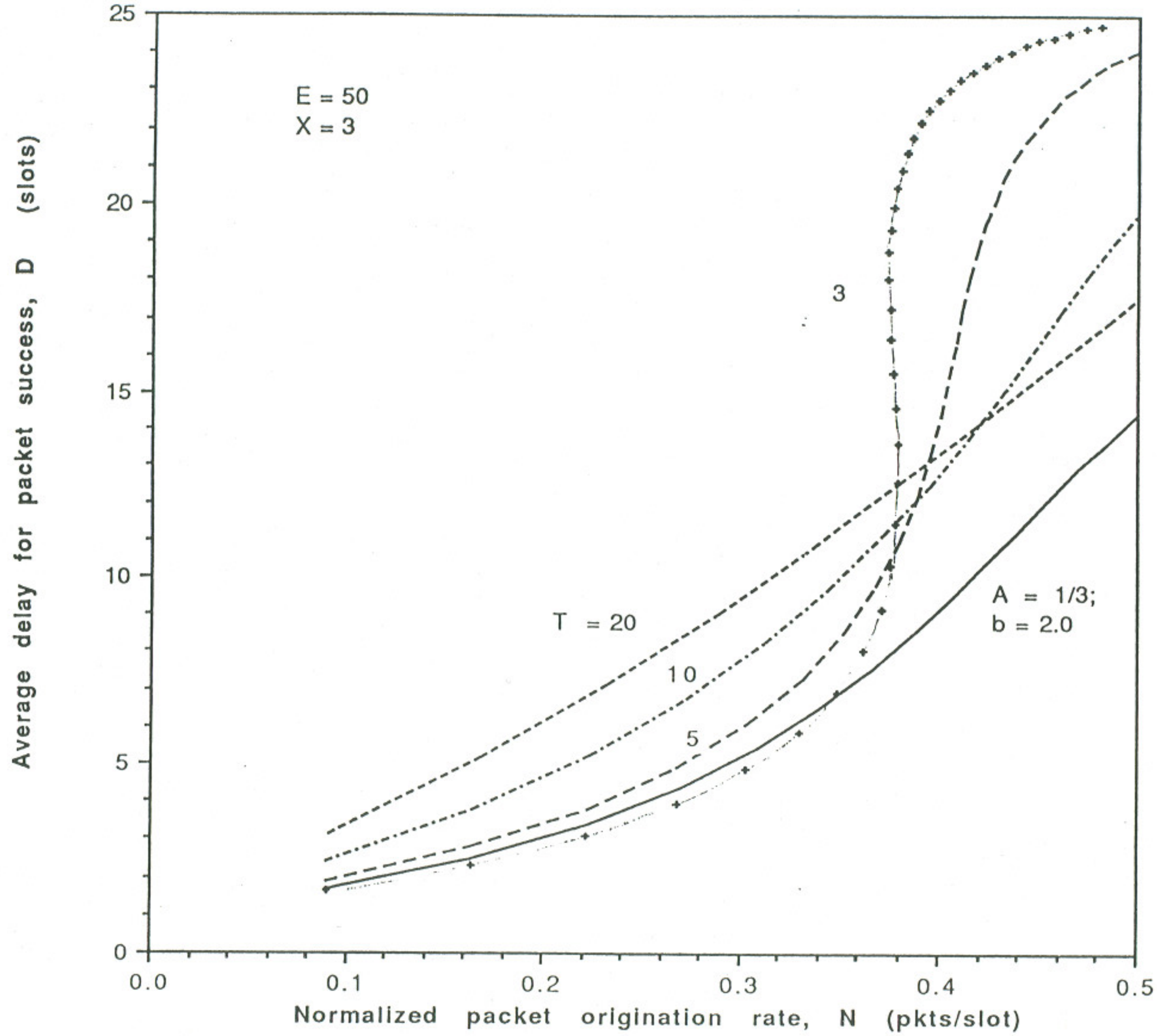

Fig. 6: Uniform traffic - fixed & adaptive retransmission delays

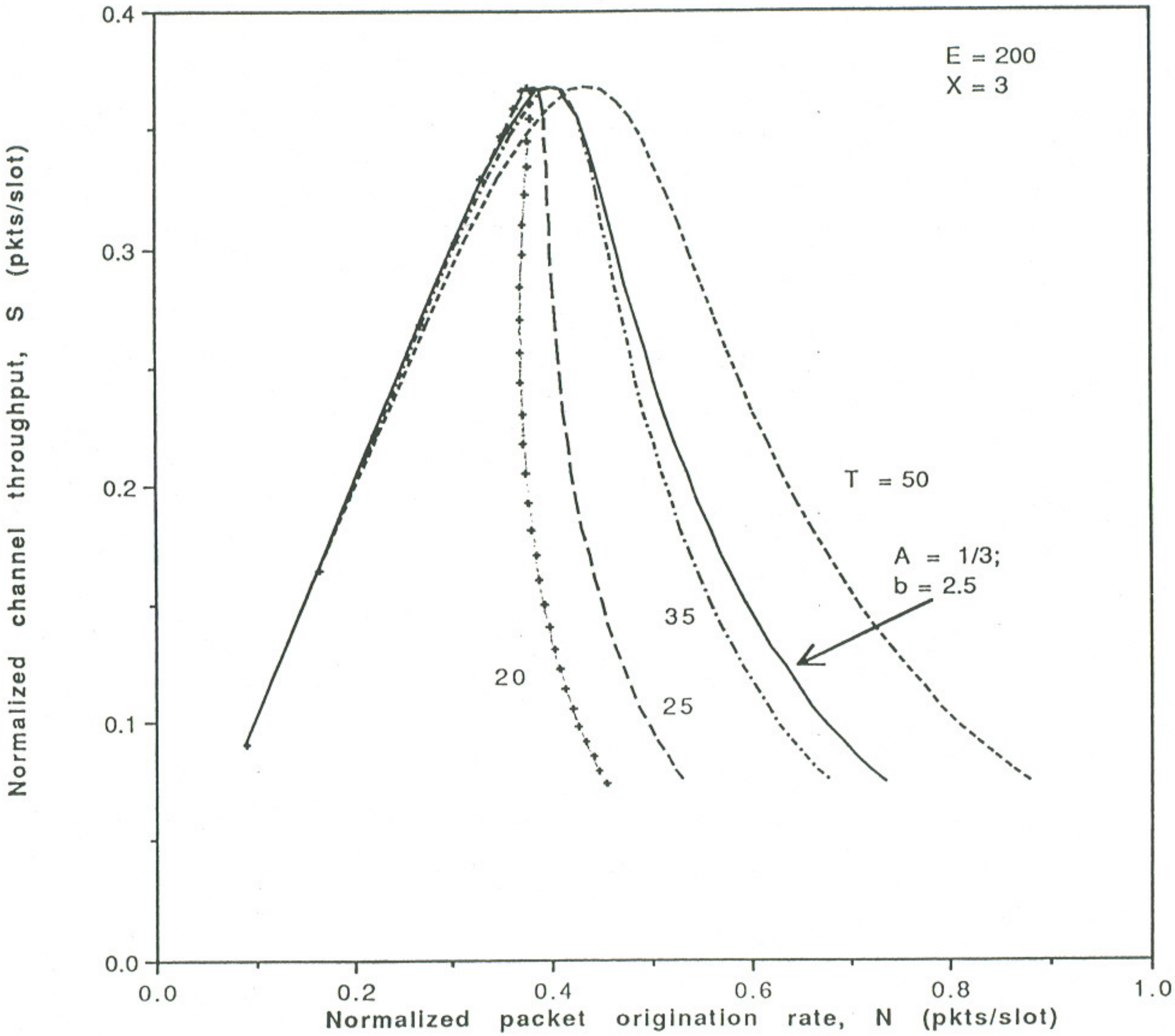

Fig. 7: Uniform traffic - fixed & adaptive retransmission delays

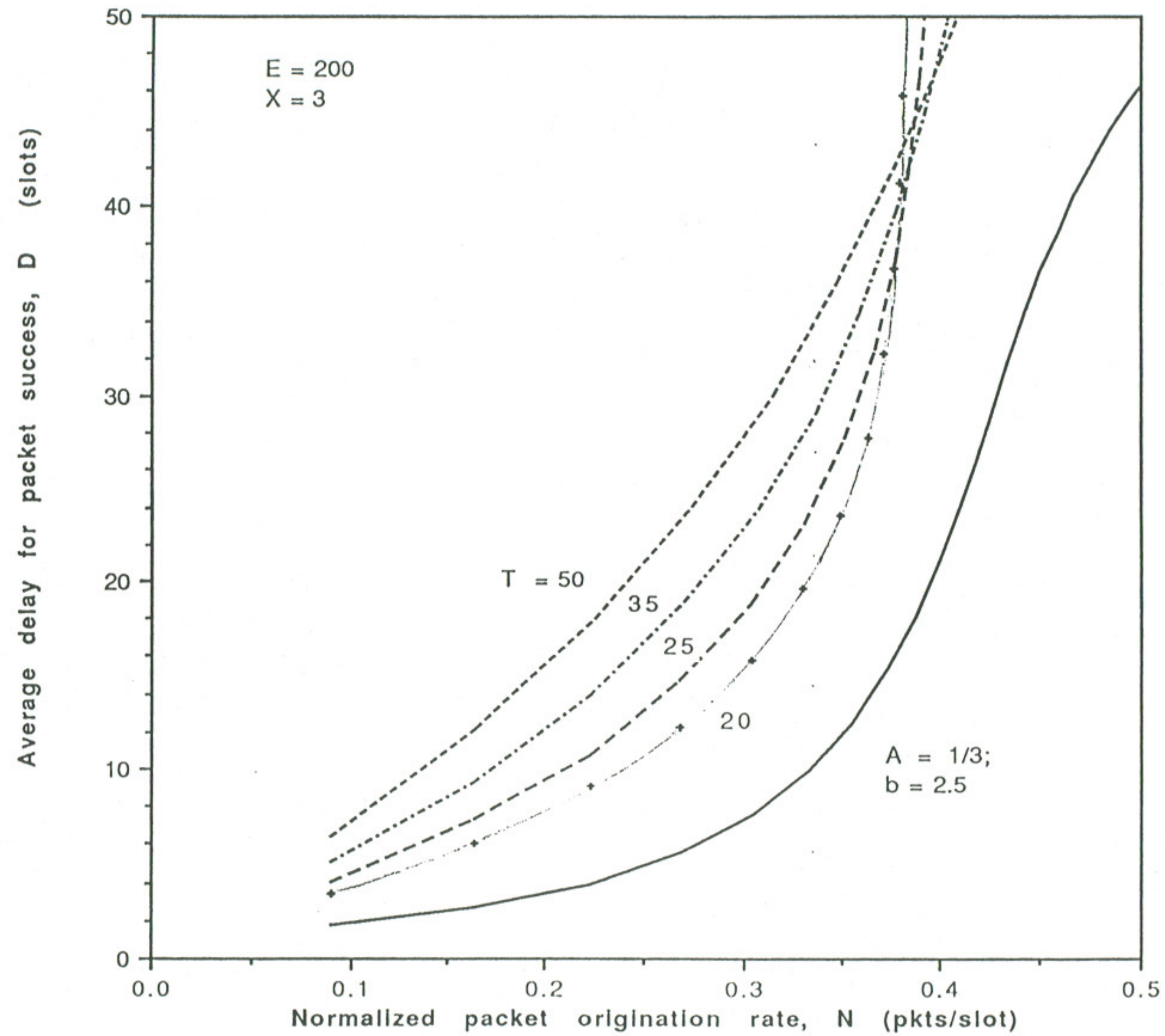

Fig. 8: Uniform traffic - fixed & adaptive retransmission delays

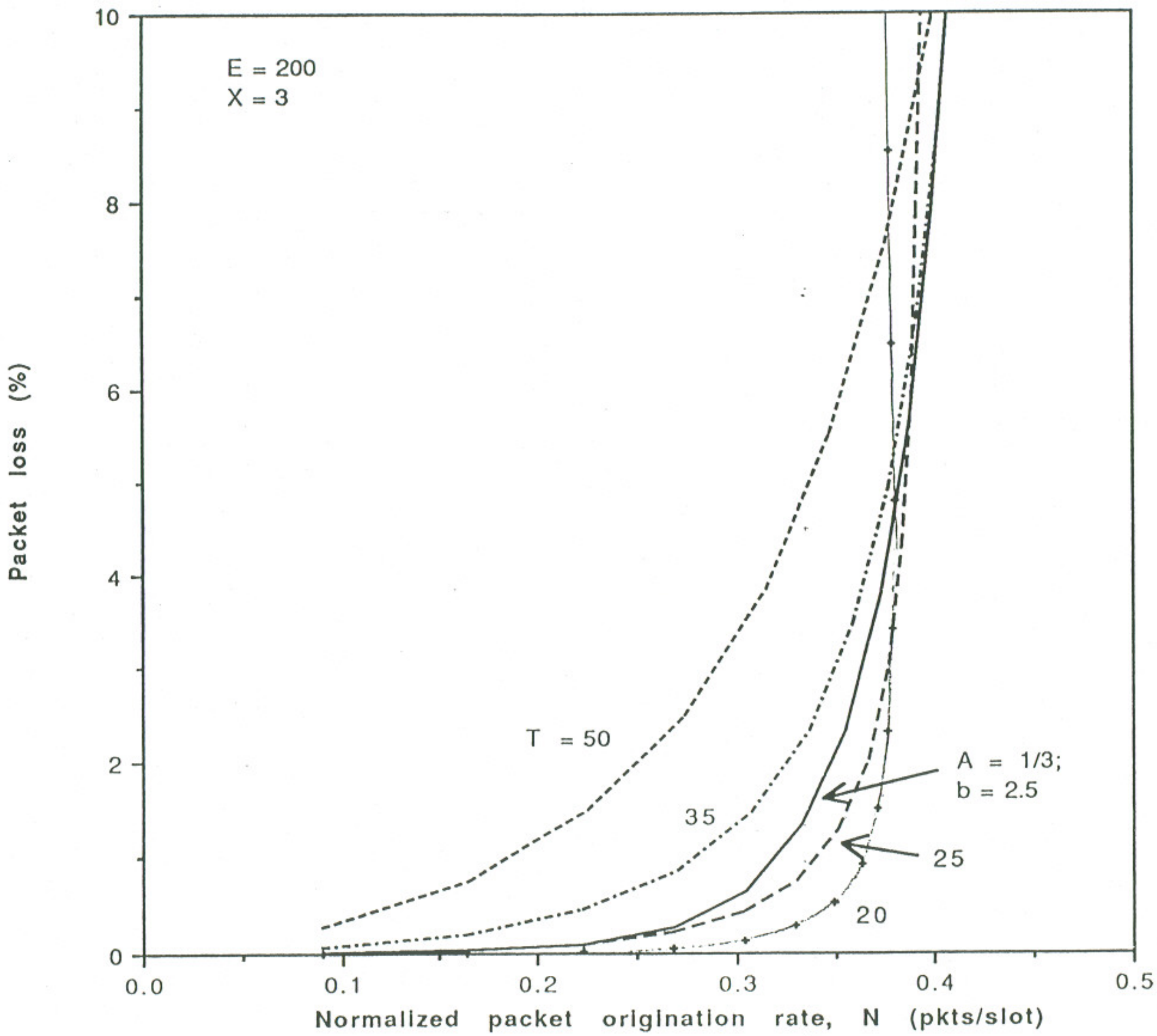

Fig. 9: Mixed traffic - fixed & adaptive retransmission delays

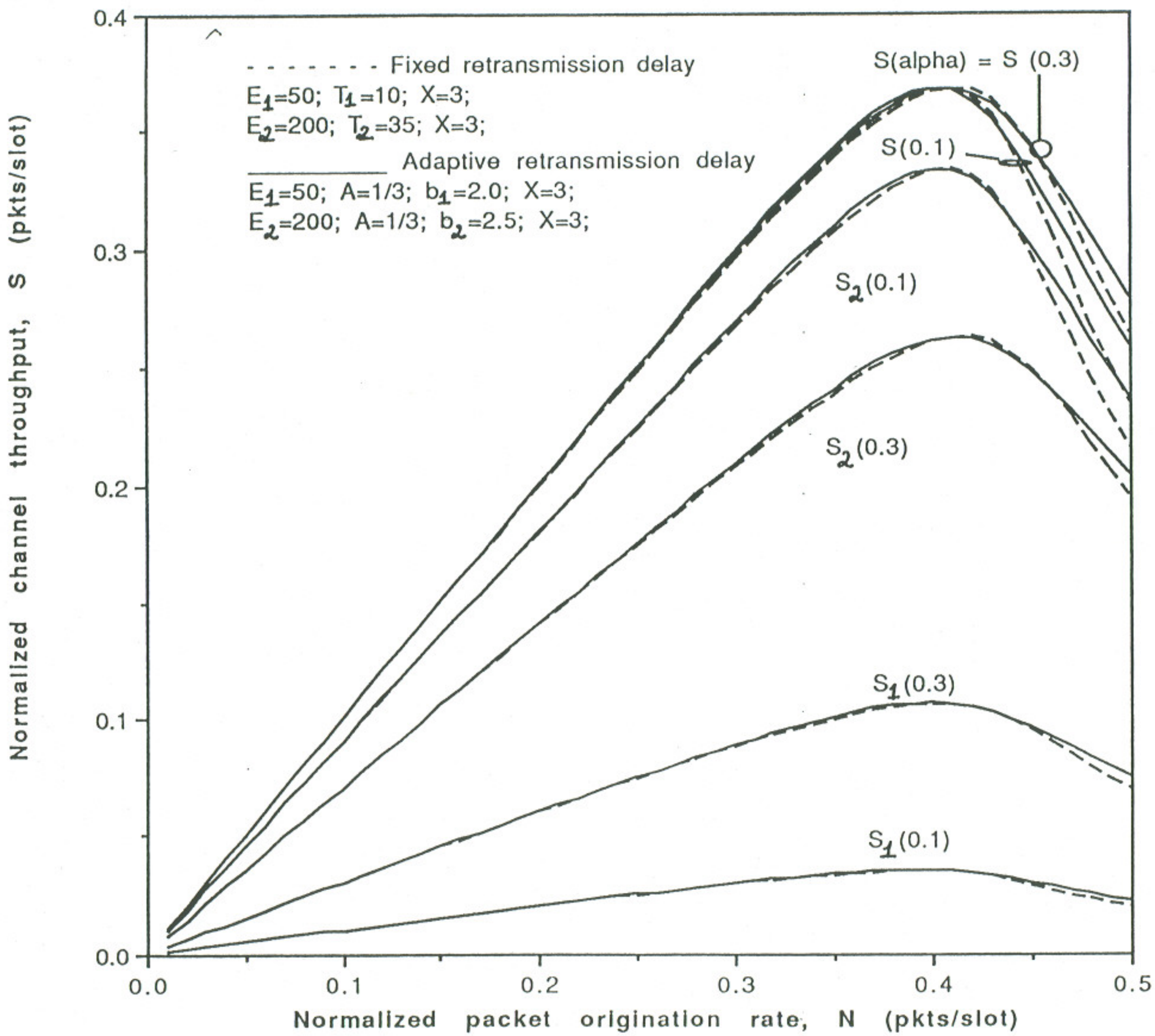

Fig. 10: Mixed traffic - fixed & adaptive retransmission delays

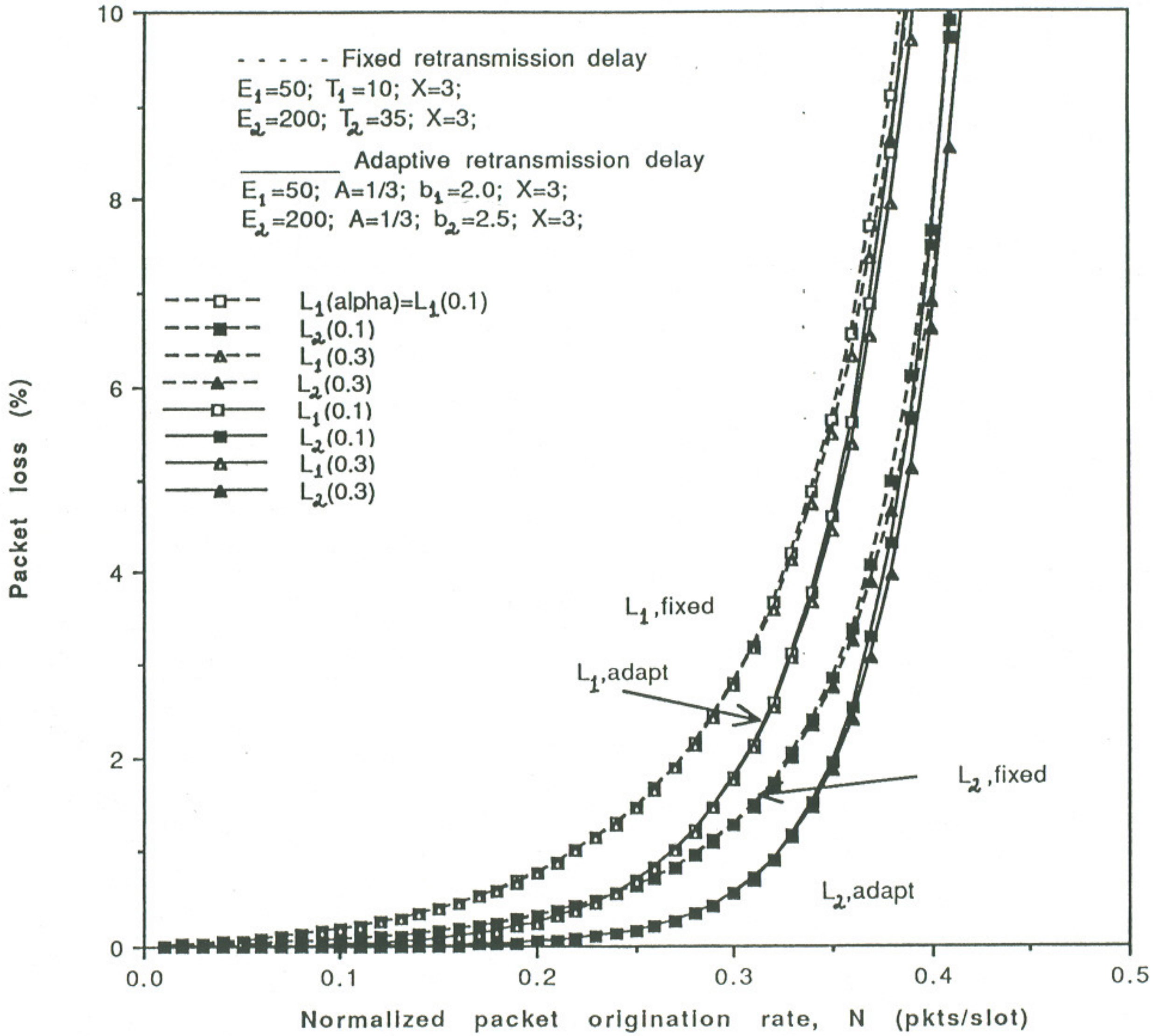

Fig. 11: Mixed traffic - fixed & adaptive retransmission delays

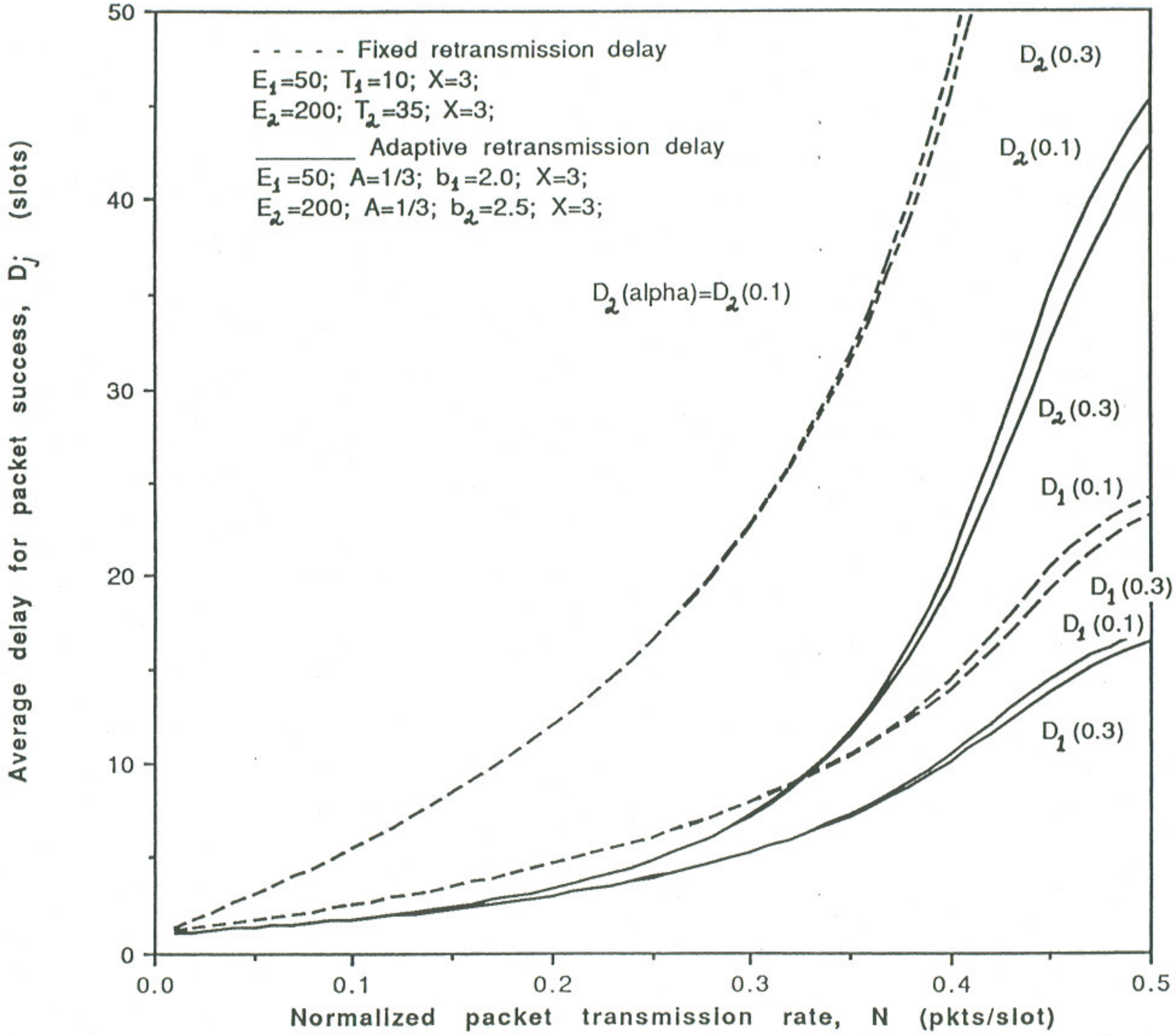

Fig. 12: Mixed traffic - adaptive retransmission delays

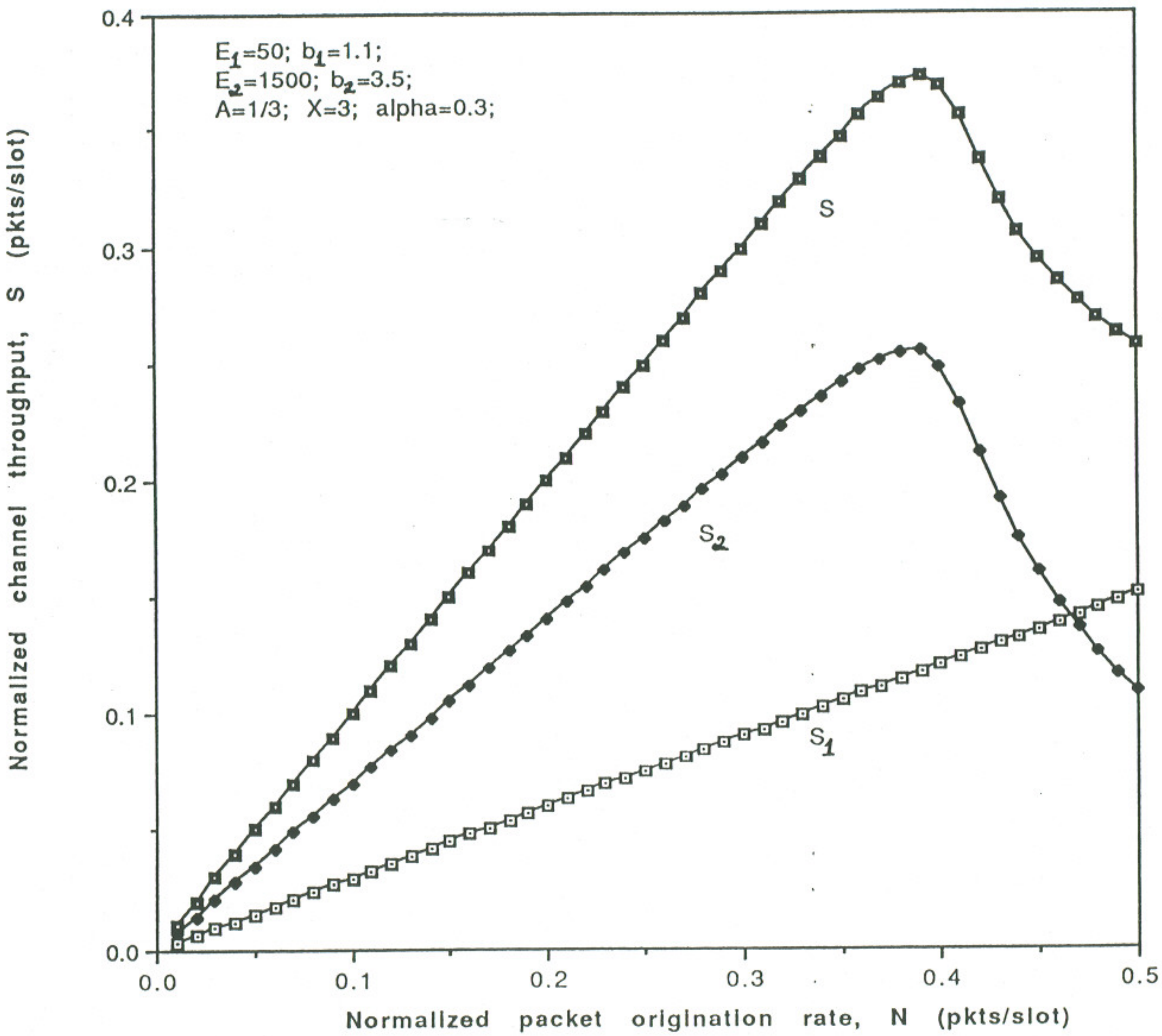

Fig. 13: Mixed traffic - adaptive retransmission delays

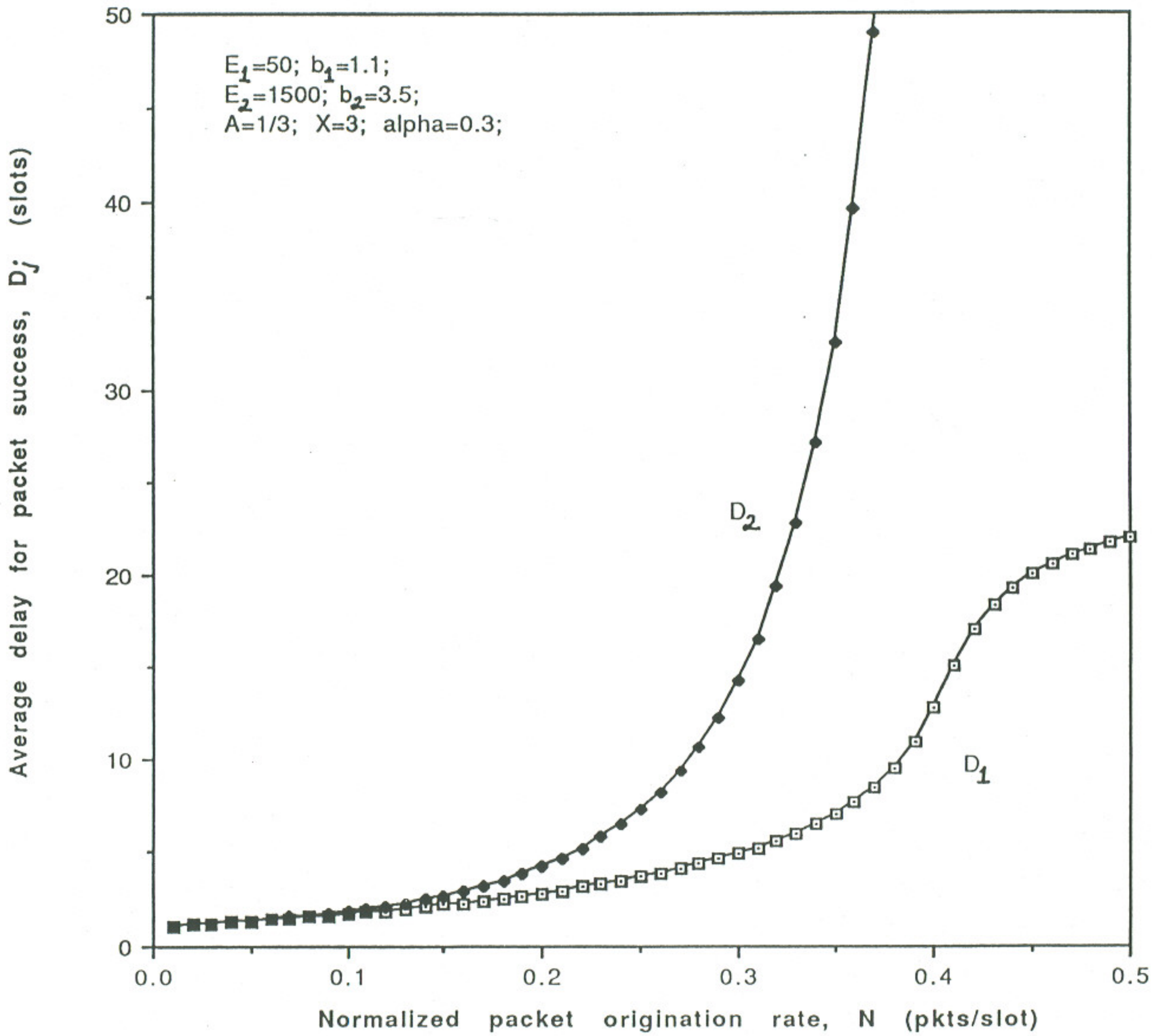

Fig. 14: Mixed traffic - adaptive retransmission delay